INTRODUCTION TO ELEMENTARY PARTICLES

David Griffiths
Reed College



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

INTRODUCTION TO ELEMENTARY PARTICLES

Copyright © 1987 John Wiley & Sons, Inc.

All rights reserved. Published simultaneously in Canada.

Reproduction or translation of any part of this work beyond that permitted by Section 107 or 108 of the 1976 United States Copyright Act without the permission of the copyright owner is unlawful. Requests for permission or further information should be addressed to the Permissions Department, John Wiley & Sons, Inc.

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 vii

Introduction 1

Elementary Particle Physics 1
How Do You *Produce* Elementary Particles? 4
How Do You *Detect* Elementary Particles? 7
Units 8
References and Notes 10

1 Historical Introduction to the Elementary Particles 11

1.1 The Classical Era (1897–1932) 11
1.2 The Photon (1900–1924) 14
1.3 Mesons (1934–1947) 17
1.4 Antiparticles (1930–1956) 18
1.5 Neutrinos (1930–1962) 22
1.6 Strange Particles (1947–1960) 28
1.7 The Eightfold Way (1961–1964) 33
1.8 The Quark Model (1964) 37
1.9 The November Revolution and Its Aftermath (1974–1983) 41
1.10 Intermediate Vector Bosons (1983) 44
1.11 The Standard Model (1978–?) 46
References and Notes 49
Problems 51

2 Elementary Particle Dynamics 55

2.1 The Four Forces 55
2.2 Quantum Electrodynamics (QED) 56
2.3 Quantum Chromodynamics (QCD) 60
2.4 Weak Interactions 65
2.5 Decays and Conservation Laws 72
2.6 Unification Schemes 76
References and Notes 78
Problems 78

3 Relativistic Kinematics 81

3.1 Lorentz Transformations 81
3.2 Four-Vectors 84
3.3 Energy and Momentum 87
3.4 Collisions 91
3.5 Examples and Applications 93
References and Notes 99
Problems 99

4 Symmetries 103

4.1 Symmetries, Groups, and Conservation Laws 103
4.2 Spin and Orbital Angular Momentum 107
4.3 Addition of Angular Momenta 109
4.4 Spin $\frac{1}{2}$ 113
4.5 Flavor Symmetries 116
4.6 Parity 122
4.7 Charge Conjugation 128
4.8 *CP* Violation 130
4.9 Time Reversal and the *TCP* Theorem 134
References and Notes 135
Problems 137

5 Bound States 143

5.1 The Schrödinger Equation for a Central Potential 143
5.2 The Hydrogen Atom 148
5.3 Fine Structure 151
5.4 The Lamb Shift 154
5.5 Hyperfine Structure 156
5.6 Positronium 159
5.7 Quarkonium 164
5.8 Light Quark Mesons 168
5.9 Baryons 172
5.10 Baryon Masses and Magnetic Moments 180
References and Notes 184
Problems 186

6 The Feynman Calculus 189

6.1 Lifetimes and Cross Sections 189
6.2 The Golden Rule 194
6.3 The Feynman Rules for a Toy Theory 201
6.4 Lifetime of the *A* 204
6.5 Scattering 204

6.6 Higher-Order Diagrams 206
References and Notes 210
Problems 211

7 Quantum Electrodynamics 213

7.1 The Dirac Equation 213
7.2 Solutions to the Dirac Equation 216
7.3 Bilinear Covariants 222
7.4 The Photon 225
7.5 The Feynman Rules for Quantum Electrodynamics 228
7.6 Examples 231
7.7 Casimir's Trick and the Trace Theorems 236
7.8 Cross Sections and Lifetimes 240
7.9 Renormalization 246
References and Notes 250
Problems 251

8 Electrodynamics of Quarks and Hadrons 257

8.1 Electron-Quark Interactions 257
8.2 Hadron Production in e^+e^- Scattering 258
8.3 Elastic Electron-Proton Scattering 262
8.4 Inelastic Electron-Proton Scattering 266
8.5 The Parton Model and Bjorken Scaling 269
8.6 Quark Distribution Functions 273
References and Notes 277
Problems 277

9 Quantum Chromodynamics 279

9.1 Feynman Rules for Chromodynamics 279
9.2 The Quark-Quark Interaction 284
9.3 Pair Annihilation in QCD 289
9.4 Asymptotic Freedom 292
9.5 Applications of QCD 295
References and Notes 296
Problems 296

10 Weak Interactions 301

10.1 Charged Leptonic Weak Interactions 301
10.2 Decay of the Muon 304
10.3 Decay of the Neutron 309
10.4 Decay of the Pion 314

10.5 Charged Weak Interactions of Quarks 317
10.6 Neutral Weak Interactions 322
10.7 Electroweak Unification 330
References and Notes 338
Problems 339

11 Gauge Theories 343

11.1 Lagrangian Formulation of Classical Particle Mechanics 343
11.2 Lagrangians in Relativistic Field Theory 344
11.3 Local Gauge Invariance 348
11.4 Yang–Mills Theory 350
11.5 Chromodynamics 355
11.6 Feynman Rules 357
11.7 The Mass Term 360
11.8 Spontaneous Symmetry-Breaking 362
11.9 The Higgs Mechanism 365
References and Notes 368
Problems 368

APPENDIX A. The Dirac Delta Function 372

APPENDIX B. Decay Rates and Cross Sections 376

APPENDIX C. Pauli and Dirac Matrices 378

APPENDIX D. Feynman Rules 380

Index 384

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:

Fast ↓ \ Small →	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

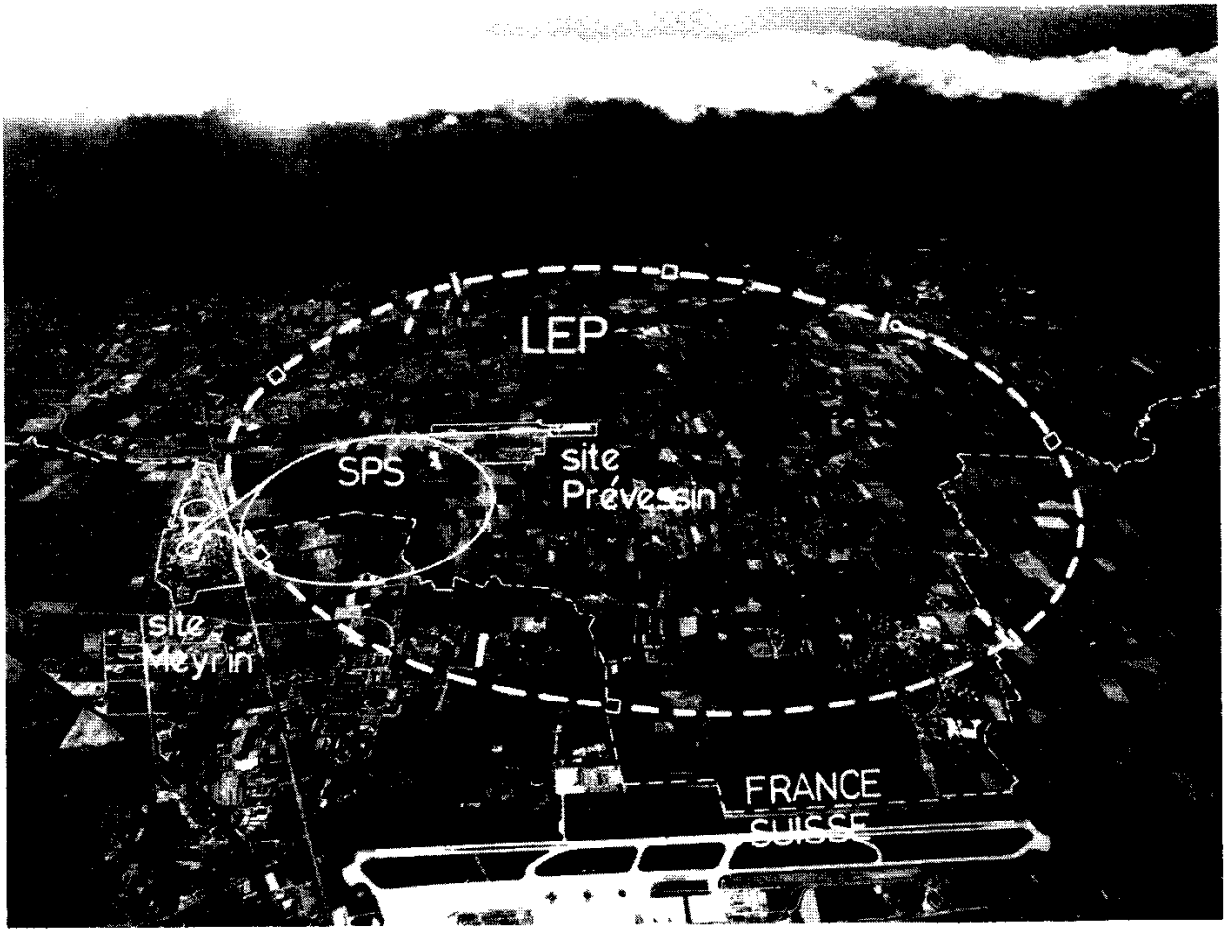


Figure I.2 CERN, outside Geneva, Switzerland. SPS is the 450 GeV Super Proton Synchrotron, later modified to make a proton–antiproton collider; LEP is a 50 GeV electron–positron storage ring now under construction. (Photo courtesy of CERN.)

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 ($\Delta x\, \Delta p \geq h/4\pi$)—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

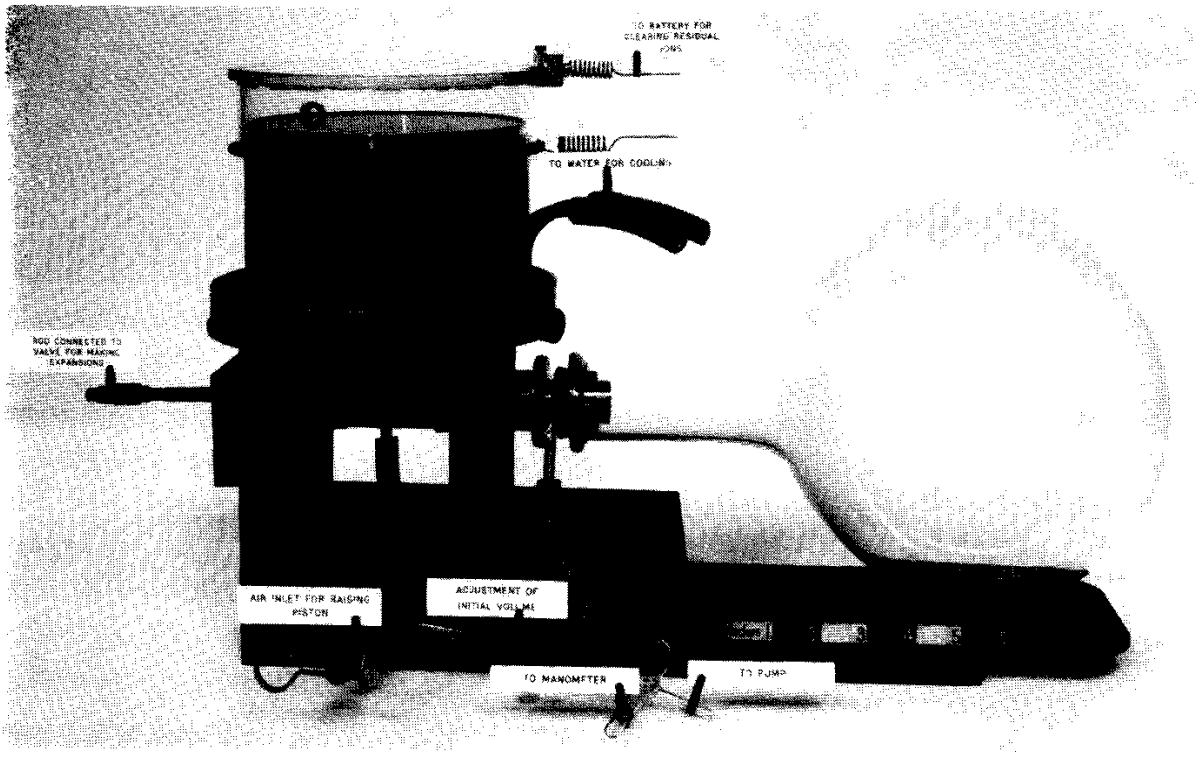


Figure I.3 An early particle detector: Wilson's cloud chamber (ca. 1900). (Photo courtesy Science Museum, London.)

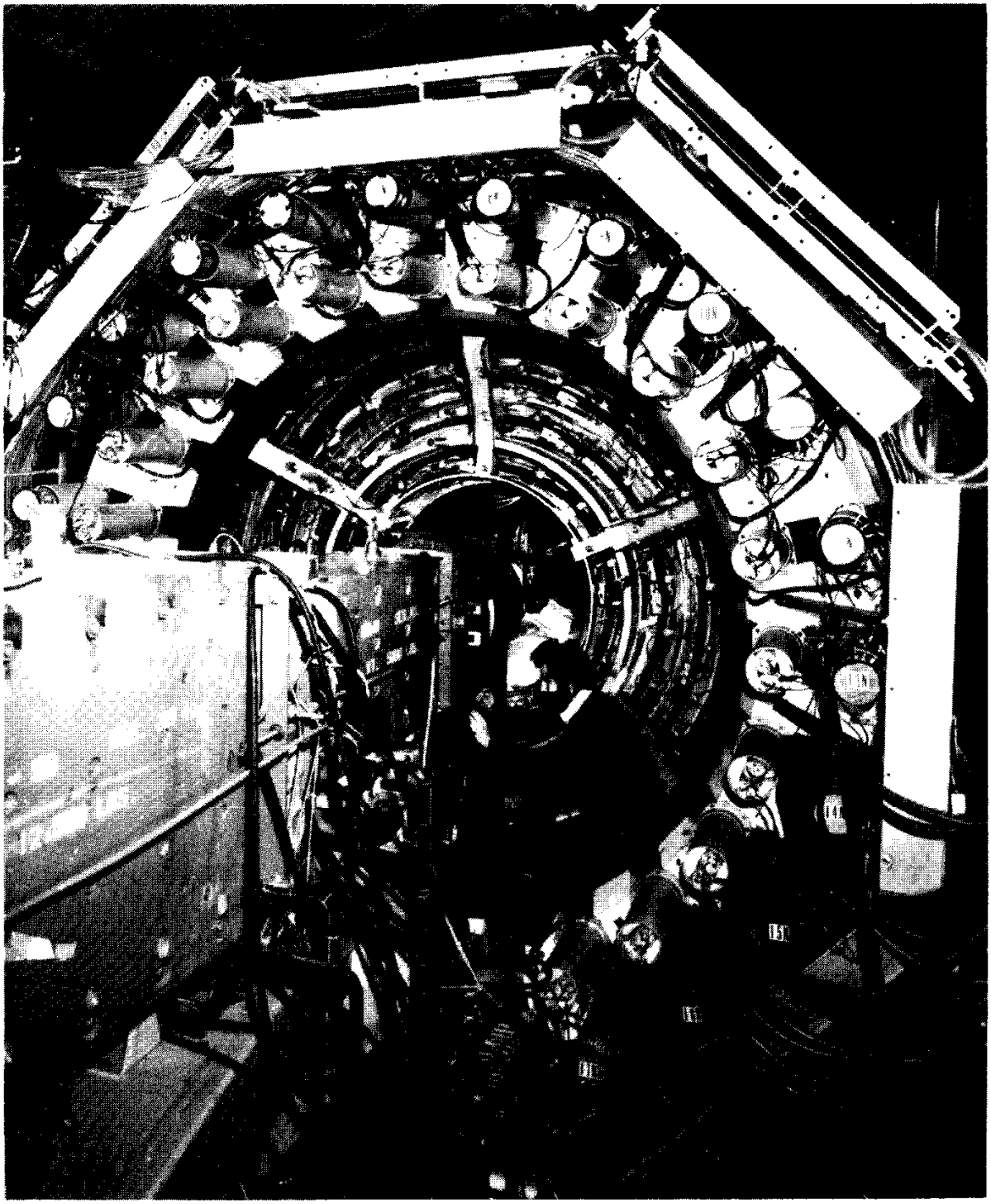


Figure I.4 A modern particle detector: The Mark I, at SLAC. (Photo courtesy SLAC.)

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

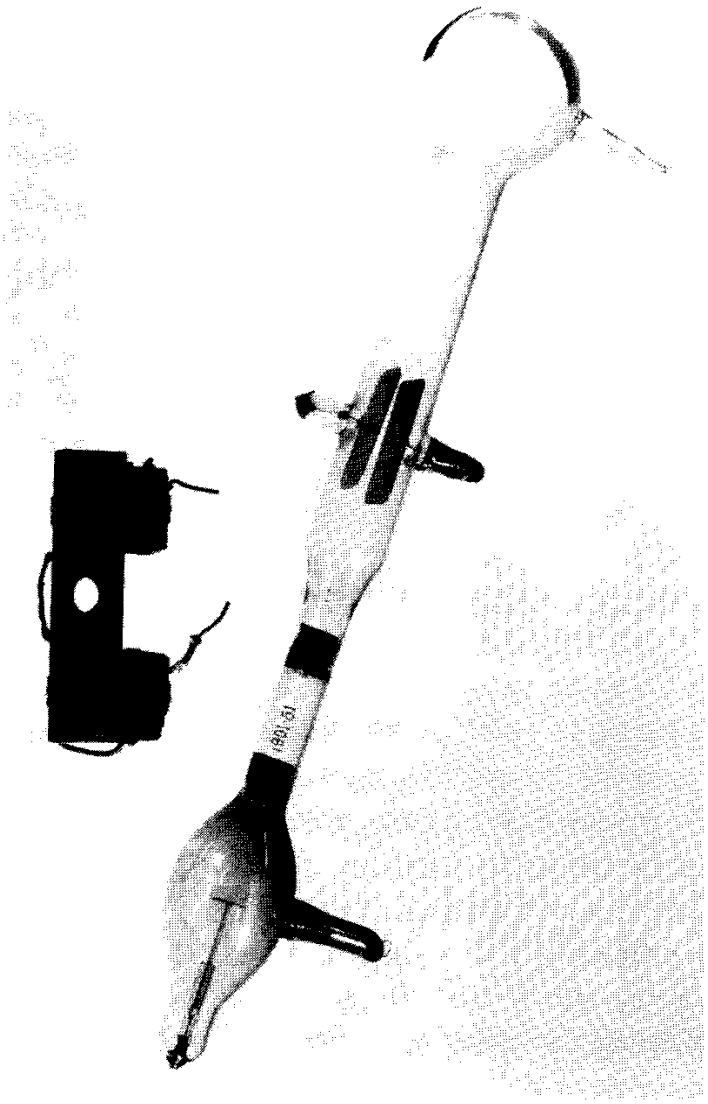


Figure 1.1 The apparatus with which J. J. Thomson discovered the electron. (Photo courtesy Science Museum, London.)

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

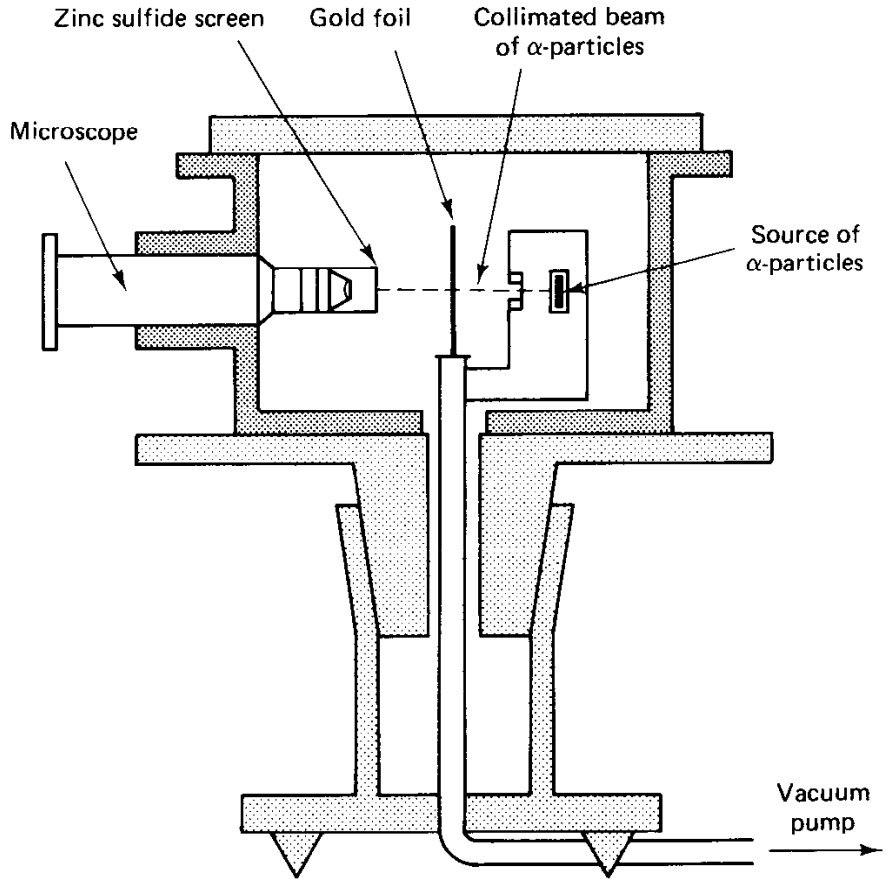


Figure 1.2 Schematic diagram of the apparatus used in the Rutherford scattering experiment. Alpha particles scattered by the gold foil strike a fluorescent screen, giving off a flash of light, which is observed visually through a microscope.

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

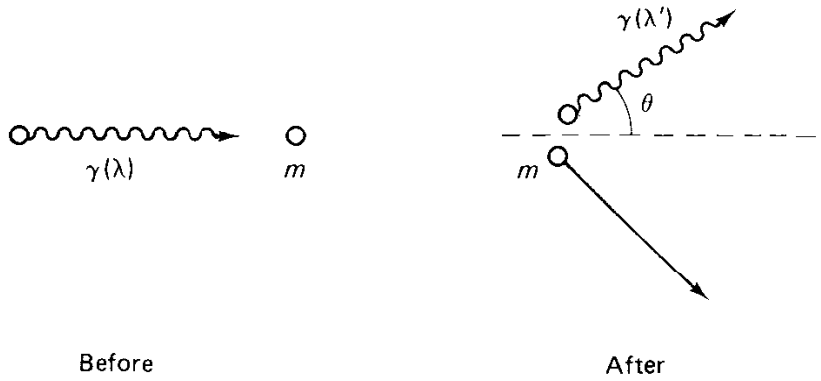


Figure 1.3 Compton scattering. A photon of wavelength λ scatters off a particle, initially at rest, of mass m. The scattered photon carries wavelength λ' given by equation (1.4).

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

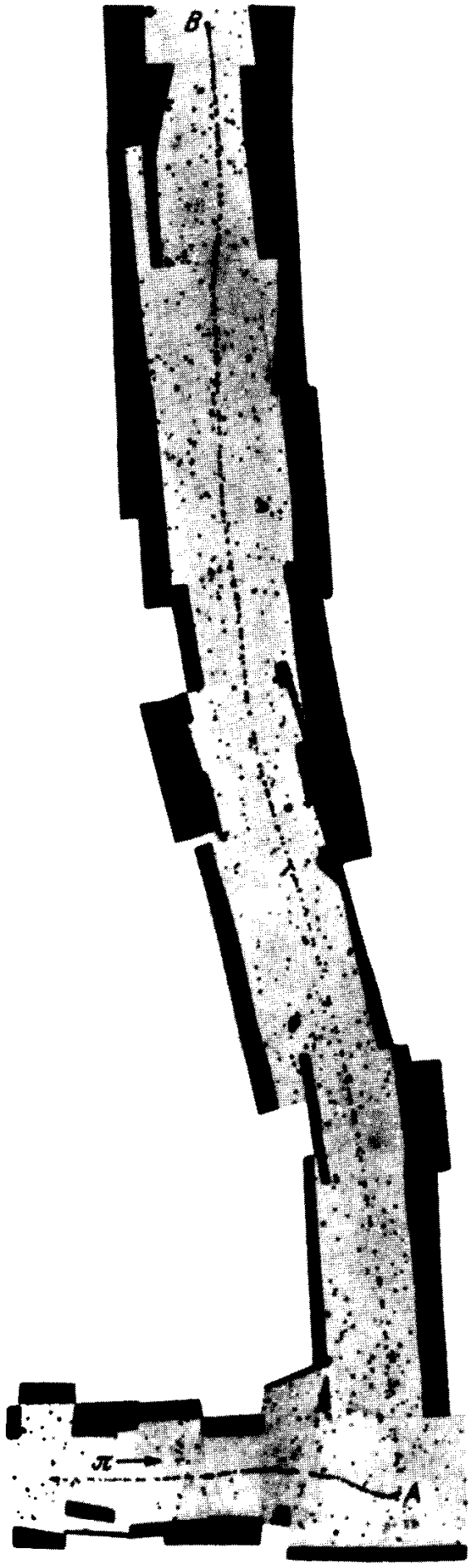


Figure 1.4 One of Powell's earliest pictures showing the track of a pion in a photographic emulsion exposed to cosmic rays at high altitude. The pion (entering from the left) decays into a muon and a neutrino (the latter is electrically neutral, and leaves no track). Reprinted by permission from C. F. 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

$$A + B \to C + D$$

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,

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

In addition, the *reverse* reaction occurs $C + D \to A + B$, 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 ($^{40}_{19}\mathrm{K} \rightarrow {}^{40}_{20}\mathrm{Ca}$), copper goes to zinc ($^{64}_{29}\mathrm{Cu} \rightarrow {}^{64}_{30}\mathrm{Zn}$), tritium goes to helium ($^{3}_{1}\mathrm{H} \rightarrow {}^{3}_{2}\mathrm{He}$), and so on. [The upper number is the *atomic*

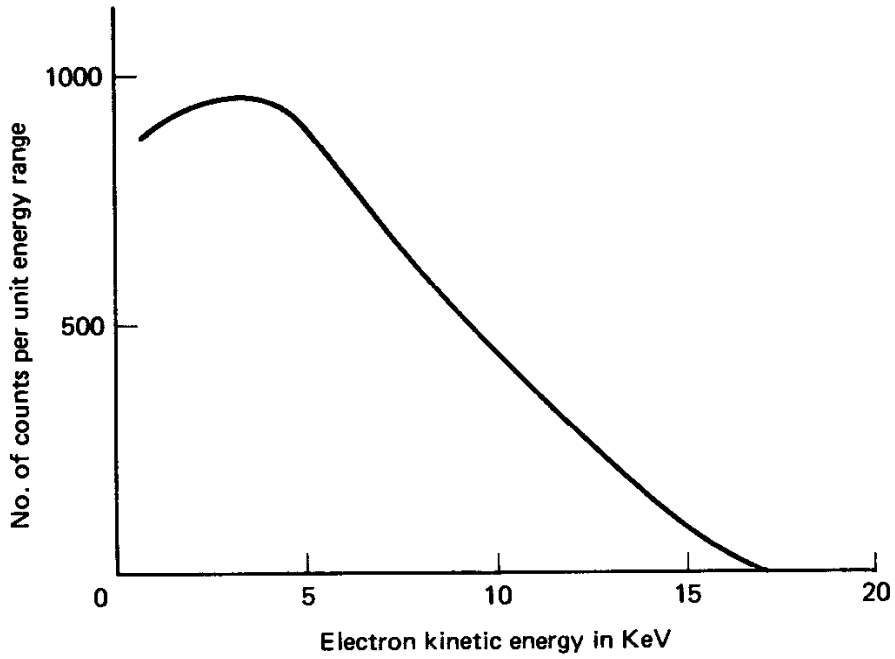


Figure 1.6 The beta decay spectrum of tritium (${}_1^3\mathrm{H} \rightarrow {}_2^3\mathrm{He}$). (*Source:* G. M. Lewis, *Neutrinos* (London: Wykeham, 1970), p. 30.)

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

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

(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:

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

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:

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

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.

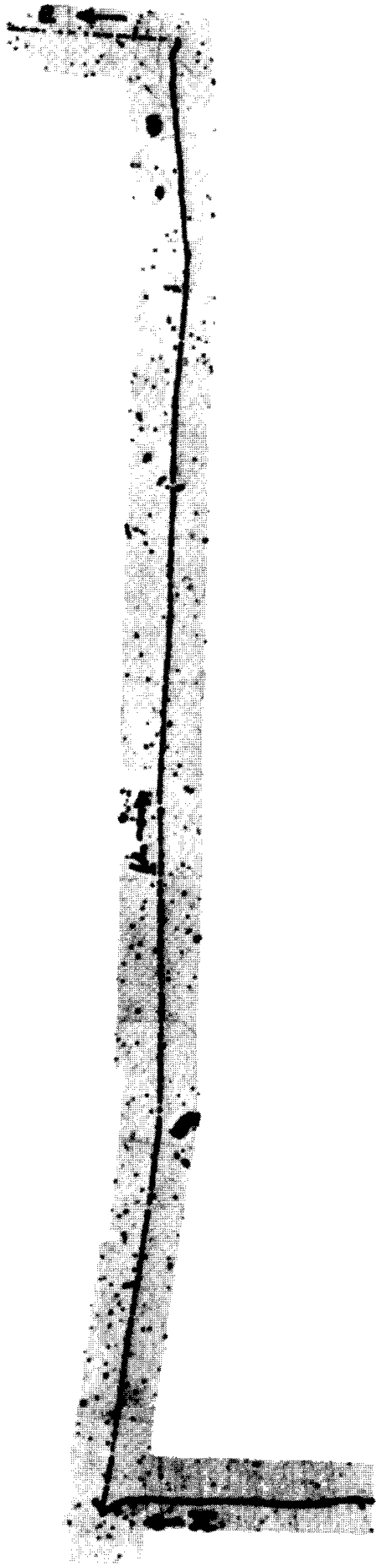


Figure 1.7 Here, a pion decays into a muon (plus a neutrino); the muon subsequently decays into an electron (and two neutrinos). Reprinted by permission from C. F. Powell, P. H. Fowler, and D. H. Perkins, *The Study of Elementary Particles by the Photographic Method* (New York: Pergamon, 1959). First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

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

and the muon decays (1.10) are really

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

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:

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

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 $\mu \longrightarrow e + \gamma$ suggests a law of conservation of "mu-ness"; but then how are we to explain the observed decays $\mu \longrightarrow e + \nu + \bar{\nu}$? 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

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

the pion decays are

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

and the muon decays take the form

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

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

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

and no cases of the forbidden process

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

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.

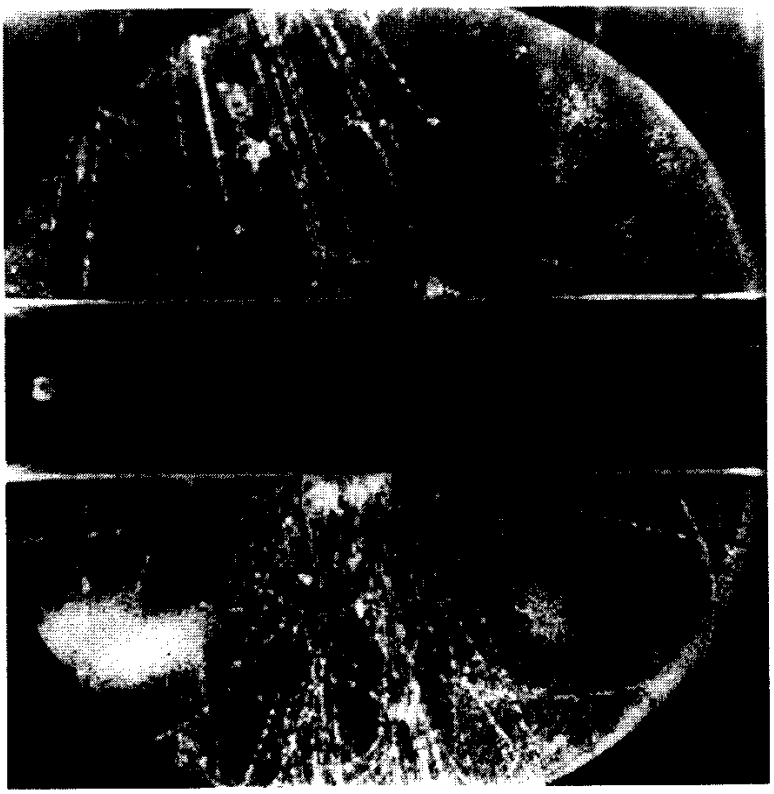
Meanwhile, in 1950 another neutral "V" particle was found, this time by Anderson's group at Cal Tech. The photographs were similar to Rochester's (Fig. 1.8), but this time the products were a p^+ and a π^-. Evidently this particle is substantially heavier than the proton; we call it the Λ:

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

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 $\bar{n}$)



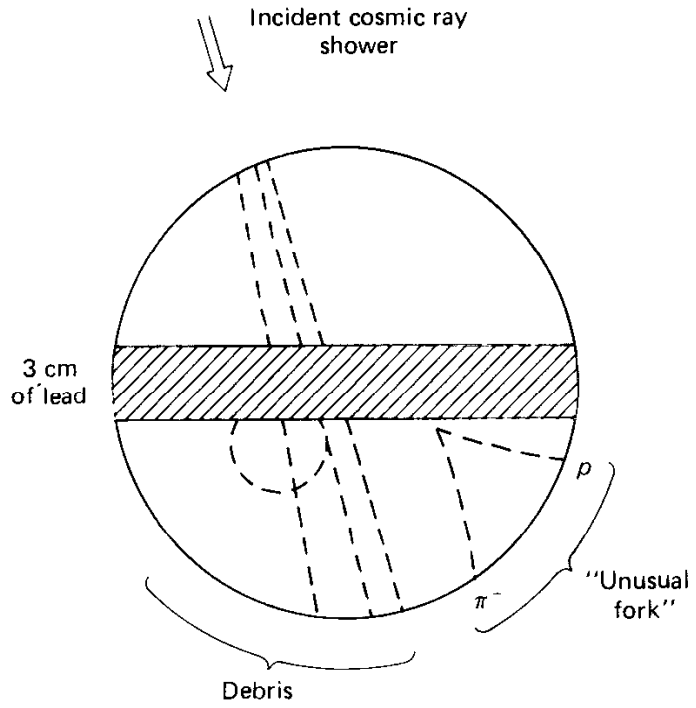


Figure 1.8 The first strange particle. Cosmic rays strike a lead plate, producing a K^0, which subsequently decays into a pair of charged pions. (Photo courtesy of Prof. G. D. Rochester. Reprinted by permission from *Nature* **160,** 855. Copyright © 1947, Macmillan Journals Limited.)

$A = -1$; 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:

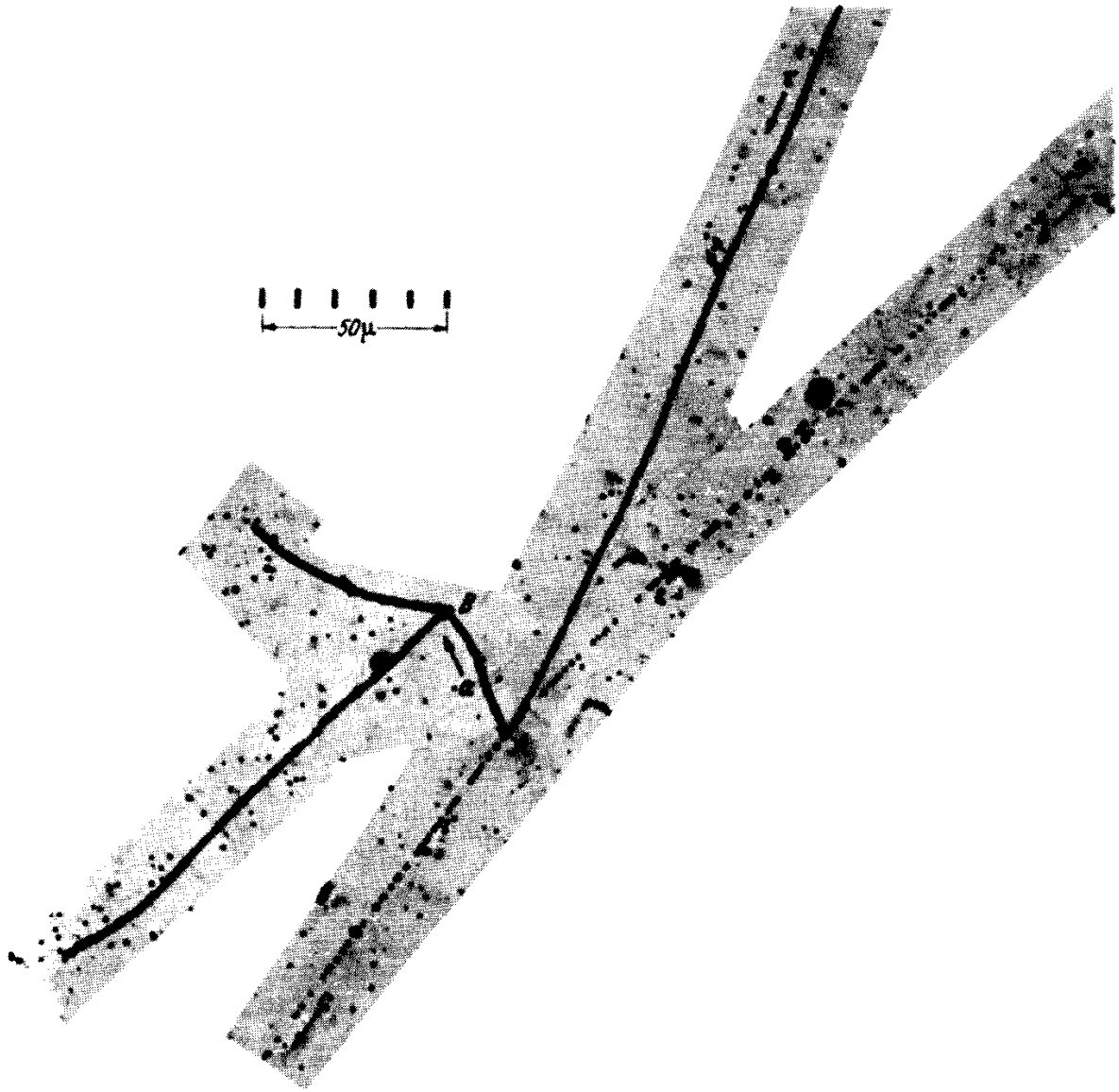


Figure 1.9 K^+, entering from above, decays at A: $K^+ \rightarrow \pi^+ + \pi^+ + \pi^-$. (The π^- subsequently causes a nuclear disintegration at B). [Reprinted by permission from C. F. Powell, P. H. Fowler, and D. H. Perkins, *The Study of Elementary Particles by the Photographic Method* (New York: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

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

($A = 2$ 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 ($\pi^- \rightarrow \mu^- + \bar{\nu}_\mu$) a meson disappears, and in lambda decay ($\Lambda \rightarrow p^+ + \pi^-$) 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:

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

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:

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

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

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

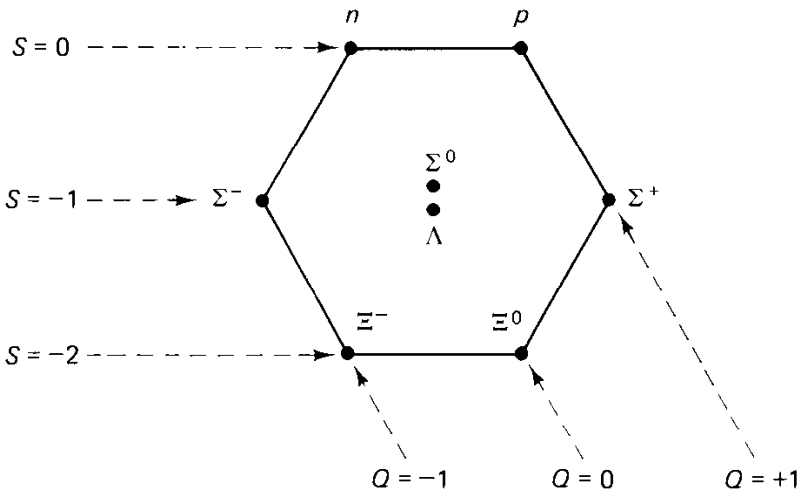
for these are *weak* processes, which do not respect conservation of strangeness.

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:

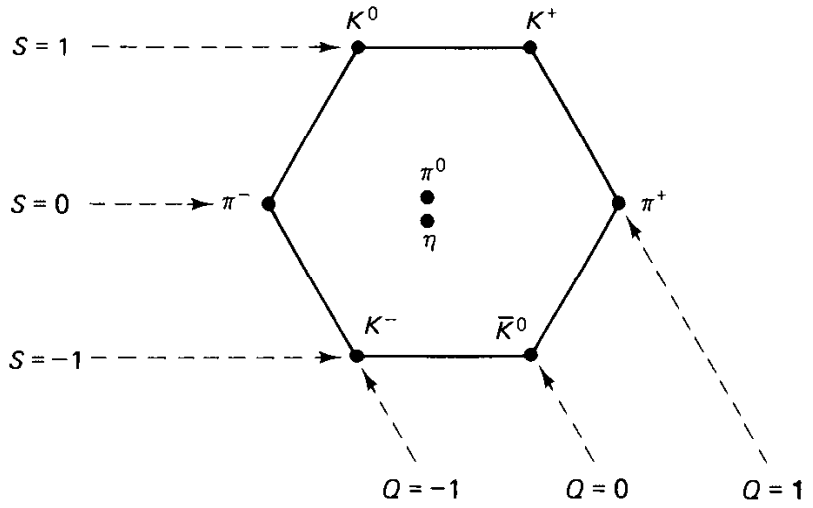


The Baryon Octet

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.

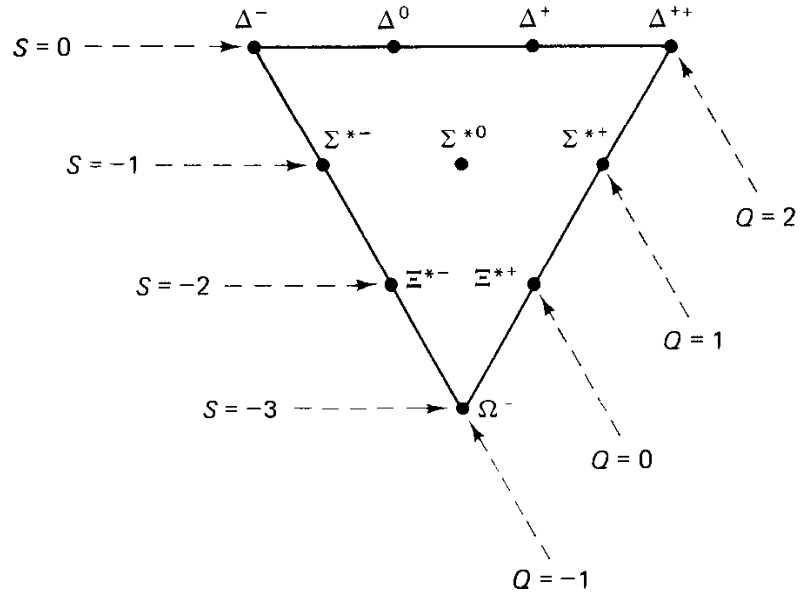
The eight lightest mesons fill a similar hexagonal pattern, forming the (*pseudo-scalar*) *meson octet:*



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

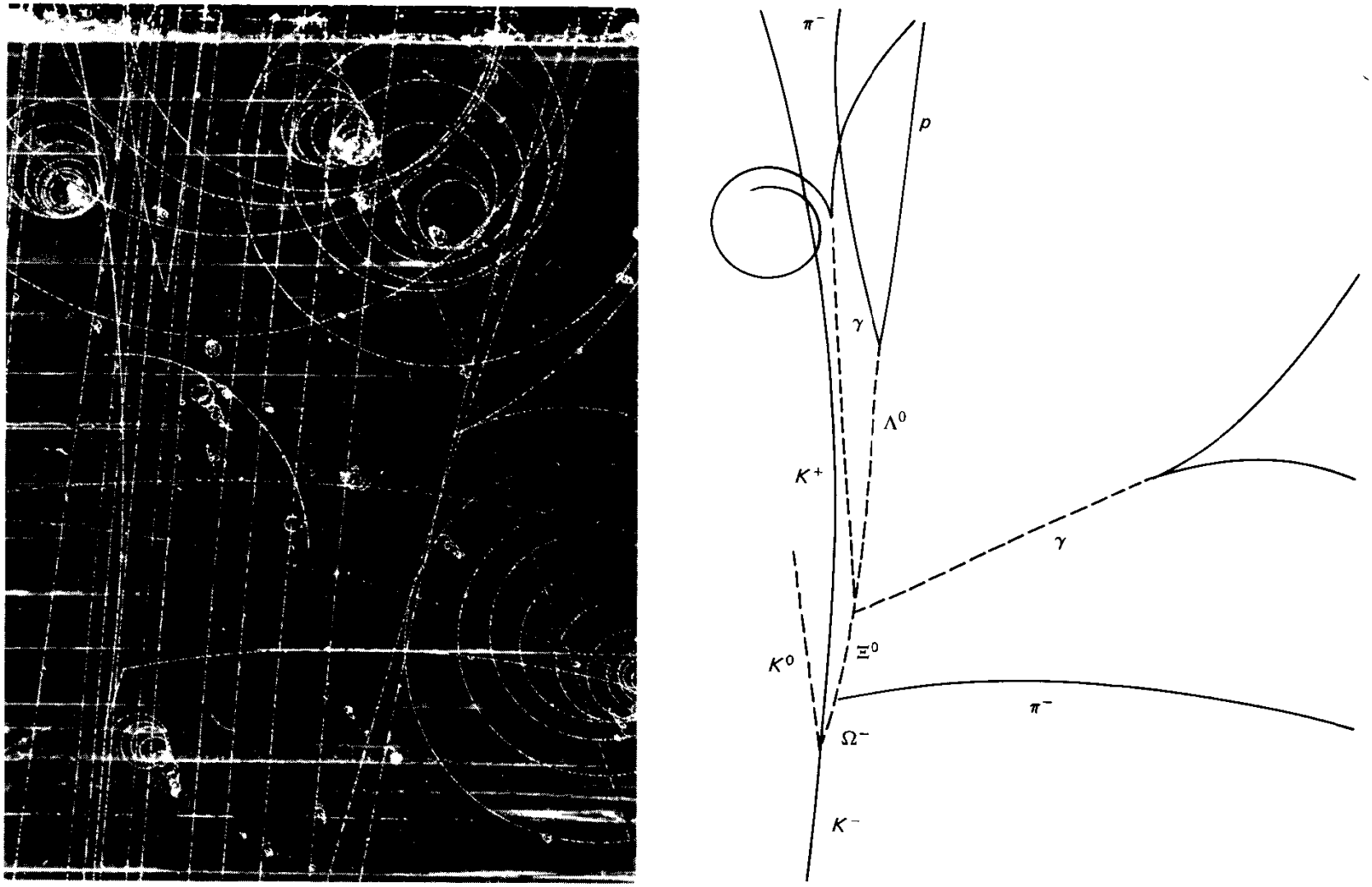


Figure 1.10 The discovery of the Ω^-. The actual bubble chamber photograph is shown on the left; a line diagram of the relevant tracks on the right. (Photo courtesy Brookhaven National Laboratory.)

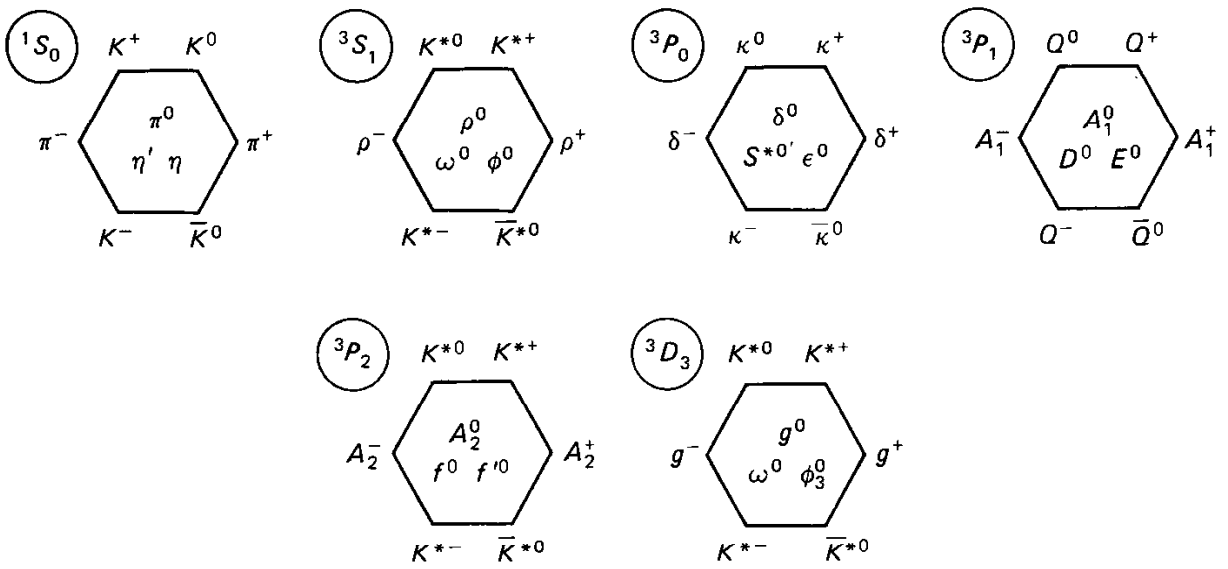


Figure 1.11 Established meson nonets. Obviously, we are running out of letters. It is customary to distinguish different particles represented by the same letter by indicating the mass parenthetically (in MeV/c^2), thus $K^*(892)$, $K^*(1430)$, $K^*(1650)$, and so on. In this figure the supermultiplets 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

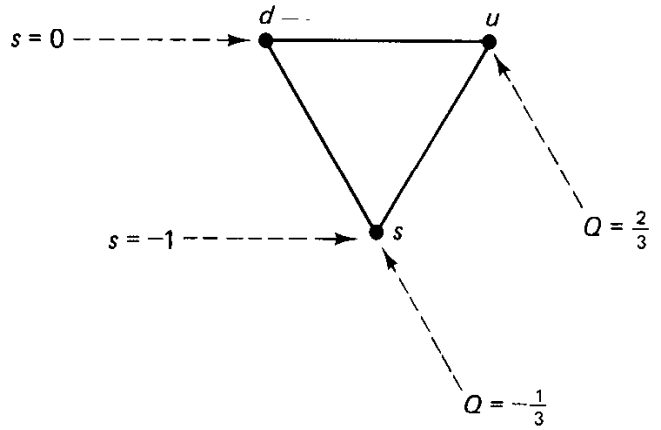
* A similar thing happened in the case of the Periodic Table. There were three famous "holes" (missing elements) on Mendeleev's chart, and he predicted that new elements would be discovered to fill in the gaps. Like Gell-Mann, he confidently described their properties, and within 20 years all three—gallium,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

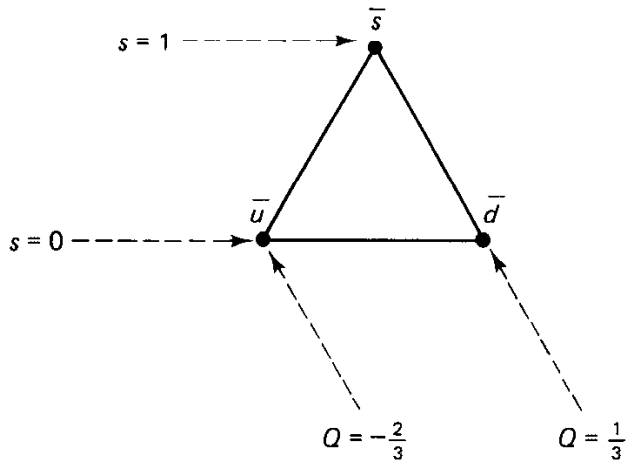
1.8 THE QUARK MODEL (1964)

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,

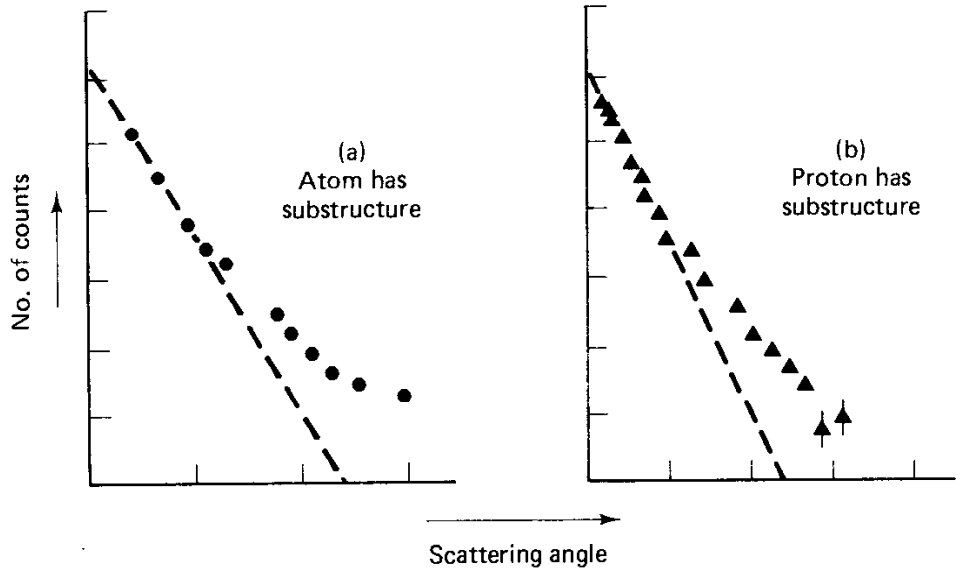


Figure 1.12 (a) In Rutherford scattering the number of particles deflected through large angles indicates that the atom has internal structure (a nucleus). (b) In deep inelastic scattering the number of particles deflected through large angles indicates that the proton has internal structure (quarks). 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, ν_e, μ, ν_μ
Quarks: d, u, s

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

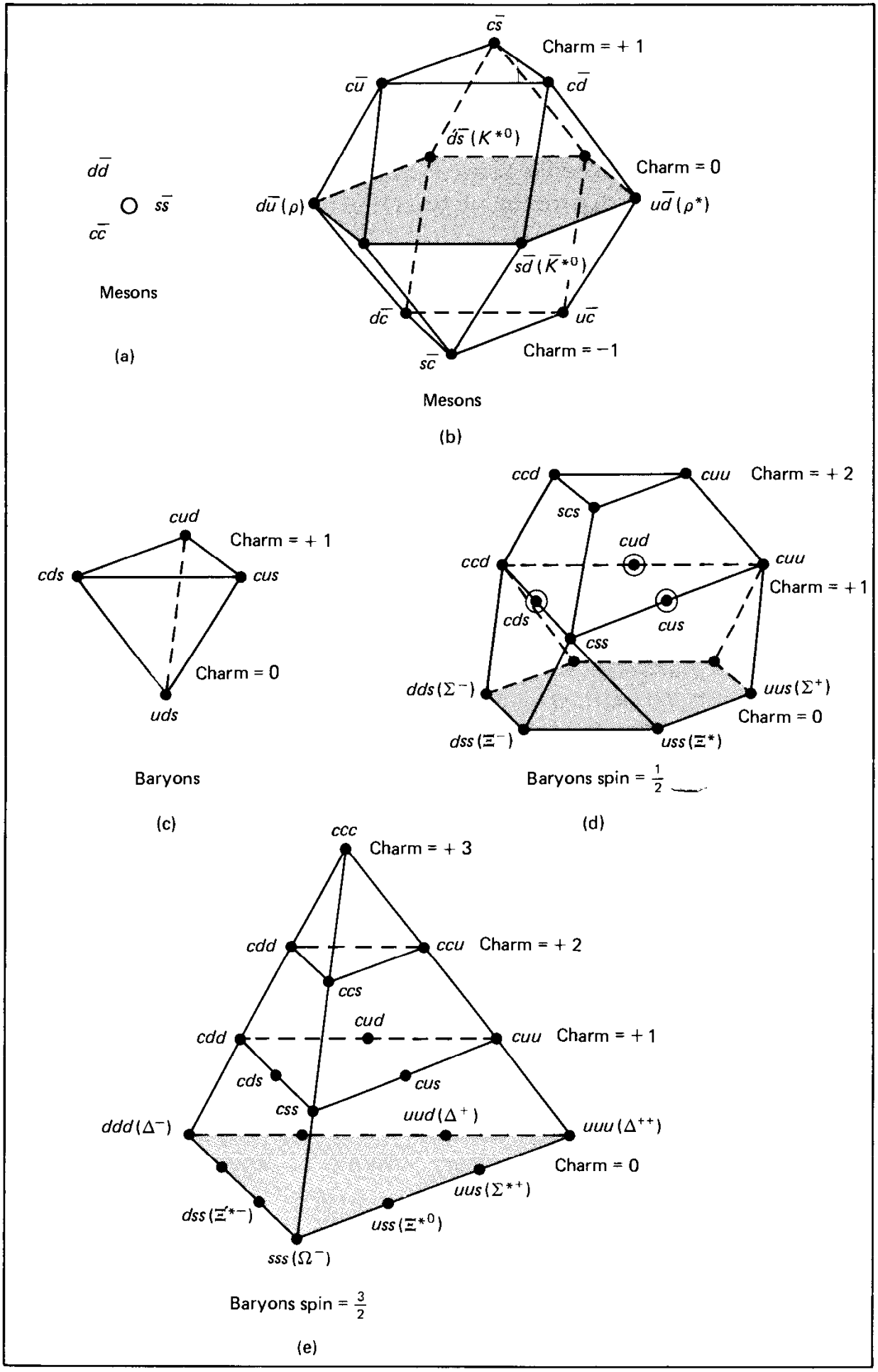


Figure 1.13 Supermultiplets constructed with four quarks. (From "Quarks with Color and Flavor," by S. Glashow. Copyright © Oct. 1975 by Scientific American, Inc. All rights reserved.)

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 ± 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 ± 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 ac-

cording 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

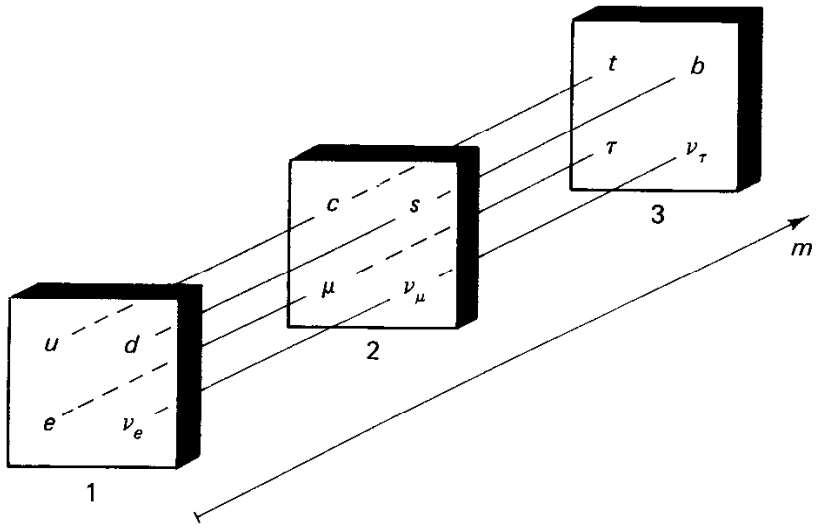


Figure 1.15 The three generations of quarks and leptons, in order of increasing mass.

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 $M_W \geq 37$ GeV/c^2 and $M_Z \geq 75$ 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,* $\Delta E\, \Delta t \geq \hbar$, 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 $\Sigma \longrightarrow \pi$, $\Lambda \longrightarrow \eta$, 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, $\Delta^{++} \longrightarrow p^+ + \pi^+$. 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 ($cc\bar{n}\bar{n}\bar{n}$, 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

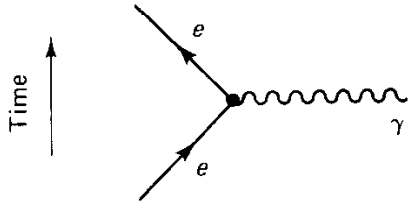
* The "strength" of a force is an intrinsically ambiguous notion—after all, it depends on the nature of the source and on how far away you are. So the numbers in this table should not be taken too literally, and (especially in the case of the weak force) you will see quite different figures quoted elsewhere.

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.

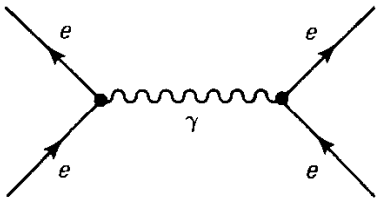
2.2 QUANTUM ELECTRODYNAMICS (QED)

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:*



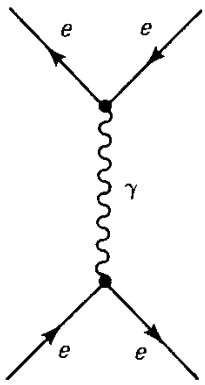
This diagram reads: Charged particle e enters, emits (or absorbs) a photon, γ, and exits.* For the sake of argument, I'll assume the charged particle is an electron; it could just as well be a quark, or any lepton except a neutrino (the latter is neutral, of course, and does not experience an electromagnetic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

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



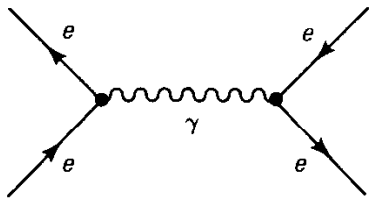
Here, two electrons enter, a photon passes between them (I need not say which one emits the photon and which one absorbs it; the diagram represents *both* orderings), and the two then exit. This diagram, then, describes the interaction between two electrons; in the classical theory we would call it the 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:



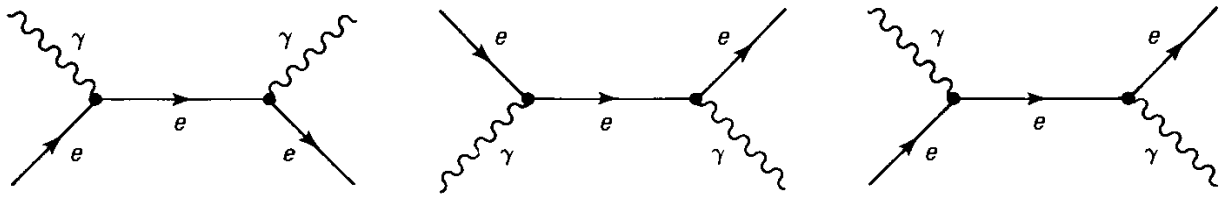
The rule of the game is that a particle line running "backward in time" (as indicated by the arrow) is to be interpreted as the corresponding *antiparticle* going *forward* (the photon is its own antiparticle, that's why I didn't need an arrow on the photon line). So in this process an electron and a 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.

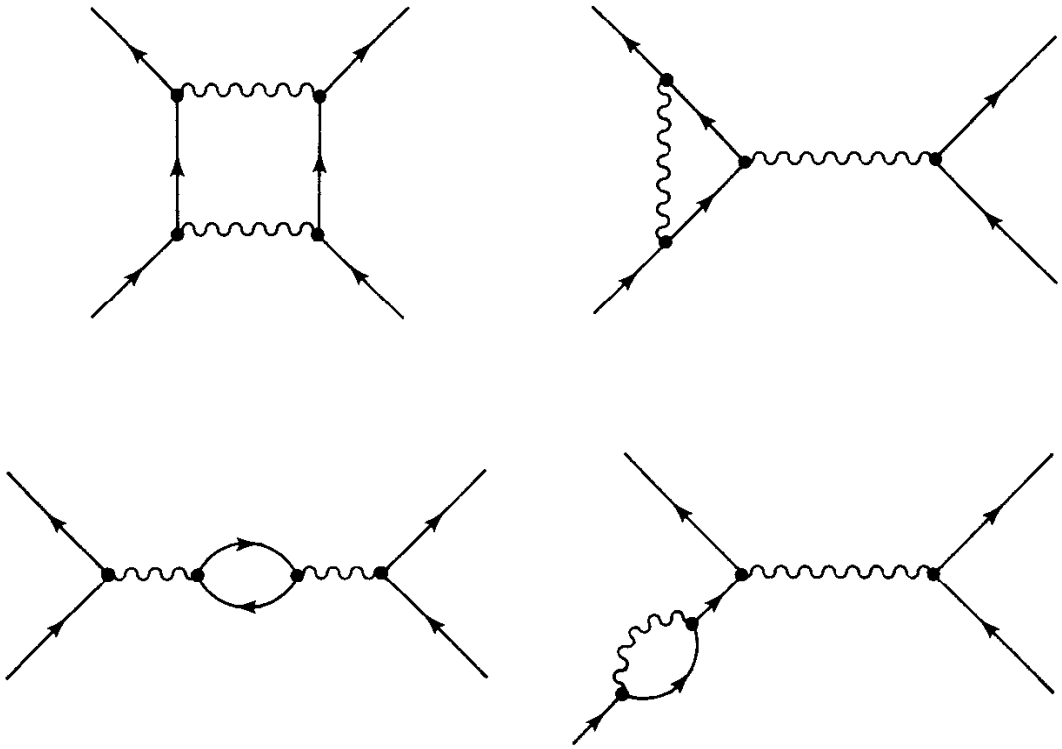


Both diagrams must be included in the analysis of Bhabha scattering.

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:

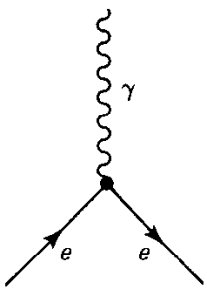


In each of these figures two electrons went in and two electrons came out. They too describe the repulsion of like charges (Møller scattering). The "innards" of the diagram are irrelevant as far as the observed process is concerned. Internal lines (those which begin and end within the diagram) represent 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: $e^- \rightarrow e^- + \gamma$ 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 $e^- + e^+ \rightarrow \gamma$ kinematically possible, although it is easy enough to draw the diagram:



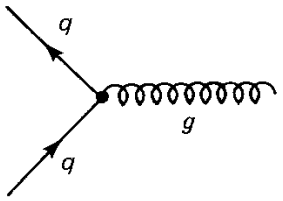
In the center-of-mass system the electron and positron enter symmetrically with equal and opposite velocities, so the total momentum before the collision is obviously zero. But the *final* momentum *cannot* be zero, since photons always

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.

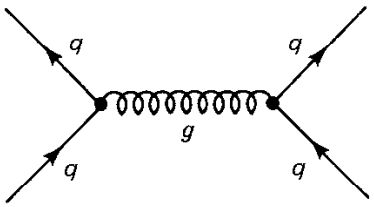
[Actually, the *physical* distinction between real and virtual particles is not quite as sharp as I have implied. If a photon is emitted on Alpha Centauri, and absorbed in your eye, it is technically a virtual photon, I suppose. However, in general, the farther a virtual particle is from its mass shell 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 $e^- \longrightarrow e^- + \gamma$) is quark $\longrightarrow$ quark-plus-gluon (since leptons do not carry color, they do not participate in the strong interactions):



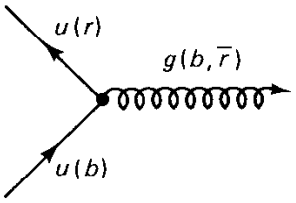
As before, we combine two or more such "primitive vertices" to represent more complicated processes. For example, the force between two quarks (which is responsible in the first instance for binding quarks together to make baryons, and indirectly for holding the neutrons and protons together to form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



* In special relativity, the energy E, momentum, $\mathbf{p}$, and mass $\mathbf{m}$ 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.

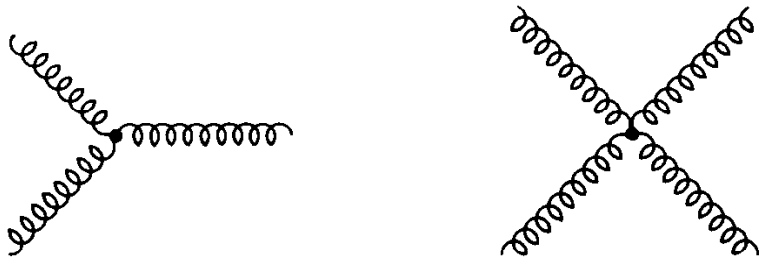
We say that the force between two quarks is "mediated" by the exchange of gluons.

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

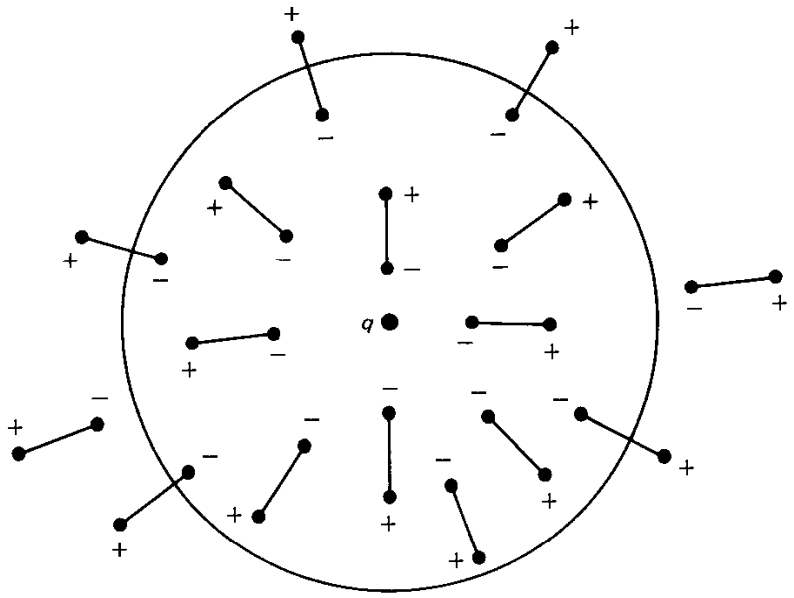


Figure 2.1 Screening of a charge q by a dielectric medium.

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

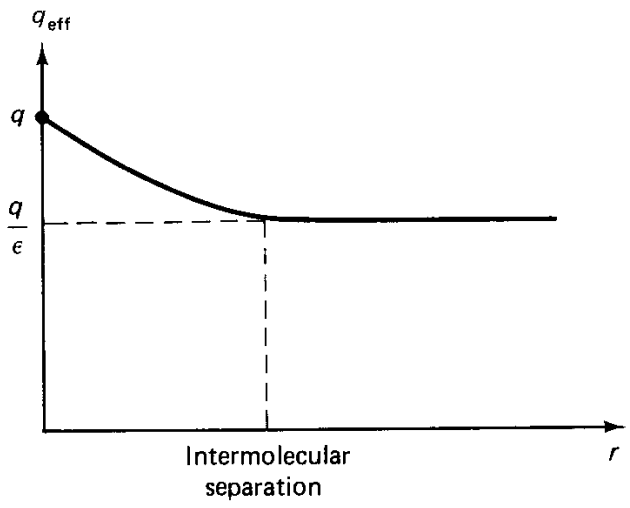


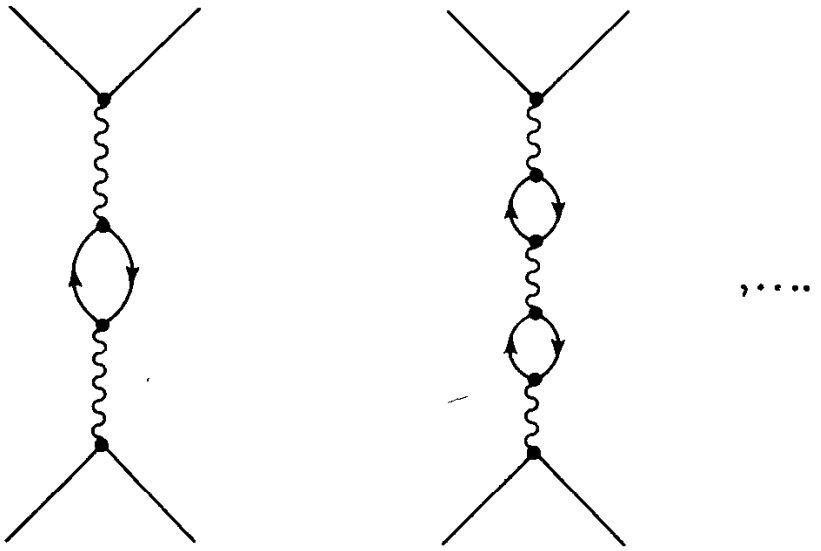
Figure 2.2 Effective charge as a function of distance.

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

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

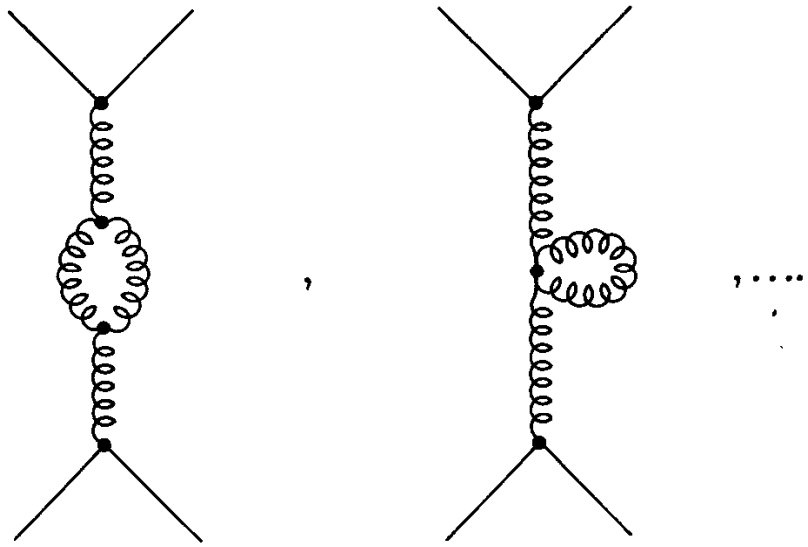
(The factor ϵ by which the field is reduced is called the *dielectric constant* of the material; it is a measure of the ease with which the substance can be polarized.[2]) Of course, if you are in closer than the nearest molecule, then there is no such screening, and you "see" the full charge q. Thus 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:

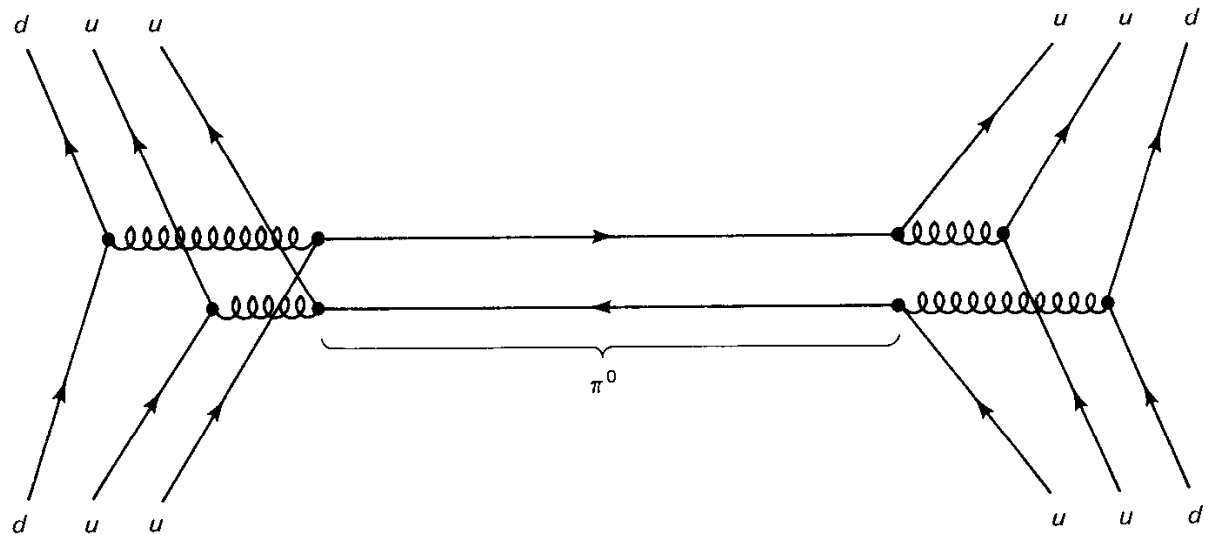


It is not clear a priori what influence these diagrams will have on the story;[3] as it turns out, their effect is the *opposite:* There occurs a kind of competition between the quark polarization diagrams (which drive α_s *up* at short distances) and gluon polarization (which drives it *down*). Since 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:



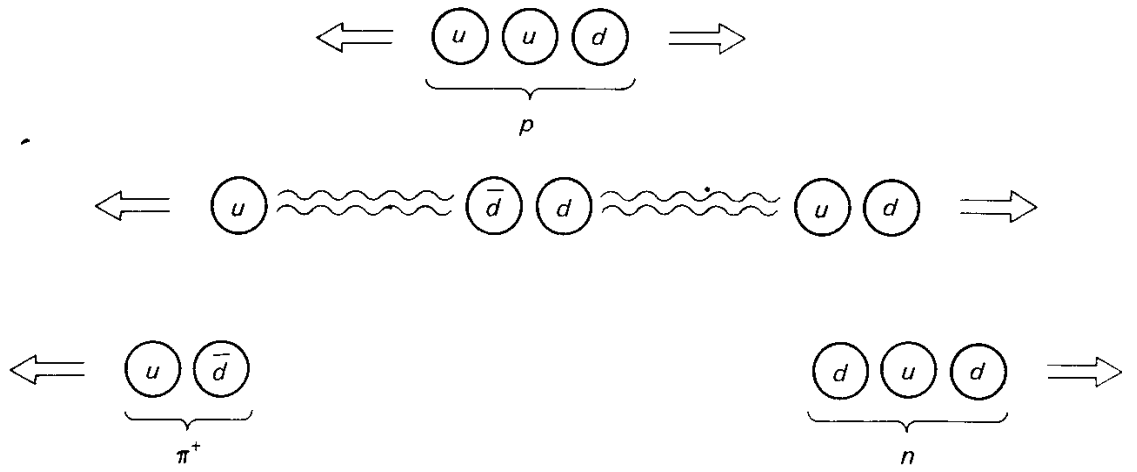


Figure 2.3 A possible scenario for quark confinement: As we pull a u quark out of the proton a pair of quarks is created, and instead of a free quark, we are left with a pion and a neutron.

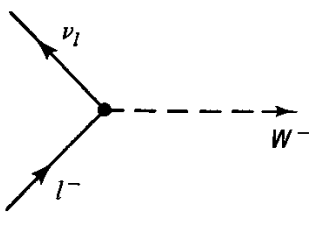
You will recognize here the remnants of Yukawa's pion-exchange model, but the entire process is enormously more complex that Yukawa ever imagined. If QCD is correct, it must contain the explanation for quark confinement; that is, it must be possible to *prove,* as a consequence of this theory, that quarks 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]

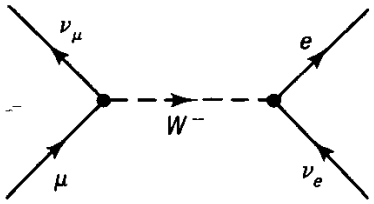
There is no particular name for the "stuff" that produces weak forces, in the sense that electric charge produces electromagnetic forces and color produces strong forces. Some people call it "weak charge." Whatever word you use, all quarks and all leptons carry it. (Leptons have no color, so they do 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

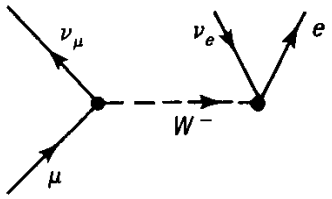
The fundamental charged vertex looks like this:



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:

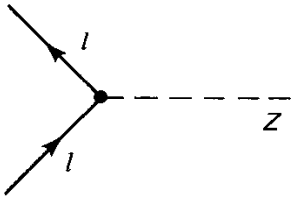


Such a neutrino-muon scattering event would be hard to set up in the laboratory, but with a slight twist essentially the same diagram describes the decay of the muon, $\mu^- \rightarrow e^- + \nu_\mu + \bar{\nu}_e$, which happens all the time:

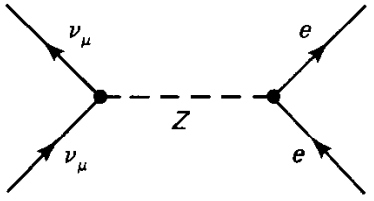


(Technically, this is only the lowest-order contribution to muon decay, but in weak interaction theory one almost never needs to consider higher-order corrections.)

The fundamental neutral vertex is:



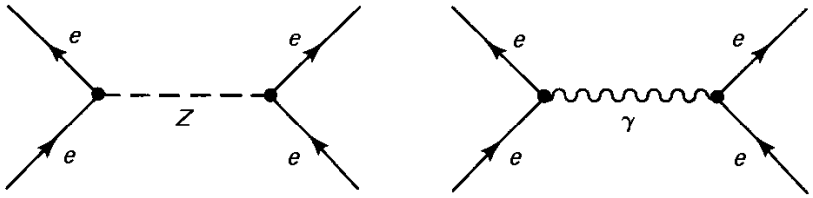
In this case l can be *any* lepton (including neutrinos). The Z mediates such processes as neutrino-electron scattering ($\nu_\mu + e^- \rightarrow \nu_\mu + e^-$):



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.

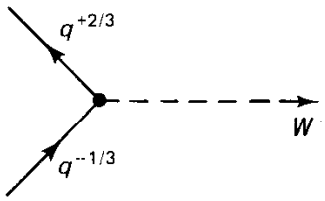
includes neutral weak processes as essential ingredients, and their existence was confirmed experimentally at CERN in 1973.[5] The reason it took so long for neutral weak processes to be discovered is twofold: (1) nobody was looking for them and (2) they tend to be masked by much stronger electromagnetic 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 $e^- + e^+ \longrightarrow \mu^- + \mu^+$ 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.

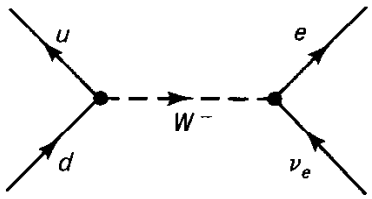
2.4.2 Quarks

Notice that the leptonic weak vertices connect members of the *same generation:* e^- converts to ν_e (with emission of W^-), or $\mu^- \longrightarrow \mu^-$ (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:

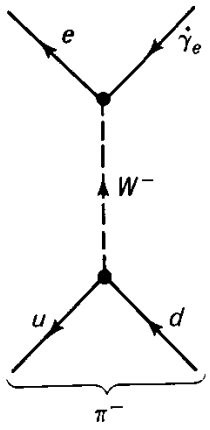


A quark with charge $-\frac{1}{3}$ (which is to say: d, s, or b) converts into the corresponding quark with charge $+\frac{2}{3}$ (u, c, or t, respectively), with the emission of a W^-. The outgoing quark carries the same color as the ingoing one, but a different flavor. It's not that the W carries off 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.")

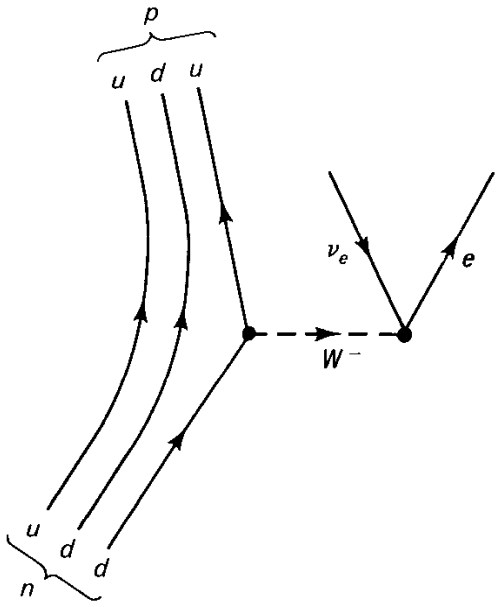
The far end of the W line can couple to leptons (a "semileptonic" process), or to other quarks (a purely hadronic process). The most important semileptonic process is undoubtedly $d + \nu_e \longrightarrow u + e$:



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, $\pi^- \longrightarrow e^- + \bar{\nu}_e$:

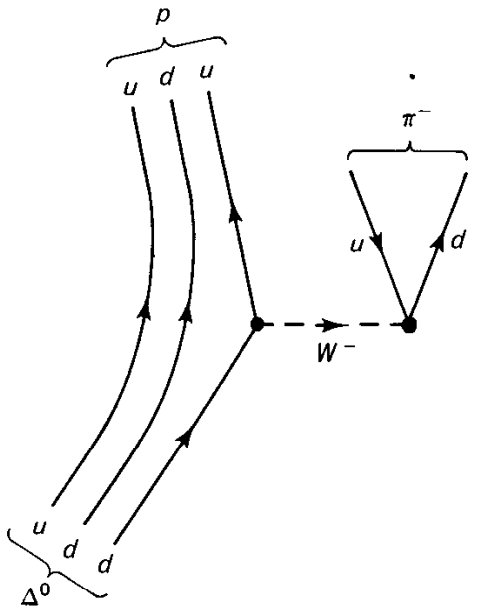


(For reasons to be discussed later, the more common decay is actually $\pi^- \longrightarrow \mu^- + \bar{\nu}_\mu$, but the diagram is the same, with e replaced by μ.) Moreover, essentially the same diagram accounts for the beta decay of the neutron ($n \longrightarrow p^+ + e^- + \bar{\nu}_e$):

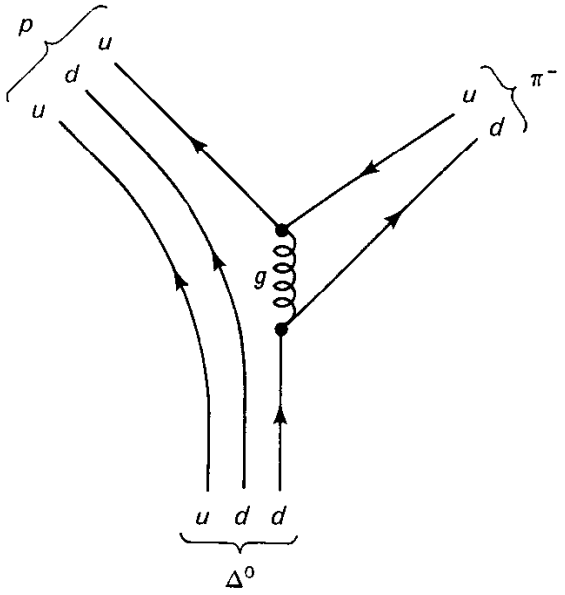


Thus, apart from strong interaction contamination (in the form of the "spectator" u and d quarks), the decay of the neutron is identical in structure to the decay of the muon, and closely related to the decay of the pion. In the days before the quark model, these appeared to be three very different processes.

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

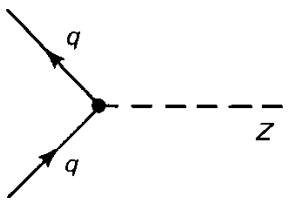


Actually, this particular decay also proceeds by the strong interaction:



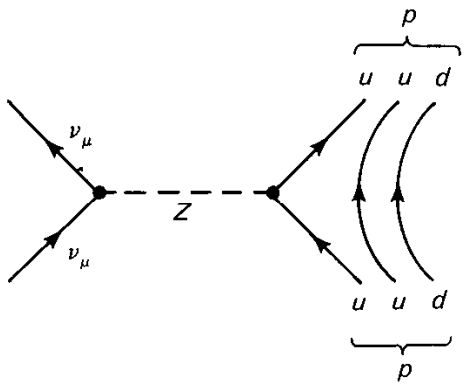
The weak mechanism is an immeasurably small contribution. We'll see more realistic examples of nonleptonic weak interactions in a moment.

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:



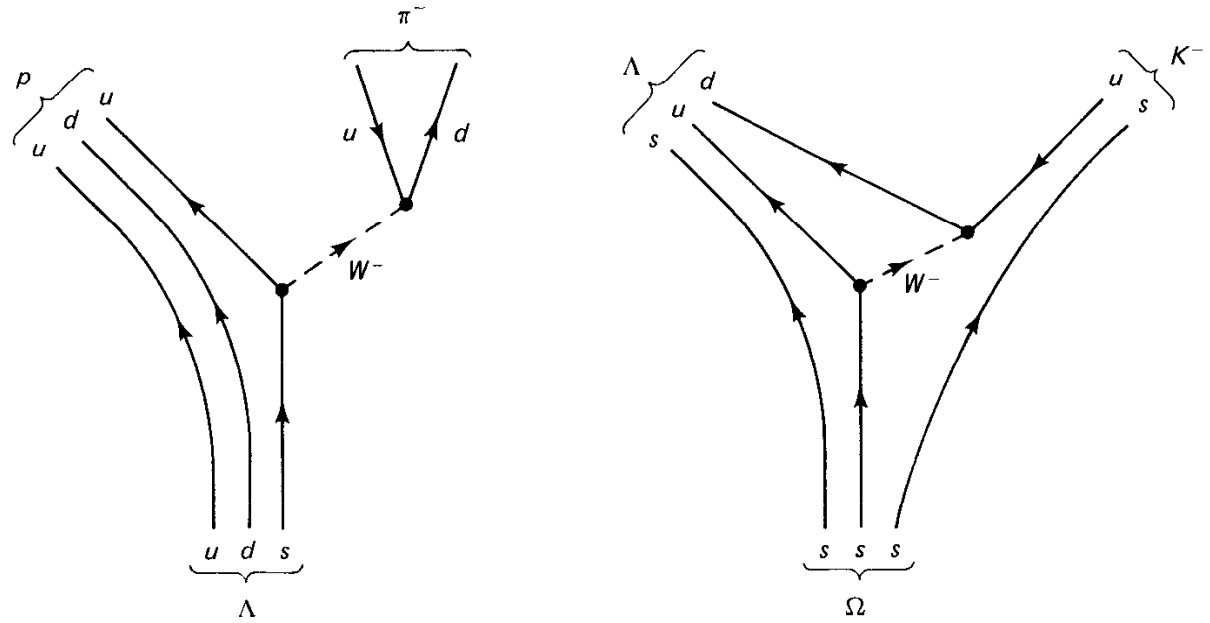
* The Δ^0 has the same quark content as the neutron, but this decay is not possible for neutrons because they are not heavy enough to make a proton and a pion.

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



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 ($\Lambda \rightarrow p^+ + \pi^-$) or the omega-minus ($\Omega^- \rightarrow \Lambda + K^-$), 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:

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

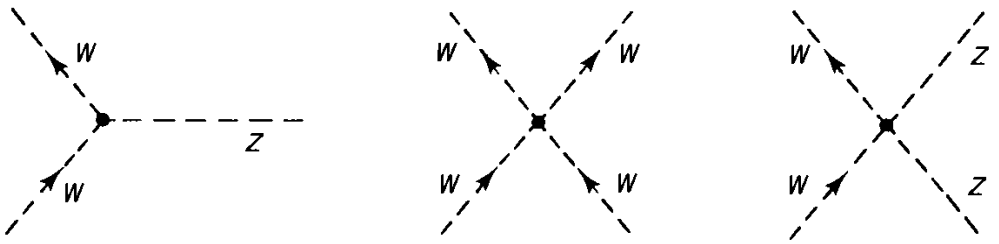
If this 3×3 *Kobayashi–Maskawa matrix* were the *unit* matrix, then d', s', and b' would be the same as d, s, and b, and no "cross-generational" transitions could occur. "Upness-plus-downness" would be absolutely conserved (just as the electron number is); "strangeness-plus-charm" would be conserved (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]

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

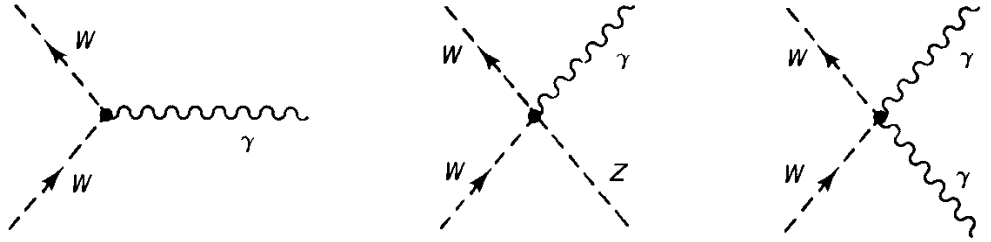
V_{ud} measures the coupling of u to d, V_{us} the coupling of u to s, and so on. The fact that the latter is nonzero is what permits strangeness-changing processes, such as the decay of the Λ and the Ω^-, to occur.

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

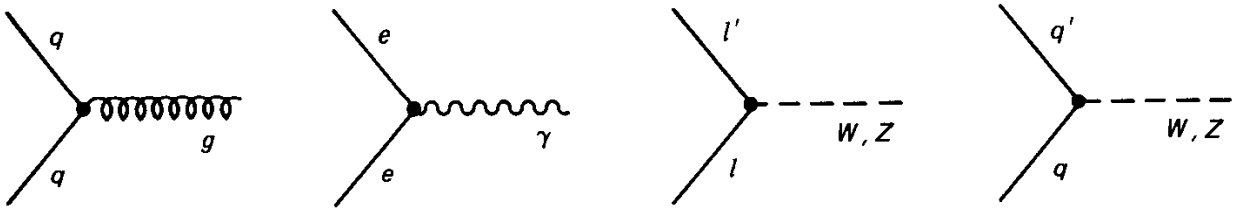
One of the most striking general properties of elementary particles is their tendency to disintegrate; we might almost call it a universal principle that *every particle decays into lighter particles, unless prevented from doing so by some conservation law.* The neutrinos and the photon are stable (having 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 $q + \bar{q} \rightarrow g$ 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:

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

for instance, but

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

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*

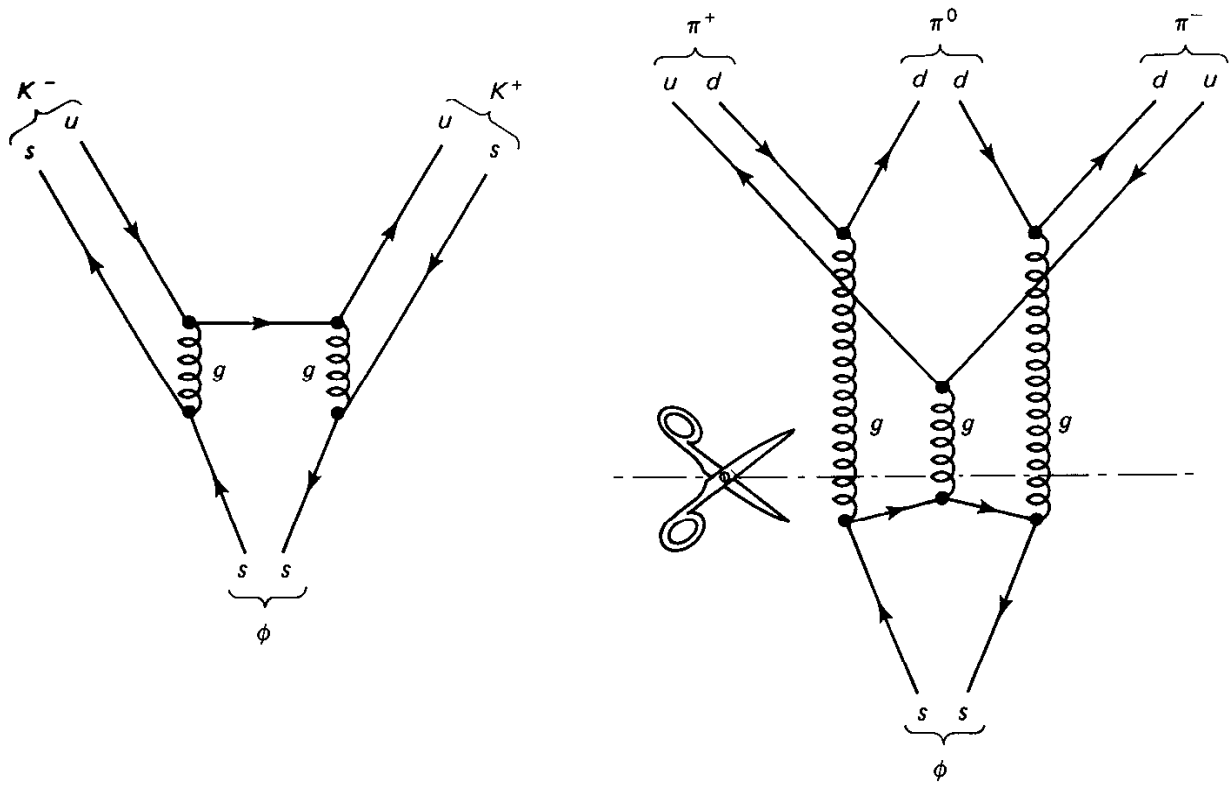


Figure 2.4 The OZI rule: If the diagram can be cut in two by slicing only gluon lines (and not cutting open any external particles), the process is suppressed.

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

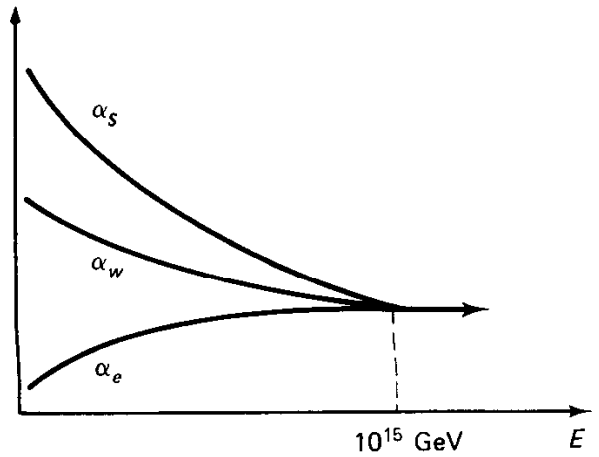


Figure 2.5 Evolution of the three fundamental coupling constants.

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,

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

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

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

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

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 $\psi' \to \psi + \pi^+ + \pi^-$. 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

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

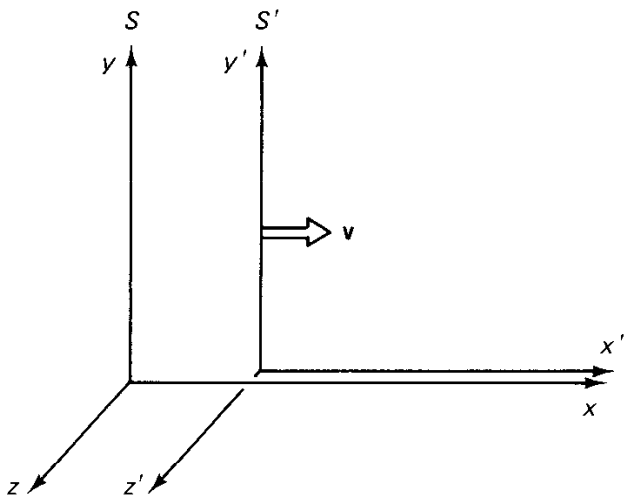


Figure 3.1 The inertial systems S and S'.

(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 $\boldsymbol{\Lambda}$:

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

(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 $\boldsymbol{\Lambda}$ 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 $\mathbf{v}$.]

* 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 **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

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

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, $\mathbf{v}$, 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

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

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:

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

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, η, which is the distance traveled (again, measured in the lab frame) divided by the *proper* time:*

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

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

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

However, η 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 $m\eta$, 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, $T_A + T_B = T_C + T_D$.

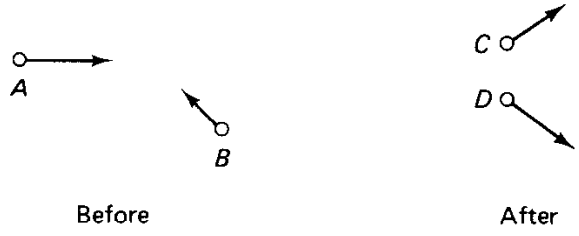


Figure 3.2 A collision in which $A + B \rightarrow C + D$.

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

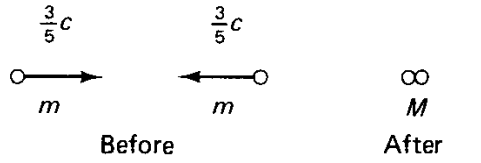


Figure 3.3 Sticky collision of two equal masses (Example 3.1).

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

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

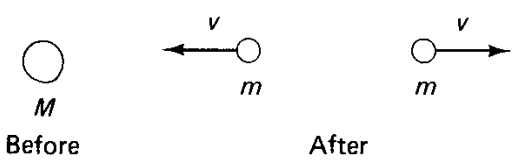


Figure 3.4 A particle decays into two equal pieces. (Example 3.2).

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 $M \rightarrow 2m$ 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:

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

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

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

or

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



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

Solving for $|\mathbf{p}_\mu|$, we find

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

Meanwhile, the energy of the muon [from eq. (3.49)] is

$$E_\mu = \frac{m_\pi^2 + m_\mu^2}{2m_\pi} c^2$$

Once we know the energy and momentum of a particle, it is easy to find its velocity. If $E = \gamma mc^2$ and $\mathbf{p} = \gamma m\mathbf{v}$, dividing gives

$$\mathbf{p}/E = \mathbf{v}/c^2$$

Suggestion 2. If you know the energy and momentum of a particle, and you want to determine its velocity, use

$$\mathbf{v} = \mathbf{p}c^2/E \tag{3.50}$$

So the answer to our problem is

$$v_\mu = \frac{m_\pi^2 - m_\mu^2}{m_\pi^2 + m_\mu^2} c$$

Putting in the actual masses, I get $v_\mu = 0.271c$.

There is nothing wrong with that calculation; it was a straightforward and systematic exploitation of the conservation laws. But I want to show you now a faster way to get the energy and momentum of the muon, by using four-vector notation. [I should put a superscript μ on all the four-vectors, but I don't want you to confuse the spacetime index μ with the particle identifier μ, so here, and often in the future, I will suppress the spacetime indices, and use a dot to indicate the scalar product.] Conservation of energy and momentum requires

$$p_\pi = p_\mu + p_\nu, \quad \text{or} \quad p_\nu = p_\pi - p_\mu$$

Taking the scalar product of each side with itself, we obtain

$$p_\nu^2 = p_\pi^2 + p_\mu^2 - 2p_\pi \cdot p_\mu$$

But

$$p_\nu^2 = 0; \qquad p_\pi^2 = m_\pi^2 c^2, \qquad p_\mu^2 = m_\mu^2 c^2; \quad \text{and} \quad p_\pi \cdot p_\mu = \frac{E_\pi}{c}\frac{E_\mu}{c} = m_\pi E_\mu$$

Therefore $$0 = m_\pi^2 c^2 + m_\mu^2 c^2 - 2m_\pi E_\mu$$

from which E_μ follows immediately. By the same token

$$p_\mu = p_\pi - p_\nu$$

Squaring yields

$$m_\mu^2 c^2 = m_\pi^2 c^2 - 2m_\pi E_\nu$$

But $E_\nu = |\mathbf{p}_\nu|c = |\mathbf{p}_\mu|c$, so

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

which gives us $|\mathbf{p}_\mu|$. In this case the problem was simple enough that the savings afforded by four-vector notation are meager, but in more complicated problems the benefits can be enormous.

Suggestion 3. Use four-vector notation, and exploit the invariant dot product.

One reason the use of invariants is so powerful in this business is that we are free to evaluate them in any inertial system we like. Frequently the laboratory frame is not the simplest one to work with. In a typical scattering experiment, for instance, a beam of particles is fired at a stationary target. The reaction under study might be, say, $p + p \rightarrow$ whatever, but in the laboratory the situation is asymmetrical, since one proton is moving and the other is at rest. Kinematically, the process is much simpler when viewed from a system in which the two protons approach one another with equal speeds. We call this the *center-of-momentum* (CM) frame, because in this system the total (three-vector) momentum is *zero.*

EXAMPLE 3.4

The Bevatron at Berkeley was built with the idea of producing antiprotons, by the reaction $p + p \rightarrow p + p + p + \bar{p}$. That is, a high-energy proton strikes a proton at rest, creating (in addition to the original particles) a proton-antiproton pair. Question: What is the threshold energy for this reaction (i.e., the minimum energy of the incident proton)?

Solution. In the laboratory the process looks like Figure 3.6a; in the CM frame, it looks like Figure 3.6b. Now, what is the condition for threshold? Answer: just barely enough incident energy to create the extra two particles. In the lab frame it is hard to see how we would formulate this condition, but in the CM it is easy: *All four final particles must be at rest,* with no energy "wasted" in the form of kinetic energy. (We can't have that in the

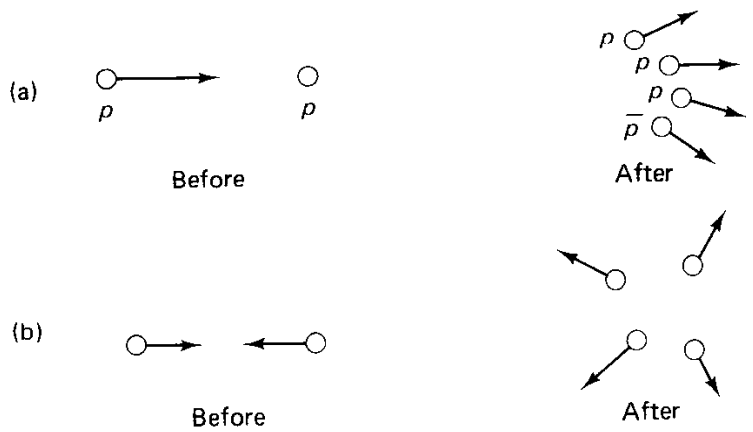


Figure 3.6 $p + p \rightarrow p + p + p + \bar{p}$. (a) In the lab frame; (b) in the CM frame.

lab frame, of course, since conservation of momentum requires that there be some residual motion.)

Let p^{μ}_{TOT} be the *total* energy-momentum four-vector in the lab; it is conserved, so it doesn't matter whether we evaluate it before or after the collision. We'll do it *before:*

$$p^{\mu}_{\text{TOT}} = \left(\frac{E + mc^2}{c}, |\mathbf{p}|, 0, 0\right)$$

where E and $\mathbf{p}$ are the energy and momentum of the incident proton, and m is the proton mass. Let p'^{μ}_{TOT} be the total energy-momentum four-vector in the CM. Again, we can evaluate it before or after the collision, this time we'll do it *after:*

$$p'^{\mu}_{\text{TOT}} = (4mc, 0, 0, 0)$$

since (at threshold) all four particles are at rest. Now $p^{\mu}_{\text{TOT}} \neq p'^{\mu}_{\text{TOT}}$, obviously, but the invariant products $p_{\mu\text{TOT}}p^{\mu}_{\text{TOT}}$ and $p'_{\mu\text{TOT}}p'^{\mu}_{\text{TOT}}$ *are* equal:

$$\left(\frac{E}{c} + mc\right)^2 - \mathbf{p}^2 = (4mc)^2$$

Using the standard relation (3.49) to eliminate $\mathbf{p}^2$, and solving for E, we find

$$E = 7mc^2$$

Evidently, the incident proton must carry a kinetic energy at least six times its rest energy, for this process to occur. (And in fact the first antiprotons were discovered when the machine reached about 6000 MeV.)

This is perhaps a good place to emphasize the distinction between a *conserved* quantity and an *invariant* quantity. Energy is *conserved*—the same value *after* the collision as *before*—but it is not invariant. Mass is *invariant*—the same in all inertial systems—but it is not conserved. Some quantities are both invariant and conserved; many are neither. As Example 3.4 indicates, the clever exploitation of conserved and invariant quantities can save you a lot of messy algebra. It also demonstrates that some problems are easier to analyse in the CM system, whereas others may be simpler in the lab frame.

> *Suggestion 4.* If a problem seems cumbersome in the lab frame, try analyzing it in the CM system.

Even if you're dealing with something more complicated than a collision of two identical particles, the center-of-momentum (in which $\mathbf{p}_{\text{TOT}} = 0$) is still a useful reference frame, for in this system conservation of momentum is trivial: zero before, zero after. But you might wonder whether there *is* always a CM frame. In other words, given a swarm of particles with masses $m_1, m_2, m_3, \ldots$, and velocities $\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3, \ldots$, does there necessarily exist an inertial system in

which the total (three-vector) momentum is zero? The answer is *yes;* I will prove it by finding the velocity of that frame and demonstrating that this velocity is less than c. The total energy and momentum in the lab frame (S) are

$$E_{\text{TOT}} = \sum_i \gamma_i m_i c^2; \qquad \mathbf{p}_{\text{TOT}} = \sum_i \gamma_i m_i \mathbf{v}_i \tag{3.51}$$

Since p^μ_{TOT} is a four-vector, we can use the Lorentz transformations to get the momentum in system S', moving in the direction of $\mathbf{p}_{\text{TOT}}$ with speed v

$$|\mathbf{p}'_{\text{TOT}}| = \gamma\left(|\mathbf{p}_{\text{TOT}}| - \beta \frac{E_{\text{TOT}}}{c}\right)$$

In particular, this momentum is *zero* if v is chosen such that

$$\frac{v}{c} = \frac{|\mathbf{p}_{\text{TOT}}|c}{E_{\text{TOT}}} = \frac{|\sum \gamma_i m_i \mathbf{v}_i|}{\sum \gamma_i m_i c}$$

Now, the length of the sum of three-vectors cannot exceed the sum of their lengths (this geometrically evident fact is known as the *triangle inequality*), so

$$\frac{v}{c} \leqslant \frac{\sum \gamma_i m_i (v_i/c)}{\sum \gamma_i m_i}$$

and since $v_i < c$, we can be sure that $v < c$.* Thus, the CM system always exists, and its velocity relative to the lab frame is given by

$$\mathbf{v}_{\text{CM}} = \frac{\mathbf{p}_{\text{TOT}} c^2}{E_{\text{TOT}}} \tag{3.52}$$

It seems odd, looking back at the answer to Example 3.4, that it takes an incident kinetic energy *six times* the proton rest energy to produce a p–$\bar{p}$ pair. After all, we're only creating $2mc^2$ of new rest energy. This example illustrates the inefficiency of scattering off a stationary target; conservation of momentum forces you to waste a lot of energy as kinetic energy in the final state. Suppose we could have fired the two protons at one another, making the laboratory itself the CM system. Then it would suffice to give each proton a kinetic energy of only mc^2, one-sixth of what the stationary-target experiment requires. This realization led, in the early 1970s, to the development of so-called *colliding-beam* machines (see Fig. 3.7). Today, virtually every new machine in high-energy physics is a collider.

EXAMPLE 3.5

Suppose two identical particles, each with mass m and kinetic energy T, collide head-on. Question: What is their *relative* kinetic energy, T' (i.e., the kinetic energy of one in the rest system of the other)?

* I am tacitly assuming that at least one of the particles is massive. If *all* of them are massless, we *may* obtain $v = c$, in which case there is no CM system. For example, there is no CM frame for a single photon.



(a) (b)

Figure 3.7 Two experimental arrangements: (a) Colliding beams; (b) fixed target.

Solution. There are many ways to do this one. A quick method is to write down the total four-momentum in the CM and in the lab

$$p^{\mu}_{\text{TOT}} = \left(\frac{2E}{c}, \mathbf{0}\right), \qquad p'^{\mu}_{\text{TOT}} = \left(\frac{E' + mc^2}{c}, \mathbf{p}'\right)$$

set $(p_{\text{TOT}})^2 = (p'_{\text{TOT}})^2$:

$$\left(\frac{2E}{c}\right)^2 = \left(\frac{E' + mc^2}{c}\right)^2 - \mathbf{p}'^2$$

use equation (3.49) to eliminate $\mathbf{p}'$

$$2E^2 = mc^2(E' + mc^2)$$

and express the answer in terms of $T = E - mc^2$ and $T' = E' - mc^2$

$$T' = 4T\left(1 + \frac{T}{2mc^2}\right) \tag{3.53}$$

The *classical* answer would have been $T' = 4T$, to which this reduces when $T \ll mc^2$. (In the rest system of B, A has, classically, twice the velocity, and hence four times as much kinetic energy as in the CM.) Now, a factor of 4 is *some* benefit, to be sure, but the *relativistic* gain can be greater by far. Colliding electrons with a laboratory kinetic energy of 1 GeV, for example, would have a *relative* kinetic energy of 4000 GeV!

REFERENCES AND NOTES

1. There are many excellent textbooks on Special Relativity. I recommend J. H. Smith, *Introduction to Special Relativity* (New York: Benjamin, 1967). For a fascinating (but unorthodox) approach, see E. F. Taylor and J. A. Wheeler, *Spacetime Physics* (San Francisco: Freeman, 1966).
2. If you want to go into this more deeply, I recommend R. Hagedorn, *Relativistic Kinematics* (New York: Benjamin, 1964).

PROBLEMS

3.1. Solve equation (3.1) for x, y, z, t in terms of x', y', z', t', and check that you recover equation (3.3).

3.2. **(a)** Derive equation (3.4).

(b) According to clocks on the ground (system S), streetlights A and B (situated 4 km apart) were both turned on at precisely 8:00 P.M. Which one went on first according to an observer on a train (system S'), which moves from A toward

B at $\frac{3}{5}$ the speed of light? How much later (in seconds) did the other light go on? [*Note:* As always in relativity, we are talking here about what S' *observed, after* correcting for the time it took the light to reach her, *not* what she actually *saw* (which would depend on where she was located on the train).]

3.3. **(a)** How do *volumes* transform? Specifically, if a container has volume V' in its own rest frame, S', what is its volume as measured by an observer in S, with respect to which it is moving at speed v?

(b) How do *densities* transform? (If a container holds ρ' molecules per unit volume in its own rest frame, S', how many molecules per unit volume does it carry in S?)

3.4. Cosmic ray muons are produced high in the atmosphere (at 8000 m, say) and travel toward the earth at very nearly the speed of light (0.998 c, say).

(a) Given the lifetime of the muon (2.2×10^{-6} sec), how far would it go before disintegrating, according to prerelativistic physics? Would the muons make it to ground level?

(b) Now answer the same question using *relativistic* physics. (Because of time dilation, the muons last longer, so they travel farther.)

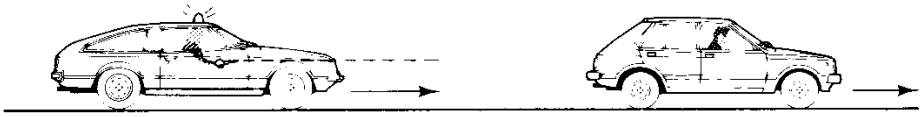
(c) Now analyze the same process from the perspective of the *muon.* (In *its* reference frame it only lasts 2.2×10^{-6} sec; how, then, does it make it to ground?)

(d) Pions are also produced in the upper atmosphere. [In fact, the sequence is proton (from outer space) hits proton (in atmosphere) $\longrightarrow p + p +$ pions. The pions then decay into muons: $\pi^- \longrightarrow \mu^- + \bar{\nu}_\mu$; $\pi^+ \longrightarrow \mu^+ + \nu_\mu$.] But the lifetime of the pion is much shorter, a hundredth that of the muon. Should the pions reach ground level? (Assume that the pions also have a speed of 0.998 c.)

3.5. As the outlaws escape in their getaway car, which goes $\frac{3}{4}c$, the cop fires a bullet from the pursuit car, which only goes $\frac{1}{2}c$. The muzzle velocity (speed relative to gun) of the bullet is $\frac{1}{3}c$. Does the bullet reach its target

(a) According to prerelativistic physics?

(b) According to relativity?



3.6. Find the matrix $\mathbf{M}$ that inverts equation (3.11): $x^\mu = M^\mu_\nu x'^\nu$ [use eq. (3.3)]. Show that $\mathbf{M}$ is the matrix inverse of $\mathbf{\Lambda}$: $\mathbf{\Lambda M} = 1$.

3.7. Show that the quantity I [in eq. (3.12)] is invariant under the Lorentz transformation (3.7).

3.8. A second-rank tensor is called *symmetric* if it is unchanged when you switch the indices ($s^{\nu\mu} = s^{\mu\nu}$); it is called *antisymmetric* if it changes sign ($a^{\nu\mu} = -a^{\mu\nu}$).

(a) How many independent elements are there in a symmetric tensor? (Since $s^{12} = s^{21}$, these would count as only *one* independent element.)

(b) How many independent elements are there in an antisymmetric tensor?

(c) If $s^{\mu\nu}$ is symmetric, show that $s_{\mu\nu}$ is also symmetric. If $a^{\mu\nu}$ is antisymmetric, show that $a_{\mu\nu}$ is antisymmetric.

(d) If $s^{\mu\nu}$ is symmetric and $a^{\mu\nu}$ is antisymmetric, show that $s^{\mu\nu}a_{\mu\nu} = 0$.

(e) Show that any second-rank tensor ($t^{\mu\nu}$) can be written as the sum of an antisymmetric part ($a^{\mu\nu}$) and a symmetric part ($s^{\mu\nu}$): $t^{\mu\nu} = a^{\mu\nu} + s^{\mu\nu}$. Construct $s^{\mu\nu}$ and $a^{\mu\nu}$ explicitly, given $t^{\mu\nu}$.

3.9. A particle is traveling at $\frac{3}{5}c$ in the x direction. Determine its proper velocity, η^μ (all four components).

3.10. Consider a collision in which particle A (with mass m_A and proper velocity η_A) hits particle B (mass m_B, proper velocity η_B), producing particle $C(m_C, \eta_C)$ and particle $D(m_D, \eta_D)$. Suppose that (relativistic) energy and momentum are conserved in system S (i.e., $p_A^\mu + p_B^\mu = p_C^\mu + p_D^\mu$). Using the Lorentz transformations (3.7), show that (relativistic) energy and momentum are *also* conserved in S'. (Do *not* assume that *mass* is conserved—in general, it is *not*: $m_A + m_B \neq m_C + m_D$.)

3.11. Is p^μ timelike, spacelike, or lightlike, for a (real) particle of mass m? How about a massless particle? How about a virtual particle?

3.12. How much more does a hot potato weigh than a cold one (in kg)?

3.13. A pion traveling at speed v decays into a muon and a neutrino, $\pi^- \to \mu^- + \bar{\nu}_\mu$. If the neutrino emerges at 90° to the original pion direction, at what angle does the μ come off? [*Answer:* $\tan\theta = (1 - m_\mu^2/m_\pi^2)/(2\beta\gamma^2)$]

3.14. Particle A (energy E) hits particle B (at rest), producing particles $C_1, C_2, \ldots$: $A + B \to C_1 + C_2 + \cdots + C_n$. Calculate the threshold (i.e., minimum E) for this reaction, in terms of the various particle masses.

$$\left[\textit{Answer:} \quad E = \frac{M^2 - m_A^2 - m_B^2}{2m_B}\,c^2, \qquad \text{where } M \equiv m_1 + m_2 + \cdots + m_n\right]$$

3.15. Use the result of Problem 3.14 to find the threshold energies for the following reactions, assuming the target proton is stationary:

(a) $p + p \to p + p + \pi^0$

(b) $p + p \to p + p + \pi^+ + \pi^-$

(c) $\pi^- + p \to p + \bar{p} + \mathrm{n}$

(d) $\pi^- + p \to K^0 + \Sigma^0$

(e) $p + p \to p + \Sigma^+ + K^0$

3.16. Particle A, at rest, decays into particles B and C ($A \to B + C$).

(a) Find the energy of the outgoing particles, in terms of the various masses.

$$\left[\textit{Answer:} \qquad E_B = \frac{m_A^2 + m_B^2 - m_C^2}{2m_A}\,c^2\right]$$

(b) Find the magnitudes of the outgoing momenta.

$$\left[\textit{Answer:} \qquad |\mathbf{p}_B| = |\mathbf{p}_C| = \frac{\sqrt{\lambda(m_A^2, m_B^2, m_C^2)}}{2m_A}\,c,\right.$$

where λ is the so-called *triangle function:*

$$\left.\lambda(x, y, z) \equiv x^2 + y^2 + z^2 - 2xy - 2xz - 2yz.\right]$$

(c) Note that λ factors: $\lambda(a^2, b^2, c^2) = (a + b + c)(a + b - c)(a - b + c)(a - b - c)$. Thus $|\mathbf{p}_B|$ goes to zero when $m_A = m_B + m_C$, and runs imaginary if $m_A < (m_B + m_C)$. Explain.

3.17. Use the result of Problem 3.16 to find the CM energy of each decay product in the following reactions:

(a) $\pi^- \to \mu^- + \bar{\nu}_\mu$

(b) $\pi^0 \to \gamma + \gamma$

(c) $K^+ \to \pi^+ + \pi^0$

(d) $\Lambda \to p + \pi^-$

(e) $\Omega^- \to \Lambda + K^-$

3.18. **(a)** A pion at rest decays into a muon and a neutrino ($\pi^- \to \mu^- + \bar{\nu}_\mu$). On the average, how far will the muon travel (in vacuum) before disintegrating? [*Answer:* $d = [(m_\pi^2 - m_\mu^2)/(2m_\pi m_\mu)]c\tau = 186$ m.]

(b) The length of the muon track in Figure 1.7 is about 0.6 mm (the photograph has been enlarged). How do you explain this?

3.19. Particle A, at rest, decays into three or more particles: $A \to B + C + D + \cdots$.

(a) Determine the maximum and minimum energies that B can have in such a decay, in terms of the various masses.

(b) Find the maximum and minimum electron energies in muon decay, $\mu^- \to e^- + \bar{\nu}_e + \nu_\mu$.

3.20. **(a)** A particle traveling at speed v approaches an identical particle at rest. What is the speed of each particle in the CM frame? (*Classically,* of course, it would just be $v/2$. Why *isn't* this true relativistically?)

(b) Use your result in part (a) to compute the kinetic energy of each particle in the CM frame, and thus rederive equation (3.53).

3.21. In reactions of the type $A + B \to A + C_1 + C_2 + \cdots$ (in which particle A scatters off particle B, producing C_1, C_2, ...), there is another inertial frame [besides the lab (B at rest) and the CM ($\mathbf{p}_{\text{TOT}} = 0$)] which is sometimes useful. It is called the *Breit,* or "brick wall," frame, and it is the system in which A recoils with its momentum reversed ($\mathbf{p}_{\text{after}} = -\mathbf{p}_{\text{before}}$), as though it had bounced off a brick wall. Take the case of elastic scattering ($A + B \to A + B$); if particle A carries energy E, and scatters at an angle θ, in the CM, what is its energy in the Breit frame? Find the velocity of the Breit frame (magnitude and direction) relative to the CM.

3.22. In a two-body scattering event, $A + B \to C + D$, it is convenient to introduce the *Mandelstam variables*

$$s \equiv (p_A + p_B)^2/c^2$$
$$t \equiv (p_A - p_C)^2/c^2$$
$$u = (p_A - p_D)^2/c^2$$

(a) Show that $s + t + u = m_A^2 + m_B^2 + m_C^2 + m_D^2$.

The *theoretical* virtue of the Mandelstam variables is that they are Lorentz invariants, with the same value in any inertial system. *Experimentally,* though, the more accessible parameters are energies and scattering angles.

(b) Find the CM energy of A, in terms of s, t, u and the masses. [*Answer:* $E_A^{\text{CM}} = (s + m_A^2 - m_B^2)c^2/2\sqrt{s}$]

(c) Find the Lab (B at rest) energy of A. [*Answer:* $E_A^{\text{lab}} = (s - m_A^2 - m_B^2)c^2/2m_B$]

(d) Find the total CM energy ($E_{\text{TOT}} = E_A + E_B = E_C + E_D$). [*Answer:* $E_{\text{TOT}}^{\text{CM}} = \sqrt{s}c^2$]

3.23. For elastic scattering of identical particles, $A + A \to A + A$, show that the Mandelstam variables (Problem 3.22) become

$$s = 4(\mathbf{p}^2 + m^2c^2)/c^2$$
$$t = -2\mathbf{p}^2(1 - \cos\theta)/c^2$$
$$u = -2\mathbf{p}^2(1 + \cos\theta)/c^2$$

where $\mathbf{p}$ is the CM momentum of the incident particle, and θ is the scattering angle.

3.24. (Compton scattering.) A photon of wavelength λ collides elastically with a charged particle of mass m. If the photon scatters at angle θ, find its outgoing wavelength, λ'. [*Answer:* $\lambda' = \lambda + (h/mc)(1 - \cos\theta)$]

Chapter 4

Symmetries

This chapter is a grab bag of special topics having to do with symmetry. The first section contains some general remarks about the mathematical description of symmetry ("group theory") and the relation between symmetry and conservation laws (Noether's theorem). We then take up the case of rotational symmetry and its relation to angular momentum and spin. This leads in turn to the "internal" symmetries—isospin, SU(3), and flavor SU(6). Finally, we consider "discrete" symmetries—parity, charge conjugation, and time reversal. Except for the theory of spin (Sections 4.2, 4.3, and 4.4)—which will be used extensively in subsequent chapters—and the material on parity (Section 4.6)—which is useful background for Chapter 10—this chapter can be studied as superficially (or as deeply) as the reader desires. I recommend a quick pass at this stage and a return to specific sections later, as warranted. Some knowledge of matrix theory is presupposed here; readers familiar with quantum mechanics will find the sections on angular momentum an easy review (those unacquainted with quantum mechanics may find them obscure, in which case they should study the relevant chapter of an introductory quantum text). Group theory is touched on here in a scandalously cursory fashion (my main purpose is to introduce some standard terminology), but a serious student of elementary particle physics should plan eventually to study this subject in far greater detail.

4.1 SYMMETRIES, GROUPS, AND CONSERVATION LAWS

Examine the graph in Figure 4.1. I have no idea what the functional form of $f(x)$ might be, but this much I *can* say: it's an *odd* function, $f(-x) = -f(x)$.

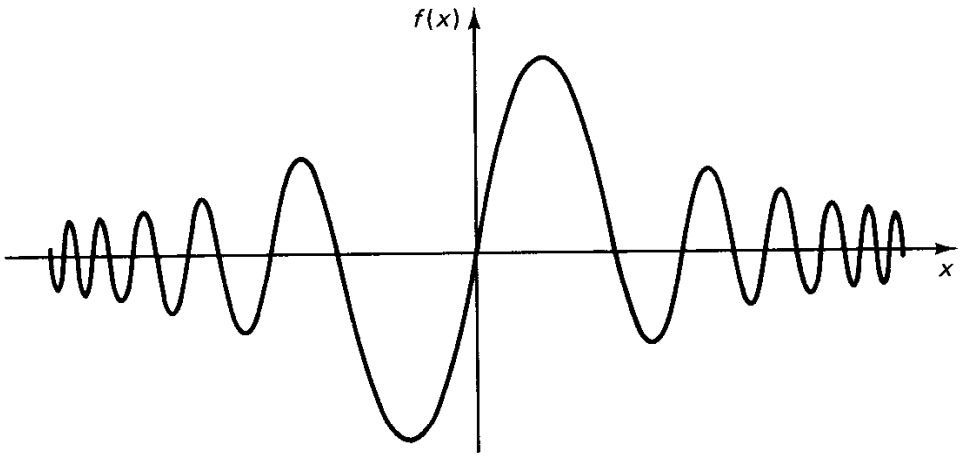


Figure 4.1 An odd function.

(If you don't believe me, trace the curve, rotate the tracing by 180°, and check that it perfectly fits the original graph.) It follows, for instance, that

$$[f(-x)]^6 = [f(x)]^6, \qquad \int_{-3}^{+3} f(x)dx = 0,$$

$$\left.\frac{df}{dx}\right|_{+2} = \left.\frac{df}{dx}\right|_{-2}, \qquad \int_{-7}^{+7} [f(x)]^2\,dx = 2\int_{0}^{+7} [f(x)]^2\,dx \tag{4.1}$$

I know that no cosines appear in the Fourier expansion of $f(x)$, and that its Taylor series contains only odd powers. In fact, I know quite a lot about $f(x)$, even though I don't know its functional form, just from my observation that it has a particular symmetry—oddness, in this case. In physics, intuition or a general principle often suggests symmetries in a problem, and their systematic exploitation can be an extremely powerful tool. [In some respects the appeal to symmetry is characteristic of an incomplete theory. For example, if we somehow discovered the explicit form of $f(x)$, say, $f(x) = e^{-x^2}\sin(x^3)$, then the theorems in equation (4.1) would lose their luster. Why bother with *partial* information when we can have it *all?* But even in a mature theory, symmetry considerations often lead to deeper understanding and calculational simplification; for instance, if you're integrating $f(x)$ from -3 to $+3$, it pays to notice that $f(x)$ is odd, even if you *do* know its functional form!]

The most obvious examples of symmetry in physics are, I suppose, crystals. But we're not so much interested here in static symmetries of *shape* as in dynamical symmetries of *motion.* The Greeks apparently believed that the symmetries of nature should be directly reflected in the motion of objects: Stars must move in circles because those are the most symmetrical trajectories. Of course, planets don't, and that was embarrassing (it was not the last time that naïve intuitions about symmetry ran into trouble with experiment). Newton recognized that fundamental symmetries are revealed not in the motions of individual objects, but in the *set of all possible* motions—symmetries are manifest in the *equations of motion* rather than in particular solutions to those equations. Newton's law of universal gravitation, for instance, exhibits spherical symmetry—the force is the same in all directions—yet planetary orbits are elliptical. Thus the underlying symmetry of the system is only indirectly revealed to us; indeed,

TABLE 4.1 SOME SYMMETRIES AND THE ASSOCIATED CONSERVATION LAWS

Symmetry		Conservation law
Translation in time	↔	Energy
Translation in space	↔	Momentum
Rotation	↔	Angular momentum
Gauge transformation	↔	Charge

you might wonder how we would ever have discovered it from the observed planetary trajectories if we didn't have a pretty strong hunch that the gravitational field of the sun "ought" to be spherically symmetrical.

It was not until 1917 that the dynamical implications of symmetry were completely understood. In that year Emmy Noether published her famous theorem relating symmetries and conservation laws:

NOETHER'S THEOREM: SYMMETRIES ↔ CONSERVATION LAWS

Every symmetry of nature yields a conservation law; conversely, every conservation law reveals an underlying symmetry. For example, the laws of physics are symmetrical with respect to *translations in time;* they work the same today as they did yesterday. Noether's theorem relates this invariance to conservation of energy. If a system is invariant under translations in space, then momentum is conserved; if it is symmetrical under rotations about a point, then angular momentum is conserved. Similarly, the invariance of electrodynamics under gauge transformations leads to conservation of charge (we call this an *internal* symmetry, in contrast to the space–time symmetries). I'm not going to *prove* Noether's theorem; the details are not terribly enlightening.[1] The important thing is the profound and beautiful *idea* that symmetries are associated with conservation laws (see Table 4.1).

I have been speaking rather casually about symmetries and I cited some examples, but what precisely *is* a symmetry? It is an operation you can perform (at least conceptually) on a system that leaves it *invariant*—that carries it into a configuration indistinguishable from the original one. In the case of the function in Figure 4.1, changing the sign of the argument, $x \rightarrow -x$, and multiplying the whole thing by -1 [$f(x) \rightarrow -f(-x)$] is a symmetry operation. For a meatier example, consider the equilateral triangle (Fig. 4.2). It is carried into itself by a clockwise rotation through 120° (R_+), and by a counterclockwise rotation through

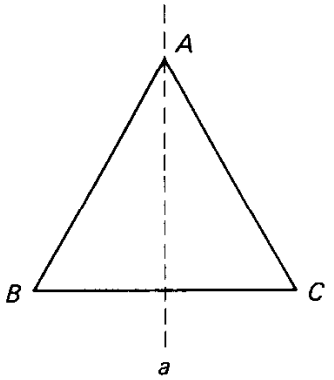


Figure 4.2 Symmetries of the equilateral triangle.

120° (R_-), by flipping it about the axis Aa (R_a), or around the corresponding axis through B (R_b), or through C (R_c). Is that all? Well, doing *nothing at all* (I) obviously leaves it invariant, so this too is a symmetry operation, albeit a pretty trivial one. And then we could combine operations, for example, rotate clockwise through 240°. But that's the same as rotating *counter*clockwise by 120° (i.e., $R_+^2 = R_-$). As it turns out, we have already identified all the distinct symmetry operations on the equilateral triangle (see Problem 4.1).

The set of symmetry operations on any system must have the following properties:

1. *Closure.* If R_i and R_j are in the set, then the *product*, R_iR_j—meaning: first perform R_j, then perform R_i—is also in the set; that is, there exists some R_k such that $R_iR_j = R_k$.
2. *Identity.* There is an element I such that $IR_i = R_iI = R_i$ for all elements R_i.
3. *Inverse.* For every element R_i there is an *inverse*, R_i^{-1}, such that $R_iR_i^{-1} = R_i^{-1}R_i = I$.
4. *Associativity.* $R_i(R_jR_k) = (R_iR_j)R_k$.

These are precisely the defining properties of a *group.* Indeed, the mathematical theory of groups may be regarded as the systematic study of symmetries. Notice that group elements need not *commute:* $R_iR_j \neq R_jR_i$, in general; if all the elements *do* commute, the group is called *Abelian.* Translations in space and time form an Abelian group; rotations do *not.*[2] Groups can be *finite* (like the triangle group, which has just six elements) or *infinite* (for example, the set of integers, with addition playing the role of group "multiplication"). We shall encounter *continuous* groups (such as the group of all rotations in a plane) in which the elements depend on one or more continuous parameters (the angle of rotation, in this case) and *discrete* groups, in which the element may be labeled by an index that takes on only integer values (all finite groups are, of course, discrete).

As it turns out, most of the groups of interest in physics are groups of *matrices.* The Lorentz group, for instance, consists of the set of 4×4 Λ matrices introduced in Chapter 3. In elementary particle physics the most common groups are of the type mathematicians call $U(n)$: the collection of all unitary $n \times n$ matrices (see Table 4.2). (A *unitary matrix* is one whose inverse is equal to its transpose conjugate: $U^{-1} = \tilde{U}^*$.) If we restrict further to unitary matrices with determinant 1, the group is called $SU(n)$. (The S stands for "special," which just

TABLE 4.2 THE MOST IMPORTANT GROUPS IN ELEMENTARY PARTICLE PHYSICS

Group name	Matrices in group	
$U(n)$	$n \times n$	unitary ($\tilde{U}^*U = 1$)
$SU(n)$	$n \times n$	unitary with determinant 1
$O(n)$	$n \times n$	orthogonal ($\tilde{O}O = 1$)
$SO(n)$	$n \times n$	orthogonal with determinant 1

means "determinant 1".) If we restrict ourselves to *real* unitary matrices, the group is $O(n)$. (O stands for "orthogonal"; an orthogonal matrix is one whose inverse is equal to its transpose: $O^{-1} = \tilde{O}$.) Finally, the group of real, orthogonal, $n \times n$ matrices of determinant 1 is $SO(n)$. $SO(n)$ may be thought of as the group of all *rotations* in a space of n dimensions. Thus $SO(3)$ describes the rotational symmetry of our world, a symmetry that is related by Noether's theorem to the conservation of angular momentum. Indeed, the entire quantum theory of angular momentum is really closet group theory. It so happens that $SO(3)$ is almost identical in mathematical structure to $SU(2)$, which is the most important *internal* symmetry in elementary particle physics. So the theory of angular momentum, to which we turn next, will actually serve us twice.

One final thing. Every group G can be *represented* by a group of *matrices:* For every group element a there is a corresponding matrix M_a, and the correspondence respects group multiplication, in the sense that if $ab = c$, then $M_a M_b = M_c$. A representation need not be "faithful": there may be many distinct group elements represented by the same matrix. (Mathematically, the group of matrices is *homo*morphic, but not necessarily *iso*morphic, to G.) Indeed, there is a trivial case, in which we represent every element by the 1×1 *unit* matrix (which is to say, the number 1). If G is a group of matrices, such as $SU(6)$ or $O(18)$, then it is a (faithful) representation of *itself*—we call it the *fundamental* representation. But there will in general be many other representations, by matrices of various dimensions. For example, $SU(2)$ has representations of dimension 1 (the trivial one), 2 (the fundamental one), 3, 4, 5, and in fact *every* positive integer. A major problem in group theory is the enumeration of all the representations of a given group. Of course, you can always construct a new representation by combining two old ones, thus

$$M_a = \begin{pmatrix} \boxed{M_a^{(1)}} & \text{(zeros)} \\ \text{(zeros)} & \boxed{M_a^{(2)}} \end{pmatrix}$$

But we don't count this separately; when we list the representations of a group, we are talking about the so-called *irreducible* representations, which *cannot* be decomposed into block-diagonal form. Actually, you have already encountered several examples of group representations, probably without realizing it: An ordinary scalar belongs to the one-dimensional representation of the rotation group, $SO(3)$, and a vector belongs to the three-dimensional representation; four-vectors belong to the four-dimensional representation of the Lorentz group; and the curious geometrical arrangements of Gell-Mann's Eightfold Way correspond to irreducible representations of the group $SU(3)$.

4.2 SPIN AND ORBITAL ANGULAR MOMENTUM

The earth, in its motion, carries two kinds of angular momentum: *orbital* angular momentum, $\boldsymbol{r} \boldsymbol{m} \boldsymbol{v}$, associated with its annual revolution around the sun, and

spin angular momentum, $I\omega$, associated with its daily rotation about the north–south axis. The same goes for the electron in a hydrogen atom: It too carries both orbital and spin angular momentum. In the macroscopic case the distinction is not terribly profound; after all, the spin angular momentum of the earth is nothing but the sum total of the "orbital" angular momenta of all the rocks and dirt clods that make it up, in their daily "orbit" around the axis. In the case of the electron this interpretation is not open to us: The electron, as far as we know, is a true point particle; its spin angular momentum is not attributable to constituent parts revolving about an axis, but is simply an intrinsic property of the particle itself (see Problem 4.8).

Classically, we are free to measure all three components of the orbital angular momentum vector, $\mathbf{L} = \mathbf{r} \times m\mathbf{v}$, to any desired accuracy, and these components can assume any values whatever. In quantum-mechanics, however, it is *impossible in principle* to measure all three components simultaneously; a measurement of L_x, say, inevitably alters the value of L_y, by an unpredictable amount. The best we can do is to measure the *magnitude* of $\mathbf{L}$, (or rather, its *square:* $L^2 = \mathbf{L} \cdot \mathbf{L}$) together with one component (which we customarily take to be the z component, L_z). Furthermore, these measurements can only return certain "allowed" values.* Specifically, a measurement of L^2 always yields a number of the form

$$l(l + 1)\hbar^2 \tag{4.2}$$

where l is a nonnegative integer:

$$l = 0, 1, 2, 3, \ldots \tag{4.3}$$

For a given value of l, a measurement of L_z always gives a result of the form

$$m_l\hbar \tag{4.4}$$

where m_l is an integer in the range $[-l, +l]$:

$$m_l = -l, -l + 1, \ldots, -1, 0, +1, \ldots, l - 1, l \tag{4.5}$$

[$(2l + 1)$ possibilities in all]. Figure 4.3 may help you to visualize the situation. Here $l = 2$, so the magnitude of $\mathbf{L}$ is $\sqrt{6}\hbar = 2.45\hbar$; L_z can assume the values $2\hbar$, $\hbar$, 0, $-\hbar$, or $-2\hbar$. Notice that the angular momentum vector cannot be oriented purely in the z direction.

The same goes for spin angular momentum: A measurement of $S^2 = \mathbf{S} \cdot \mathbf{S}$ can only return values of the form

$$s(s + 1)\hbar^2 \tag{4.6}$$

In the case of spin, however, the quantum number s can be a *half*-integer as well as an integer:

$$s = 0, \tfrac{1}{2}, 1, \tfrac{3}{2}, 2, \tfrac{5}{2}, \ldots \tag{4.7}$$

* I am not going to *prove* the quantization rules for angular momentum, and if this material is new to you, I suggest that you consult a textbook on quantum mechanics. All I propose to do here is summarize the essential results we will need in what follows.

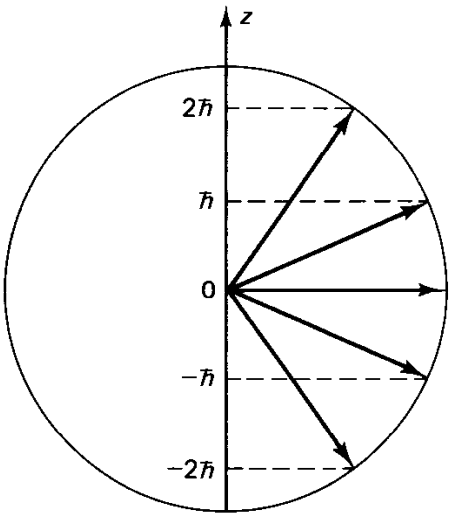


Figure 4.3 Possible orientations of the angular momentum vector for $l = 2$.

For a given value of s, a measurement of S_z must yield an answer of the form

$$m_s\hbar \tag{4.8}$$

where m_s is an integer or half-integer (whichever s is) in the range $[-s, s]$:

$$m_s = -s, -s + 1, \ldots, s - 1, s \tag{4.9}$$

[$(2s + 1)$ possibilities].

Now, a given particle can be put into any *orbital* angular momentum state you wish, but for each type of particle the value of s is fixed. Every pion or kaon, for example, has $s = 0$; every electron, proton, neutron, and quark carries $s = \frac{1}{2}$; for the ρ, the ψ, the photon, and the gluon, $s = 1$; for the Δ's and the Ω^-, $s = \frac{3}{2}$; and so on. We call s the "spin" of the particle. Particles with half-integer spin are known as *fermions*—all baryons, leptons, and quarks are fermions; particles with integer spin are known as *bosons*—all mesons and mediators are bosons (see Table 4.3).

4.3 ADDITION OF ANGULAR MOMENTA

Angular momentum states are labeled with a "ket": $|lm_l\rangle$ or $|sm_s\rangle$. Thus if I say the electron in a hydrogen atom occupies the orbital state $|3\ -1\rangle$ and the spin state $|\frac{1}{2}\ \frac{1}{2}\rangle$, I mean that $l = 3$, $m_l = -1$, $s = \frac{1}{2}$ (which is unnecessary, of course; if it's an electron, s *must* be $\frac{1}{2}$), and $m_s = \frac{1}{2}$. Now, it may happen that we are not interested in the spin and orbital angular momenta *separately,* but rather in

TABLE 4.3 CLASSIFICATION OF PARTICLES BY SPIN

Bosons (integer spin)		Fermions ($\frac{1}{2}$-integer spin)		
Spin 0	Spin 1	Spin $\frac{1}{2}$	Spin $\frac{3}{2}$	
—	Mediators	Quarks, leptons	—	←Elementary
Pseudo-scalar mesons	Vector mesons	Baryon octet	Baryon decuplet	←Composite

the *total* angular momentum, $\mathbf{J} = \mathbf{L} + \mathbf{S}$. (In the presence of coupling between $\mathbf{L}$ and $\mathbf{S}$—tidal, if it's the earth–sun system; magnetic, for the electron–proton system—it is $\mathbf{J}$, and not $\mathbf{L}$ and $\mathbf{S}$ individually, that will be conserved.) Or perhaps we are studying the two quarks that go to make a ψ meson; in this case. as we shall see, the *orbital* angular momentum is zero, but we are confronted with the problem of combining the two quark spins to get the total spin of the ψ: $\mathbf{S} = \mathbf{S}_1 + \mathbf{S}_2$. In either case the question arises: *How do we add two angular momenta*

$$\mathbf{J} = \mathbf{J}_1 + \mathbf{J}_2 \tag{4.10}$$

Classically, of course, we just add the components. But in quantum mechanics we do not have access to all three components; we are obliged to work with *one* component and the *magnitude.* So the question becomes: If we combine states $|j_1 m_1\rangle$ and $|j_2 m_2\rangle$, what *total* angular momentum state(s) $|jm\rangle$ do we get? The z components still add, naturally, so

$$m = m_1 + m_2 \tag{4.11}$$

but the magnitudes do *not;* it all depends on the relative orientation of $\mathbf{J}_1$ and $\mathbf{J}_2$ (Fig. 4.4). If they are parallel the magnitudes add, but if they are antiparallel the magnitudes subtract; in general, the magnitude of the vector sum is somewhere between these extremes. As it turns out, we get *every j from* $(j_1 + j_2)$ down to $|j_1 - j_2|$, in integer steps:[3]

$$j = |j_1 - j_2|, |j_1 - j_2| + 1, \ldots, (j_1 + j_2) - 1, (j_1 + j_2) \tag{4.12}$$

For instance, a particle of spin 1 in an orbital state $l = 3$ could have total angular momentum $j = 4$ (i.e., $J^2 = 20\hbar^2$), or $j = 3$ ($J^2 = 12\hbar^2$), or $j = 2$ ($J^2 = 6\hbar^2$).

EXAMPLE 4.1

A quark and an antiquark are bound together, in a state of zero orbital angular momentum, to form a meson. What are the possible values of the meson's spin?

Solution. Quarks carry spin $\frac{1}{2}$, so we can get $\frac{1}{2} + \frac{1}{2} = 1$ or $\frac{1}{2} - \frac{1}{2} = 0$. The spin-0 combination gives us the "pseudo-scalar" mesons, π's, K's, η, η'; "scalar" means spin 0; "pseudo-" will be explained shortly. The spin-1 combination gives the "vector" mesons, ρ's, K^*'s, ϕ, ω; "vector" means spin 1.

To add *three* angular momenta, we combine two of them first, using equation (4.12), and then add on the third. Thus if we allow the quarks in Example 4.1 an orbital angular momentum $l > 0$, we get mesons with spin $l + 1$, l, and

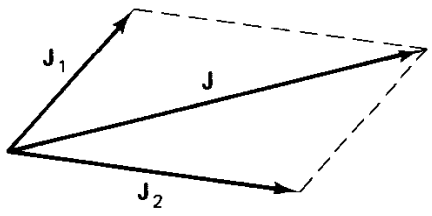


Figure 4.4 Addition of angular momenta.

$l - 1$. Because the orbital quantum number has to be an integer, all mesons carry integer spin (they're bosons). By the same token, all baryons (made up of *three* quarks) must have half-integer spin (they're fermions).

EXAMPLE 4.2

Suppose you combine three quarks in a state of zero orbital angular momentum. What are the possible spins of the resulting baryon?

Solution. From *two* quarks, each spin $\frac{1}{2}$, we get a total angular momentum of $\frac{1}{2} + \frac{1}{2} = 1$ or $\frac{1}{2} - \frac{1}{2} = 0$. Adding in the third quark yields $1 + \frac{1}{2} = \frac{3}{2}$ or $1 - \frac{1}{2} = \frac{1}{2}$ (when the first two add to 1), and $0 + \frac{1}{2} = \frac{1}{2}$ (when the first two add to zero). Thus the baryon can have a spin of $\frac{3}{2}$ or $\frac{1}{2}$, and the latter can be achieved in two different ways. In practice, $s = \frac{3}{2}$ is the decuplet, $s = \frac{1}{2}$ is the octet, and evidently the quark model would allow for another family with $s = \frac{1}{2}$. If we permit the quarks to revolve around one another, throwing in some *orbital* angular momentum, the number of possibilities increases accordingly.

Well, equation (4.12) tells us what total angular momenta j we can obtain by combining j_1 and j_2, but occasionally we require the explicit decomposition of $|j_1 m_1\rangle|j_2 m_2\rangle$ into states of total angular momentum $|jm\rangle$:

$$|j_1 m_1\rangle|j_2 m_2\rangle = \sum_{j=|j_1-j_2|}^{(j_1+j_2)} C^{j\ j_1\ j_2}_{m\,m_1\,m_2}|jm\rangle, \qquad \text{with } m = m_1 + m_2 \tag{4.13}$$

The numbers $C^{j\ j_1\ j_2}_{m\,m_1\,m_2}$ are known as *Clebsch–Gordan coefficients.* A book on group theory or advanced quantum-mechanics will explain how to calculate them. In practice, we normally look them up in a table. (There is one in the Particle Data Booklet, and the case $j_1 = 2$, $j_2 = \frac{1}{2}$ is reproduced in Figure 4.5). The Clebsch–Gordan coefficients give the probability of getting $j(j+1)\hbar^2$, if we measure J^2 (the total angular momentum squared) on a system consisting of two angular momentum states $|j_1 m_1\rangle$ and $|j_2 m_2\rangle$: The probability is the *square* of the corresponding Clebsch–Gordan coefficient.

EXAMPLE 4.3

The electron in a hydrogen atom occupies the orbital state $|2\ -1\rangle$ and the spin state $|\frac{1}{2}\ \frac{1}{2}\rangle$. If we measure J^2, what values might we get, and what is the probability of each?

Solution. The possible values of j are $l + s = 2 + \frac{1}{2} = \frac{5}{2}$ and $l - s = 2 - \frac{1}{2} = \frac{3}{2}$. The z components add: $m = -1 + \frac{1}{2} = -\frac{1}{2}$. We go to the Clebsch–Gordan table (Fig. 4.5) labeled $2 \times \frac{1}{2}$, which indicates that we are combining $j_1 = 2$ with $j_2 = \frac{1}{2}$, and look for the horizontal row, labeled -1, $\frac{1}{2}$; these are the values of m_1 and m_2. Reading off the two entries, we find $|2\ -1\rangle|\frac{1}{2}\ \frac{1}{2}\rangle = \sqrt{2/5}|5/2\ -\frac{1}{2}\rangle - \sqrt{3/5}|3/2\ -\frac{1}{2}\rangle$. So the probability of getting $j = \frac{5}{2}$ is $\frac{2}{5}$, and the probability of getting $j = \frac{3}{2}$ is $\frac{3}{5}$. Notice that the probabilities add to 1, as, of course, they *should.*

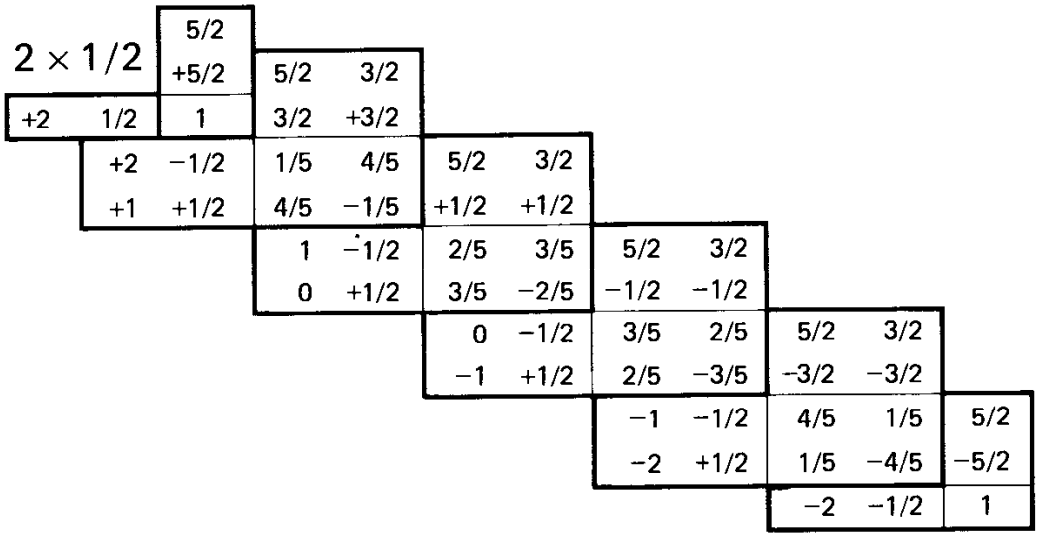


Figure 4.5 Clebsch–Gordan coefficients for $j_1 = 2, j_2 = \frac{1}{2}$. (A square root sign over each number is implied.)

EXAMPLE 4.4

We know from Example 4.1 that two spin-$\frac{1}{2}$ states combine to give spin 1 and spin 0. Find the explicit Clebsch–Gordan decomposition for these states.

Solution. Consulting the $\frac{1}{2} \times \frac{1}{2}$ table, we find

$$\begin{aligned} |\tfrac{1}{2}\tfrac{1}{2}\rangle|\tfrac{1}{2}\tfrac{1}{2}\rangle &= |11\rangle \\ |\tfrac{1}{2}\tfrac{1}{2}\rangle|\tfrac{1}{2}\,-\tfrac{1}{2}\rangle &= (1/\sqrt{2})|10\rangle + (1/\sqrt{2})|00\rangle \\ |\tfrac{1}{2}\,-\tfrac{1}{2}\rangle|\tfrac{1}{2}\tfrac{1}{2}\rangle &= (1/\sqrt{2})|10\rangle - (1/\sqrt{2})|00\rangle \\ |\tfrac{1}{2}\,-\tfrac{1}{2}\rangle|\tfrac{1}{2}\,-\tfrac{1}{2}\rangle &= |1\,-1\rangle \end{aligned} \tag{4.14}$$

Thus the three spin 1 states are

$$\left.\begin{aligned} |11\rangle &= |\tfrac{1}{2}\tfrac{1}{2}\rangle|\tfrac{1}{2}\tfrac{1}{2}\rangle \\ |10\rangle &= (1/\sqrt{2})[|\tfrac{1}{2}\tfrac{1}{2}\rangle|\tfrac{1}{2}\,-\tfrac{1}{2}\rangle + |\tfrac{1}{2}\,-\tfrac{1}{2}\rangle|\tfrac{1}{2}\tfrac{1}{2}\rangle] \\ |1\,-1\rangle &= |\tfrac{1}{2}\,-\tfrac{1}{2}\rangle|\tfrac{1}{2}\,-\tfrac{1}{2}\rangle \end{aligned}\right\} \tag{4.15}$$

whereas the spin 0 state is

$$|00\rangle = (1/\sqrt{2})[|\tfrac{1}{2}\tfrac{1}{2}\rangle|\tfrac{1}{2}\,-\tfrac{1}{2}\rangle - |\tfrac{1}{2}\,-\tfrac{1}{2}\rangle|\tfrac{1}{2}\tfrac{1}{2}\rangle] \tag{4.16}$$

[By the way, equations (4.15) and (4.16) can be read directly off the Clebsch–Gordan table; the coefficients work both directions:

$$|jm\rangle = \sum_{j_1, j_2} C^{j\ j_1\ j_2}_{m\,m_1\,m_2} |j_1 m_1\rangle |j_2 m_2\rangle \tag{4.17}$$

This time we read down the columns, instead of along the rows.] The spin-1 combination is called the "triplet," for obvious reasons, and spin 0 is called the "singlet." For future reference, notice that the triplet is *symmetric* under interchange of the particles, $1 \leftrightarrow 2$, whereas the singlet is antisymmetric (that is, it changes sign). Incidentally, in a singlet state the spins are oppositely aligned (antiparallel); however, it is *not* the case that

in a triplet state the spins are necessarily parallel; they are for $m = 1$ and $m = -1$, but not for $m = 0$.

4.4 SPIN $\frac{1}{2}$

The most important spin system is $s = \frac{1}{2}$; the proton, neutron, electron, all quarks, and all leptons carry spin $\frac{1}{2}$. Furthermore, once you understand the formalism for $s = \frac{1}{2}$, any other case is a relatively simple matter to work out. So I'll pause here to develop the theory of spin $\frac{1}{2}$ in some detail. A particle with spin $\frac{1}{2}$ can have $m_s = \frac{1}{2}$ ("spin up") or $m_s = -\frac{1}{2}$ ("spin down"). Informally, we represent these two states by arrows: $\uparrow$ and $\downarrow$. But a better notation is afforded by two-component column vectors, or *spinors:*

$$|\tfrac{1}{2}\,\tfrac{1}{2}\rangle = \begin{pmatrix}1\\0\end{pmatrix}, \qquad |\tfrac{1}{2}\,-\tfrac{1}{2}\rangle = \begin{pmatrix}0\\1\end{pmatrix} \tag{4.18}$$

It is often said that a particle of spin $\frac{1}{2}$ can only exist in one or the other of these two states, but that is quite false. The most general state of a spin-$\frac{1}{2}$ particle is the *linear combination*

$$\begin{pmatrix}\alpha\\\beta\end{pmatrix} = \alpha\begin{pmatrix}1\\0\end{pmatrix} + \beta\begin{pmatrix}0\\1\end{pmatrix} \tag{4.19}$$

where α and β are two complex numbers. It is true that a *measurement* of S_z can only return the value $+\frac{1}{2}\hbar$ or $-\frac{1}{2}\hbar$, but the first outcome, say, does not prove that the particle was in the state $\begin{pmatrix}1\\0\end{pmatrix}$ prior to the measurement. In the general case $\begin{pmatrix}\alpha\\\beta\end{pmatrix}$, $|\alpha|^2$ is the *probability* that a measurement of S_z would yield the value $+\frac{1}{2}\hbar$, and $|\beta|^2$ is the probability of getting $-\frac{1}{2}\hbar$. Since these are the only allowed results, it follows that

$$|\alpha|^2 + |\beta|^2 = 1 \tag{4.20}$$

Apart from this "normalization" condition, there is no a priori constraint on the numbers α and β.

Suppose now that we are to measure S_x or S_y on a particle in the state $\begin{pmatrix}\alpha\\\beta\end{pmatrix}$. What results might we get, and what is the probability of each? Symmetry dictates that the allowed values be $\pm\frac{1}{2}\hbar$; after all, it's perfectly arbitrary which direction we choose to call z in the first place. But determining the probabilities is not so simple. To each component of $\mathbf{S}$ we associate a 2×2 matrix:*

* Again, the *derivation* of these matrices will be found in any quantum-mechanics text. My purpose here is to show you *how* angular momentum is handled in particle physics, not to explain *why* it is done this way.

$$\hat{S}_x = \frac{\hbar}{2}\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \qquad \hat{S}_y = \frac{\hbar}{2}\begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \qquad \hat{S}_z = \frac{\hbar}{2}\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \tag{4.21}$$

The eigenvalues of $\hat{S}_x$ are $\pm\hbar/2$, and corresponding normalized eigenvectors are*

$$\chi_{\pm} = \begin{pmatrix} 1/\sqrt{2} \\ \pm 1/\sqrt{2} \end{pmatrix} \tag{4.22}$$

(see Problem 4.15). An arbitrary spinor $\begin{pmatrix} \alpha \\ \beta \end{pmatrix}$ can be written as a linear combination of these eigenvectors:

$$\begin{pmatrix} \alpha \\ \beta \end{pmatrix} = a\begin{pmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \end{pmatrix} + b\begin{pmatrix} 1/\sqrt{2} \\ -1/\sqrt{2} \end{pmatrix} \tag{4.23}$$

where $$a = (1/\sqrt{2})(\alpha + \beta); \qquad b = (1/\sqrt{2})(\alpha - \beta) \tag{4.24}$$

The probability that a measurement of S_x will yield the value $\frac{1}{2}\hbar$ is $|a|^2$; the probability of getting $-\frac{1}{2}\hbar$ is $|b|^2$. Evidently, $|a|^2 + |b|^2 = 1$ (see Problem 4.16).

The general procedure, of which this was a particular instance, is as follows:

1. Construct the matrix, $\hat{A}$, representing the observable A in question.
2. The allowed values of A are the eigenvalues of $\hat{A}$.
3. Write the state of the system as a linear combination of eigenvectors of $\hat{A}$; the absolute square of the coefficient of the ith eigenvector is the probability that a measurement of A would yield the ith eigenvalue.

EXAMPLE 4.5

Suppose we measure S_x^2 on a particle in the state $\begin{pmatrix} \alpha \\ \beta \end{pmatrix}$. What values might we get, and what is the probability of each?

Solution. The matrix representing S_x^2 would be the square of the matrix representing S_x:

$$\hat{S}_x^2 = \frac{\hbar^2}{4}\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \tag{4.25}$$

* A nonzero column matrix

$$\chi = \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{pmatrix}$$

is called an *eigenvector* of a given $n \times n$ matrix M if

$$M\chi = \lambda\chi$$

for some number λ (the *eigenvalue*). Notice that any *multiple* of χ is still an eigenvector, with the same eigenvalue.

Since $$\frac{\hbar^2}{4}\begin{pmatrix}1 & 0\\ 0 & 1\end{pmatrix}\begin{pmatrix}\alpha\\ \beta\end{pmatrix} = \frac{\hbar^2}{4}\begin{pmatrix}\alpha\\ \beta\end{pmatrix}$$

every spinor is an eigenvector of $\hat{S}_x^2$, with eigenvalue $\hbar^2/4$. Thus we would be *certain* to get $\hbar^2/4$ (probability 1). The same goes for $\hat{S}_y^2$ and $\hat{S}_z^2$, so every spinor is an eigenstate of $\hat{S}^2 = \hat{S}_x^2 + \hat{S}_y^2 + \hat{S}_z^2$, with eigenvalue $3\hbar^2/4$. This should come as no surprise—in general, for spin s we must have $S^2 = s(s+1)\hbar^2$.

For mathematical purposes the factor of $\hbar/2$ in equation (4.21) is ugly, and it is customary to introduce the *Pauli spin matrices:*

$$\sigma_x = \begin{pmatrix}0 & 1\\ 1 & 0\end{pmatrix}, \qquad \sigma_y = \begin{pmatrix}0 & -i\\ i & 0\end{pmatrix}, \qquad \sigma_z = \begin{pmatrix}1 & 0\\ 0 & -1\end{pmatrix} \tag{4.26}$$

so that $\hat{\mathbf{S}} = (\hbar/2)\boldsymbol{\sigma}$. The Pauli matrices have many interesting properties, some of which are explored in Problems 4.19 and 4.20. We shall encounter them repeatedly in the course of this book.

In a sense, spinors (two-component objects) occupy an intermediate position between scalars (one component) and vectors (three components). Now, when you rotate your coordinate axes, the components of a vector change, in a prescribed manner (see Problem 4.6), and we might inquire how the components of a *spinor* transform, under the same circumstances. The answer[4] is provided by the following rule:

$$\begin{pmatrix}\alpha'\\ \beta'\end{pmatrix} = U(\boldsymbol{\theta})\begin{pmatrix}\alpha\\ \beta\end{pmatrix} \tag{4.27}$$

where $U(\boldsymbol{\theta})$ is the 2×2 matrix

$$U(\boldsymbol{\theta}) = e^{-i\boldsymbol{\theta}\cdot\boldsymbol{\sigma}/2} \tag{4.28}$$

The vector $\boldsymbol{\theta}$ points along the axis of rotation, and its magnitude is the angle of rotation, in the right-hand sense, about that axis. Notice that the exponent here is itself a matrix (!). An expression of this form is to be interpreted as shorthand for the power series:

$$e^A \equiv 1 + A + \tfrac{1}{2}A^2 + \tfrac{1}{3!}A^3 + \cdots \tag{4.29}$$

(see Problem 4.21).* As you can check for yourself (Problem 4.22), $U(\boldsymbol{\theta})$ is a unitary matrix of determinant 1; in fact, the set of all such rotation matrices constitutes the group $SU(2)$. Thus spin-$\frac{1}{2}$ particles transform under rotations according to the fundamental, two-dimensional representation of $SU(2)$. Similarly, particles of spin 1, described by *vectors,* belong to the three-dimensional representation of $SU(2)$; spin-$\frac{3}{2}$ particles, described by a four-component object,

* *Beware:* For matrices it is *not* the case that $e^A e^B = e^{A+B}$, in general. You might want to check this by using the matrices in Problem 4.21. However, the usual rule *does* apply if A and B *commute* (i.e., if $AB = BA$).

transform under the four-dimensional representation of $SU(2)$; and so on. (The construction of these higher-dimensional representations is explored in Problem 4.23.) You're probably wondering what $SU(2)$ has to do with rotations; well, as I mentioned earlier, $SU(2)$ is essentially* the same group as $SO(3)$, the group of rotations in three dimensions. Particles of different spin, then, belong to different *representations* of the rotation group.

4.5 FLAVOR SYMMETRIES

There's an extraordinary thing about the neutron, which Heisenberg observed shortly after its discovery in 1932: apart from the obvious fact that it carries no charge, it is almost identical to the proton. In particular, their masses are astonishingly close, $m_p = 938.28\ \text{MeV}/c^2$, $m_n = 939.57\ \text{MeV}/c^2$). Heisenberg[5] proposed that we regard them as two "states" of a *single* particle, the *nucleon.* Even the small difference in mass might be attributed to the fact that the proton is charged. since the energy stored in its electric field contributes, according to Einstein's formula ($E = mc^2$) to its inertia. Unfortunately, this argument suggests that the *proton* should be the heavier of the two, which is not only untrue, but would be disastrous for the stability of matter. More on this in a moment. If we could somehow "turn off" all electric charge, the proton and neutron would, according to Heisenberg, be indistinguishable. Or, to put it more prosaically, the *strong* forces experienced by protons and neutrons are identical.

To implement Heisenberg's idea, we write the nucleon as a two-component column matrix

$$N = \begin{pmatrix} \alpha \\ \beta \end{pmatrix} \tag{4.30}$$

with

$$p = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \quad \text{and} \quad n = \begin{pmatrix} 0 \\ 1 \end{pmatrix} \tag{4.31}$$

This is nothing but notation, of course, but it is notation seductively reminiscent of the spinors we encountered in the theory of angular momentum. By direct analogy with spin, **S**, we are led to introduce *isospin,* **I**.† However, **I** is not a vector in *ordinary* space, with components along the coordinate directions x. y, and z, but rather in an abstract "isospin space," with components we'll call I_1, I_2, and I_3. On this understanding, we may borrow the entire apparatus of

* There is actually a subtle distinction between $SU(2)$ and $SO(3)$. According to Problem 4.21. the matrix U for rotation through an angle of 2π is -1; a spinor *changes sign* under such a rotation. And yet, *geometrically,* a rotation through 2π is equivalent to no rotation at all. $SU(2)$ is a kind of "doubled" version of $SO(3)$, in which you don't come back to the beginning until you've turned through 720°. In this sense spinor representations of $SU(2)$ are *not* true representations of the rotation group, and that's why they do not appear in classical physics. In quantum mechanics only the *square* of the wave function carries physical significance, and in the squaring the minus sign goes away.

† The word derives from the older term *isotopic* spin, which was misleading, since two isotopes of a given nucleus have different numbers of nucleons, whereas isospin rotations preserve the number of nucleons. Nuclear physicists use the (better) term *isobaric* spin.

angular momentum, as developed earlier in the chapter. The nucleon carries isospin $\frac{1}{2}$, and the third component, I_3, has the eigenvalues* $+\frac{1}{2}$ (the proton) and $-\frac{1}{2}$ (the neutron):

$$p = |\tfrac{1}{2}\,\tfrac{1}{2}\rangle, \qquad n = |\tfrac{1}{2}\,-\tfrac{1}{2}\rangle \tag{4.32}$$

The proton is "isospin up"; the neutron is "isospin down." This is *still* just notation; the *physics* comes in Heisenberg's proposition that the strong interactions are invariant under rotations in isospin space, just as, for example, electrical forces are invariant under rotations in ordinary configuration space. We call this an "internal" symmetry, because it has nothing to do with space and time, but rather with the relations between different particles. A rotation through 180° about axis number 1 in isospin space converts protons into neutrons, and vice versa. If the strong force is invariant under rotations in isospin space, it follows, by Noether's theorem, that *isospin is conserved in all strong interactions,* just as angular momentum is conserved in processes with rotational invariance in ordinary space.†

In the language of group theory, Heisenberg asserted that the strong interactions are invariant under an internal symmetry group $SU(2)$, and the nucleons belong to the two-dimensional representation (isospin $\frac{1}{2}$). In 1932 this was a bold suggestion; today the evidence is all around us, most conspicuously in the "multiplet" structure of the hadrons. Recall the Eightfold Way diagrams in Chapter 1: The horizontal rows all display exactly the feature that caught Heisenberg's eye in the case of the nucleons; they have very similar masses but different charges. To each of these multiplets we assign a particular isospin I, and to each member of the multiplet we assign a particular I_3. For the pions, $I = 1$:

$$\pi^+ = |11\rangle, \qquad \pi^0 = |10\rangle, \qquad \pi^- = |1\,-1\rangle \tag{4.33}$$

for the Λ, $I = 0$:

$$\Lambda = |00\rangle \tag{4.34}$$

for the Δ's, $I = \frac{3}{2}$:

$$\Delta^{++} = |\tfrac{3}{2}\,\tfrac{3}{2}\rangle, \qquad \Delta^{+} = |\tfrac{3}{2}\,\tfrac{1}{2}\rangle, \qquad \Delta^{0} = |\tfrac{3}{2}\,-\tfrac{1}{2}\rangle, \qquad \Delta^{-} = |\tfrac{3}{2}\,-\tfrac{3}{2}\rangle \tag{4.35}$$

and so on. To determine the isospin of a multiplet, just count the number of particles it contains; since I_3 ranges from $-I$ to $+I$, in integer steps, the number of particles in the multiplet is $2I + 1$:

$$\text{multiplicity} = 2I + 1 \tag{4.36}$$

* There is no factor of $\hbar$ in this case; isospin is dimensionless, by convention.

† By the way, it is tempting to overstate the so-called "charge independence" of the strong forces (the fact that they are the same for protons as for neutrons). It does *not* say that you'll get the same result if you substitute an *individual* proton for a neutron, only if you interchange *all* protons and neutrons. [For example, there exists a bound state of the proton and the neutron (to wit, the deuteron), but there is *no* bound state of two protons or two neutrons.] Indeed, any such assertion would be incompatible with the Pauli exclusion principle, since a proton and a neutron can be in the same quantum state, but two neutrons (or two protons) cannot.

The third component of isospin, I_3, is determined by the charge, Q, of the particle. We assign the maximum value, $I_3 = I$, to the member of the multiplet with the highest charge, and fill in the rest in order of decreasing Q. For the "pre-1974" hadrons—those composed of u, d, and s quarks only—the explicit relation between Q and I_3 is the *Gell-Mann–Nishijima formula:*

$$Q = I_3 + \tfrac{1}{2}(A + S) \tag{4.37}$$

where A is the baryon number and S is the strangeness.* Originally, this equation was a purely empirical observation, but in the context of the quark model it follows simply from the isospin assignments for quarks: u and d form a "doublet" (like the proton and the neutron):

$$u = |\tfrac{1}{2}\,\tfrac{1}{2}\rangle, \qquad d = |\tfrac{1}{2}\,-\tfrac{1}{2}\rangle \tag{4.38}$$

and all the other flavors carry isospin zero† (see Problems 4.25 and 4.26).

But classification is not all that isospin does for us. It also has important *dynamical* implications. For example, suppose we have *two* nucleons. From the rules for addition of angular momenta we know that the combination gives a total isospin of 1 or 0. Specifically (using Example 4.4), we obtain a symmetric isotriplet:

$$\begin{aligned} &\textbf{(a)} \quad |11\rangle = pp \\ &\textbf{(b)} \quad |10\rangle = (1/\sqrt{2})(pn + np) \\ &\textbf{(c)} \quad |1\,{-1}\rangle = nn \end{aligned} \tag{4.39}$$

and an antisymmetric isosinglet:

$$|00\rangle = (1/\sqrt{2})(pn - np) \tag{4.40}$$

Experimentally, the neutron and proton form a single bound state, the deuteron (d); there is no bound state of two protons or of two neutrons. Thus the deuteron must be an isosinglet. If it were a triplet, all three states would have to occur since they differ only by a rotation in isospin space. Evidently there is a strong attraction in the $I = 0$ channel, but not in the $I = 1$ channel. Presumably the potential describing the interaction between two nucleons contains a term of the form $\mathbf{I}^{(1)} \cdot \mathbf{I}^{(2)}$, which takes the value $\frac{1}{4}$ in the triplet configuration and $-\frac{3}{4}$ in the singlet (see Problem 4.27).

Isospin invariance has implications, too, for nucleon-nucleon scattering. Consider the processes

$$\begin{aligned} &\textbf{(a)} \quad p + p \rightarrow d + \pi^+ \\ &\textbf{(b)} \quad p + n \rightarrow d + \pi^0 \\ &\textbf{(c)} \quad n + n \rightarrow d + \pi^- \end{aligned} \tag{4.41}$$

* Since Q, A, and S are all conserved by the electromagnetic forces, it follows that I_3 is also conserved. However, the other two components, and hence also I itself, are *not* conserved in electromagnetic interactions. For example, in the decay $\pi^0 \rightarrow \gamma + \gamma$, I goes from 1 to 0. As for the weak interactions, they don't even conserve S, so I_3 is not conserved in weak processes (for example, $\Lambda \rightarrow p + \pi^-$).

† Since isospin pertains only to the strong forces, it is not a relevant quantity for leptons. If you insist, all leptons and mediators carry isospin zero.

Since the deuteron carries $I = 0$, the isospin states on the right are $|11\rangle$, $|10\rangle$, and $|1\,-1\rangle$, respectively, whereas those on the left are $|11\rangle$, $(1/\sqrt{2})(|10\rangle + |00\rangle)$, and $|1\,-1\rangle$. Only the $I = 1$ combination contributes (since the final state in each case is pure $I = 1$, and isospin is conserved), so the *scattering amplitudes* are in the ratio

$$\mathcal{M}_a : \mathcal{M}_b : \mathcal{M}_c = 1 : (1/\sqrt{2}) : 1 \tag{4.42}$$

As we shall see,* the *cross section*, σ, goes like the absolute square of the amplitude; thus

$$\sigma_a : \sigma_b : \sigma_c = 2 : 1 : 2 \tag{4.43}$$

Process (c) would be hard to set up in the laboratory, but (a) and (b) have been measured, and (when corrections are made for electromagnetic effects) they are found to be in the predicted ratio.[6]

As a final example, let's consider pion-nucleon scattering, $\pi N \to \pi N$. There are six elastic processes:

$$\begin{aligned}
&\textbf{(a)}\quad \pi^+ + p \to \pi^+ + p \qquad &&\textbf{(b)}\quad \pi^0 + p \to \pi^0 + p \\
&\textbf{(c)}\quad \pi^- + p \to \pi^- + p \qquad &&\textbf{(d)}\quad \pi^+ + n \to \pi^+ + n \\
&\textbf{(e)}\quad \pi^0 + n \to \pi^0 + n \qquad &&\textbf{(f)}\quad \pi^- + n \to \pi^- + n
\end{aligned} \tag{4.44}$$

and four charge-exchange processes:

$$\begin{aligned}
&\textbf{(g)}\quad \pi^+ + n \to \pi^0 + p \qquad &&\textbf{(h)}\quad \pi^0 + p \to \pi^+ + n \\
&\textbf{(i)}\quad \pi^0 + n \to \pi^- + p \qquad &&\textbf{(j)}\quad \pi^- + p \to \pi^0 + n
\end{aligned} \tag{4.45}$$

Since the pion carries $I = 1$, and the nucleon $I = \frac{1}{2}$, the total isospin can be $\frac{3}{2}$ or $\frac{1}{2}$. So there are just two distinct amplitudes here: $\mathcal{M}_3$, for $I = \frac{3}{2}$, and $\mathcal{M}_1$, for $I = \frac{1}{2}$. From the Clebsch–Gordan tables we find the following decompositions:

$$\left.\begin{aligned}
\pi^+ + p\text{:}\ |11\rangle|\tfrac{1}{2}\tfrac{1}{2}\rangle &= |\tfrac{3}{2}\tfrac{3}{2}\rangle \\
\pi^0 + p\text{:}\ |10\rangle|\tfrac{1}{2}\tfrac{1}{2}\rangle &= \sqrt{2/3}|\tfrac{3}{2}\tfrac{1}{2}\rangle - (1/\sqrt{3})|\tfrac{1}{2}\tfrac{1}{2}\rangle \\
\pi^- + p\text{:}\ |1\,-1\rangle|\tfrac{1}{2}\tfrac{1}{2}\rangle &= (1/\sqrt{3})|\tfrac{3}{2}\,-\tfrac{1}{2}\rangle - \sqrt{2/3}|\tfrac{1}{2}\,-\tfrac{1}{2}\rangle \\
\pi^+ + n\text{:}\ |11\rangle|\tfrac{1}{2}\,-\tfrac{1}{2}\rangle &= (1/\sqrt{3})|\tfrac{3}{2}\tfrac{1}{2}\rangle + \sqrt{2/3}|\tfrac{1}{2}\tfrac{1}{2}\rangle \\
\pi^0 + n\text{:}\ |10\rangle|\tfrac{1}{2}\,-\tfrac{1}{2}\rangle &= \sqrt{2/3}|\tfrac{3}{2}\,-\tfrac{1}{2}\rangle + (1/\sqrt{3})|\tfrac{1}{2}\,-\tfrac{1}{2}\rangle \\
\pi^- + n\text{:}\ |1\,-1\rangle|\tfrac{1}{2}\,-\tfrac{1}{2}\rangle &= |\tfrac{3}{2}\,-\tfrac{3}{2}\rangle
\end{aligned}\right\} \tag{4.46}$$

Reactions (a) and (f) are pure $I = \frac{3}{2}$:

$$\mathcal{M}_a = \mathcal{M}_f = \mathcal{M}_3 \tag{4.47}$$

The others are all mixtures; for example

$$\mathcal{M}_c = \tfrac{1}{3}\mathcal{M}_3 + \tfrac{2}{3}\mathcal{M}_1, \qquad \mathcal{M}_j = (\sqrt{2}/3)\mathcal{M}_3 - (\sqrt{2}/3)\mathcal{M}_1 \tag{4.48}$$

(I'll let you work out the rest, see Problem 4.28). The cross sections, then, stand in the ratio

* The theory of scattering amplitudes and cross sections will be developed in Chapter 6. In this and the following paragraph I anticipate later results, but I hope it is clear from the context how the calculation proceeds. If you wish, skip these two paragraphs for now.

$$\sigma_a : \sigma_c : \sigma_j = 9|\mathcal{M}_3|^2 : |\mathcal{M}_3 + 2\mathcal{M}_1|^2 : 2|\mathcal{M}_3 - \mathcal{M}_1|^2 \tag{4.49}$$

At a *CM* energy of 1232 MeV there occurs a famous and dramatic bump in pion-nucleon scattering, first discovered by Fermi in 1951;[7] here the pion and nucleon join to form a short-lived "resonance" state—the Δ. We know the Δ carries $I = \frac{3}{2}$, so we expect that at this energy $\mathcal{M}_3 \gg \mathcal{M}_1$, and hence

$$\sigma_a : \sigma_c : \sigma_j = 9 : 1 : 2 \tag{4.50}$$

Experimentally, it is easier to measure the *total* cross sections, so (c) and (j) are combined:

$$\frac{\sigma_{\text{tot}}(\pi^+ + p)}{\sigma_{\text{tot}}(\pi^- + p)} = 3 \tag{4.51}$$

As you can see in Figure 4.6, this prediction is well satisfied by the data.

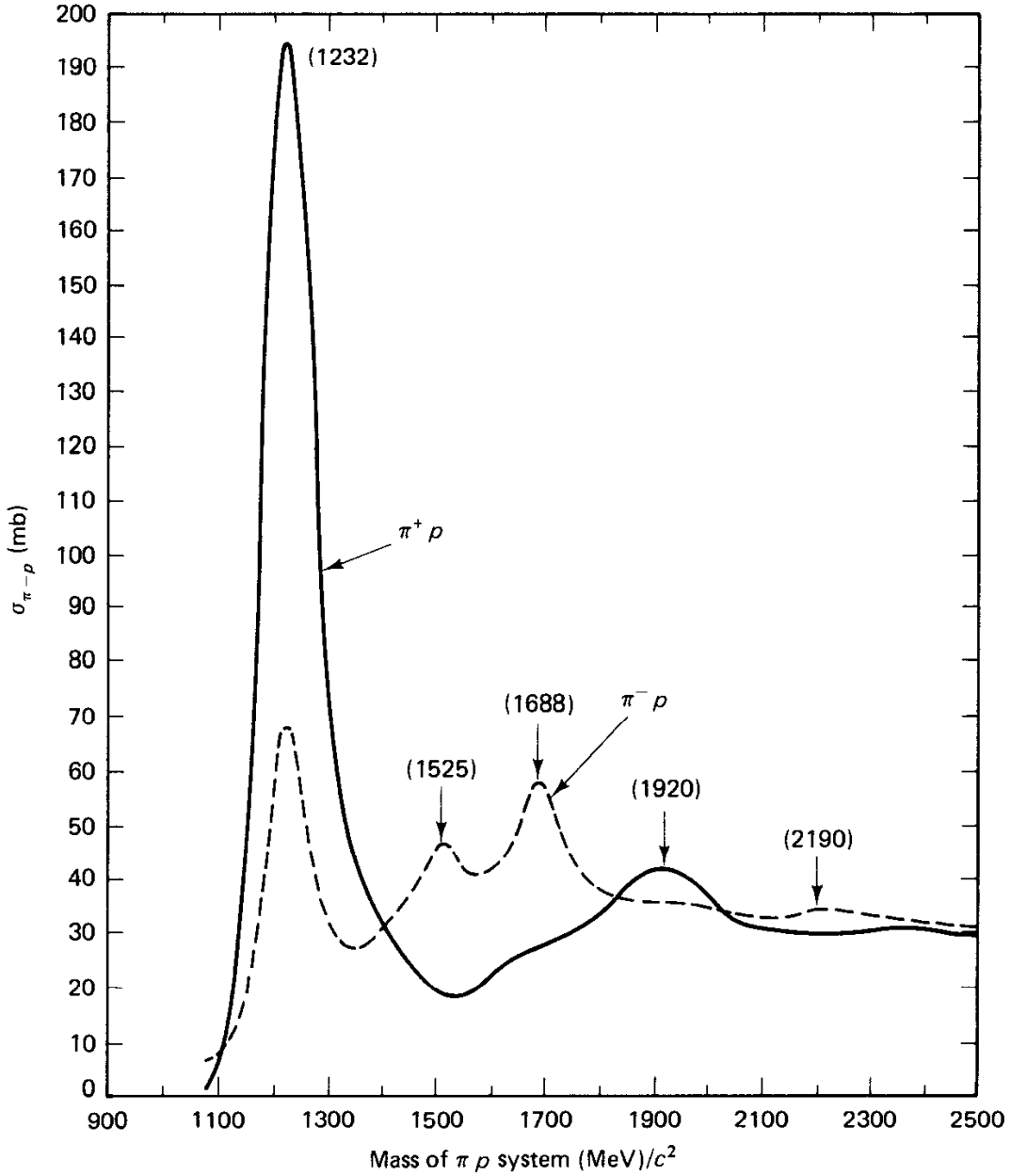


Figure 4.6 Total cross sections for $\pi^+ p$ (solid line) and $\pi^- p$ (dashed line) scattering. (*Source:* S. Gasiorowicz, *Elementary Particle Physics* (New York: Wiley, copyright © 1966, page 294. Reprinted by permission of John Wiley and Sons, Inc.)

In the late fifties history repeated itself. Just as in 1932 the proton and neutron were seen to form a pair, it was now increasingly clear that the nucleons, the Λ, the Σ's, and the Ξ's together, constituted a natural grouping within the baryon family. They all carry spin $\frac{1}{2}$, and their masses are similar. It is true that the latter range from 940 MeV/c^2, for the nucleons, up to 1320 MeV/c^2, for the Ξ's, so it would be stretching things a bit to argue that they are all different states of one particle, as Heisenberg had implied for the proton and neutron. Nevertheless, it was tempting to regard these eight baryons as a *supermultiplet,* and this presumably meant that they belonged in the same representation of some enlarged symmetry group, in which the $SU(2)$ of isospin would be incorporated as a *subgroup.* The critical question became: What is the larger group? (The "Eight Baryon Problem," as it was called, was not always phrased this way; at the time most physicists were surprisingly ignorant of group theory. Gell-Mann worked out most of the formalism he needed from scratch, and only later learned that it was well known to mathematicians.) The Eightfold Way was Gell-Mann's solution to the Eight Baryon Problem. The symmetry group is $SU(3)$; the octets constitute eight-dimensional representations of $SU(3)$, the decuplet a ten-dimensional representation, and so on. One thing that made this case more difficult than Heisenberg's was that no naturally occurring particles fall into the *fundamental* (three-dimensional) representation of $SU(3)$, as the nucleons, and later the K's, the Ξ's, and so on, do for $SU(2)$. This role was reserved for the quarks: u, d, and s together form a three-dimensional representation of $SU(3)$, which breaks down into an isodoublet (u, d) and an isosinglet (s) under $SU(2)$.

Of course, when the charmed quark came along, the flavor symmetry group of the strong interactions expanded once again—this time to $SU(4)$ (some $SU(4)$ supermultiplets are shown in Fig. 1.13). But things barely paused there before the arrival of the bottom quark, taking us to $SU(5)$, and the putative top quark, $SU(6)$. However, there is an important *caveat* in this neat hierarchy: Isospin, $SU(2)$, is a very "good" symmetry; the members of an isospin multiplet differ in mass by at most 2 or 3%, which is about the level at which electromagnetic corrections would be expected.* But the Eightfold Way, SU(3), is a badly "broken" symmetry; mass splittings within the baryon octet are around 40%. The symmetry breaking is even worse when we include charm; the Λ_c^+ (udc) weighs more than twice the Λ (uds), although they are in the same $SU(4)$ supermultiplet. It is worse still with bottom, and absolutely terrible with top.

Why is isospin such a good symmetry, the Eightfold Way fair, and flavor $SU(6)$ so poor? The Standard Model blames it all on the quark masses. Now, the theory of quark masses is a slippery business, given the fact that they are not accessible to direct experimental measurement. Various arguments[8] suggest that the u and d quarks are *intrinsically* very light, about ten times the mass of the electron. However, within the confines of a hadron their *effective* mass is much greater. The precise value, in fact, depends on the context; it tends to be a little

* Indeed, it used to be thought that isospin was an *exact* symmetry of the strong interactions, and the *whole* of the symmetry breaking was attributable to electromagnetic contamination. The fact that the n–p mass splitting is in the wrong direction to be purely electromagnetic was troubling, however, and we now believe that $SU(2)$ is only an *approximate* symmetry of the strong interactions.

higher in baryons than in mesons (more on this in Chapter 5). In somewhat the same way, the effective inertia of a teaspoon is greater when you're stirring honey than when you're stirring tea, and in either case it exceeds the true mass of the spoon. Generally speaking, the effective mass of a quark in a hadron is about 350 MeV/c^2 greater than its bare mass (see Table 4.4). Compared to this, the quite different *bare* masses of up and down quarks are practically irrelevant; they *function* as though they had identical masses. But the *s* quark is distinctly heavier, and the *c*, *b*, and *t* quarks are widely separated. Apart from the differences in quark masses, the strong interactions treat all flavors equally. Thus isospin is a good symmetry because the effective *u* and *d* masses are so nearly equal (which is to say, on a more fundamental level, because their *bare* masses are so small); the Eightfold Way is a *fair* symmetry because the effective mass of the strange quark is not *too* far from that of the *u* and *d*. But the heavy quarks are so far apart that their flavor symmetry is severely broken. Of course, this "explanation" raises two further questions: (1) *Why* does the binding of quarks into hadrons increase their effective mass by about 350 MeV/c^2? The answer presumably lies within QCD, although the details are not yet understood.[9] (2) Why do the *bare* quarks have the particular masses they do? Is there some pattern here? To this question the Standard Model offers no answer; the six bare quark masses, and also the six lepton masses, are simply input parameters, for now, and it is the business of theories *beyond* the Standard Model to say where they come from.

4.6 PARITY

Prior to 1956 it was taken for granted that the laws of physics are ambidextrous; that is, the mirror image of any physical process also represents a perfectly possible physical process. To be sure, we drive on the right (at least, Americans do) and our hearts are on the left, but these are obviously historical or evolutionary accidents; it could just as well have been the other way around. Indeed, most physicists regarded the mirror symmetry (or "parity invariance") of the laws of nature as self-evident. But in 1956 Lee and Yang[10] were led to wonder (for reasons we will come back to at the end of this section) whether there had been any *experimental* test of this assumption. Searching the literature, they were

TABLE 4.4 QUARK MASSES (MeV/c^2)

	Quark flavor	Bare mass	Effective mass	
			in mesons	in baryons
Light quarks	*u*	4.2	310	363
	d	7.5		
	s	150	483	538
Heavy quarks	*c*	1,100	1,500	
	b	4,200	4,700	
	t	>23,000		

The second column lists the bare (or "current") mass; the third column lists the effective (or "constituent") mass. *Warning:* These numbers are somewhat speculative and model-dependent.

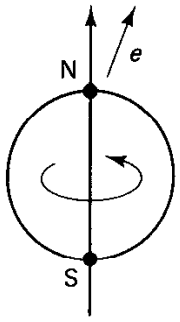


Figure 4.7 In the beta decay of cobalt 60, most electrons are emitted in the direction of the nuclear spin.

surprised to discover that although there was ample evidence for parity invariance in strong and electromagnetic processes, there was no confirmation in the case of *weak* interactions. They proposed a test, which was carried out later that year by C. S. Wu,[11] to settle the issue. In this famous experiment, radioactive cobalt 60 nuclei were carefully aligned, so that their spins pointed in, say, the z direction (Fig. 4.7). Cobalt 60 undergoes beta decay, and Wu recorded the direction of the emitted electrons. What she found was that most of them came out in the "northerly" direction, that is, in the direction of the nuclear spin.

That's all there was to it. But that simple observation had astonishing implications. For suppose we examine the mirror image of that same process (Fig. 4.8). The image nucleus rotates in the opposite direction; its spin points *downward.* And yet the electrons (in the mirror) still came off upward. In the mirror, then, the electrons are emitted preferentially in the direction *opposite* to the nuclear spin. Here, then, is a physical process whose mirror image does *not* occur in nature; evidently parity is *not* an invariance of the weak interactions. If it *were,* the electrons in Wu's experiment would have to come out in equal distribution ("north" and "south"), but they don't.

The overthrow of parity had a profound effect on physicists—devastating to some, exhilarating to others.[12] The violation is not a small effect; as we shall see in Chapter 10, it is in fact "maximal." Nor is it limited to beta decay in cobalt; once you *look* for it, parity violation is practically the *signature* of the weak force. It is most dramatically revealed in the behavior of the neutrino. Let me explain. In the theory of angular momentum the axis of quantization is, by

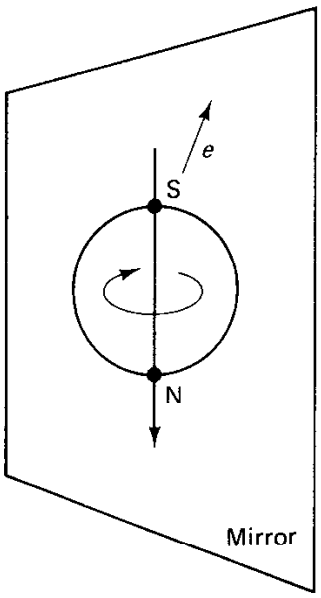


Figure 4.8 Mirror image of Figure 4.7: Most electrons are emitted *opposite* to nuclear spin.



(a) Right-handed (b) Left-handed

Figure 4.9 Helicity. In (a) the spin and velocity are parallel (helicity +1); in (b) they are antiparallel (helicity −1).

convention, the z axis. Of course, the orientation of the z axis is completely up to us, but if we are dealing with a particle traveling through the laboratory at velocity **v**, a natural choice suggests itself: Why not pick the *direction of motion* as the z axis? The value of (m_s/s) for this axis is called the *helicity* of the particle. Thus a particle of spin $\frac{1}{2}$ can have a helicity of $+1$ ($m_s = \frac{1}{2}$) or -1 ($m_s = -\frac{1}{2}$): we call the former "right-handed" and the latter "left-handed."* The difference is not terribly profound, however, because it is not Lorentz-invariant. Suppose I have a right-handed electron going to the right (Fig. 4.9a), and someone else looks at it from an inertial system traveling to the right at a speed *greater* than v. From *his* perspective the electron is going to the *left* (Fig. 4.9b); but it is still spinning the same way, so this observer will say it's a *left*-handed electron. In other words, you can convert a right-handed electron into a left-handed one simply by changing your frame of reference. That's what I mean when I say the distinction is not Lorentz-invariant.

But what if we applied that same reasoning to a *neutrino,* instead of an electron? The neutrino is massless, so it travels at the speed of light, and hence there *is* no observer traveling faster. It is *impossible* to "reverse the direction of motion" of a neutrino by getting into a faster-moving reference system, and therefore the helicity of a neutrino (or any other massless particle†) is Lorentz-invariant—a fixed and fundamental property, which is *not* an artifact of the observer's reference frame. It becomes an important *experimental* matter to determine the helicity of a given neutrino. Until the mid-fifties everyone assumed that half of all neutrinos would be left-handed, and half right-handed, just like photons. What they, in fact, *discovered* was that

ALL NEUTRINOS ARE LEFT-HANDED,
AND ALL ANTINEUTRINOS ARE RIGHT-HANDED.

Of course, it's tough to measure the helicity of a neutrino directly; they're hard enough to detect at *all.* There is, however, a relatively easy indirect method. using the decay of the pion: $\pi^- \rightarrow \mu^- + \bar{\nu}_\mu$. If the pion is at rest, the muon and the antineutrino come out back to back (Fig. 4.10). Moreover, since the pion

* In Chapter 10 I shall introduce a technical distinction between "handedness" and *helicity* but for the moment we shall use the terms interchangeably.

† For massless particles only the maximal value of $|m_s|$ occurs. For example, the photon can have $m_s = +1$ or $m_s = -1$, but not $m_s = 0$. So the helicity of a massless particle is always ± 1. In the case of the photon these represent states of left- and right-circular polarization. The absence of $m_s = 0$ corresponds to the absence of longitudinal polarization in classical optics.

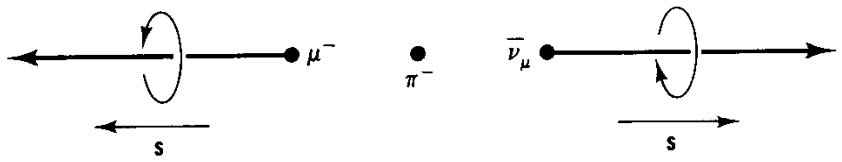


Figure 4.10 Decay of π^- at rest.

has spin 0, the muon and the antineutrino spins must be oppositely aligned.* Therefore if the antineutrino is right-handed, the muon must be right-handed too (in the pion rest frame)—and this is precisely what is found experimentally.[13] Measurement of the *muon* helicity, then, enables us to determine the *antineutrino* helicity. By the same token, in π^+ decay the antimuon is always left-handed, and this indicates that the neutrino is left-handed. For contrast, consider the decay of the *neutral* pion, $\pi^0 \rightarrow \gamma + \gamma$. Once again, in any given decay the two photons must have the same helicity. But this is an *electromagnetic* process, which respects parity, and thus, on the average, we get just as many right-handed photon pairs as left-handed pairs. Not so for neutrinos; they only interact weakly, and every one is left-handed; the mirror image of a neutrino does not exist.† That is about the starkest violation of mirror symmetry you could ask for.‡

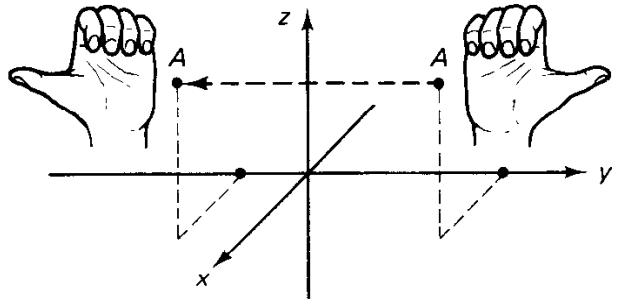
In spite of its violation in weak processes, parity invariance remains a valid symmetry of the strong and electromagnetic interactions. It is useful, therefore, to develop some formalism and terminology. First a minor technical point: instead of *reflections,* which oblige us to choose arbitrarily the plane of the "mirror", we'll talk about *inversions,* in which every point is carried through the origin to the diametrically opposite location (Fig. 4.11). Both transformations have the property of turning a right hand into a left hand; in fact, an inversion is nothing but a reflection followed by a rotation (180° about the y axis, in the figure). Thus in the cases of interest (which *also* possess rotational symmetry) it is a matter of indifference which one is used. Let P denote inversion; we call it the "parity operator." If the system in question is a right hand, P turns it into an upside-

* The orbital angular momentum (if there is any) points perpendicular to the outgoing velocities, so it does not affect this argument.

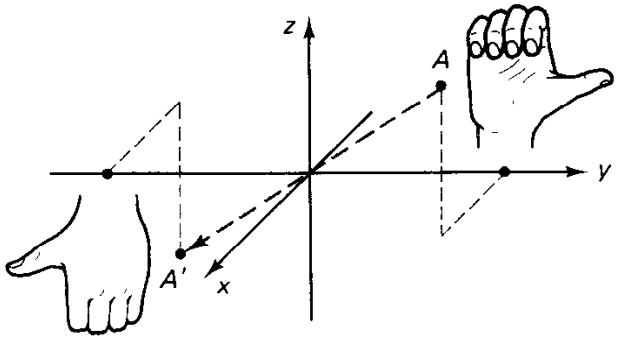
† This is perhaps too strong a statement. There could, I suppose, be right-handed neutrinos around, but they do not interact with ordinary matter by any mechanism presently known. If it turns out that neutrinos have a small but nonzero mass, then, of course, right-handed neutrinos *must* exist. None of this, however, can alter the fact that when a π^- decays, the emerging μ^- is right-handed in the *CM* frame and that by itself destroys mirror symmetry.

By the way, back in 1929, shortly after the publication of Dirac's equation, Weyl presented a beautifully simple theory of massless particles of spin $\frac{1}{2}$, which had the feature that they carried a fixed "handedness." At the time Weyl's theory aroused limited interest, since there were no massless particles known, except for the photon, which carries spin 1. When Pauli introduced the neutrino, in 1931, you might suppose that he would dust off Weyl's theory and put it to use. He did not. Pauli rejected Weyl's theory out of hand, on the ground that it violated mirror symmetry. He lived to regret this mistake, and in 1957 Weyl's theory was triumphantly vindicated.

‡ It may occur to you, as it did to many physicists at the time, that if we simultaneously convert all particles into their antiparticles, then a kind of mirror symmetry is restored; the image of $\pi^- \rightarrow \mu^- + \bar{\nu}_\mu$ (with a right-handed antineutrino) becomes $\pi^+ \rightarrow \mu^+ + \nu_\mu$ (with a left-handed neutrino), which is perfectly okay. This realization was some comfort, until 1964, when it, too, was shown to fail. More on this in the following sections.



(a) Reflection (in the x-z plane)
$(x, y, z) \rightarrow (x, -y, z)$



(b) Inversion $(x, y, z) \rightarrow (-x, -y, -z)$

Figure 4.11 Reflections and inversions.

down and backward *left* hand (Fig. 4.11b). When applied to a vector, **a**, P produces a vector pointing in the opposite direction: $P(\mathbf{a}) = -\mathbf{a}$. How about the cross product of two vectors: $\mathbf{c} = \mathbf{a} \times \mathbf{b}$? Well, if P changes the sign of **a** and of **b**, then evidently **c** itself does *not* change sign: $P(\mathbf{c}) = \mathbf{c}$. Very strange! Apparently there are two kinds of vectors—"ordinary" ones, which change sign under the parity transformation, and this other type, of which the cross product is the classic example, which do *not*. We call the former "polar" vectors, when the distinction must be drawn, and the latter "pseudo" (or "axial") vectors. Notice that the cross product of a polar vector with a *pseudo*vector would be a *polar* vector.

You will have encountered pseudovectors before, perhaps without using this language; angular momentum is one, and so is the magnetic field. In a theory with parity invariance, you must never add a vector to a pseudovector, just as in a theory with rotational symmetry, you cannot add a vector to a scalar. For example, in the Lorentz force law, $\mathbf{F} = q(\mathbf{E} + \mathbf{v}/c \times \mathbf{B})$, it is the cross product (a polar vector) that enters; **B** itself could never be added to **E**. As we shall see, it is precisely the addition of a vector to a pseudovector in the theory of weak interactions that leads to the breakdown of parity. Finally, the *dot* product of two polar vectors does not change sign under P, but the dot product of a polar vector and a pseudovector [or the triple product of three vectors: $\mathbf{a} \cdot (\mathbf{b} \times \mathbf{c})$] *does* change sign. So there are two kinds of *scalars*, too: the "ordinary" kind, which don't change sign, and "pseudoscalars," which do. All this is summarized in Table 4.5.*

* The terminology extends very simply to special relativity: $a^\mu = (a^0, \mathbf{a})$ is called a pseudovector if its *spatial* components constitute a pseudovector $P(\mathbf{a}) = \mathbf{a}$; p is a pseudoscalar if it goes into minus itself under spatial inversions $P(p) = -p$.

TABLE 4.5 BEHAVIOR OF SCALARS AND VECTORS UNDER THE PARITY TRANSFORMATION, P

Scalar	:	$P(s) = s$
Pseudoscalar	:	$P(p) = -p$
Vector (or polar vector)	:	$P(\mathbf{v}) = -\mathbf{v}$
Pseudovector (or axial vector)	:	$P(\mathbf{a}) = \mathbf{a}$

If you apply the parity operator *twice,* of course, you're right back where you started:

$$P^2 = I \tag{4.52}$$

(The parity group, then, consists of just two elements: I and P.) It follows that the eigenvalues of P are ± 1 (see Problem 4.34). For example, scalars and pseudovectors have eigenvalue $+1$, whereas vectors and pseudoscalars have eigenvalue -1. The hadrons are eigenstates of P and can be classified according to their eigenvalue, just as they are classified by spin, charge, isospin, strangeness, and so on. According to Quantum Field Theory the parity of a fermion (half-integer spin) must be *opposite* to that of the corresponding antiparticle, while the parity of a boson (integer spin) is the *same* as its antiparticle. We take the *quarks* to have *positive* intrinsic parity, so the antiquarks are negative.* The parity of a composite system in its ground state is the *product* of the parities of its constituents (we say that parity is a "multiplicative" quantum number, in contrast to charge, strangeness, and so on, which are "additive".† Thus the baryon octet and decuplet have positive parity, $[(+1)^3]$, whereas the pseudoscalar and vector meson nonets have negative parity $[(-1)(+1)]$. (The prefix "pseudo" tells you the *parity* of the particles.) For excited states there is an extra factor of $(-1)^l$, where l is the orbital angular momentum.[14] Thus, in general, the mesons carry a parity of $(-1)^{l+1}$ (see Table 4.6). Meanwhile, the *photon* is a *vector* particle (it is represented by the vector potential A^μ). Its spin is 1 and its intrinsic parity is -1.‡

The mirror symmetry of strong and electromagnetic interactions means that parity is conserved in all such processes. Originally, everyone took it for

* This choice is completely arbitrary; we could just as well do it the other way around. Indeed, in principle we could assign positive parity to some quark flavors and negative to others. This would lead to a different set of hadronic parities, but the *conservation* of parity would still hold. The rule stated in the text is obviously the *simplest,* and it leads to the conventional assignments.

† There is less to this distinction than meets the eye; in a sense, it results from a notational anomaly. Scrupulous consistency would require that we write the parity operator in exponential form, $P = e^{i\pi K}$, with the operator K playing a role analogous to, say, spin [see eq. (4.28)]. The eigenvalues of K would be 0 and 1, corresponding to $+1$ and -1 for P, and multiplication of parities would correspond to addition, mod 2, of K.

‡ Incidentally, you will notice here a certain parallel between parity and angular momentum. Just as angular momentum comes in two varieties—*orbital* (associated with the particle's motion) and *spin* (an intrinsic property of the particle itself), so too does parity. There is the factor $(-1)^l$ (associated with the spatial configuration) and the **intrinsic** parities of the constituents. But whereas the net angular momentum of a composite system is the *sum* of the individual terms, the net *parity* is the *product*.

TABLE 4.6 QUANTUM NUMBERS OF MESONS COMPOSED OF *u*, *d*, AND *s* QUARKS

Orbital ang. mom.	Net spin	J^{PC}	Observed Nonet $I = 1$	$I = \frac{1}{2}$	$I = 0$	Typical mass MeV/c^2
$l = 0$	$s = 0$	0^{-+}	π	K	η, η'	500
	$s = 1$	1^{--}	ρ	K^*	ω, ϕ	800
$l = 1$	$s = 0$	1^{+-}	B	Q_2	$H, ?$	1250
	$s = 1$	0^{++}	δ	κ	ϵ, S^*	1150
		1^{++}	A_1	Q_1	D, E	1300
		2^{++}	A_2	K^*	f, f'	1400

granted that the same goes for the weak interactions as well. But a disturbing paradox arose in the early fifties, known as the "tau–theta puzzle." Two strange mesons, called at the time τ and θ, appeared to be identical in every respect—same mass, same spin, same charge, and so on—except that one of them decayed into two pions and the other into three pions, states of opposite parity:

$$\theta^+ \longrightarrow \pi^+ + \pi^0 \qquad (P = +1)$$

$$\tau^+ \longrightarrow \begin{Bmatrix} \pi^+ + \pi^0 + \pi^0 \\ \pi^+ + \pi^+ + \pi^- \end{Bmatrix} \qquad (P = -1) \tag{4.53}$$

It seemed most peculiar that two otherwise identical particles should carry different parity. The alternative, suggested by Lee and Yang in 1956 was that τ and θ are really the *same particle* (now known as the K^+), and parity is simply not conserved in one of the decays. This idea prompted their search for evidence of parity invariance in the weak interactions and, when they found none, to their proposal for an experimental test.

4.7 CHARGE CONJUGATION

Classical electrodynamics is invariant under a change in the sign of all electric charges; the potentials and fields reverse their signs, but there is a compensating charge factor in the Lorentz law, so the forces still come out the same. In elementary particle physics we introduce an operation that generalizes this notion of "changing the sign of the charge"—it's called *charge conjugation, C,* and it converts each particle into its antiparticle:

$$C|p\rangle = |\bar{p}\rangle \tag{4.54}$$

"Charge conjugation" is something of a misnomer, for C can be applied to a neutral particle, such as the neutron (yielding an antineutron), and it changes the sign of *all* the "internal" quantum numbers—charge, baryon number, lepton number, strangeness, charm, beauty, truth—while leaving mass, energy, momentum, and spin untouched.

As with P, application of C *twice* brings us back to the original state:

$$C^2 = I \tag{4.55}$$

and hence the eigenvalues of C are ± 1. Unlike P, however, most of the particles in nature are clearly not eigenstates of C. For if $|p\rangle$ is an eigenstate of C, it follows from equation (4.54) that

$$C|p\rangle = \pm|p\rangle = |\bar{p}\rangle \tag{4.56}$$

so $|p\rangle$ and $|\bar{p}\rangle$ differ at most by a sign, which means that they represent the same physical state. Thus *only those particles that are their own antiparticles can be eigenstates of* C. This leaves us the photon, as well as all those mesons which lie at the center of their Eightfold Way diagrams: π^0, η, η', ρ^0, ϕ, ω, ψ, and so on. Because the photon is the quantum of the electromagnetic field, which changes sign under C, it makes sense that the photon's "charge conjugation number" is -1. It can be shown[15] that a system consisting of a spin-$\frac{1}{2}$ particle and its antiparticle in a configuration with orbital angular momentum l and total spin s constitutes an eigenstate of C, with eigenvalue $(-1)^{l+s}$. According to the quark model, the mesons in question are of precisely this form: For the pseudo-scalars, $l = 0$ and $s = 0$, so $C = +1$; for the vectors, $l = 0$ and $s = 1$, so $C = -1$. (Often, as in Table 4.6, C is listed as though it were a valid quantum number for the entire supermultiplet; in *fact* it pertains only to the central members.)

C is a multiplicative quantum number, and, like parity, it is conserved in the strong and electromagnetic interactions. Thus, for example, the π^0 decays into two photons:

$$\pi^0 \rightarrow \gamma + \gamma \tag{4.57}$$

($C = +1$ before and after), but it cannot decay into three photons. (For a system of n photons, $C = (-1)^n$.) Similarly, the ω goes to $\pi^0 + \gamma$, but never to $\pi^0 + 2\gamma$. In the strong interactions, charge conjugation invariance requires, for example, that the energy distribution of the charged pions in the reaction

$$p + \bar{p} \rightarrow \pi^+ + \pi^- + \pi^0 \tag{4.58}$$

should (on the average) be identical.[16] On the other hand, charge conjugation is *not* a symmetry of the weak interactions: when applied to a neutrino (left-handed, remember), C gives a left-handed *anti*neutrino, which does not exist. So the charge-conjugated version of any process involving neutrinos is certainly *not* a possible physical process. And purely hadronic weak interactions also show violations of C as well as P.

Because so few particles are eigenstates of C, its direct application in elementary particle physics is rather limited. Its power can be somewhat extended, if we confine our attention to the strong interactions, by combining it with an appropriate isospin transformation. Rotation by 180° about the number 2 axis in isospin space* will carry I_3 into $-I_3$, converting, for instance, a π^+ into a π^-.

* Some authors use the number 1 axis. Obviously, any axis in the 1–2 plane will do the job.

If we then apply the charge conjugation operator, we come back to π^+. Thus the charged pions are eigenstates of this combined operator, even though they are not eigenstates of C alone. For some reason the product transformation is called "G-parity":

$$G = CR_2, \quad \text{where} \quad R_2 = e^{i\pi I_2} \tag{4.59}$$

All mesons that carry no strangeness (or charm, beauty, or truth) are eigenstates of G;* for a multiplet of isospin I the eigenvalue is given (see Problem 4.36) by

$$G = (-1)^I C \tag{4.60}$$

where C is the charge conjugation number of the neutral member. For a single pion, $G = -1$, and for a state with n pions

$$G = (-1)^n \tag{4.61}$$

This is a very handy result, for it tells you how many pions can be emitted in a particular decay. For example, the ρ mesons, with $I = 1$, $C = -1$, and hence $G = +1$, can go to *two* pions, but not to three, whereas the ϕ, the ω, and the ψ can go to *three,* but not to two.

4.8 *CP* VIOLATION

As we have seen, the weak interactions are not invariant under the parity transformation P; the cleanest evidence for this is the fact that the antimuon emitted in pion decay

$$\pi^+ \longrightarrow \mu^+ + \nu_\mu \tag{4.62}$$

always comes out left-handed. Nor are the weak interactions invariant under C, for the charge-conjugated version of reaction (4.62) would be

$$\pi^- \longrightarrow \mu^- + \bar{\nu}_\mu \tag{4.63}$$

with a left-handed muon, whereas in fact the muon always comes out *right*-handed. However, if we *combine* the two operations we're back in business: CP turns the left-handed antimuon into a *right*-handed muon, which is exactly what we observe in nature. Many people who had been shocked by the fall of parity were consoled by this realization; perhaps it was the combined operation that our intuition had been talking about all along—maybe what we should have *meant* by the "mirror image" of a right-handed electron was a left-handed *positron.*† If we had defined parity from the start to be what we now call CP, the

* K^+, for example, is not an eigenstate of G, for R_2 takes it to K^0, and C takes that to $\bar{K}^0$. The idea could be extended to the K's, by using an appropriate $SU(3)$ transformation in place of R_2, but since $SU(3)$ is not a very good symmetry of the strong forces, there is little advantage in doing so.

† Incidentally, we could perfectly well take electric charge to be a pseudoscalar in classical electrodynamics; **E** becomes a pseudovector and **B** a vector, but the *results* are all the same. It is really a matter of taste whether you say the mirror image of a plus charge is positive or negative.

trauma of parity violation might have been avoided (or at any rate postponed). It is too late to change the terminology now, but at least this pacifies our visceral sense that the world "ought" to be left–right symmetric.

CP invariance has bizarre implications for the neutral K mesons, as was first pointed out in a classic paper by Gell-Mann and Pais.[17] They noted that the K^0, with strangeness $+1$, can turn into its antiparticle $\bar{K}^0$, strangeness -1

$$K^0 \rightleftarrows \bar{K}^0 \tag{4.64}$$

through a second-order weak interaction we now represent by the diagrams in Figure 4.12. (The possibility of such an interconversion between two particles is almost unique to the neutral kaon system; among the "stable" particles only $D^0\bar{D}^0$ and $B^0\bar{B}^0$ share the property. See Problem 4.38.) As a result, the particles we normally observe in the laboratory are not K^0 and $\bar{K}^0$, but rather some linear combination of the two. In particular, we can form eigenstates of *CP*, as follows. Because the K's are pseudoscalars

$$P|K^0\rangle = -|K^0\rangle, \qquad P|\bar{K}^0\rangle = -|\bar{K}^0\rangle \tag{4.65}$$

On the other hand, from equation (4.54)

$$C|K^0\rangle = |\bar{K}^0\rangle, \qquad C|\bar{K}^0\rangle = |K^0\rangle \tag{4.66}$$

Accordingly

$$CP|K^0\rangle = -|\bar{K}^0\rangle, \qquad CP|\bar{K}^0\rangle = -|K^0\rangle \tag{4.67}$$

and hence the (normalized) eigenstates of *CP* are

$$|K_1\rangle = (1/\sqrt{2})(|K^0\rangle - |\bar{K}^0\rangle) \quad \text{and} \quad |K_2\rangle = (1/\sqrt{2})(|K^0\rangle + |\bar{K}^0\rangle) \tag{4.68}$$

with

$$CP|K_1\rangle = |K_1\rangle \quad \text{and} \quad CP|K_2\rangle = -|K_2\rangle \tag{4.69}$$

Assuming CP is conserved in the weak interactions, K_1 can only decay into a state with $CP = +1$, whereas K_2 must go to a state with $CP = -1$. Typically,

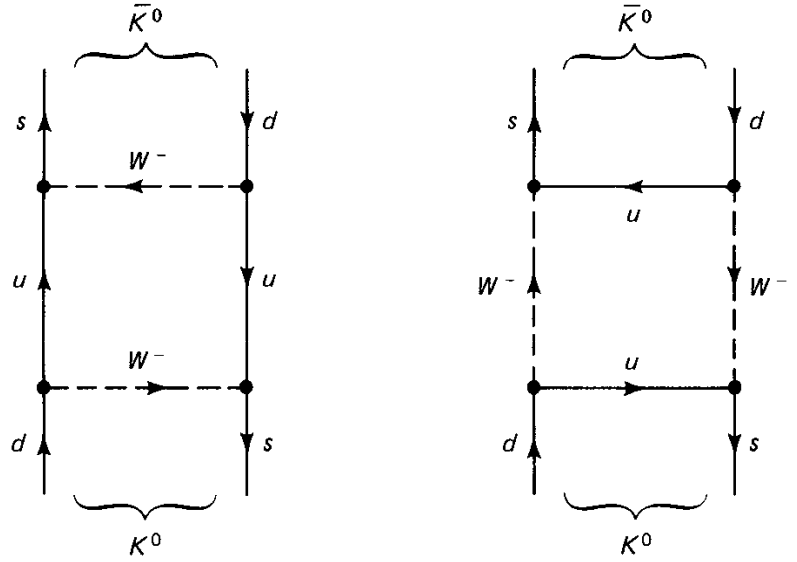


Figure 4.12 Feynman diagrams contributing to $K^0 \rightleftarrows \bar{K}^0$. (There are others, including those with one or both u quarks replaced by c or t.)

neutral kaons decay into two or three pions. But we have already seen that the two-pion configuration carries a parity of $+1$, and the three-pion system has $P = -1$ [eq. (4.53)]; both have $C = +1$. Conclusion: K_1 decays into two pions (never three); K_2 decays into three pions (never two):

$$K_1 \to 2\pi, \qquad K_2 \to 3\pi \tag{4.70}$$

Now, the 2π decay is much faster, because the energy released is greater. So if we start with a beam of K^0's

$$|K^0\rangle = (1/\sqrt{2})(|K_1\rangle + |K_2\rangle) \tag{4.71}$$

the K_1 component will quickly decay away, and down the line we shall have a beam of pure K_2's. Near the source we should see a lot of 2π events, but farther along we expect only 3π decays.

Well ... that's a lot to swallow. As Cronin put it, in a delightful memoir:[18]

> So these gentlemen, Gell-Mann and Pais, predicted that in addition to the short-lived K mesons, there should be long-lived K mesons. They did it beautifully, elegantly and simply. I think theirs is a paper one should read sometime just for its pure beauty of reasoning. It was published in the *Physical Review* in 1955. A very lovely thing! You get shivers up and down your spine, especially when you find you understand it. At the time, many of the most distinguished theoreticians thought this prediction was really baloney.

But it wasn't baloney, and in 1956 Lederman and his collaborators discovered the K_2 meson at Brookhaven.[19] Experimentally, the two lifetimes are

$$\begin{aligned} \tau_1 &= 0.89 \times 10^{-10} \text{ sec} \\ \tau_2 &= 5.2 \times 10^{-8} \text{ sec} \end{aligned} \tag{4.72}$$

so the K_1's are mostly gone after a few centimeters, whereas the K_2's can travel many meters. Notice that K_1 and K_2 are not antiparticles of one another, like K^0 and $\bar{K}^0$; rather, each is its own antiparticle ($C = -1$ for K_1 and $C = +1$ for K_2). They differ ever-so-slightly in mass; experiments give[20]

$$m_2 - m_1 = 3.5 \times 10^{-6} \text{ eV} \tag{4.73}$$

The neutral kaon system adds a subtle twist to the old question, "What is a particle?" Kaons are typically *produced* by the strong interactions, in eigenstates of strangeness (K^0 and $\bar{K}^0$), but they *decay* by the *weak* interactions, as eigenstates of CP (K_1 and K_2). Which, then, is the "real" particle? If we hold that a "particle" must have a unique lifetime, then the "true" particles are K_1 and K_2.* But we need not be so dogmatic. In practice, it is sometimes more convenient to use one set, and sometimes the other. The situation is in many ways analogous to polarized light. Linear polarization can be regarded as a superposition of left-circular polarization and right-circular polarization. If you imagine a medium that preferentially absorbs right-circularly polarized light, and shine on it a linearly polarized beam, it will become progressively more left-circularly polarized as it

* This, incidentally, was the position advocated by Gell-Mann and Pais.

passes through the material, just as a K^0 beam turns into a K_2 beam. But whether you choose to analyze the process in terms of states of linear or circular polarization is largely a matter of taste.

The neutral kaons provide a perfect experimental system for testing *CP* invariance. By using a long enough beam, we can produce an arbitrarily pure sample of the long-lived species. If at this point we observe a 2π decay, we shall know that *CP* has been violated. Such an experiment was reported by Cronin and Fitch in 1964.[21] At the end of a beam 57 feet long, they found 45 two-pion events in a total of 22,700 decays. That's a tiny fraction (roughly one in 500), but unmistakable evidence of *CP* violation. Evidently the long-lived neutral kaon is *not* a perfect eigenstate of *CP* after all, but contains a small admixture of K_1:

$$|K_L\rangle = \frac{1}{\sqrt{1 + |\epsilon|^2}} (|K_2\rangle + \epsilon|K_1\rangle) \tag{4.74}$$

The coefficient ϵ is a measure of nature's departure from perfect *CP* invariance; experimentally its magnitude is about 2.3×10^{-3}.

Although the effect is small, and has never been observed outside the neutral kaon system, *CP* violation poses a far deeper problem than parity ever did. The nonconservation of parity was quickly incorporated into the theory of weak interactions (in fact, *part* of the "new" theory—Weyl's equation for the neutrino—had been "waiting in the wings" for many years). Parity violation was easier to handle precisely because it was such a *large* effect: *all* neutrinos are left-handed, not just 50.01% of them. Parity is, in this sense, *maximally* violated, in the weak interactions. By contrast, *CP* violation is a small effect on any scale, and no one has yet found a "natural" way to accommodate it.*

The Fitch–Cronin experiment destroyed the last hope for any form of exact mirror symmetry in nature. And subsequent study of the semileptonic decays of K_L has revealed even more dramatic evidence for *CP* violation. Although 34% of all K_L's decay by the 3π mode we have discussed, some 39% go to

$$\begin{aligned} &\text{(a)} \quad \pi^+ + e^- + \bar{\nu}_e \\ \text{or} \quad &\text{(b)} \quad \pi^- + e^+ + \nu_e \end{aligned} \tag{4.75}$$

Notice that *CP* takes (a) into (b), so if *CP* were conserved, and K_L were a pure eigenstate, (a) and (b) would be equally probable. But experiments show[22] that K_L decays more often into a positron than into an electron, by a fractional amount 3.3×10^{-3}. Here for the first time is a process that makes an absolute distinction between matter and antimatter, and provides an unambiguous, convention-free definition of positive charge: *it is the charge carried by the lepton preferentially produced in the decay of the long-lived neutral K meson.* The fact that *CP* violation permits unequal treatment of particles and antiparticles suggests

* A complex phase factor δ in the Kobayashi–Maskawa matrix is a convenient vehicle for introducing *CP* violation into the Standard Model. Indeed, it was this that led Kobayashi and Maskawa to propose a third generation of quarks, before even charm was discovered. At present, however, δ is an arbitrary input parameter, which (like the other elements in the KM matrix) nobody knows how to *calculate.*

that it may be responsible for the dominance of matter over antimatter in the universe.[23]

4.9 TIME REVERSAL AND THE TCP THEOREM

Suppose we made a movie of some physical process, say, an elastic collision of two billiard balls. If we ran the movie backward, would it depict a possible physical process, or would the viewer be able to say with certainty "No, no. that's impossible; the film must be running in reverse"? In the case of classical elastic collisions the "time-reversed" process *is* perfectly possible. (To be sure. if we put a *lot* of billiard balls in the picture the backward version may be highly *improbable;* we would be *surprised* to see the balls gather themselves together into a perfect triangle, with a single cue ball rolling away, and we would strongly *suspect* that the film had been reversed. But that's just because we know it would be extraordinarily difficult to set up the necessary starting conditions, such that all the balls would roll together at just the right speeds and in just the right directions. Thus the initial conditions may give us a clue to the "arrow of time," but the laws governing the collisions themselves work just as well forward as backward.) Until fairly recently it was taken for granted that all elementary particle interactions share this time-reversal invariance. But with the downfall of parity it was natural to wonder whether time reversal was really so sacred.[24]

As it turns out, time reversal is a lot harder to test than P or C. In the first place, whereas all particles are eigenstates of P, and many are eigenstates of C. none is an eigenstate of T (the "time-reversal operator," which runs the movie backward).* So we cannot check the "conservation of T" simply by multiplying numbers, the way we can for P and C. The most direct test would be to take a particular reaction (say, $n + p \rightarrow d + \gamma$), and run it in reverse ($d + \gamma \rightarrow n + p$). For corresponding conditions of momentum, energy, and spin, the reaction rate should be the same in either direction. (This is called the "principle of detailed balance," and it follows directly from time-reversal invariance.) Such tests work fine for the strong and electromagnetic interactions, and a variety of processes have been checked. The results have always been negative (no evidence of T violation), but this is hardly surprising. On the basis of our experience with P and C we expect to see a failure of time reversal in the weak interactions, if anywhere. Unfortunately, inverse-reaction experiments are tough to do in the weak interactions. Take, for instance, the typical weak decay $\Lambda \rightarrow p^+ + \pi^-$. The inverse reaction would be $p^+ + \pi^- \rightarrow \Lambda$, but we are never going to see such a process, because the strong interaction of the proton and the pion will totally swamp the feeble weak interaction. To avoid strong and electromagnetic contamination, we might go to a neutrino process. But it is notoriously difficult to do accurate measurements on neutrinos, and here we are presumably looking for a very small effect. In practice, therefore, the critical tests of T invariance involve careful measurements of quantities that should be precisely zero if T is a perfect symmetry. The classic example is a static electric dipole moment on

* A particle can be identical to its mirror image, and, if it's neutral, to its own antiparticle. but it can't very well be identical to itself-going-backward-in-time (at least, not if anything ever happens to it).

an elementary particle.* Probably the most sensitive experiment to date is Ramsey's upper limit on the electric dipole moment of the neutron:[25]

$$|\mathbf{d}| < e \cdot (6 \times 10^{-25} \text{ cm}) \tag{4.76}$$

where e is the charge of the proton. As of 1985, *no* experiment has shown direct evidence of T violation.

Nevertheless, there is a compelling reason to believe that time reversal cannot be a perfect symmetry of nature. It comes from the so-called *TCP* theorem, one of the deepest results of quantum field theory.[26] Based only on the most general assumptions—Lorentz invariance, quantum mechanics, and the idea that interactions are carried by fields—the *TCP* theorem states that the combined operation of time reversal, charge conjugation, and parity (in any order) is an exact symmetry of *any* interaction. It is simply *impossible to construct* a quantum field theory in which the product *TCP* is violated. If, as the Fitch–Cronin experiment demonstrated, *CP* is violated, there must be a compensating violation of T. Of course, like any assertion of impossibility, the *TCP* theorem may just be a measure of our lack of imagination; it must be tested in the laboratory, and that is one reason it is so important to look for independent evidence of T violation. But the *TCP* theorem has other implications that are also subject to experimental verification: If the theorem is correct, every particle must have precisely the same mass and lifetime as its antiparticle.† Measurements have been made on a number of particle–antiparticle pairs; the most sensitive test to date is the $K^0 - \bar{K}^0$ mass difference, which, as a fraction of the K^0 mass, is known to be less than 6×10^{-19}. So the *TCP* theorem is on extremely firm ground theoretically, and it is relatively secure experimentally. Indeed, as one prominent theorist has put it, if a departure is ever found, "all hell breaks loose."

REFERENCES AND NOTES

1. See, for example, F. Halzen and A. D. Martin, *Quarks and Leptons* (New York: Wiley, 1984) Section 14.2. For a more complete (if somewhat dated) discussion, see E. L. Hill, *Rev. Mod. Phys.* **23**, 253 (1951).
2. If this fact is not familiar, you might want to look at D. Halliday and R. Resnick, *Physics,* 3d Ed. Part I, (New York: Wiley, 1977), p. 221.
3. See, for example, E. Merzbacher, *Quantum Mechanics,* 2d Ed. (New York: Wiley, 1970), Chapter 16, Sections 6 and 7.
4. See ref. 3, Chapter 12, Sections 5 and 6.
5. W. Heisenberg, *Z. Phys.* **77**, 1 (1932). An English translation of this classic paper appears in D. M. Brink, *Nuclear Forces* (Elmsford, NY: Pergamon, 1965).
6. V. B. Fliagin et al., *Sov. Phys. JETP* **35**(8), 592 (1959).

* For an elementary particle, the dipole moment **d**, would have to point along the axis of the spin, **s**; there is no other direction available. But **d** is a vector, whereas **s** is a pseudovector, so a nonzero dipole moment would imply violation of P. Similarly, **s** changes sign under time reversal, but **d** does not, so a nonzero **d** would also (and more interestingly) mean violation of T. For further details see ref. 25.

† This would also follow from C invariance. However, since we know that the latter is violated, it is significant that the equality of masses and lifetimes (also magnetic moments, incidentally, although they have opposite *signs*) follows from the far weaker assumption of *TCP* symmetry.

7. H. L. Anderson, E. Fermi, E. A. Long, and D. E. Nagle, *Phys. Rev.* **85,** 936 (1952).
8. A. De Rujula, M. Georgi, and S. L. Glashow, *Phys. Rev.* **D12,** 147 (1975); S. Weinberg, *Trans. N. Y. Acad. Sci., Series II,* **38,** 185 (1977); S. Gasiorowicz and J. L. Rosner, *Am. J. Phys.* **49,** 962 (1981); J. Gasser and H. Leutwyler, *Phys. Rep.* **87,** 78 (1982).
9. A qualitatively plausible mechanism is suggested by the "MIT Bag Model". Free quarks of mass m, confined within a spherical shell of radius R, are found to have an effective mass $m_{\text{eff}} = \sqrt{m^2 + (\hbar x/Rc)^2}$, where x is a dimensionless number around 2.5. Using the radius of the proton (say, 1.5×10^{-13} cm) for R, we obtain $m_{\text{eff}} = 330$ MeV/c^2 for the up and down quarks. See F. E. Close, *An Introduction to Quarks and Partons* (London: Academic, 1979), Section 18.1.
10. T. D. Lee and C. N. Yang, *Phys. Rev.* **104,** 254 (1956).
11. C. S. Wu et al., *Phys. Rev.* **105,** 1413 (1957). In the interest of clarity I am ignoring the formidable technical difficulties involved in this experiment. To keep the cobalt nuclei aligned, the sample had to be maintained at a temperature of less than 1° K for 10 minutes. Small wonder that no earlier experiments had stumbled on evidence of parity violation.
12. See, for example, Pauli's letter to Weisskopf in W. Pauli, *Collected Scientific Papers,* Vol. I, R. Kronig and V. F. Weisskopf, eds. (New York: Wiley-Interscience, 1964). p. xii; *Scientific American* (March 1957), p. 154; P. Morrison, *Scientific American* (April 1957).
13. G. Backenstoss et al., *Phys. Rev. Lett.* **6,** 415 (1961); M. Bardon et al., *Phys. Rev. Lett.* **7,** 23 (1961).
14. This comes from the angular part of the spatial wave function, $Y_l^m(\theta, \phi)$. See, for example, ref. 3, p. 186 (or Problem 5.3 below).
15. D. H. Perkins, *Introduction to High-Energy Physics,* 2d Ed. (Reading, MA: Addison-Wesley, 1982), p. 99.
16. C. Baltay et al., *Phys. Rev. Lett.* **15,** 591 (1965).
17. M. Gell-Mann and A. Pais, *Phys. Rev.* **97,** 1387 (1955). This paper was written before the overthrow of parity, but the essential idea remains unchanged if we substitute *CP* for their *C*. Of course, they didn't draw a quark diagram like Figure 4.12; they based their argument for equation (4.64) on the fact that both K^0 and $\bar{K}^0$ can decay into $\pi^+ + \pi^-$, so that $K^0 \leftrightarrow \pi^+ + \pi^- \leftrightarrow \bar{K}^0$.
18. J. W. Cronin and M. S. Greenwood, *Phys. Today* (July 1982), p. 38. Cronin uses an unorthodox sign convention, putting a -1 into our equation (4.66), but the physics is still the same.
19. K. Lande et al., *Phys. Rev.* **103,** 1901 (1956).
20. The detection of so minute a mass difference is itself a fascinating story. See, for example, ref. 15, Sect. 16.13.1.
21. J. H. Christenson et al., *Phys. Rev. Lett.* **13,** 138 (1964).
22. S. Gjesdal et al., *Phys. Lett.* **52B,** 113 (1974); S. Bennett et al., *Phys. Rev. Lett.* **19,** 993 (1967); D. Dorfan et al., *Phys. Rev. Lett.* **19,** 987 (1967).
23. F. Wilczek, *Scientific American* (December 1980).
24. Excellent informal treatments of time reversal are O. E. Overseth, *Scientific American* (October 1969) and R. G. Sachs, *Science,* **176,** 587 (1972).
25. N. F. Ramsey, *Rep. Prog. Phys.* **45,** 95 (1982).
26. The *TCP* Theorem was discovered by J. Schwinger and G. Lüders, and perfected by W. Pauli, *Niels Bohr and the Development of Physics,* W. Pauli, ed., (New York: McGraw-Hill, 1955). At first no one paid much attention to the *TCP* theorem, because at that time everyone thought *T, C,* and *P* were all perfect symmetries individually. It was only with the fall of parity, and especially with the failure of *CP,* that the importance of this theorem was fully appreciated.

PROBLEMS

4.1. Prove that I, R_+, R_-, R_a, R_b, and R_c are *all* the symmetries of the equilateral triangle. [*Hint:* One way to do this is to label the three corners, as in Figure 4.2. A given symmetry operation carries A into the position formerly occupied by A, B, or C. If $A \rightarrow A$, then either $B \rightarrow B$ and $C \rightarrow C$, or else $B \rightarrow C$ and $C \rightarrow B$. Take it from there.]

4.2. Construct a "multiplication table" for the triangle group, filling in the blanks on the following diagram:

	I	R_+	R_-	R_a	R_b	R_c
I						
R_+						
R_-						
R_a						
R_b						
R_c						

[In row i, column j, put the product R_iR_j.] Is this an Abelian group? How can you tell, just by looking at the multiplication table?

4.3. The triangle group, like any other group, has a trivial one-dimensional representation. It *also* has a *non*trivial, one-dimensional representation, in which the elements are *not* all represented by 1. Work out this second one-dimensional representation. That is, figure out what number (1×1 matrix) each group element is represented by. Is this representation faithful?

4.4. Work out the symmetry group of a square. How many elements does it have? Construct the multiplication table, and determine whether or not the group is Abelian.

4.5. **(a)** Show that the set of all unitary $n \times n$ matrices constitutes a group. (To prove closure, for instance, you must show that the product of two unitary matrices is itself unitary.)

(b) Show that the set of all $n \times n$ unitary matrices with determinant 1 constitutes a group.

(c) Show that $O(n)$ is a group.

(d) Show that $SO(n)$ is a group.

4.6. Consider a vector **A** in two dimensions. Suppose its components with respect to Cartesian axes x, y, are (a_x, a_y). What are its components (a'_x, a'_y) in a system x', y' which is rotated, counterclockwise, by an angle θ, with respect to x, y? Express your answer in the form of a 2×2 matrix $R(\theta)$:

$$\begin{pmatrix} a'_x \\ a'_y \end{pmatrix} = R \begin{pmatrix} a_x \\ a_y \end{pmatrix}$$

Show that R is an orthogonal matrix. What is its determinant? The set of *all* such rotations constitutes a group; what is the name of this group? By multiplying the matrices, show that $R(\theta_1)R(\theta_2) = R(\theta_1 + \theta_2)$; is this an Abelian group?

4.7. Consider the matrix $\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$. Is it in the group $O(2)$? How about $SO(2)$? What is its effect on the vector A of Problem 4.6? Does it describe a possible rotation of the plane?

4.8. Suppose we interpret the electron literally as a classical solid sphere of radius r, mass m, spinning with angular momentum $\frac{1}{2}\hbar$. What is the speed, v, of a point on its "equator"? Experimentally, it is known that r is less than 10^{-16} cm. What is the corresponding equatorial speed? What do you conclude from this?

4.9. When you're adding angular momenta, using equation (4.12), it is useful to check your results by counting the number of states before and after the addition. For instance, in Example 4.1 we had two quarks to begin with, each could have $m_s = +\frac{1}{2}$ or $m_s = -\frac{1}{2}$, so there were four possibilities in all. *After* adding the spins, we had one combination with spin 1 (hence $m_s = 1$, 0, or -1) and one with spin 0 ($m_s = 0$)—again, four states in all.

(a) Apply this check to Example 4.2.

(b) Add angular momenta 2, 1, and $\frac{1}{2}$: list the possible values of the total angular momentum, and check your answer by counting states.

4.10. Show that the "original" beta-decay reaction $n \rightarrow p + e$ would violate conservation of angular momentum (all three particles have spin $\frac{1}{2}$). If you were Pauli, proposing that the reaction is really $n \rightarrow p + e + \bar{\nu}_e$, what spin would you assign to the neutrino?

4.11. In the decay $\Delta^{++} \rightarrow p + \pi^+$, what are the possible values of the orbital angular momentum quantum number, l, in the final state?

4.12. An electron in a hydrogen atom is in a state with orbital angular momentum quantum number $l = 1$. If the *total* angular momentum quantum number j is $\frac{3}{2}$, and the z component of total angular momentum is $\frac{1}{2}\hbar$, what is the probability of finding the electron with $m_s = +\frac{1}{2}$?

4.13. Suppose you had two particles of spin 2, each in a state with $S_z = 0$. If you measured the *total* angular momentum of this system, given that the *orbital* angular momentum is zero, what values might you get, and what is the probability of each? Check that they add up to 1.

4.14. Suppose you had a particle of spin $\frac{3}{2}$, and another of spin 2. If you knew that their orbital angular momentum was zero, and that the *total* spin of the composite system was $\frac{5}{2}$, and its z component was $-\frac{1}{2}$, what values might you get for a measurement of S_z on the spin-2 particle? What is the probability of each? Check that they add up to 1.

4.15. Check that $\chi_\pm$, equation (4.22), are normalized eigenvectors of $\hat{S}_x$, [equation (4.21)]. and find the associated eigenvalues.

4.16. Show that $|a|^2 + |b|^2 = 1$, [equation (4.24)], provided the spinor in question is normalized [equation (4.20)].

4.17. **(a)** Find the eigenvalues and normalized eigenspinors of $\hat{S}_y$, [equation (4.21)].

(b) If you measured S_y on an electron in the state $\begin{pmatrix} \alpha \\ \beta \end{pmatrix}$, what values might you get, and what is the probability of each?

4.18. Suppose an electron is in the state $\begin{pmatrix} 1/\sqrt{5} \\ 2/\sqrt{5} \end{pmatrix}$.

(a) If you measured S_x, what values might you get, and what is the probability of each?

(b) If you measured S_y, what values might you get, and what is the probability of each?
(c) If you measured S_z, what values might you get, and what is the probability of each?

4.19. (a) Show that $\sigma_x^2 = \sigma_y^2 = \sigma_z^2 = 1$. ("1" here really means the 2×2 unit matrix; if no matrix is specified, the unit matrix is understood.)
(b) Show that $\sigma_x\sigma_y = i\sigma_z$, $\sigma_y\sigma_z = i\sigma_x$, $\sigma_z\sigma_x = i\sigma_y$.
These results are neatly summarized in the formula

$$\sigma_i\sigma_j = \delta_{ij} + i\epsilon_{ijk}\sigma_k$$

(summation over k implied), where δ_{ij} is the Kronecker delta:

$$\delta_{ij} = \begin{cases} 1, & \text{if } i = j \\ 0, & \text{otherwise} \end{cases}$$

and ϵ_{ijk} is the Levi-Civita symbol:

$$\epsilon_{ijk} = \begin{cases} 1, & \text{if } ijk = 123, 231, \text{ or } 312 \\ -1, & \text{if } ijk = 132, 213, \text{ or } 321 \\ 0, & \text{otherwise} \end{cases}$$

4.20. Use the results of Problem 4.19 to show that
(a) The commutator, $[A, B] \equiv AB - BA$, of two Pauli matrices is $[\sigma_i, \sigma_j] = 2i\epsilon_{ijk}\sigma_k$.
(b) The *anticommutator*, $\{A, B\} \equiv AB + BA$, is $\{\sigma_i, \sigma_j\} = 2\delta_{ij}$.
(c) For any two vectors $\mathbf{a}$ and $\mathbf{b}$, $(\boldsymbol{\sigma}\cdot\mathbf{a})(\boldsymbol{\sigma}\cdot\mathbf{b}) = \mathbf{a}\cdot\mathbf{b} + i\boldsymbol{\sigma}\cdot(\mathbf{a}\times\mathbf{b})$.

4.21. (a) Show that $e^{i\pi\sigma_z/2} = i\sigma_z$.
(b) Find the matrix U representing a rotation by 180° about the y axis, and show that it converts "spin up" into "spin down", as we would expect.
(c) More generally, show that

$$U(\theta) = \cos\frac{\theta}{2} - i(\hat{\theta}\cdot\boldsymbol{\sigma})\sin\frac{\theta}{2}$$

where $U(\theta)$ is given by equation (4.28), θ is the magnitude of $\boldsymbol{\theta}$, and $\hat{\theta} = \boldsymbol{\theta}/\theta$. [*Hint:* Use Problem 4.20, part (c).]

4.22. (a) Show that U, in equation (4.28), is *unitary*.
(b) Show that $\det U = 1$.
[*Hint:* You can either do this directly (however, see footnote on page 115), or else use the results of Problem 4.21.]

4.23. The extension of everything in Section 4.4 to higher spin is relatively straightforward. For spin 1 we have three states ($m_s = +1, 0, -1$), which we may represent by column vectors:

$$\begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \quad \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, \quad \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$$

respectively. The only problem is to construct the 3×3 matrices $\hat{S}_x$, $\hat{S}_y$ and $\hat{S}_z$. The latter is easy:
(a) Construct $\hat{S}_z$ for spin 1.

To obtain $\hat{S}_x$ and $\hat{S}_y$ it is easiest to start with the "raising" and "lowering" operators, $S_\pm \equiv S_x \pm iS_y$, which have the property

$$S_\pm|sm\rangle = \hbar\sqrt{s(s+1) - m(m \pm 1)}\,|s(m \pm 1)\rangle$$

(b) Construct the matrices $\hat{S}_+$ and $\hat{S}_-$, for spin 1.
(c) Using (b), determine the spin-1 matrices $\hat{S}_x$ and $\hat{S}_y$.
(d) Carry out the same construction for spin $\frac{3}{2}$.

4.24. Determine the isospin assignments $|I\,I_3\rangle$ for each of the following particles (refer to the Eightfold Way diagrams in Chapter 1): Ω^-, Σ^+, Ξ^0, ρ^+, η, $\bar{K}^0$.

4.25. **(a)** Check that the Gell-Mann–Nishijima formula works for the quarks u, d, and s.
(b) What are the appropriate isospin assignments, $|I\,I_3\rangle$, for the *anti*quarks, $\bar{u}$, $\bar{d}$, and $\bar{s}$? Check that your assignment is consistent with the Gell-Mann–Nishijima formula.
[Since Q, I_3, A, and S all *add*, when we combine quarks together, it follows that the Gell-Mann–Nishijima formula holds for all hadrons made out of u, d, s, $\bar{u}$, $\bar{d}$, and $\bar{s}$.]

4.26. **(a)** The Gell-Mann–Nishijima formula, equation (4.37), was proposed in the early fifties, which is to say long before the discovery of charm, beauty, or truth. Using the table of quark properties (on page 47), and the quark isospin assignments, equation (4.38), deduce the *general* formula expressing Q in terms of A, I_3, S, C, B, and T.
(b) Because u and d are the only quarks with nonzero isospin, it should be possible to express I_3 in terms of U ("upness") and D ("downness"). What's the formula? Likewise, express A in terms of the flavor numbers U, D, S, C, B, and T.
(c) Putting it all together, obtain the formula for Q in terms of the flavor numbers (that is, eliminate A and I_3 from your formula in part (a)). This final version represents the cleanest statement of the Gell-Mann–Nishijima formula, in the three-generation quark model.

4.27. For two isospin-$\frac{1}{2}$ particles, show that $\mathbf{I}^{(1)} \cdot \mathbf{I}^{(2)} = \frac{1}{4}$ in the triplet state and $-\frac{3}{4}$ in the singlet. [*Hint:* $\mathbf{I}_{\text{tot}} = \mathbf{I}^{(1)} + \mathbf{I}^{(2)}$; square both sides.]

4.28. **(a)** Referring to equations (4.47) and (4.48), work out all the πN scattering amplitudes, $\mathcal{M}_a$ through $\mathcal{M}_j$, in terms of $\mathcal{M}_1$ and $\mathcal{M}_3$.
(b) Generalize equation (4.49) to include all 10 cross sections.
(c) In the same way, generalize equation (4.50).

4.29. Find the ratio of the cross sections for the following reactions, when the total *CM* energy is 1232 MeV: (a) $\pi^- + p \to K^0 + \Sigma^0$; (b) $\pi^- + p \to K^+ + \Sigma^-$; (c) $\pi^- + p \to K^+ + \Sigma^+$.

4.30. What are the possible total isospins for the following reactions: (a) $K^+ + p \to \Sigma^+ + \pi^+$; (b) $K^- + p \to \Sigma^+ + \pi^-$. Find the ratio of the two cross sections, assuming one or the other isospin channel dominates.

4.31. On the graph in Figure 4.6 we see "resonances" as 1525, 1688, 1920, and 2190 (as well as the one at 1232). By comparing the two curves, determine the isospin of each resonance. The nomenclature is N (followed by the mass) for any state with $I = \frac{1}{2}$, and Δ for any state with $I = \frac{3}{2}$. Thus the nucleon is $N(939)$, and the "original" Δ is $\Delta(1232)$. Name the other resonances, and confirm your answers by looking in the Particle Data Booklet.

4.32. The Σ^{*0} can decay into $\Sigma^+ + \pi^-$, $\Sigma^0 + \pi^0$, or $\Sigma^- + \pi^+$ (also $\Lambda + \pi^0$, but we're not

concerned with that here). Suppose you observed 100 such disintegrations, how many would you expect to see of each type?

4.33. **(a)** The α particle is a bound state of two protons and two neutrons, that is, a ^{4}He nucleus. There is no isotope of hydrogen with an atomic weight of four (^4H), nor of lithium ^{4}Li. What do you conclude about the isospin of an α particle?
(b) The reaction $d + d \rightarrow \alpha + \pi^0$ has never been observed. Explain why not.
(c) Would you expect ^{4}Be to exist? How about a bound state of four neutrons?

4.34. **(a)** Using equation (4.52), prove that the eigenvalues of P are ± 1.
(b) Show that any function $f(x, y, z)$ can be expressed as the sum of an eigenfunction $f_+(x, y, z)$ with eigenvalue $+1$ and an eigenfunction $f_-(x, y, z)$ with eigenvalue -1. Construct the functions f_+ and f_-, in terms of f.

4.35. **(a)** Is the neutrino an eigenstate of P? If so, what is its intrinsic parity?
(b) Now that we know τ^+ and θ^+ are actually both the K^+, which of the decays in equation (4.53) actually violates parity conservation?

4.36. **(a)** Using the information in Table 4.6, determine the G parity of the following mesons: $\pi(140)$, $\rho(770)$, $\omega(783)$, $\eta(549)$, $\eta'(958)$, $\phi(1020)$, $f(1270)$.
(b) Show that $R_2|I0\rangle = (-1)^I|I0\rangle$, and use this result to justify equation (4.60).

4.37. The dominant decays of the η meson are

$$\eta \rightarrow 2\gamma \ (39\%), \qquad \eta \rightarrow 3\pi \ (56\%), \qquad \eta \rightarrow \pi\pi\gamma \ (5\%)$$

and it's classified as a "stable" particle, so evidently none of these is a purely strong interaction. Offhand, this seems odd, since at 549 MeV/c^2 the η has plenty of mass to decay strongly into 2π or 3π.
(a) Explain why the 2π mode is forbidden, for both strong and electromagnetic interactions.
(b) Explain why the 3π mode is forbidden as a *strong* interaction, but allowed as an *electromagnetic* decay.

4.38. For two particles to interconvert, $A \rightleftarrows B$, it is necessary that they have the same mass (which in practice means that they must be antiparticles of one another), the same charge, the same baryon and lepton numbers. In the Standard Model, with the usual three generations, show that A and B would have to be neutral mesons, and identify their possible quark contents. Which of these particles have been found, so far? Why doesn't the neutron mix with the antineutron, in the same way as the K^0 and $\bar{K}^0$ mix to produce K_1 and K_2? Why don't we see mixing of the neutral strange *vector* mesons K^{0*} and $\bar{K}^{0*}$?

4.39. Suppose you wanted to inform someone on a distant galaxy that humans have their hearts on the left side. How could you communicate this unambiguously, without sending an actual "handed" object (such as a corkscrew, a circularly polarized light beam, or a neutrino). For all you know their galaxy may be made of antimatter. You cannot afford to wait for any replies, but you are allowed to use English.

Chapter 5

Bound States

Most of this chapter is devoted to the nonrelativistic theory of two-particle bound states, with emphasis on hydrogen (e^-p^+), positronium (e^-e^+), charmonium ($c\bar{c}$), and bottomonium ($b\bar{b}$). This material is not used in subsequent chapters and may be skimmed, saved for later, or skipped entirely. Two tools from elementary quantum mechanics are essential: the Schrödinger equation and perturbation theory; readers unacquainted with these subjects should refer to the appropriate sections of an introductory quantum text (though the essential points are reviewed here, as they arise). The final two sections (5.8 and 5.9) concern relativistic light quark systems—the familiar mesons and baryons—about which far less can be said with confidence. I concentrate on the spin/flavor/color structure of the wave functions and develop a model for estimating masses and magnetic moments. This material does not involve the Schrödinger equation or perturbation theory and can be read independently of Sections 5.1 through 5.7. It will be used briefly later on, in the latter part of Chapter 9.

5.1 THE SCHRÖDINGER EQUATION FOR A CENTRAL POTENTIAL

Until recently, the theory of bound states played a rather minor role in particle physics. There was, of course, hydrogen (proton plus electron)—but this (the simplest atom) fell in the domain of atomic physics. And there was the deuteron (proton plus neutron)—but this (the simplest nontrivial nucleus) belonged more properly to nuclear physics. The quark model, though, changed everything. Suddenly the hadrons themselves were bound states—all mesons were two-quark systems, and all baryons were three-quark systems. With this discovery the theory of bound states became an important component of elementary particle physics.

The analysis of a bound state is simplest when the constituents travel at speeds substantially less than c, for then the apparatus of nonrelativistic quantum mechanics can be brought to bear. Such is the case for hydrogen and for hadrons made out of heavy quarks (c, b, and t). The more familiar light-quark states (made out of u, d, and s) are much more difficult to handle, because they are intrinsically relativistic, and quantum field theory (as currently practiced) is not well suited to bound-state problems. (Most of the techniques available start from the premise that the particles are initially free, and free again after some brief interaction (a collision, typically), whereas in a bound state, by its nature, the particles interact continuously over a long period.) At present, therefore, there exists a very rich theory of "charmonium" ($c\bar{c}$, the ψ meson system) and "bottomonium" ($b\bar{b}$, the Υ system), but comparatively little can be said about the excited states of $u\bar{u}$ (say) or $d\bar{d}$. How can we tell whether a given bound state is relativistic or not? The simplest criterion is the following: If the binding energy is small compared to the rest energies of the constituents, then the system is nonrelativistic.* For example, the binding energy of hydrogen is 13.6 eV, whereas the rest energy of an electron is 511,000 eV—this is clearly a nonrelativistic system. On the other hand, quark-quark binding energies are on the order of a few hundred MeV, which is about the same as the effective rest energy of u, d or s quarks, but substantially less than c, b, and t (see Table 4.4). So the light quark hadrons are relativistic, but the heavy quark systems are not.

For most of this chapter we shall restrict our attention to nonrelativistic bound states of two particles. To establish the framework for this discussion I must first review some basic quantum mechanics.[1] The principal foundation for nonrelativistic quantum theory is Schrödinger's equation. I cannot pretend to *derive* the Schrödinger equation—any more than one can *derive* Newton's laws of motion—it is, after all, an *axiom* of the theory. But I can perhaps make it *plausible,* as follows. In classical mechanics the sum of the kinetic energy ($\frac{1}{2}mv^2 = \mathbf{p}^2/2m$) and the potential energy ($V(x, y, z)$) is a constant—the *total* energy (E):†

$$\frac{1}{2m}\mathbf{p}^2 + V = E \tag{5.1}$$

In quantum mechanics the momentum $\mathbf{p}$ is replaced by the *momentum operator*

$$\mathbf{p} \to \frac{\hbar}{i}\nabla \qquad \left(p_x \to \frac{\hbar}{i}\frac{\partial}{\partial x}, p_y \to \frac{\hbar}{i}\frac{\partial}{\partial y}, p_z \to \frac{\hbar}{i}\frac{\partial}{\partial z}\right) \tag{5.2}$$

* In general, the total energy of a composite system is the sum of three terms: (i) the rest energy of the constituents, (ii) the kinetic energy of the constituents, and (iii) the potential energy of the configuration. The latter two are typically comparable in size (the precise relation is given by the virial theorem, which you can look up on any mechanics or quantum text). If the binding energy is much less than the constituent rest energies, so too is their kinetic energy, and hence the system is nonrelativistic. On the other hand, if the mass of the composite structure is substantially different from the sum of the rest masses of the constituents, then the kinetic energy is large and the system is relativistic.

† In a dissipative system (say, one with friction), the mechanical energy is not conserved. But on the microscopic level there is no such thing as a dissipative force, and what may have looked like nonconservation of energy on the macroscopic scale is simply a conversion of energy into an unseen form (typically heat).

and energy by the operator

$$E \rightarrow i\hbar \frac{\partial}{\partial t} \tag{5.3}$$

Equation (5.1) becomes a collection of derivatives, which we take to act on the "wave function," $\Psi(x, y, z, t)$:

$$\left(-\frac{\hbar^2}{2m}\nabla^2 + V\right)\Psi = i\hbar \frac{\partial}{\partial t}\Psi \tag{5.4}$$

This is the (time-dependent) Schrödinger equation. The solution (Ψ) describes a particle of mass m in the presence of a specified potential energy function V. Specifically, $|\Psi(x, y, z, t)|^2\, dx\, dy\, dz$ is the probability of finding the particle in the volume element $d^3x \equiv dx\, dy\, dz$, at time t. Since the particle must be *somewhere*, the integral of $|\Psi|^2$ over *all* space has to be 1:

$$\int |\Psi|^2\, d^3x = 1 \tag{5.5}$$

We say that the wave function is "normalized."*

The Schrödinger equation can be solved by "separation of variables." To begin with, we look for solutions that are simple *products* of a function of position, $\psi(x, y, z)$, and a function of time, $f(t)$:†

$$\Psi(x, y, z, t) = \psi(x, y, z)f(t) \tag{5.6}$$

With this ansätz, equation (5.4) can be written

$$\frac{1}{\psi}\left(-\frac{\hbar^2}{2m}\nabla^2 + V\right)\psi = \frac{i\hbar}{f}\frac{df}{dt} \tag{5.7}$$

The left side depends only on position, the right side depends only on time; the only way this equation can hold for all x, y, z, and t is if both sides are in fact *constant*. Physically, this "separation constant" represents the total energy of the particle, so we call it E:

$$\left(-\frac{\hbar^2}{2m}\nabla^2 + V\right)\psi = E\psi \tag{5.8}$$

$$i\hbar \frac{df}{dt} = Ef \tag{5.9}$$

The second of these is easy to solve:‡

* Notice that a solution to the Schrödinger equation (5.4) can be multiplied by any constant and still remain a solution. In practice, we fix this constant by demanding that equation (5.5) be satisfied; this process is called "normalizing" the wave function.

† Of course, *most* solutions to the Schrödinger equation do *not* have this form. For a justification of the method of separation of variables, see, for example, reference 2, Section 3.3.

‡ The general solution to equation (5.9) includes an overall multiplicative constant. However, since we're going to multiply f by ψ [eq. (5.6)], we may as well absorb the constant into ψ and keep f as simple as possible.

$$f(t) = e^{-iEt/\hbar} \tag{5.10}$$

The first [eq. (5.8)] is called the *time-independent* Schrödinger equation. The operator on the left is known as the "Hamiltonian":

$$H \equiv -\frac{\hbar^2}{2m}\nabla^2 + V \tag{5.11}$$

and the (time-independent) Schrödinger equation has the form of an eigenvalue equation:

$$H\psi = E\psi \tag{5.12}$$

ψ is an eigenfunction of H, and E is the eigenvalue. Evidently, the complete wave function for a particle of mass m and energy E, under the influence of a potential energy $V(x, y, z)$, is

$$\Psi(x, y, z, t) = \psi(x, y, z)e^{-iEt/\hbar} \tag{5.13}$$

where ψ satisfies equation (5.8).*

In the case of a spherically symmetrical (or "central") potential, V is a function only of the distance from the origin, and we adopt the usual spherical coordinates (r, θ, ϕ), in which the *Laplacian*, $\nabla^2 \equiv \partial^2/\partial x^2 + \partial^2/\partial y^2 + \partial^2/\partial z^2$ takes the form[2]

$$\nabla^2 = \frac{1}{r^2}\frac{\partial}{\partial r}\left(r^2\frac{\partial}{\partial r}\right) + \frac{1}{r^2\sin\theta}\frac{\partial}{\partial\theta}\left(\sin\theta\frac{\partial}{\partial\theta}\right) + \frac{1}{r^2\sin^2\theta}\frac{\partial^2}{\partial\phi^2} \tag{5.14}$$

The time independent Schrodinger equation can now be solved by further separation of variables. Writing

$$\psi(r, \theta, \phi) = R(r)\Theta(\theta)\Phi(\phi) \tag{5.15}$$

equation (5.8) reduces to three (ordinary) differential equations for the functions R, Θ, and Φ:

$$\frac{1}{r^2}\frac{d}{dr}\left(r^2\frac{dR}{dr}\right) = \left[\frac{l(l+1)}{r^2} + \frac{2m}{\hbar^2}(V(r) - E)\right]R \tag{5.16}$$

$$\sin\theta\frac{d}{d\theta}\left(\sin\theta\frac{d\Theta}{d\theta}\right) = \left[m_l^2 - l(l+1)\sin^2\theta\right]\Theta \tag{5.17}$$

$$\frac{d^2\Phi}{d\phi^2} = -m_l^2\Phi \tag{5.18}$$

The separation constants, l and m_l, have been aptly chosen, for they correspond precisely to the orbital angular momentum quantum numbers introduced in Chapter 4. The solution to equation (5.18) is easy:†

* Notice that $|\Psi|^2 = |\psi|^2$. For most purposes it is only the absolute square of the wave function that matters, and we shall work almost exclusively with ψ. Casually, we often refer to ψ as "the wave function," but remember that the *complete* wave function carries the exponential time dependence.

† The second (linearly independent) solution, $e^{-im_l\phi}$, is covered by letting m_l run to negative values. We could use $\sin(m_l\phi)$ and $\cos(m_l\phi)$ instead, and in electrostatics (where the same problem

$$\Phi(\phi) = e^{im_l\phi} \tag{5.19}$$

Solutions to equation (5.17) are less familiar; they are the so-called *associated Legendre functions:*[3]

$$\Theta(\theta) = P_l^{m_l}(\cos\theta) \tag{5.20}$$

where (in case you're interested)*

$$P_l^{m_l}(z) = \frac{1}{2^l l!}(1 - z^2)^{m_l/2}\left(\frac{d}{dz}\right)^{(l+m_l)}(z^2 - 1)^l \tag{5.21}$$

Actually, it is customary to combine Θ and Φ, with a conventional normalization factor, to form *spherical harmonics:*†

$$Y_l^{m_l}(\theta, \phi) = \sqrt{\frac{(2l+1)}{4\pi}\frac{(l-m_l)!}{(l+m_l)!}}(-1)^{m_l}P_l^{m_l}(\cos\theta)e^{im_l\phi} \tag{5.22}$$

Spherical harmonics are tabulated in many places (including the Particle Data Booklet); a few of the more useful ones are given in Table 5.1.

Please don't let the complicated-looking formulas distract you. I include them mainly for completeness, but also to demonstrate that there is nothing *mysterious* about the angular part of the wave function; it's just that it involves functions that are not terribly familiar. Notice, however, that in spite of the fancy *names,* they're only combinations of sines, cosines, and exponentials. Our real concern is not with the angular dependence at all, but rather with equation (5.16), which carries the sole reference to the specific potential. We can simplify the situation slightly by introducing a new function:

$$u(r) \equiv rR(r) \tag{5.23}$$

in terms of which equation (5.16) becomes

$$-\frac{\hbar^2}{2m}\frac{d^2u}{dr^2} + \left[V(r) + \frac{\hbar^2}{2m}\frac{l(l+1)}{r^2}\right]u = Eu \tag{5.24}$$

We call this the "radial Schrödinger equation"; curiously enough, it has exactly the same form as equation (5.8) for one dimension, except that the potential is augmented by the so-called "centrifugal barrier," $(\hbar^2/2m)l(l+1)/r^2$.

Equation (5.24) is about as far as we can pursue the matter in general terms; at this point we have to put in the particular potential $V(r)$ for the problem at hand. The strategy then will be to solve the radial equation for $u(r)$ and combine

---

arises in solving Laplace's equation) we *would,* since the potential must be real. But there is no such constraint on ψ, and in quantum mechanics the exponential form (5.19) is preferable, because, unlike the sine and cosine, it is an eigenstate of L_z. Notice that m_l must be an integer, in order that $\Phi(\phi + 2\pi)$ will equal $\Phi(\phi)$; after all, ϕ and $\phi + 2\pi$ describe the same geometrical point.

* Notice that equation (5.21) makes sense only if l, like m_l, is an integer, and gives zero if $m_l > l$. There exist solutions to equation (5.17) for other values of l and m_l, but they do not yield normalizable wave functions.

† This assumes $m_l \geqq 0$; for $m_l < 0$, we use $Y_l^{m_l} = (-1)^{m_l}(Y_l^{-m_l})^*$.

TABLE 5.1 SPHERICAL HARMONICS FOR $l = 0, 1, 2$, AND 3

$$Y_0^0 = \frac{1}{\sqrt{4\pi}}, \quad Y_1^0 = \sqrt{\frac{3}{4\pi}}\cos\theta, \quad Y_2^0 = \sqrt{\frac{5}{16\pi}}(3\cos^2\theta - 1),$$

$$Y_3^0 = \sqrt{\frac{7}{16\pi}}(5\cos^3\theta - 3\cos\theta),$$

$$Y_1^1 = -\sqrt{\frac{3}{8\pi}}\sin\theta\, e^{i\phi}, \quad Y_2^1 = -\sqrt{\frac{15}{8\pi}}\sin\theta\cos\theta\, e^{i\phi}$$

$$Y_3^1 = -\sqrt{\frac{21}{64\pi}}\sin\theta(5\cos^2\theta - 1)\, e^{i\phi}$$

$$Y_2^2 = \sqrt{\frac{15}{32\pi}}\sin^2\theta\, e^{2i\phi}, \quad Y_3^2 = \sqrt{\frac{105}{32\pi}}\sin^2\theta\cos\theta\, e^{i\phi}$$

$$Y_3^3 = -\sqrt{\frac{35}{64\pi}}\sin^3\theta\, e^{3i\phi}$$

the result with the appropriate spherical harmonic to get the full wave function ψ. In the course of solving the radial equation, however, we shall discover that only certain special values of E lead to acceptable results. For *most* values of E the solution to equation (5.24) blows up at large r, and yields a nonnormalizable wave function. Such a solution does not represent a possible physical state. This rather technical detail is the source of the most striking and important feature of quantum mechanics: a bound system cannot have just any old energy (as it could classically); instead, the energy can take on only certain specific values, the so-called "allowed energies" of the system. Indeed, our real concern is not with the wave function itself, but with the *spectrum* of allowed energies.

5.2 THE HYDROGEN ATOM

The hydrogen atom consists, of course, of an electron and a proton. The proton, however, is so heavy (relatively) that it essentially just sits at the origin; the wave function in question is that of the *electron*. Its potential energy, due to the electrical attraction of the nucleus, is (in Gaussian units)

$$V(r) = -\frac{e^2}{r} \tag{5.25}$$

When this potential is put into the radial equation, it is found (see Problem 5.5) that normalizable solutions occur only when E assumes one of the special values

$$E_n = -\frac{me^4}{2\hbar^2 n^2} = -\alpha^2 mc^2\left(\frac{1}{2n^2}\right) = -13.6\ \text{eV}/n^2 \quad (n = 1, 2, 3, \ldots) \tag{5.26}$$

where

$$\alpha \equiv \frac{e^2}{\hbar c} = \frac{1}{137.036} \tag{5.27}$$

is the fine structure constant. The corresponding (normalized) wave function is[4]

$$\psi_{n,l,m_l}(r, \theta, \phi) = \left\{\left(\frac{2}{na}\right)^3 \frac{(n-l-1)!}{2n[(n+l)!]^3}\right\}^{1/2} e^{-r/na}\left(\frac{2r}{na}\right)^l L_{n-l-1}^{2l+1}\left(\frac{2r}{na}\right) Y_l^{m_l}(\theta, \phi) \quad (5.28)$$

where

$$a \equiv \frac{\hbar^2}{me^2} = 0.529 \times 10^{-8} \text{ cm} \quad (5.29)$$

is the "Bohr radius" (roughly speaking, the radius of a hydrogen atom) and

$$L_{q-p}^p(z) \equiv (-1)^p \left(\frac{d}{dz}\right)^p \left[e^z \left(\frac{d}{dz}\right)^q (e^{-z} z^q)\right] \quad (5.30)$$

is an *associated Laguerre polynomial.* Obviously, the wave function itself is a bit of a mess, but that's not really what concerns us. The crucial thing is the formula of the allowed energies, equation (5.26). [This result was first obtained by Bohr in 1913 (more than a decade before the Schrödinger equation was introduced) by a brilliant (although, in retrospect, extraordinarily lucky) amalgam of classical ideas and primitive quantum theory, a blend, as Rabi put it, of "artistry and effrontery."] Observe that the wave function is labeled by three numbers: n, (the "principal quantum number"), which can be any positive integer—it determines the energy of the state [eq. (5.26)]; l, an integer which ranges from 0 up to $n-1$ and specifies the total orbital angular momentum [eq. (4.2)]; and m_l, an integer which can assume any value between $-l$ and $+l$, giving the z component of the angular momentum [eq. (4.4)]. Evidently, there are $2l+1$ different m_l's, for each l, and n different l's, for each n. The total number of distinct states that share the same principal quantum number n, and hence the same energy, is, therefore

$$\sum_{l=0}^{n-1} (2l+1) = n^2 \quad (5.31)$$

This is called the *degeneracy* of the nth energy level. Hydrogen is a surprisingly degenerate system; spherical symmetry alone dictates that the $2l+1$ states with a given value of the total angular momentum should be degenerate, since they differ only in the orientation of $\mathbf{L}$, but this suggests a sequence 1, 3, 5, 7, . . . , whereas the energy levels of hydrogen have much higher degeneracies: 1, 4, 9, 16, ... So far, however, we have neglected the electron's spin, as well as a small relativistic effect. Their inclusion, as we shall see in the next section, leads to corrections that "lift" the "extra" degeneracy, splitting the Bohr energies into groups of closely spaced levels.

Before coming to that, however, I should describe how the quantized energy levels of hydrogen reveal themselves in the laboratory, for in practice we do not measure the energies themselves, but rather the wavelength of the light emitted when the electron makes a transition from a higher level to a lower one (or the light *absorbed* when it goes the other way).[5] The photon carries the *difference*

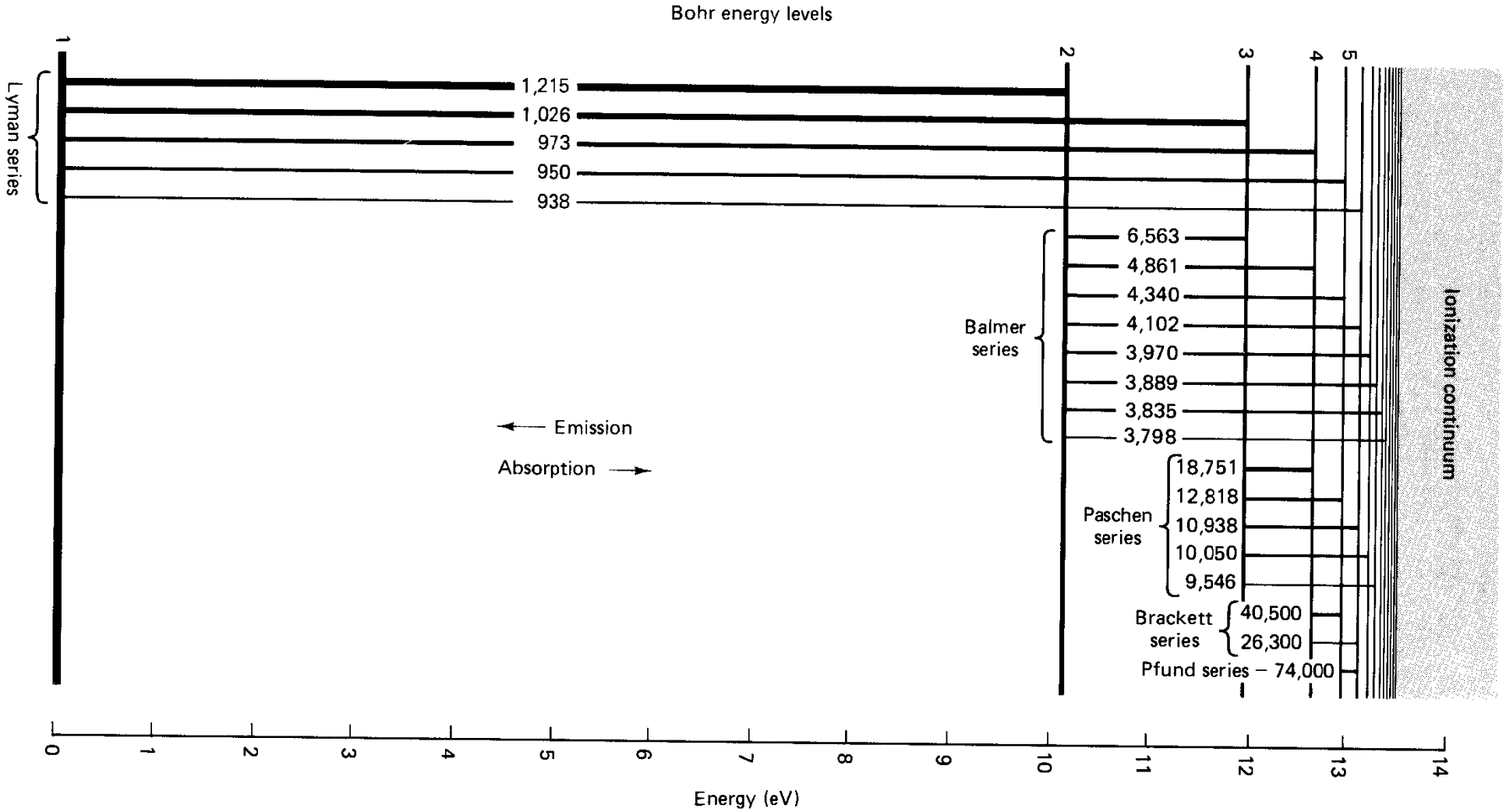


Figure 5.1 The spectrum of hydrogen. When an atom changes from one state to another, the difference in energy appears as a quantum of radiation. The energy of the photon is directly proportional to the frequency of the radiation and inversely proportional to the wavelength. Absorption of radiation stimulates a transition to a state of higher energy; an atom falling to a state of lower energy emits radiation. The spectrum is organized into series of lines that share a common lower level. Wavelengths are given in angstroms; the relative intensity of the lines is indicated by thickness. (From "The Spectrum of Atomic Hydrogen," by T. W. Hänsch, A. L. Schawlow, and G. W. Series. Copyright © March 1979 by Scientific American, Inc. All rights reserved.)

in energy between the initial and final states. According to the Planck formula [eq. (1.1)], then

$$E_{\text{photon}} = h\nu = E_{\text{initial}} - E_{\text{final}} = -\frac{me^4}{2\hbar^2}\left(\frac{1}{n_i^2} - \frac{1}{n_f^2}\right) \tag{5.32}$$

The emitted wavelength, therefore, is given by

$$\frac{1}{\lambda} = R\left(\frac{1}{n_f^2} - \frac{1}{n_i^2}\right) \tag{5.33}$$

where

$$R \equiv \frac{me^4c}{4\pi\hbar^3} \tag{5.34}$$

This is the famous *Rydberg formula* for the spectrum of hydrogen. It was discovered experimentally by nineteenth-century spectroscopists, for whom R was simply an empirical constant. The greatest triumph of Bohr's theory was its derivation of the Rydberg formula, and its calculation of R in terms of the fundamental constants m, e, c, and $\hbar$ (see Fig. 5.1).

5.3 FINE STRUCTURE

As the precision of experimental spectroscopy improved, small departures from the Rydberg formula were detected. Spectral lines were resolved into doublets, triplets, and even larger families of closely spaced peaks. This *fine structure* is actually attributable to two distinct mechanisms: one a small relativistic correction, and the other a magnetic coupling between the electron's spin and its orbital motion. We shall analyze the two effects separately, and then combine them.

But first a word about perturbation theory. Suppose we have solved the Schrödinger equation for some Hamiltonian H, in the sense that we know the allowed energies E_j and the corresponding wave functions ψ_j (j being whatever collection of indices is used to label the states; in the case of hydrogen, it stands for n, l, and m_l). Now suppose we *change* the Hamiltonian slightly, adding a small *perturbation* ΔH, so that the new Hamiltonian is $H + \Delta H$. This will presumably displace the energy levels a bit, and we may ask the question: By how much, ΔE_j, is the jth energy level shifted? The answer, to good approximation, is that ΔE_j is the *average* value of ΔH, as computed in the state ψ_j (see Problem 5.8):*

$$\Delta E_j = \langle \Delta H \rangle_j = \int \psi_j^* (\Delta H) \psi_j \, d^3x \tag{5.35}$$

* I think this is a reasonably plausible result, but if you want to see it derived, look in any introductory quantum text (such as those in ref. 1) under "perturbation theory."

5.3.1 The Relativistic Correction

In developing the Schrödinger equation, I used the classical formula for kinetic energy: $T = \mathbf{p}^2/2m$. The relativistic formula (3.45), expressed in terms of $\mathbf{p}$, is

$$\begin{aligned} T_{\text{rel}} &= \sqrt{\mathbf{p}^2c^2 + m^2c^4} - mc^2 = mc^2[\sqrt{1 + (\mathbf{p}/mc)^2} - 1] \\ &= mc^2[1 + \tfrac{1}{2}(\mathbf{p}/mc)^2 - \tfrac{1}{8}(\mathbf{p}/mc)^4 + \cdots - 1] \\ &= \frac{\mathbf{p}^2}{2m} - \frac{\mathbf{p}^4}{8m^3c^2} + \cdots \end{aligned} \tag{5.36}$$

The lowest-order relativistic correction to the Hamiltonian is therefore

$$\Delta H_{\text{rel}} = -\left(\frac{1}{8m^3c^2}\right)\mathbf{p}^4 \tag{5.37}$$

Now, to apply equation (5.35), we need the expectation value of $\mathbf{p}^4$ (using the quantum replacement (5.2) for $\mathbf{p}$) in the state $\psi_{n,l,m_l} \ldots$, and that does not look easy. Fortunately, we can get around it by using the fact that in such a state

$$\frac{\mathbf{p}^2}{2m} = E_n - V \tag{5.38}$$

Thus

$$\Delta E_{\text{rel}} = -\frac{1}{2mc^2}(E_n^2 - 2E_n\langle V\rangle + \langle V^2\rangle) \tag{5.39}$$

Since $V = -e^2/r$, we now require only the expectation values of $1/r$ and $1/r^2$. These are worked out in the more advanced quantum texts:[6]

$$\left\langle \frac{1}{r} \right\rangle = \frac{1}{n^2a} \tag{5.40}$$

$$\left\langle \frac{1}{r^2} \right\rangle = \frac{1}{(l + \frac{1}{2})n^3a^2} \tag{5.41}$$

where a, again, is the Bohr radius (5.29). Using these expressions in equation (5.39), we find that

$$\Delta E_{\text{rel}} = -\alpha^4mc^2\frac{1}{4n^2}\left[\frac{2n}{(l + \frac{1}{2})} - \frac{3}{2}\right] \tag{5.42}$$

The Bohr energy levels go like α^2mc^2 [eq. (5.26)]; the relativistic correction carries two more powers of α, and hence is smaller by a factor of about 10^{-4}. So we're talking about a very small perturbation indeed. The *fine structure constant*, α, owes its name to the fact that it (or rather, α^2) sets the relative scale of the fine structure in hydrogen.*

* However, one might say equally well that α^2 sets the scale of the Bohr levels themselves, in the sense that $\alpha^2 = -2E_1/mc^2$.

5.3.2 Spin-Orbit Coupling

From the electron's point of view it is the *proton* that circles around, and this orbital motion creates a magnetic field at the center given by*

$$B = \frac{ev}{cr^2} \tag{5.43}$$

or, in terms of the electron's orbital angular momentum, $L = rmv$:

$$\mathbf{B} = \frac{e}{mcr^3}\mathbf{L} \tag{5.44}$$

The spinning electron constitutes a tiny magnetic dipole, with dipole moment†

$$\boldsymbol{\mu} = -\frac{e}{mc}\mathbf{S} \tag{5.45}$$

From classical electrodynamics we know that the energy of a magnetic dipole $\boldsymbol{\mu}$ in the presence of a magnetic field $\mathbf{B}$ is[7]

$$W = -\boldsymbol{\mu}\cdot\mathbf{B} \tag{5.46}$$

But the electron is not in an inertial frame;‡ to make all this rigorous we should really speak of the "instantaneously comoving frame" of the electron, the inertial system whose velocity coincides with that of the electron at a given moment. Following the motion of the electron, then, involves a continuous succession of infinitesimal Lorentz transformations, as we step from one comoving frame to the next. This procedure leads to the so-called *Thomas precession,*[8] which in the present context simply introduces a factor of $\frac{1}{2}$. The Hamiltonian for spin-orbit coupling is thus

$$\Delta H_{so} = \frac{e^2}{2m^2c^2r^3}(\mathbf{L}\cdot\mathbf{S}) \tag{5.47}$$

* You can get this most simply by looking up the formula for the magnetic field at the center of a circular ring, $B = 2\pi I/rc$, and using $I = e/t$ for the "proton current," where $t = 2\pi r/v$ is the time it takes to complete one revolution. Alternatively, exploit the fact that if $B = 0$ in one inertial system (the proton's), then $\mathbf{B} = -\mathbf{v}/c \times \mathbf{E}$ in a system moving with velocity $\mathbf{v}$; in this case $\mathbf{E}$ is the field of the proton: e/r^2.

† The proportionality factor between $\boldsymbol{\mu}$ and $\mathbf{S}$ is known as the *gyromagnetic ratio.* Classically, it should have the value $-e/2mc$, and this is correct for *orbital* angular momentum. But for relativistic reasons spin is "twice as effective as it ought to be" in producing a magnetic dipole. One of the major successes of Dirac's original theory of the electron was its explanation of this extra 2. As it happens, however, even *this* is not quite right; there are minute corrections introduced by quantum electrodynamics that were first calculated by Schwinger in the late forties. By now both experimental and theoretical determinations of the *anomalous magnetic moment* of the electron have been carried out to fantastic precision, and stunning agreement. The Particle Data Booklet lists 13 significant digits! For our present purpose, though, the magnetic moment of the electron is $e\hbar/2mc$, the *Bohr magneton.*

‡ One can, of course, analyze the whole problem in the rest frame of the *proton.* From this perspective there is *no* magnetic field, and spin-orbit coupling is attributable to the torque $\boldsymbol{\tau}$ on a magnetic dipole $\mathbf{m}$ that moves with velocity $\mathbf{v}$ through an *electric* field $\mathbf{E}$: $\boldsymbol{\tau} = -(1/c)\mathbf{m} \times (\mathbf{v} \times \mathbf{E})$.

In the presence of spin-orbit coupling neither **L** nor **S** is separately conserved; the conserved quantity is the *total* angular momentum, $\mathbf{J} = \mathbf{L} + \mathbf{S}$.[9] Squaring, we have

$$J^2 = L^2 + S^2 + 2\mathbf{L}\cdot\mathbf{S} \tag{5.48}$$

and hence, using expressions (4.2) and (4.6), we have

$$\mathbf{L}\cdot\mathbf{S} = \tfrac{1}{2}\hbar^2[j(j+1) - l(l+1) - s(s+1)] \tag{5.49}$$

where j, l, and s are the quantum numbers describing total, orbital, and spin angular momenta, respectively (for the electron, $s = \frac{1}{2}$, of course). This time we need the expectation value of r^{-3}:*

$$\left\langle \frac{1}{r^3} \right\rangle = \frac{1}{l(l+\frac{1}{2})(l+1)n^3a^3} \tag{5.50}$$

Putting this into equation (5.35), we find

$$\Delta E_{so} = \alpha^4 mc^2 \frac{[j(j+1) - l(l+1) - \frac{3}{4}]}{4n^3 l(l+\frac{1}{2})(l+1)} \tag{5.51}$$

Notice that this is of the same order as the relativistic correction, $\alpha^4 mc^2$, even though the physical agencies involved are (in this treatment) entirely different. The combination of the two effects, equations (5.42) and (5.51), yields the total fine structure for hydrogen. Using the fact that j can only be $l + \frac{1}{2}$ or $l - \frac{1}{2}$ [eq. (4.12)], we obtain (Problem 5.9)

$$\Delta E_{fs} = -\alpha^4 mc^2 \frac{1}{4n^2}\left(\frac{2n}{(j+\frac{1}{2})} - \frac{3}{2}\right) \tag{5.52}$$

Curiously, this formula is identical to that for the relativistic correction alone, equation (5.42), except that l is replaced by j. The energies are all depressed (ΔE_{fs} is negative). Since l can take on any integer value from 0 to $n - 1$, j can be any half-integer from $\frac{1}{2}$ to $n - \frac{1}{2}$, so the nth energy level, E_n, splits into n sublevels (see Fig. 5.2).

5.4 THE LAMB SHIFT

A striking feature of the fine structure formula (5.52) is that it depends only on j, not on l; in general, two different values of l share the same energy. For example, the $2S_{1/2}$ ($n = 2$, $l = 0$, $j = \frac{1}{2}$) and $2P_{1/2}$ ($n = 2$, $l = 1$, $j = \frac{1}{2}$) states should remain perfectly degenerate. In 1947 Lamb and Retherford performed a classic experiment[10] which demonstrated that this is not, in fact, the case; the S state is slightly

* See reference 6. When $l = 0$, $\langle r^{-3}\rangle$ blows up. Really, this case should be handled separately. Fortunately, the l cancels out, and our final result, equation (5.52), is correct even when $l = 0$. The spin-orbit coupling by itself is zero, of course, when the orbital angular momentum vanishes, but there is a compensating correction to the relativistic part of the fine structure, having to do with the non-Hermiticity of $\mathbf{p}^4$ when $l = 0$.

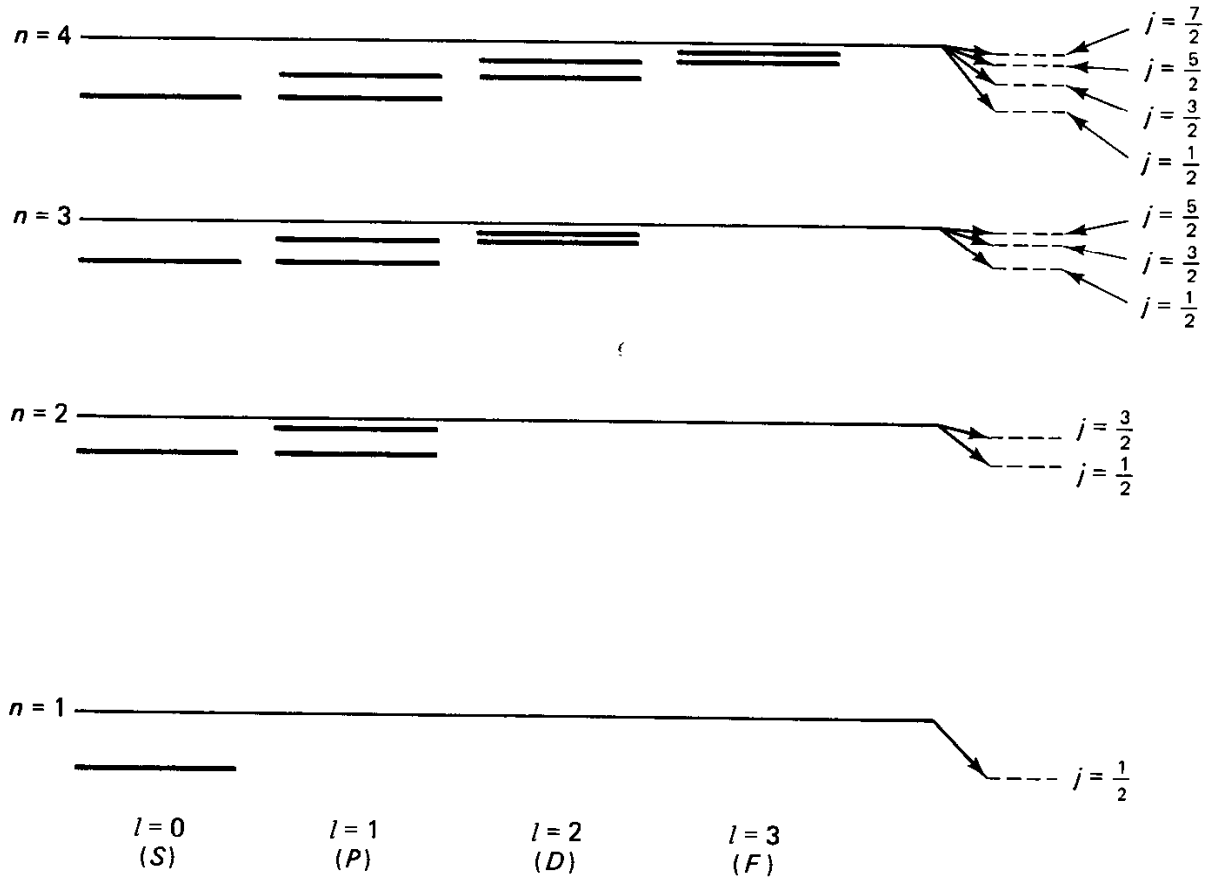
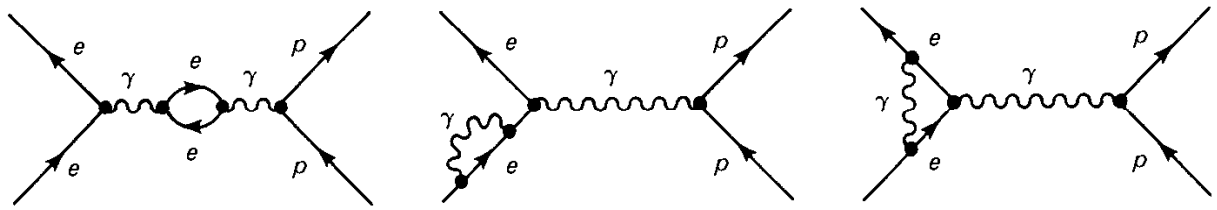


Figure 5.2 Fine structure in hydrogen. The nth Bohr level (fine line) splits into n sublevels (dashed lines), characterized by $j = \frac{1}{2}, \frac{3}{2}, \ldots, (n - \frac{1}{2})$. Except for the last of these, two different values of l contribute to each level: $l = j - \frac{1}{2}$ and $l = j + \frac{1}{2}$. Spectroscopists' nomenclature—S for $l = 0$, P for $l = 1$, D for $l = 2$, F for $l = 3$—is indicated. All levels are shifted downward, as shown (the diagram is not to scale, however).

higher in energy than the P state. The explanation of this *Lamb shift* was provided by Bethe, Feynman, Schwinger, Tomonaga, and others: it is due to the *quantization of the electromagnetic field itself.* Everywhere else in this discussion—in the calculation of the Bohr levels, in the derivation of the fine structure formula, and even in the analysis of hyperfine structure in the next section—the electromagnetic field is treated entirely classically. Coulomb's law is the basis for equation (5.25); the Biot-Savart Law yields equation (5.43); the Lorentz force law is responsible for equation (5.46). The Lamb shift, by contrast, is an example of a *radiative correction* in quantum electrodynamics, to which the semiclassical* theory is insensitive. In the Feynman formalism, it results from *loop diagrams,* such as those in Figure 5.3, which we shall discuss quantitatively later on.

Qualitatively, the first diagram in Fig. 5.3 describes spontaneous production of electron-positron pairs in the neighborhood of the nucleus, leading to a partial screening of the proton's charge (Fig. 2.1). The second diagram reflects the fact that the ground state of the electromagnetic field is not zero;[11] as the electron moves through the "vacuum fluctuations" in the field, it jiggles slightly, and this alters its energy. The third diagram leads to a tiny modification of the electron's

* I call it *semi*classical because the *electron* is treated quantum mechanically, whereas the *electrodynamic field* is treated classically.



Vacuum polarization Electron mass renormalization Anomalous magnetic moment

Figure 5.3 Some loop diagrams contributing to the Lamb shift.

magnetic dipole moment; equation (5.45) picks up a factor $(1 + \alpha/2\pi) = 1.00116$. We are not in a position to calculate these effects now, but for completeness, here are the results:[12]

For $l = 0$

$$\Delta E_{\text{Lamb}} = \alpha^5 mc^2 \frac{1}{4n^3} \{k(n, 0)\} \tag{5.53}$$

where $k(n, 0)$ is a numerical factor that varies slightly with n, from 12.7 (for $n = 1$) to 13.2 (for $n \to \infty$).

For $l \neq 0$

$$\Delta E_{\text{Lamb}} = \alpha^5 mc^2 \frac{1}{4n^3} \left\{k(n, l) \pm \frac{1}{\pi(j + \frac{1}{2})(l + \frac{1}{2})}\right\}, \qquad \text{for } j = l \pm \tfrac{1}{2} \tag{5.54}$$

where $k(n, l)$ is a very small number (less than 0.05) which varies slightly with n and l. Evidently the Lamb shift tiny, except for states with $l = 0$, for which it amounts to about 10% of the fine structure. However, because it depends on l, it lifts the degeneracy of the pairs of states with common n and j, on Figure 5.2, and in particular it splits the $2S_{1/2}$ and $2P_{1/2}$ levels (See Problem 5.11).

5.5 HYPERFINE STRUCTURE

The fine structure and the Lamb shift are minute corrections to the Bohr energy levels, but they are not the end of the story; there is a refinement that is smaller still (by a factor of a thousand), due to the spin of the *nucleus*. The proton, like the electron, constitutes a tiny magnet, but because it is so much heavier, the same angular momentum ($\frac{1}{2}\hbar$) corresponds to a much smaller dipole moment:*

$$\boldsymbol{\mu}_p = \gamma_p \frac{e}{m_p c} \mathbf{S}_p \tag{5.55}$$

(The proton is a composite object, and its magnetic moment is not simply $e\hbar/2m_p c$, as it would be for a truly elementary particle of spin $\frac{1}{2}$. Hence the

* In fancier language, the proton's gyromagnetic ratio is much smaller than the electron's.

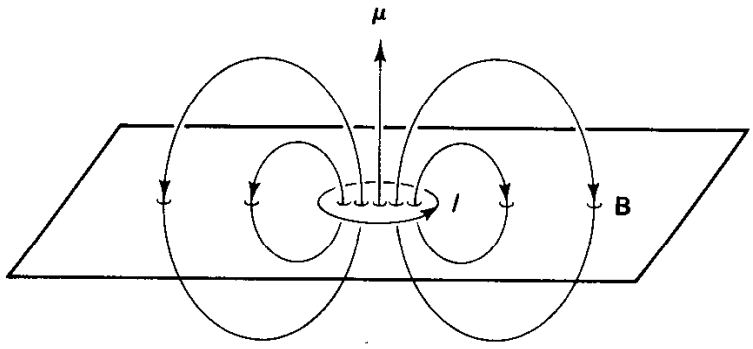


Figure 5.4 Field of a magnetic dipole. In the plane of the current loop, **B** points *down* outside the ring, and *up* inside the ring.

factor γ_p, whose experimental value is 2.7928. Later on we shall see how to calculate this quantity in the quark model.) The nuclear spin interacts with the electron's orbital motion by the same mechanism as the spin-orbit contribution to the fine structure, only this time there is no hanky-panky about changing reference systems, and no Thomas precession to worry about. Referring to our earlier calculation [eq. (5.47)], we find that the nuclear spin-orbit interaction is*

$$\Delta H_{\text{pso}} = \frac{\gamma_p e^2}{m m_p c^2 r^3} (\mathbf{L} \cdot \mathbf{S}_p) \tag{5.56}$$

At the same time, the proton spin also interacts directly with the electron spin. In many books on electricity and magnetism[13] it is shown that a magnetic dipole $\boldsymbol{\mu}$ sets up a field

$$\mathbf{B}(\mathbf{r}) = \frac{1}{r^3}\left[3\frac{(\boldsymbol{\mu}\cdot\mathbf{r})\mathbf{r}}{r^2} - \boldsymbol{\mu}\right] \tag{5.57}$$

Actually, this formula is not quite right; if we picture the dipole as a tiny current loop, equation (5.57) says that in the plane of the loop (where $\boldsymbol{\mu}\cdot\mathbf{r} = 0$) **B** always points in the direction opposite to $\boldsymbol{\mu}$, whereas *inside* the loop **B** is in fact *parallel* to $\boldsymbol{\mu}$ (see Fig. 5.4). It is true that as we shrink the loop down, to make a "perfect" point dipole, the region where the field has the "wrong" direction gets smaller and smaller. On the other hand, its *strength* gets larger and larger, since all those field lines must pass through the ring. The correct handling of this "anomalous" field is a rather delicate problem in classical electrodynamics;[14] the conclusion is that

$$\mathbf{B}(\mathbf{r}) = \frac{1}{r^3}\left[3\frac{(\boldsymbol{\mu}\cdot\mathbf{r})\mathbf{r}}{r^2} - \boldsymbol{\mu}\right] + \frac{8\pi}{3}\boldsymbol{\mu}\,\delta^3(\mathbf{r}) \tag{5.58}$$

where $\delta^3(\mathbf{r})$ is the Dirac delta function†—infinite at $r = 0$ and zero elsewhere. As long as you stay away from the origin, the "naïve" formula (5.57) is perfectly

* Note that equation (5.44) picks up a minus sign this time, since we're talking about the field of an orbiting *electron*, not a *proton*.

† The delta function is discussed in Appendix A. For now, all you need to know is that $\int f(\mathbf{r})\delta^3(\mathbf{r})d^3r = f(0)$. Technically, the first term in equation (5.58) applies to the region outside a tiny sphere of radius ϵ, and the second term to the region *inside*. At the end of a calculation ϵ is set equal to zero.

correct. However, as we shall soon see, the delta function term makes a crucial contribution to hyperfine structure. In the present case the proton sets up the field **B**, and the energy of the electron in its presence is given by equation (5.46). Thus the spin-spin Hamiltonian is

$$\Delta H_{ss} = \frac{\gamma_p e^2}{m m_p c^2}\left\{\frac{1}{r^3}\left[3(\mathbf{S}_p \cdot \hat{r})(\mathbf{S}_e \cdot \hat{r}) - (\mathbf{S}_p \cdot \mathbf{S}_e)\right] + \frac{8\pi}{3}(\mathbf{S}_p \cdot \mathbf{S}_e)\delta^3(\mathbf{r})\right\} \tag{5.59}$$

where $\mathbf{S}_p$ and $\mathbf{S}_e$ are the spin angular momenta of the proton and electron, respectively.

Together, the nuclear spin-orbit interaction (5.56) and the proton-electron spin-spin coupling (5.59) are responsible for the hyperfine structure of hydrogen. To calculate the resulting energy shifts, we again use the standard formula from perturbation theory (5.35). At this point the treatment is quite different for states with $l = 0$, so we'll consider them first. Since the orbital angular momentum is zero, there is no spin-orbit coupling. Moreover, because the wave function is spherically symmetrical ($Y_0^0(\theta, \phi) = 1/\sqrt{4\pi}$) the expectation value of the first term in equation (5.59) vanishes (see Problem 5.12). This leaves only the delta function contribution

$$\Delta E_{hf} = \frac{8\pi\gamma_p e^2}{3 m m_p c^2}(\mathbf{S}_p \cdot \mathbf{S}_e)|\psi_{n00}(0)|^2 \tag{5.60}$$

From equation (5.28) we find (Problem 5.13) that

$$|\psi_{n00}(0)|^2 = \frac{1}{\pi n^3 a^3} \tag{5.61}$$

Let

$$\mathbf{F} = \mathbf{L} + \mathbf{S}_e + \mathbf{S}_p = \mathbf{J} + \mathbf{S}_p \tag{5.62}$$

be the total angular momentum of the atom, with quantum numbers f and m_f; in the present case $\mathbf{L} = 0$, so $F^2 = S_e^2 + S_p^2 + 2\mathbf{S}_e \cdot \mathbf{S}_p$, and hence

$$\mathbf{S}_p \cdot \mathbf{S}_e = \frac{\hbar^2}{2}\left[f(f+1) - s_p(s_p+1) - s_e(s_e+1)\right] = \frac{\hbar^2}{2}\left[f(f+1) - \frac{3}{2}\right] \tag{5.63}$$

Thus

$$\Delta E_{hf} = \left(\frac{m}{m_p}\right)\alpha^4 m c^2 \frac{4\gamma_p}{3n^3}\left[f(f+1) - \frac{3}{2}\right] \tag{5.64}$$

Comparing the fine structure formula (5.52), we see that the difference in scale is due to the mass ratio (m/m_p) in front; it follows that hyperfine effects in hydrogen are about one thousand times smaller. Notice that f can take on two values: zero, in the singlet state (when the spins are oppositely aligned) and one, in the triplet state (when the spins are parallel). Thus each $l = 0$ level divides into two, with the singlet pushed down and the triplet pushed up (Fig. 5.5). For $n = 1$ the energy gap is

$$\epsilon = E_{\text{triplet}} - E_{\text{singlet}} = \frac{32\gamma_p E_1^2}{3 m_p c^2} \tag{5.65}$$

corresponding to a photon of wavelength

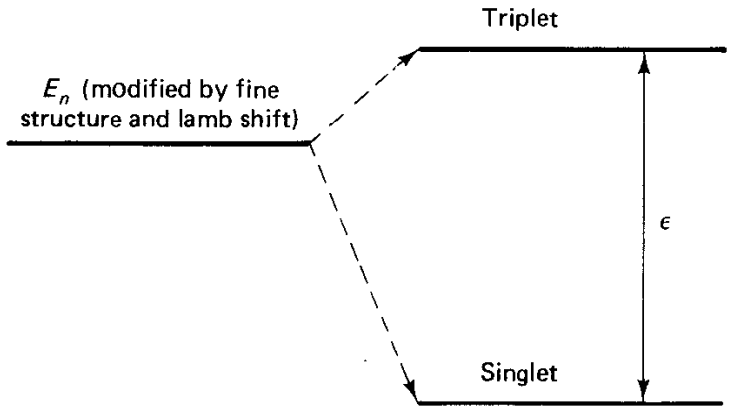


Figure 5.5 Hyperfine splitting for $l = 0$.

$$\lambda = \frac{2\pi\hbar c}{\epsilon} = 21.1 \text{ cm} \tag{5.66}$$

This is the transition that gives rise to the famous "21-centimeter line" in microwave astronomy.[15]

For $l = 0$, as we have seen, the hyperfine structure comes entirely from the "contact" term in the spin-spin coupling; for $l \neq 0$ it is exactly the reverse. This time the delta function contributes nothing; the wave function (5.28) goes like r^l at small r, so $\psi(0) = 0$ when $l > 0$. Physically, the centrifugal (pseudo-) force keeps the electron away from the nucleus. Accordingly, we now have

$$\Delta E_{\text{hf}} = \frac{\gamma_p e^2}{m m_p c^2} \left\langle \frac{1}{r^3} [(\mathbf{L}\cdot\mathbf{S}_p) + 3(\mathbf{S}_p\cdot\hat{r})(\mathbf{S}_e\cdot\hat{r}) - (\mathbf{S}_p\cdot\mathbf{S}_e)] \right\rangle \tag{5.67}$$

This expectation value is calculated, for instance, by Bethe and Salpeter;[16] the result is

$$\Delta E_{\text{hf}} = \left(\frac{m}{m_p}\right)\alpha^4 mc^2 \frac{\gamma_p}{2n^3} \frac{[f(f+1) - j(j+1) - \frac{3}{4}]}{j(j+1)(l+\frac{1}{2})} \tag{5.68}$$

As it turns out, this formula works for $l = 0$ as well, since in that case $j = \frac{1}{2}$, and we recover equation (5.64). Because the proton carries spin $\frac{1}{2}$, f can only be $j + \frac{1}{2}$ or $j - \frac{1}{2}$; with this in mind, equation (5.68) simplifies slightly:

$$\Delta E_{\text{hf}} = \left(\frac{m}{m_p}\right)\alpha^4 mc^2 \frac{\gamma_p}{2n^3} \frac{\pm 1}{(f+\frac{1}{2})(l+\frac{1}{2})}, \qquad \text{for } f = j \pm \tfrac{1}{2} \tag{5.69}$$

Each of the levels in Figure 5.2, characterized by particular values of n, l, and j, is split into two—one moving up, and the other down.

5.6 POSITRONIUM

The theory we have developed here for ordinary hydrogen carries over, with some modifications, to the so-called "exotic" atoms, in which either the proton

or the electron is replaced by some other particle. For instance, one can make muonic hydrogen ($p^+\mu^-$), pionic hydrogen ($p^+\pi^-$), positronium (e^+e^-), muonium (μ^+e^-), and so on. Of course, these exotic states are unstable, but many of them last long enough to exhibit a well-defined spectrum. In particular, positronium provides a rich testing ground for quantum electrodynamics. It was analyzed theoretically by Pirenne in 1944, and first produced in the laboratory by Deutsch in 1951.[17] In recent years positronium has assumed a special importance as the model for quarkonium.

The most conspicuous difference between positronium and hydrogen is that we are no longer dealing with a heavy, essentially stationary nucleus, around which the electron orbits, but rather with two particles of *equal* mass, both orbiting around the common center. As in classical mechanics, this two-body problem can be converted into an equivalent one-body problem with the *reduced mass*

$$m_{\text{red}} = \frac{m_1 m_2}{m_1 + m_2} \tag{5.70}$$

For if the two-body Hamiltonian has the form

$$H = \frac{\mathbf{p}_1^2}{2m_1} + \frac{\mathbf{p}_2^2}{2m_2} + V(\mathbf{r}_1, \mathbf{r}_2) \tag{5.71}$$

and if the potential depends only on the separation distance, $r = |\mathbf{r}_2 - \mathbf{r}_1|$, and if we work in the *CM* system, where $\mathbf{p}_1 = -\mathbf{p}_2 = \mathbf{p}$, then

$$H = \frac{\mathbf{p}^2}{2m_{\text{red}}} + V(r) \tag{5.72}$$

which is the Hamiltonian for a *single* particle of momentum $\mathbf{p}$ and mass m_{red}, whose "radial" coordinate, r, is the distance between 1 and 2. The unperturbed Hamiltonian for positronium is of the form (5.71), with $m_1 = m_2 = m$, so that $m_{\text{red}} = m/2$ and $V = -e^2/r$, the same as hydrogen. So we get the unperturbed energy levels for positronium by the simple substitution $m \to m/2$ in the Bohr formula (5.26):*

$$E_n^{\text{pos}} = \tfrac{1}{2}E_n = -\alpha^2 mc^2 \frac{1}{4n^2} \qquad (n = 1, 2, 3, \ldots) \tag{5.73}$$

For example, the ground-state binding energy is 13.6 eV/2 = 6.8 eV. The unperturbed wave functions are the same as before (5.28), except that the Bohr radius, which goes like $1/m$ [see eq. (5.29)], is doubled:

$$a^{\text{pos}} = 2a = 1.06 \times 10^{-8} \text{ cm} \tag{5.74}$$

The perturbations run much as before, apart from pesky numerical factors. For example, the relativistic correction picks up a factor of 2:

* In the case of ordinary hydrogen, the reduced mass differs from the electron mass by only a very small amount, about 0.05%. Nevertheless, technically the m in the Bohr formula is the reduced mass, and this does lead to observable differences between the spectra of hydrogen and deuterium.

$$\Delta H_{\text{rel}} = -\frac{1}{8m_1^3c^2}\mathbf{p}_1^4 - \frac{1}{8m_2^3c^2}\mathbf{p}_2^4 = -\frac{1}{4m^3c^2}\mathbf{p}^4 \tag{5.75}$$

On the other hand, the expectation value of $\mathbf{p}^4$ in the hydrogenic state ψ_{nlm} goes like $(mc)^4$, so for positronium it is reduced by $(\frac{1}{2})^4$. All told, then, the relativistic correction for positronium is one eighth that of hydrogen [eq. (5.42)]. More significant is the fact that in positronium the hyperfine splitting is of the same order as the fine structure (α^4mc^2), since the mass ratio (m/m_p) which suppresses proton spin effects in hydrogen is *one* in positronium.* Meanwhile, since the "nucleus" (e^+) is not stationary, we are no longer working with a truly static potential, and there is a new correction due to the finite propagation time for the electromagnetic field. This can be calculated using classical electrodynamics; it has the form

$$\Delta H_{\text{ret}} = -\frac{e^2}{2m^2c^2}\frac{1}{r}[\mathbf{p}^2 + (\mathbf{p}\cdot\hat{r})^2] \tag{5.76}$$

and its contribution is also of order (α^4mc^2). When all this is put together, the fine structure formula for positronium is found to be[18]

$$E_{\text{fs}}^{\text{pos}} = \alpha^4mc^2\frac{1}{2n^3}\left[\frac{11}{32n} - \frac{(1+\frac{1}{2}\epsilon)}{(2l+1)}\right] \tag{5.77}$$

where $\epsilon = 0$ for the singlet spin combination, whereas for the triplet

$$\epsilon = \begin{cases} \dfrac{-(3l+4)}{(l+1)(2l+3)}, & \text{for } j = l+1 \\ \dfrac{1}{l(l+1)}, & \text{for } j = l \\ \dfrac{(3l-1)}{l(2l-1)}, & \text{for } j = l-1 \end{cases} \tag{5.78}$$

[In hydrogen, where the proton spin ($\mathbf{S}_p$) contributes only at the hyperfine level, we used $\mathbf{J}$ for the sum of the electron's spin and orbital angular momentum ($\mathbf{J} = \mathbf{L} + \mathbf{S}_e$); for the *total* angular momentum we needed a new letter: $\mathbf{F} = \mathbf{J} + \mathbf{S}_p = \mathbf{L} + \mathbf{S}_e + \mathbf{S}_p$. In positronium the two spins contribute on an equal footing, and it is customary to combine them first ($\mathbf{S} = \mathbf{S}_1 + \mathbf{S}_2$) and use $\mathbf{J}$ for the total: $\mathbf{J} = \mathbf{L} + \mathbf{S} = \mathbf{L} + \mathbf{S}_1 + \mathbf{S}_2$.]

The Lamb shift, of order α^5mc^2, makes a smallish correction to this; however, since the "accidental" degeneracy is already broken at the fine structure level in positronium, the Lamb shift loses much of its interest, and I shall not consider it here. There is, however, an entirely new perturbation, with no analog

* This leads to some terminological confusion in the literature. I'll use the words "fine structure" for *all* perturbations of order α^4mc^2, except the pair annihilation term (see below) including the spin-spin and positron spin-orbit couplings, whose analogs in hydrogen would be called "hyperfine."

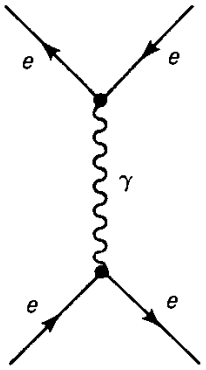


Figure 5.6 Pair annihilation diagram, which affects the spectrum of positronium but does not occur in hydrogen.

in hydrogen, resulting from the fact that e^+ and e^- can annihilate temporarily to produce a virtual photon. In the Feynman picture this process is represented by the diagram in Figure 5.6. Because it requires that the electron and positron coincide, this perturbation is proportional to $|\psi(0)|^2$, and hence occurs only when $l = 0$. [See remarks before eq. (5.67).] Moreover, since the photon carries spin 1, it takes place only in the triplet configuration. We are not in a position yet to calculate this correction, but it raises the energy of the triplet S states by an amount

$$\Delta E_{\text{annih.}} = \alpha^4 mc^2 \frac{1}{4n^3} \qquad (l = 0,\ s = 1) \tag{5.79}$$

Note that it is of the same order as the fine structure. The complete splitting of the $n = 1$ and $n = 2$ Bohr levels in positronium is indicated on Figure 5.7. [Positronium states are conventionally labeled $n^{(2s+1)}l_j$, with l given in spectroscopist's notation (S for $l = 0$, P for $l = 1$, D for $l = 2$, etc.), and s the total spin (0 for the singlet, 1 for the triplet).]

As in the case of hydrogen, positronium can make transitions from one state to another with the emission or absorption of a photon, whose wavelength is determined by the difference in energy between the two levels. *Unlike* hydrogen, positronium can disintegrate completely, the positron annihilating the electron to produce two or more real photons. (Why can't they go to a *single* real photon?) The charge conjugation number for positronium is $(-1)^{l+s}$, while for n photons $C = (-1)^n$ (see page 129). Thus charge conjugation invariance prescribes the selection rule

$$(-1)^{l+s} = (-1)^n \tag{5.80}$$

for the decay of positronium in state l, s into n photons. Since the positron and electron overlap only when $l = 0$, such decays occur only from S states.* Evidently, the singlet ($s = 0$) must go to an even number of photons (typically two), whereas the triplet ($s = 1$) must go to an odd number (typically three). As we

* Actually, positronium can in principle decay directly from a state with $l > 0$ by a higher-order process, but it is much more likely to cascade down to an S state first, and decay from there.

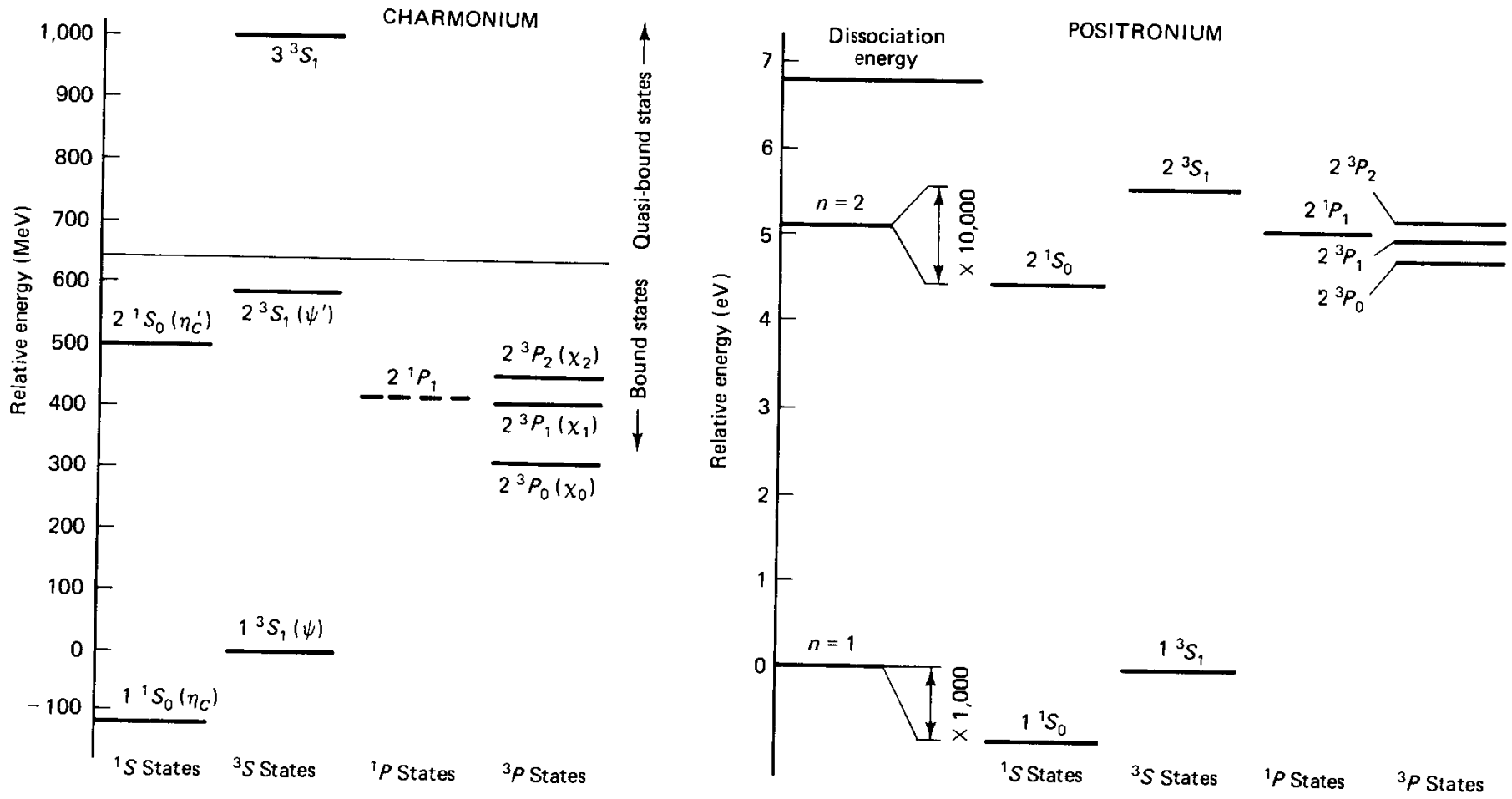


Figure 5.7 Spectrum of energy levels in positronium and charmonium. Note that the scale is greater by a factor of 100 million for charmonium. In positronium the various combinations of angular momentum cause only minuscule shifts in energy (shown by expanding the vertical scale), but in charmonium the shifts are much larger. All energies are given with reference to the 1^3S_1 state. At 6.8 electron volts positronium dissociates. At 633 MeV above the energy of the ψ charmonium becomes quasi-bound, because it can decay into D^0 and $\bar{D}^0$ mesons. (From "Quarkonium," by E. Bloom and G. Feldman. Copyright © May 1982 by Scientific American, Inc. All rights reserved.)

shall see in Chapter 7, the two-photon annihilation cross section* for low relative velocity v in the singlet configuration is

$$\sigma = 4\pi\alpha^2\left(\frac{\hbar^2}{m^2cv}\right) \tag{5.81}$$

In terms of this cross section, the decay rate is

$$\Gamma = \sigma v|\psi(0)|^2 \tag{5.82}$$

Using equation (5.61) with the Bohr radius appropriate to positronium [eq. (5.74)], we obtain

$$\Gamma = \alpha^5 mc^2 \frac{1}{2\hbar n^3} \tag{5.83}$$

In particular, the lifetime, $\tau = \Gamma^{-1}$, of the ground state is

$$\tau = \frac{2\hbar}{\alpha^5 mc^2} = 1.25 \times 10^{-10} \text{ sec} \tag{5.84}$$

5.7 QUARKONIUM

In the quark model all mesons are two-particle bound states, $q_1\bar{q}_2$, and it is natural to ask if the methods we have developed for hydrogen and positronium can be applied to mesons as well. There are two immediate difficulties with this program:

1. Unlike hydrogen and positronium, in which the forces at work are entirely electromagnetic and well understood, quarks are bound by the strong force. We don't know what potential to use, in place of Coulomb's law, or what the strong analog to magnetism might be, to obtain the spin couplings. In principle, these are derivable from chromodynamics, but no one at present knows how to do the calculation. Still, we can make some educated guesses, for chromodynamics is very similar in structure to electrodynamics, except for some nonlinear terms which, in the light of asymptotic freedom, probably don't contribute much at short distances. In the language of QCD (quantum chromodynamics), the short-distance behavior is dominated by one-gluon exchange, just as in QED (quantum electrodynamics) it is dominated by one-photon exchange. Since the gluon and the photon are both massless spin-1 particles, the interactions are, in this approximation, *identical,* apart from the overall coupling constant (α_s in place of α) and various so-called "color factors," which result from counting the number of different colors of gluons that contribute to a given process. All of this will be discussed in Chapter 9; for now, the essential point is that at short range we expect a Coulombic potential, $V \sim 1/r$, and a fine-hyperfine structure that is qualitatively similar to that of hydrogen and positronium. On the other

* This paragraph anticipates some material on cross sections, decay rates, and lifetimes that will be discussed in Chapter 6. It is included here for completeness.

hand, at large distances we have to account for quark confinement: the potential must increase without limit. The precise functional form of $V(r)$ at large r is rather speculative; some authors favor a harmonic oscillator potential, $V \sim r^2$, others a logarithmic dependence, $V \sim \ln(r)$. Perhaps the simplest is a *linear* potential, $V \sim r$, corresponding to a constant force. The fact is, *any* of these can match the data presently available reasonably well, because they do not differ substantially over the rather narrow range of distances for which we have sensitive probes. For our purposes, then, we may as well choose

$$V(r) = -\frac{4}{3}\frac{\alpha_s \hbar c}{r} + F_0 r \tag{5.85}$$

where α_s is the chromodynamic analog to the fine structure constant, and $\frac{4}{3}$ is the appropriate color factor, which we'll calculate in Chapter 9. Experimentally, F_0 is about 16 tons (!), which is to say that a quark and an antiquark attract one another with a force of at least 16 tons, regardless of how far apart they are.* This perhaps makes it easier to understand why no one has ever managed to pull a quark out of a hadron.

2. The light quark (u, d, s) mesons are intrinsically relativistic, since the binding energies (typically a few hundred MeV) are not small compared to the constituent masses. Everything we have done was based on nonrelativistic quantum mechanics, in particular, the Schrödinger equation. It is true that we included a relativistic correction, but it was only a lowest-order approximation to begin with, and the use of perturbation theory for this and the other contributions to fine and hyperfine structure was predicated on the assumption that they make very small modifications in the energy levels (the splitting is grossly exaggerated in Figures 5.2 and 5.7). Unfortunately, an exact solution of the relativistic bound-state problem is not available at this stage. For the heavy quarks (c, b, t), though, the nonrelativistic theory should work reasonably well.† Even here, however, the binding energy (E) is such a substantial fraction of the total that we are disposed to regard the various energy levels as representing *different mesons*, with masses given by

$$M = m_1 + m_2 + E/c^2 \tag{5.86}$$

Shortly before the discovery of the ψ, Politzer and Appelquist[19] suggested that if a heavy "charm" quark existed (as Glashow and others had proposed) it should form a nonrelativistic bound state $c\bar{c}$, with a spectrum of energy levels

* At extremely short distances F_0 and α_s themselves decrease, leading to asymptotic freedom, but for now we shall treat them as constants.

† In the case of a purely Coulombic potential, $E/mc^2 \sim \alpha^2$ [eq. (5.26)], so the binding energy is a fixed fraction of the rest energy of the constituents, regardless of their mass. If this were true in chromodynamics, the heavy quark mesons would be no less relativistic than the light ones. However, two things work in our favor here, both attributable to the fact that a bound state of two heavy particles is typically *smaller* than one of two light particles [the Bohr radius, for example, goes like $1/m$ (eq. (5.29)]: (i) the light quark mesons are more sensitive to the confining term in the potential, and the binding energies of a linear potential go like $m^{-1/3}$, not m, and (ii) because of asymptotic freedom, α_s itself is smaller for the heavy quark states; experimentally, it's about 0.5 for the ϕ meson, but closer to 0.2 for the ψ.

similar to positronium. They called the system "charmonium" (which does more to emphasize the parallel than to beautify the language). When the ψ was found, in 1974, it was quickly identified as the 1^3S_1 state of charmonium. (In the SLAC experiments the ψ was produced from e^+e^- annihilation through a virtual photon: $e^+e^- \rightarrow \gamma \rightarrow \psi$, so it has to carry the same quantum numbers as γ—in particular, spin 1. Thus it could not be the *ground* state of charmonium, but presumably it was the lowest-lying state with total angular momentum 1.) Consulting the positronium level diagram (Fig. 5.7), we immediately anticipate a spin-0 state at lower mass (the 1^1S_0) and six $n = 2$ configurations. Within two weeks the ψ' (2^3S_1) was found. This was easy, because it again carries the same spin—and parity—as the photon; it was produced in the same way as the ψ, simply by cranking up the beam energy. In due course all the $n = 1$ and $n = 2$ states were discovered,[20] save for the 2^1P_1 at a predicted mass of about 3500 MeV/c^2, which presents special experimental problems. The following nomenclature has been adopted: singlet S states (spin 0) are called η_c's; triplet S states (spin 1) are ψ's; and triplet P states (spin 0, 1, or 2) are designated χ_0, χ_1, χ_2. For a while the value of n was indicated by primes, but this quickly got out of hand, and the current practice is simply to list the mass parenthetically; thus for $n = 1$ we have $\psi = \psi(3100)$; for $n = 2$, $\psi' = \psi(3685)$; for $n = 3$, $\psi'' = \psi(4030)$; for $n = 4$, $\psi''' = \psi(4160)$; and so on.* The correlation between states of charmonium and those of positronium is almost perfect (Fig. 5.7). Bear in mind that the gap between the two $n = 1$ levels, which would be called hyperfine splitting in the case of hydrogen, is a factor of 10^{11} greater in charmonium than in positronium. Yet even over so huge a change of scale, the ordering of the energy levels and, for a given value of n, their relative spacing, are strikingly similar.† Unfortunately, the exact solutions to the Schrödinger equation with linear-plus-Coulomb potential are not known, and I cannot give you a simple formula for the "Bohr" energies. However, it can, of course, be done numerically (see Table 5.2), and F_0 can then be chosen so as to fit the data (that's how the value of 16 tons—or, in more sensible units, 900 MeV/fm—was obtained).[21] (See Problem 5.21.)

All the charmonium states with $n = 1$ and $n = 2$ are relatively long-lived, because the OZI rule (Chap. 2, Sect. 2.5) suppresses their strong decays. For $n \geqq 3$ the charmonium masses lie above the threshold for (OZI-allowed) production of two charmed D mesons (D^0, $\bar{D}^0$ at a mass of 1865 MeV/c^2, or $D^{\pm}$, at 1869 MeV/c^2). Their lifetimes are therefore much shorter, and we call them "quasi-bound states" (see Fig. 5.8). Quasi-bound states of charmonium have been observed going up at least as high as $n = 4$.

In the aftermath of the November Revolution there was widespread speculation about the possible existence of a third quark generation (b and t), and

* Some authors, including those of the Particle Data Booklet, number states consecutively, starting with 1 for *each* combination of s, l, and j, so that what I call a $2P$ state (in Fig. 5.7) is listed as $1P$. Sorry about that. Incidentally, the $\psi(3770)$ is a displaced 3^3D_1 state, and does not really belong in this hierarchy.

† In the early days there was some consternation when the η_c and η_c' appeared to be about 150 MeV/c^2 too light, but the experiments were wrong, and the corrected values are in accord with theory. See reference 20.

TABLE 5.2 "BOHR" ENERGY LEVELS FOR LINEAR-PLUS-COULOMB POTENTIAL (EQ. 5.85) WITH VARIOUS VALUES OF F_0. THEY ARE FOR S-STATES ($l = 0$ AND ASSUME $\alpha_s = 0.2$, $m = 1500$ MeV/c^2 (REDUCED MASS, 750 MeV/c^2).

F_0 (MeV/fm)	E_1 (MeV)	E_2 (MeV)	E_3 (MeV)	E_4 (MeV)
500	307	677	961	1210
1000	533	1100	1550	1940
1500	727	1480	2040	2550

Numerical results from unpublished tables prepared by Richard E. Crandall.

in 1976 Eichten and Gottfried[22] predicted that "bottomonium" ($b\bar{b}$) would exhibit a hierarchy of bound states even richer than charmonium (Fig. 5.9). The bottom analog to the D meson (to wit, the B) had an estimated mass large enough that not only the $n = 1$ and $n = 2$, but also the $n = 3$ levels should be bound. In 1977 the upsilon meson was discovered, and immediately interpreted as the 1^3S_1 state of bottomonium. At present, the $^3S_1$ states have been found for n up to 4, as well as the six ^3P states for $n = 2$ and $n = 3$. It happens that the level spacings in the ψ and Υ systems are remarkably similar (Fig. 5.10), in spite of the fact that the bottom quark is more than three times as heavy as the charm quark. For a purely Coulombic potential the spacing is directly proportional to mass, [eq. (5.26)], whereas for a purely linear potential it goes like $m^{-1/3}$ (Problem 5.20). If we take the interquark potential (5.85) seriously, the equal spacing must be due to a conspiracy of the two terms and an accidental feature of the particular value of F_0 (which we adjusted, remember, to fit the ψ data). The much heavier "toponium" ($t\bar{t}$) system should be more sensitive to the short-distance (Coulom-

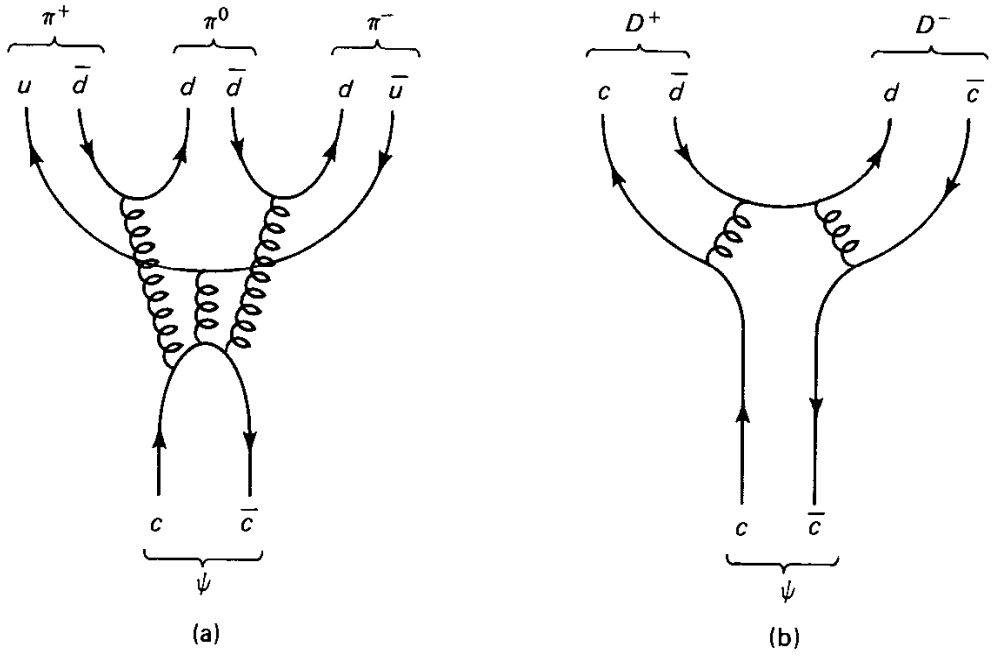


Figure 5.8 (a) OZI-suppressed decay for charmonium below the $D\bar{D}$ threshold, (b) OZI-allowed decay for charmonium above the $D\bar{D}$ threshold.

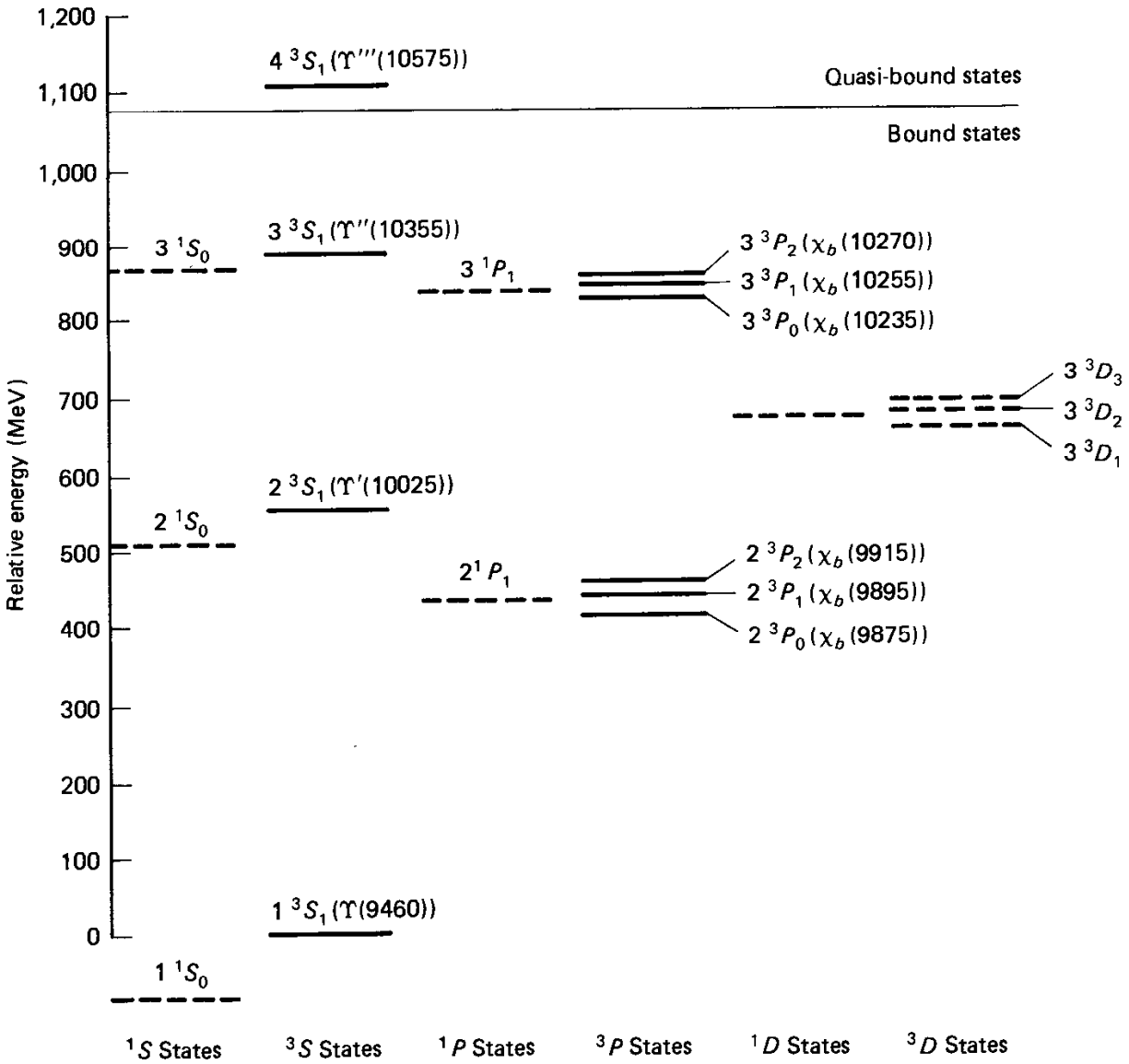


Figure 5.9 Bottomonium. Note that there are far more bound states than for charmonium—compare Fig. 5.7. (From "Quarkonium," by E. Bloom and G. Feldman. Copyright © May 1982 by Scientific American, Inc. All rights reserved.)

bic) part of the potential, and presumably will not share the equal spacing of ψ and Υ.*

Incidentally, charmonium and bottomonium are only the first two of six possible "quarkonium" systems; in addition to toponium, we can look forward eventually to the mixed combinations $c\bar{b}$, $c\bar{t}$, and $b\bar{t}$. The model for these, however, will be muonium (μ^+e^-), rather than positronium.

5.8 LIGHT QUARK MESONS

Consider now the mesons made entirely out of *light* quarks (u, d, s). These are relativistic systems, remember, so we cannot use the Schrödinger equation, and

* However, a logarithmic potential gives level separations that are independent of mass, so if the toponium system *does* match the ψ and Υ splittings, it will be an argument for using $V(r) \sim \ln(r)$.

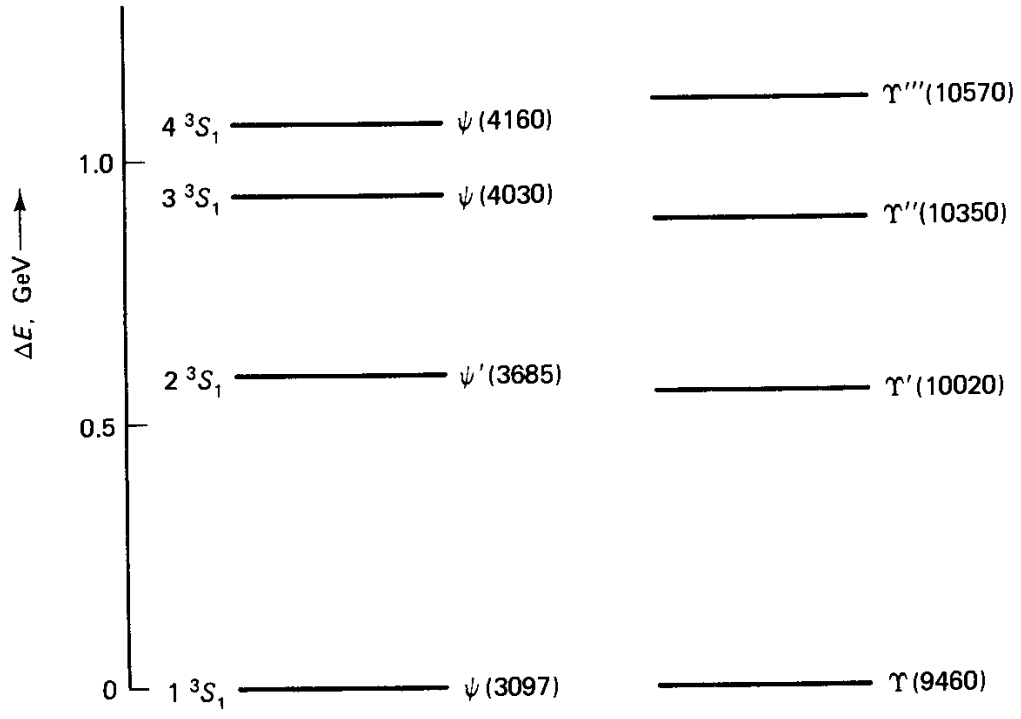


Figure 5.10 Level spacings in the ψ and Υ systems. (*Source:* D. H. Perkins, *Introduction to High-Energy Physics,* 2d Ed. (Reading, MA: Addison-Wesley, 1982), p. 214.)

the theory is rather limited.[23] In particular, we shall not concern ourselves with the spectrum of excited states (Table 4.6), as we did in the case of the heavy quarks, but will confine our attention to the ground state, with $l = 0$. The quark spins can be antiparallel (singlet state, $s = 0$) or parallel (triplet state, $s = 1$); the former configuration yields the pseudoscalar nonet, the latter gives the vector nonet (Fig. 5.11).

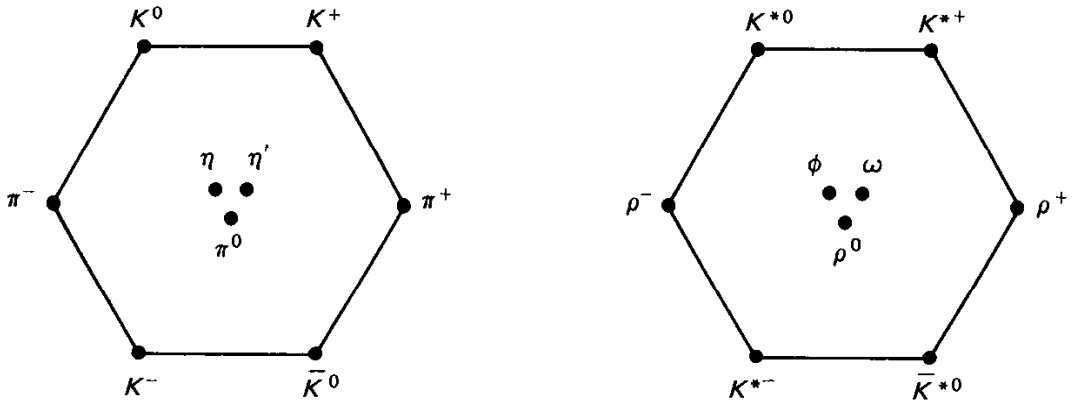
To begin with I want to clear up a problem that was not resolved in Chapter 1. We obtained nine mesons simply by combining a quark and an antiquark in all possible combinations (Chap. 1, Sect. 1.8), but this left three neutral states with strangeness 0 ($u\bar{u}$, $d\bar{d}$, and $s\bar{s}$), and it was not clear which of these was the π^0, which the η, and which the η' (or, in the vector case, the ρ^0, ω, and ϕ). We are now in a position to resolve the ambiguity. The up and down quarks constitute an isospin doublet:

$$u = |\tfrac{1}{2}\,\tfrac{1}{2}\rangle, \qquad d = |\tfrac{1}{2}\,-\tfrac{1}{2}\rangle \tag{5.87}$$

So too do the antiquarks:

$$\bar{d} = -|\tfrac{1}{2}\,\tfrac{1}{2}\rangle, \qquad \bar{u} = |\tfrac{1}{2}\,-\tfrac{1}{2}\rangle \tag{5.88}$$

(Notice that $\bar{d}$ carries $I_3 = +\frac{1}{2}$, and $\bar{u}$ has $I_3 = -\frac{1}{2}$; within a multiplet, the particle with the higher charge is assigned the greater I_3. The minus sign is a technical detail,[24] which does not affect the argument here in any essential way.) When we combine two particles with $I = \frac{1}{2}$, we obtain an isotriplet (eq. 4.15)



Pseudoscalar nonet Vector nonet

Figure 5.11 Light-quark mesons with $I = 0$.

$$\begin{cases} |1 \quad 1\rangle = -u\bar{d} \\ |1 \quad 0\rangle = (u\bar{u} - d\bar{d})/\sqrt{2} \\ |1 \quad -1\rangle = d\bar{u} \end{cases} \tag{5.89}$$

and an isosinglet (4.16)

$$|00\rangle = (u\bar{u} + d\bar{d})/\sqrt{2} \tag{5.90}$$

In the case of the pseudoscalar mesons the triplet is the pion; for the vector mesons it is the ρ. Evidently the π^0 (or the ρ^0) is neither $u\bar{u}$ nor $d\bar{d}$, but rather the linear combination

$$\pi^0, \rho^0 = (u\bar{u} - d\bar{d})/\sqrt{2} \tag{5.91}$$

If you could pull a π^0 apart, half the time you'd get a u in one hand and a $\bar{u}$ in the other, and half the time you'd get a d and a $\bar{d}$.

This leaves two $I = 0$ states (the isosinglet combination, equation (5.90), and $s\bar{s}$) which must represent η and η' (or ω and ϕ). Here the situation is not so clean, for these particles carry identical quantum numbers, and they tend in practice to "mix." In the case of the pseudoscalars the physical states appear to be

$$\eta = (u\bar{u} + d\bar{d} - 2s\bar{s})/\sqrt{6} \tag{5.92}$$

$$\eta' = (u\bar{u} + d\bar{d} + s\bar{s})/\sqrt{3} \tag{5.93}$$

whereas for the vector mesons we find

$$\omega = (u\bar{u} + d\bar{d})/\sqrt{2} \tag{5.94}$$

$$\phi = s\bar{s} \tag{5.95}$$

To the extent that the Eightfold Way is a good symmetry, the pseudoscalar combinations are more "natural", since the η', which treats u, d, and s symmetrically, is unaffected by $SU(3)$ transformations; it is a "singlet" under $SU(3)$, in exactly the same sense that the π^0 is a singlet under $SU(2)$ (isospin). The η, meanwhile, transforms as part of an $SU(3)$ "octet", whose other members are the three pions and the four K's. (This is, in fact, the original pseudoscalar octet.)

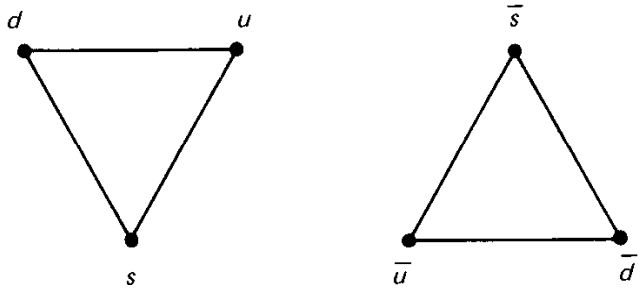


Figure 5.12 Quarks and antiquarks.

By contrast, neither the ϕ nor the ω is an $SU(3)$ singlet. They are, you might say, "maximally" mixed, since the strange quark is isolated from the other two. [Incidentally, the other meson nonets seem to follow the ϕ–ω mixing pattern, with the possible exception of the D and E (Table 4.6).]

Meanwhile, the strange mesons are constructed by combining an s quark with u or d

$$K^+ = u\bar{s}, \qquad K^0 = d\bar{s}, \qquad \bar{K}^0 = -s\bar{d}, \qquad K^- = s\bar{u} \tag{5.96}$$

In the language of group theory, the three light quarks belong to the fundamental representation (denoted 3) of $SU(3)$, whereas the antiquarks belong to the conjugate representation ($\bar{3}$) (See Fig. 5.12). What we have done is to combine these representations, obtaining an octet and a singlet:

$$3 \otimes \bar{3} = 8 \oplus 1 \tag{5.97}$$

just as in Chapter 4 we combined two two-dimensional (spin-$\frac{1}{2}$) representations of $SU(2)$ to obtain a triplet and a singlet:*

$$2 \otimes \bar{2} = 3 \oplus 1 \tag{5.98}$$

If $SU(3)$ were a perfect symmetry, all the particles in a given supermultiplet would have the same mass. But they obviously do *not;* the K weighs more than three times the π, for example. As I indicated in Chapter 4, the breaking of flavor symmetry is due to the fact that the quarks themselves have unequal masses; the u and d quarks weigh about the same (which is why isospin is such a good symmetry) but the s quark is substantially heavier. Roughly speaking, the K's weigh more than the π's because they contain an s in place of a u or d. But that cannot be the whole story, for if it were, the ρ's would weigh the same as the π's; after all, they have the same quark content and are both in the spatial ground state ($n = 1, l = 0$). Since the pseudoscalar and vector mesons differ only in the relative orientation of the quark spins, the difference in their masses must be attributed to a spin-spin interaction, the QCD analog to hyperfine splitting in the ground state of hydrogen. The QED formula, remember, is

* Unfortunately (from the point of view of notational consistency) representations of $SU(3)$ are customarily labeled by their dimensionality, whereas representations of $SU(2)$ are more often identified by their spin, so that equation (5.98) would usually be written $\frac{1}{2} \otimes \frac{1}{2} = 1 \oplus 0$. By the way, it happens that the fundamental representation of $SU(2)$ is equivalent to its conjugate; there's only one kind of spin $\frac{1}{2}$. That's why we were able in equation (5.88), to represent $\bar{u}$ and $\bar{d}$ in terms of ordinary isospin-$\frac{1}{2}$ states. For $SU(3)$ this is no longer the case.

TABLE 5.3 PSEUDOSCALAR AND VECTOR MESON MASSES (MeV/c^2)

Meson	Calculated	Observed
π	140	138
K	484	496
η	559	549
ρ	780	776
ω	780	783
K^*	896	892
ϕ	1032	1020

$$\Delta E_{\text{hf}} = \frac{8\pi\gamma_p e^2}{3mm_p c^2}(\mathbf{S}_e \cdot \mathbf{S}_p)|\psi_{100}(0)|^2 \tag{5.60}$$

It seems reasonable to suppose that the spin-spin coupling in QCD has a similar structure; that is, it should be proportional to the dot product of the spins and inversely proportional to the product of the constituent masses. On this assumption we are led to the following meson mass formula:

$$M(\text{meson}) = m_1 + m_2 + A\frac{(\mathbf{S}_1 \cdot \mathbf{S}_2)}{m_1 m_2} \tag{5.99}$$

The coefficient A is related to $|\psi(0)|^2$, which we are not in a position to calculate. One assumes it is the same for all the vector and pseudoscalar mesons, since they occupy the same quantum state.* By the usual trick of squaring $\mathbf{S} = \mathbf{S}_1 + \mathbf{S}_2$ [see eq. (5.63)], we find that

$$\mathbf{S}_1 \cdot \mathbf{S}_2 = \begin{cases} \frac{1}{4}\hbar^2, & \text{for } s = 1 \text{ (vector mesons)} \\ -\frac{3}{4}\hbar^2, & \text{for } s = 0 \text{ (pseudoscalars)} \end{cases} \tag{5.100}$$

For constituent masses $m_u = m_d = 310$ MeV/c^2, $m_s = 483$ MeV/c^2, the "best-fit" value of A is $(2m_u/\hbar)^2 160$ MeV/c^2, and we obtain the results in Table 5.3. Considering its somewhat shaky theoretical foundation, equation (5.99) works surprisingly well, matching seven independent meson masses to an accuracy of about 1%, with three adjustable input parameters. (Notice, however, that the η' is not included in the table. See Problem 5.22.)

5.9 BARYONS

Some day, presumably, we shall be able to make nonrelativistic heavy-quark baryons—*ccc*, perhaps, or even, *cbt*. These are the baryonic relatives of quar-

* In my view this is a questionable assumption: (i) For a Coulombic potential we know [eqs. (5.61 and 5.29)] that $|\psi(0)|^2$ goes like the *cube* of the reduced mass (for a linear potential it is proportional to m). Why preserve the explicit mass dependence in the denominator of equation (5.99), if we are prepared to ignore it in the numerator? (ii) The central ($I_3 = 0$) members of each nonet combine a quark with its own antiquark, and hence admit annihilation diagrams, just as positronium does, that are not possible for the other mesons. However, nothing succeeds like success, and equation (5.99) works remarkably well.

TABLE 5.4 LIGHT-QUARK BARYONS (J = SPIN, P = PARITY, S = STRANGENESS, I = ISOSPIN. THIS IS NOT A COMPLETE LIST; BARYONS WITH SPINS AS HIGH AS $\frac{11}{2}$ HAVE BEEN OBSERVED.)

$SU(3)$ Representation	J^P	$S = 0$	$S = -1$		$S = -2$	$S = -3$
			$I = 0$	$I = 1$		
8	$\frac{1}{2}^+$	$N(939)$	$\Lambda(1116)$	$\Sigma(1193)$	$\Xi(1318)$	
10	$\frac{3}{2}^+$	$\Delta(1232)$		$\Sigma(1385)$	$\Xi(1533)$	$\Omega(1672)$
1	$\frac{1}{2}^-$		$\Lambda(1405)$			
	$\frac{3}{2}^-$		$\Lambda(1520)$			
8	$\frac{1}{2}^-$	$N(1535)$	$\Lambda(1670)$	$\Sigma(1750)$		
	"	$N(1700)$	$\Lambda(1870)$			
	$\frac{3}{2}^-$	$N(1520)$	$\Lambda(1690)$	$\Sigma(1670)$	$\Xi(1820)$?	
	"	$N(1700)$		$\Sigma(1940)$?		
	$\frac{5}{2}^-$	$N(1670)$	$\Lambda(1830)$	$\Sigma(1765)$		
10	$\frac{1}{2}^-$	$\Delta(1650)$				
	$\frac{3}{2}^-$	$\Delta(1670)$				
8	$\frac{3}{2}^+$	$N(1810)$	$\Lambda(1860)$			
8	$\frac{5}{2}^+$	$N(1688)$	$\Lambda(1815)$	$\Sigma(1915)$	$\Xi(2030)$?	
10	$\frac{1}{2}^+$	$\Delta(1910)$				
	$\frac{3}{2}^+$					
	$\frac{5}{2}^+$	$\Delta(1890)$				
	$\frac{7}{2}^+$	$\Delta(1950)$		$\Sigma(2030)$		
8	$\frac{1}{2}^+$	$N(1470)$		$\Sigma(1660)$		
10	$\frac{3}{2}^+$	$\Delta(1690)$				

Source: S. Gasiorowicz and J. L. Rosner. *Am. J. Phys.* **49**, 954 (1981).

konium—"quarkelium," you might call it, since the nearest atomic analog would be helium. At present, though, it is hard enough to make a baryon with *one* heavy quark, never mind *three*, and I won't speculate here about the heavy quark baryon spectrum. On the other hand, the array of observed *light* quark baryons is immense (see Table 5.4). Baryons are harder to analyze than mesons, for several reasons. In the first place, a baryon is a three-body system. There's not just *one* orbital angular momentum to consider, but *two* (see Fig. 5.13). We'll concentrate on the ground state, for which $l = l' = 0$. In that case the angular momentum of the baryon comes entirely from the combined spins of the three quarks. Now, the quarks carry spin $\frac{1}{2}$, so each can occupy either of two states: "spin up" ($\uparrow$), or "spin down" ($\downarrow$). Thus we have *eight* possible states for the three quarks: ($\uparrow\uparrow\uparrow$), ($\uparrow\uparrow\downarrow$), ($\uparrow\downarrow\uparrow$), ($\uparrow\downarrow\downarrow$), ($\downarrow\uparrow\uparrow$), ($\downarrow\uparrow\downarrow$), ($\downarrow\downarrow\uparrow$), and ($\downarrow\downarrow\downarrow$). But these are not

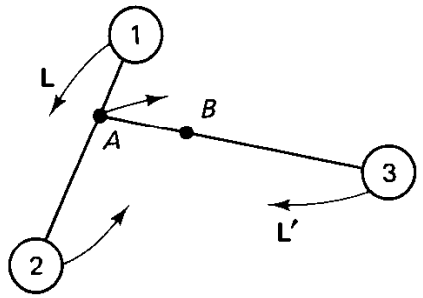


Figure 5.13 Orbital angular momenta for a three-body system. L is the angular momentum of 1 and 2 about their center of mass (A); L′ is the angular momentum of this combination and 3 about the center of mass of all three (B).

the most convenient configurations to work with, because they are not eigenstates of the total angular momentum. As we found in Example 4.2, the quark spins can combine to give a total of $\frac{3}{2}$ or $\frac{1}{2}$, and the latter can be achieved in two distinct ways. Specifically

$$\left.\begin{aligned} |\tfrac{3}{2}\ \tfrac{3}{2}\rangle &= (\uparrow\uparrow\uparrow) \\ |\tfrac{3}{2}\ \tfrac{1}{2}\rangle &= (\uparrow\uparrow\downarrow + \uparrow\downarrow\uparrow + \downarrow\uparrow\uparrow)/\sqrt{3} \\ |\tfrac{3}{2}\ -\tfrac{1}{2}\rangle &= (\downarrow\downarrow\uparrow + \downarrow\uparrow\downarrow + \uparrow\downarrow\downarrow)/\sqrt{3} \\ |\tfrac{3}{2}\ -\tfrac{3}{2}\rangle &= (\downarrow\downarrow\downarrow) \end{aligned}\right\} \quad \text{spin } \tfrac{3}{2}\ (\psi_s) \tag{5.101}$$

$$\left.\begin{aligned} |\tfrac{1}{2}\ \tfrac{1}{2}\rangle_{12} &= (\uparrow\downarrow - \downarrow\uparrow)\uparrow/\sqrt{2} \\ |\tfrac{1}{2}\ -\tfrac{1}{2}\rangle_{12} &= (\uparrow\downarrow - \downarrow\uparrow)\downarrow/\sqrt{2} \end{aligned}\right\} \quad \text{spin } \tfrac{1}{2}\ (\psi_{12}) \tag{5.102}$$

$$\left.\begin{aligned} |\tfrac{1}{2}\ \tfrac{1}{2}\rangle_{23} &= \uparrow(\uparrow\downarrow - \downarrow\uparrow)/\sqrt{2} \\ |\tfrac{1}{2}\ -\tfrac{1}{2}\rangle_{23} &= \downarrow(\uparrow\downarrow - \downarrow\uparrow)/\sqrt{2} \end{aligned}\right\} \quad \text{spin } \tfrac{1}{2}\ (\psi_{23}) \tag{5.103}$$

The spin-$\frac{3}{2}$ combinations are completely *symmetric,* in the sense that interchanging any two particles leaves the state untouched. The spin-$\frac{1}{2}$ combinations are partially *anti*symmetric: interchange of two particles switches the sign. The first set is antisymmetric in particles 1 and 2 (hence the subscript); the second is antisymmetric in 2 and 3. We could also, of course, construct a pair of states antisymmetric in 1 and 3:

$$\left.\begin{aligned} |\tfrac{1}{2}\ \ \tfrac{1}{2}\rangle_{13} &= (\uparrow\uparrow\downarrow - \downarrow\uparrow\uparrow)/\sqrt{2} \\ |\tfrac{1}{2}\ -\tfrac{1}{2}\rangle_{13} &= (\uparrow\downarrow\downarrow - \downarrow\downarrow\uparrow)/\sqrt{2} \end{aligned}\right\} \quad \text{spin } \tfrac{1}{2}\ (\psi_{13}) \tag{5.104}$$

However, these are not independent of the other two; as you can check for yourself,

$$|\ \rangle_{13} = |\ \rangle_{12} + |\ \rangle_{23} \tag{5.105}$$

In the language of group theory, the direct product of three fundamental (two-dimensional) representations of $SU(2)$ decomposes into the direct sum of a four-dimensional representation and two two-dimensional representations:*

$$2 \otimes 2 \otimes 2 = 4 \oplus 2 \oplus 2 \tag{5.106}$$

A second respect in which baryons are more complicated than mesons has to do with the Pauli exclusion principle. In its original formulation the Pauli principle stated that no two electrons can occupy the same quantum state. It was designed to explain why all the electrons in an atom don't simply cascade down to the ground state (ψ_{100}) (there wouldn't be much left of chemistry if they *did*): they cannot, because the ground state can only accommodate two of them—one spin up, one spin down. Once those positions are occupied, the next electrons are stuck in the first excited state, $n = 2, \ldots$, and so on. In this form the Pauli

* If the representations are labeled by spin, instead of dimensionality, equation (5.106) reads $\frac{1}{2} \otimes \frac{1}{2} \otimes \frac{1}{2} = \frac{3}{2} \oplus \frac{1}{2} \oplus \frac{1}{2}$. Incidentally, it is *also* possible to construct a spin-$\frac{1}{2}$ combination that is *symmetric* in particles 1 and 2: $|\ \rangle = |\ \rangle_{13} + |\ \rangle_{23}$. Some authors prefer to use $|\ \rangle_{12}$ and $|\ \rangle$, instead of $|\ \rangle_{12}$ and $|\ \rangle_{23}$.

principle seems a little ad hoc, but it is actually based on something far deeper: If two particles are *absolutely identical,* then the wave function should treat them on an equal footing. If someone secretly interchanges them, the physical state should not be altered. You might conclude from this that $\psi(1, 2) = \psi(2, 1)$, but that's a little *too* strong. Physical quantities are determined by the *square* of the wave function, so all we can say for sure is that $\psi(1, 2) = \pm\psi(2, 1)$: the wave function must either be *even*—symmetric—or *odd*—antisymmetric—under the interchange of two identical particles.* But which *is* it, even or odd? Nonrelativistic quantum-mechanics offers no answer; there are simply two classes of particles—bosons, for which the wave function is even, and fermions, for which it is odd. It is an empirical fact that all particles of integer spin are bosons, whereas those of $\frac{1}{2}$-integer spin are fermions. One of the major achievements of quantum field theory was the rigorous *proof* of this connection between "spin and statistics."

Bosons (integer spin) $\Rightarrow$ symmetric wave function: $\psi(1, 2) = \psi(2, 1)$
Fermions ($\frac{1}{2}$-integer spin) $\Rightarrow$ antisymmetric wave function: $\psi(1, 2) = -\psi(2, 1)$

Suppose that we have two particles, one in state ψ_α and the other in state ψ_β. If the particles are *distinct* (one a muon and one an electron, say) then it makes sense to ask *which* is in state ψ_α and which in state ψ_β. The wave function for the system is

$$\psi(1, 2) = \psi_\alpha(1)\psi_\beta(2)$$

if particle 1 is in ψ_α and 2 is in ψ_β, or

$$\psi(1, 2) = \psi_\beta(1)\psi_\alpha(2)$$

if it's the other way around. But if the two particles are *indistinguishable,* we cannot say which is in which state. If the particles are identical bosons, the wave function is the symmetric combination

$$\psi(1, 2) = (1/\sqrt{2})(\psi_\alpha(1)\psi_\beta(2) + \psi_\beta(1)\psi_\alpha(2)) \tag{5.107}$$

and if they are identical fermions, the wave function is the antisymmetric combination

$$\psi(1, 2) = (1/\sqrt{2})(\psi_\alpha(1)\psi_\beta(2) - \psi_\beta(1)\psi_\alpha(2)) \tag{5.108}$$

In particular, if you try to put two fermions (electrons, say) into the *same* state ($\psi_\alpha = \psi_\beta$) you get *zero;* it can't be done. That's the original Pauli exclusion principle; but we see now that it is not an ad hoc assumption, but rather a consequence of a structural requirement on the wave functions of identical particles. Notice, by the way, that the Pauli principle does not apply to bosons; you can put as many pions into the same state as you like. Nor is there any symmetry requirement for *distinguishable* particles; that's why we didn't have to worry about it when we were constructing meson wave functions (since one constituent

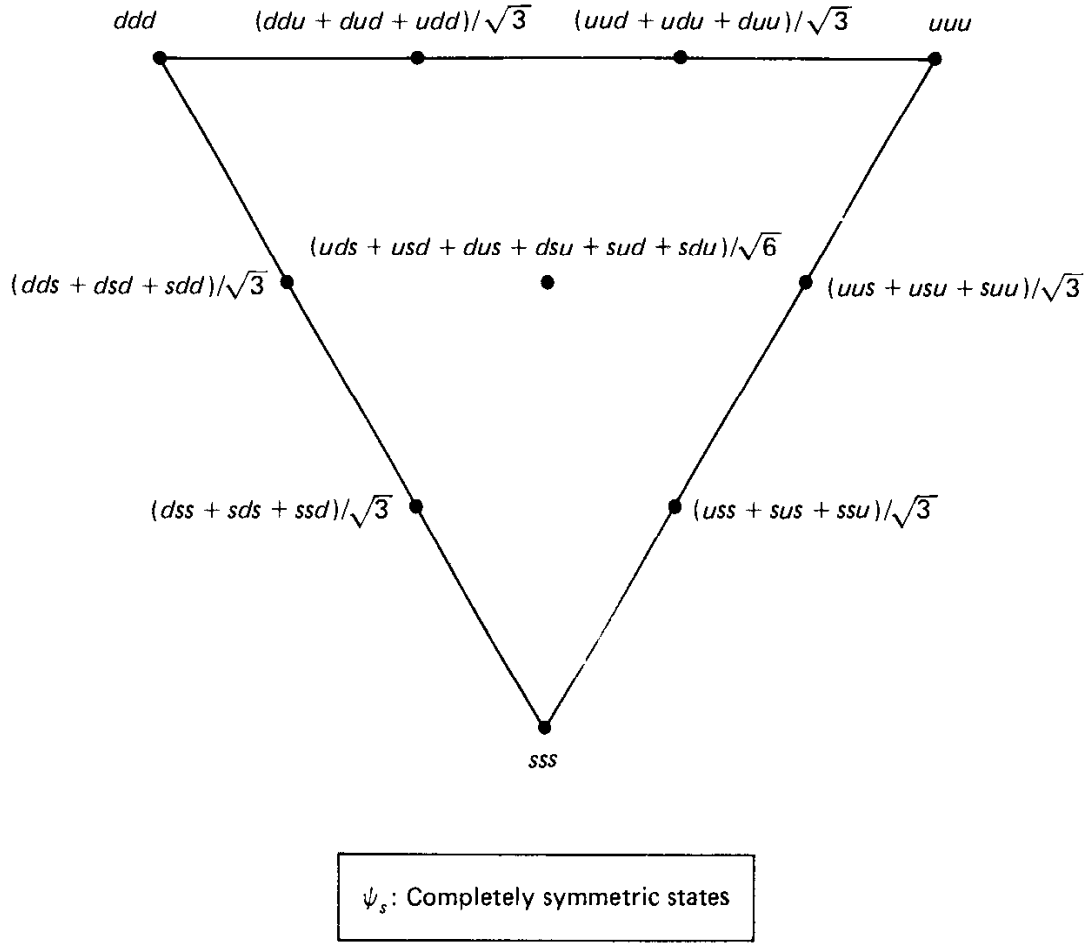
* From $|\psi(1, 2)|^2 = |\psi(2, 1)|^2$ it follows only that $\psi(1, 2) = e^{i\phi}\psi(2, 1)$. However, applying the interchange *twice* brings us back to where we started, so $e^{2i\phi} = 1$, and hence $e^{i\phi} = \pm 1$.

is a quark and the other an *antiquark,* they're *always* distinguishable). But in the case of the baryons we're putting three *quarks* together, and this time we must take the antisymmetrization requirement into account.

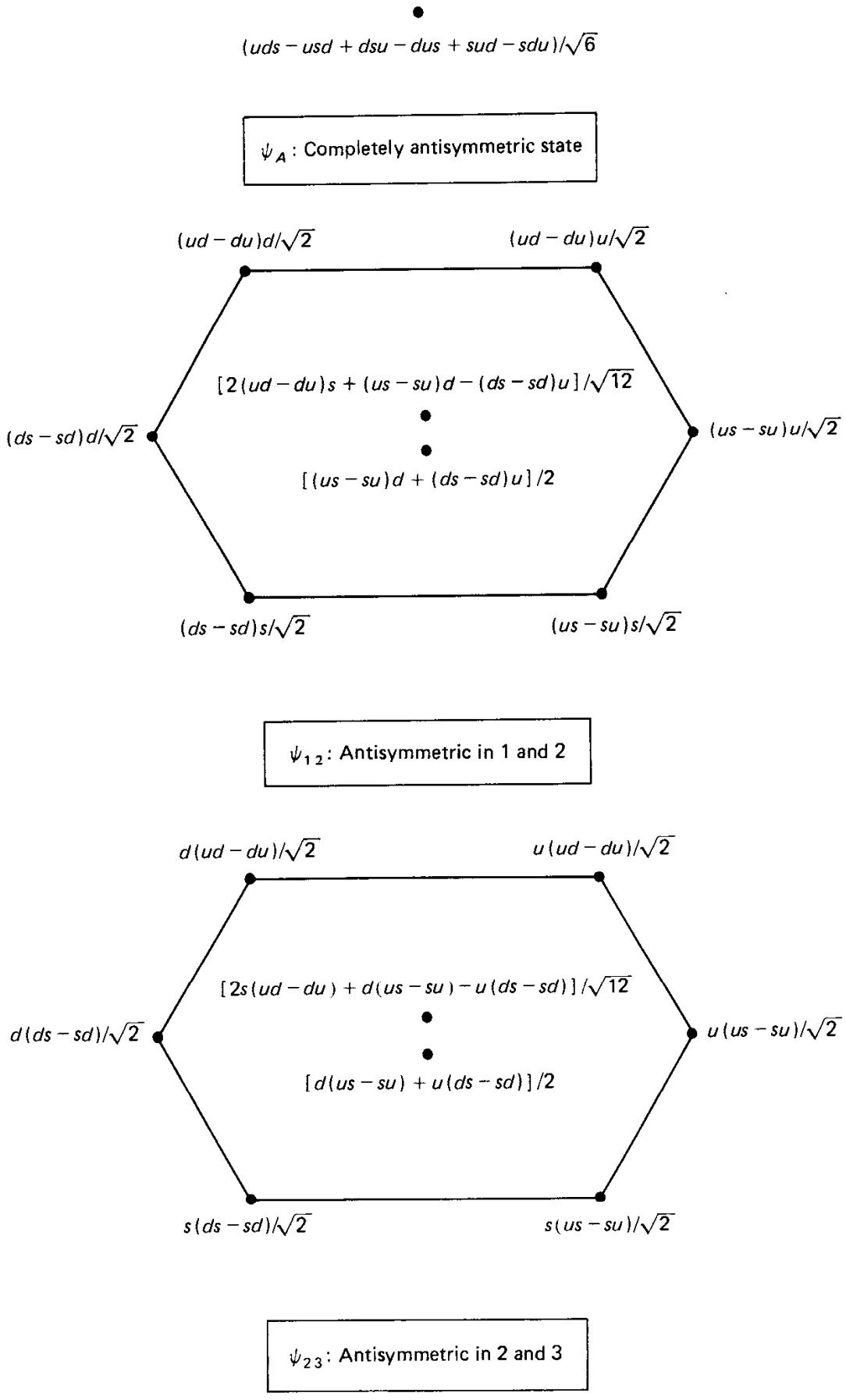
Now, the wave function of a baryon consists of several pieces; there is the *spatial* part, describing the locations of the three quarks; there is the *spin* part, representing their spins; there is a *flavor* component, indicating what combination of *u, d,* and *s* is involved; and there is a *color* term, specifying the colors of the quarks:

$$\psi = \psi(\text{space})\,\psi(\text{spin})\,\psi(\text{flavor})\,\psi(\text{color}) \tag{5.109}$$

It is the *whole works* that must be antisymmetric under the interchange of any two quarks.* We do not know the functional form of the spatial ground-state wave function, but it is surely symmetric; since $l = l' = 0$, there is no angular dependence at all. The spin state can either be completely symmetric ($j = \frac{3}{2}$) or of mixed symmetry ($j = \frac{1}{2}$). As for flavor, there are $3^3 = 27$ possibilities: *uuu, uud, udu, udd, . . . , sss,* which we reshuffle into symmetric, antisymmetric, and mixed combinations; they form irreducible representations of $SU(3)$, just as the analogous spin combinations form representations of $SU(2)$. These are conveniently displayed in eightfold-way patterns:



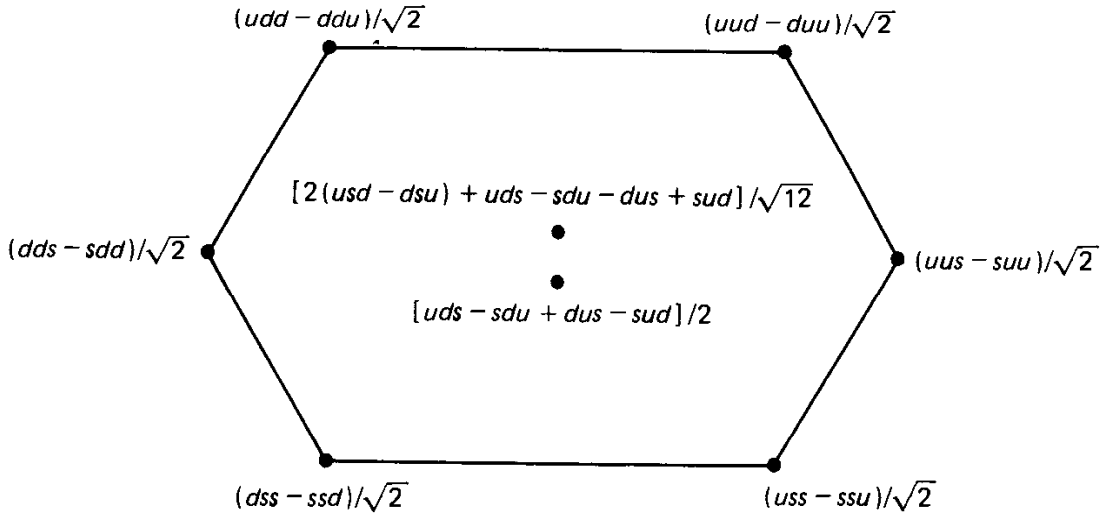
* Notice that a subtle extension of the notion of "identical particle" has implicitly been made here, for we are treating all quarks, regardless of color or even flavor, as different *states* of a single particle.



Thus the combination of three light quark flavors yields a decuplet, a singlet, and two octets; in the language of group theory, the direct product of three fundamental representations of $SU(3)$ decomposes according to the rule

$$3 \otimes 3 \otimes 3 = 10 \oplus 8 \oplus 8 \oplus 1 \tag{5.110}$$

Incidentally, we can also construct an octet which is antisymmetric in 1 and 3, but this is not independent ($\psi_{13} = \psi_{12} + \psi_{23}$); we have already used up the 27 states available in making the four representations 10, 8, 8, and 1.



ψ_{13}: Antisymmetric in 1 and 3

Finally, there is the question of color. In Chapter 1, I stated a general rule that all naturally occurring particles are colorless; if a meson contains a red quark, it must also contain an *anti*red quark, and every baryon must harbor one quark of each color. Actually, this is a naïve formulation of a deeper law:

EVERY NATURALLY OCCURRING
PARTICLE IS A COLOR SINGLET

The three *colors* generate a *color* $SU(3)$ symmetry, just as the three light quark *flavors* generate *flavor* $SU(3)$. (The former is, however, an *exact* symmetry—quarks of different colors all weigh the same—whereas the latter is only *approximate*—quarks of different flavor carry different mass.) By putting together three colors, we obtain a color decuplet, two color octets, and a color singlet (simply make the flavor → color transcription $u \rightarrow$ red, $d \rightarrow$ green, $s \rightarrow$ blue, in the diagrams above). But nature chooses the singlet, and so for baryons the color state is always

$$\psi(\text{color}) = (rgb - rbg + gbr - grb + brg - bgr)/\sqrt{6} \tag{5.111}$$

Because the color wave function is the same for *all* baryons, we generally do not bother to include it explicitly. However, it is absolutely crucial to remember that ψ(color) is *antisymmetric,* for this means the rest of the wave function must be *symmetric.* In particular, in the ground state [with ψ(space) symmetric] the product of ψ(spin) and ψ(flavor) has to be completely symmetric. Suppose we

start with the symmetric spin configuration; this must go with the symmetric flavor state, and we obtain the spin-$\frac{3}{2}$ baryon decuplet:

$$\psi(\text{baryon decuplet}) = \psi_s(\text{spin})\,\psi_s(\text{flavor}) \tag{5.112}$$

EXAMPLE 5.1

Write down the wave function for the Δ^+, in the spin state $m_j = -\frac{1}{2}$ (never mind the space and color parts).

Solution.

$$\begin{aligned}|\Delta^+\colon \tfrac{3}{2}\ -\tfrac{1}{2}\rangle &= [(uud + udu + duu)/\sqrt{3}][(\downarrow\downarrow\uparrow + \downarrow\uparrow\downarrow + \uparrow\downarrow\downarrow)/\sqrt{3}] \\ &= [u(\downarrow)u(\downarrow)d(\uparrow) + u(\downarrow)u(\uparrow)d(\downarrow) + u(\uparrow)u(\downarrow)d(\downarrow) \\ &\quad + u(\downarrow)d(\downarrow)u(\uparrow) + u(\downarrow)d(\uparrow)u(\downarrow) + u(\uparrow)d(\downarrow)u(\downarrow) \\ &\quad + d(\downarrow)u(\downarrow)u(\uparrow) + d(\downarrow)u(\uparrow)u(\downarrow) + d(\uparrow)u(\downarrow)u(\downarrow)]/3\end{aligned}$$

For instance, if you could pull such a particle apart, the probability is $\frac{1}{9}$ that the first quark would be a d with spin up, and $\frac{4}{9}$ that it would be a u with spin down.

The baryon octet is a little trickier, for here we must put together states of mixed symmetry to make a completely symmetric combination. Notice first that the product of two antisymmetric functions is itself *symmetric*. Thus $\psi_{12}(\text{spin}) \times \psi_{12}(\text{flavor})$ is symmetric in 1 and 2, for we pick up *two* minus signs when $1 \leftrightarrow 2$. Likewise, $\psi_{23}(\text{spin}) \cdot \psi_{23}(\text{flavor})$ is symmetric in 2 and 3, and $\psi_{13}(\text{spin}) \cdot \psi_{13}(\text{flavor})$ is symmetric in 1 and 3. If we now *add* these, the result will clearly be symmetric in all three (for the normalization factor, see Problem 5.26):

$$\begin{aligned}\psi(\text{baryon octet}) = (\sqrt{2}/3)[&\psi_{12}(\text{spin})\,\psi_{12}(\text{flavor}) \\ &+ \psi_{23}(\text{spin})\,\psi_{23}(\text{flavor}) + \psi_{13}(\text{spin})\,\psi_{13}(\text{flavor})]\end{aligned} \tag{5.113}$$

EXAMPLE 5.2

Write down the spin/flavor wave function for a proton with spin up.

Solution.

$$\begin{aligned}|\text{p}\colon \tfrac{1}{2}\,\tfrac{1}{2}\rangle &= \{\tfrac{1}{2}(\uparrow\downarrow\uparrow - \downarrow\uparrow\uparrow)(udu - duu) + \tfrac{1}{2}(\uparrow\uparrow\downarrow - \uparrow\downarrow\uparrow)(uud - udu) \\ &\quad + \tfrac{1}{2}(\uparrow\uparrow\downarrow - \downarrow\uparrow\uparrow)(uud - duu)\}\frac{\sqrt{2}}{3} = \{uud(2\uparrow\uparrow\downarrow - \uparrow\downarrow\uparrow - \downarrow\uparrow\uparrow) \\ &\quad + udu(2\uparrow\downarrow\uparrow - \downarrow\uparrow\uparrow - \uparrow\uparrow\downarrow) + duu(2\downarrow\uparrow\uparrow - \uparrow\downarrow\uparrow - \uparrow\uparrow\downarrow)\}\frac{1}{3\sqrt{2}} \\ &= \frac{2}{3\sqrt{2}}(u(\uparrow)u(\uparrow)d(\downarrow)) - \frac{1}{3\sqrt{2}}(u(\uparrow)u(\downarrow)d(\uparrow)) \\ &\quad - \frac{1}{3\sqrt{2}}(u(\downarrow)u(\uparrow)d(\uparrow)) + \text{permutations.}\end{aligned}$$

If nothing else, I hope you will have gathered from this exercise that the construction of baryon wave functions is a nontrivial business, in the quark model. Apart altogether from the spatial wave function, there are three spins to juggle, as well as three flavors and three colors, and it all has to be put together in a way that is consistent with the Pauli principle. Perhaps also you will forgive me for deferring the explanation of how three quarks can generate the baryon octet (the decuplet, remember, we got by naïve quark counting back in Chapter 1). The essential point is that the corners of the decuplet contain three identical quarks (*uuu, ddd,* and *sss*); they necessarily form a symmetric flavor state, and hence must go with the symmetric spin state ($j = \frac{3}{2}$). With *two* identical quarks (*uud,* say) there are three arrangements (*uud, udu, duu*); you can make a symmetric linear combination, which goes into the decuplet, and two of mixed symmetry, which belong to $SU(3)$ octets. Finally, with all three different, *uds,* there are six possibilities—the completely symmetric linear combination completes the decuplet, the completely antisymmetric combination makes an $SU(3)$ singlet, and the remaining four go into the two octets. Notice again the essential (if hidden) role of color in all this. Without it we would be looking for *anti*symmetric spin/flavor wave functions; spin $\frac{3}{2}$ (symmetric) would have to go with the flavor singlet (antisymmetric). It *is* possible to make a spin-$\frac{1}{2}$ octet without color (see Problem 5.28), but in place of the decuplet we would have just one spin-$\frac{3}{2}$ baryon. It was to avoid that disaster, without sacrificing the Pauli principle, that color was introduced in the first place.[25]

5.10 BARYON MASSES AND MAGNETIC MOMENTS

As an application of the baryon spin/flavor wave functions, we now calculate the magnetic dipole moments of the particles in the octet.* In the absence of orbital motion, the net magnetic moment of a baryon is simply the vector sum of the moments of the three constituent quarks:

$$\boldsymbol{\mu} = \boldsymbol{\mu}_1 + \boldsymbol{\mu}_2 + \boldsymbol{\mu}_3 \tag{5.114}$$

It depends on the quark flavors (because the three flavors carry different magnetic moments) and on the spin configuration (because that determines the relative orientations of the three dipoles). As I mentioned earlier (eq. (5.45)), the magnetic dipole moment of a spin-$\frac{1}{2}$ point particle of charge q and mass m is†

$$\boldsymbol{\mu} = \frac{q}{mc}\mathbf{S} \tag{5.45}$$

* No decuplet moments have been measured, so I won't bother with them. As for the mesons, the pseudoscalars have no spin, and hence have zero magnetic moment. The vector mesons do have magnetic moments, but as far as I know they have not been measured.

† This ignores radiative corrections, which should be larger for quarks than for electrons, but still small relative to the total magnetic moment.

Its magnitude, then, is

$$\mu = \frac{q\hbar}{2mc} \tag{5.115}$$

More precisely, this is the value of μ_z in the spin-up state, for which $S_z = \hbar/2$; if the charge is negative, so too is μ—this reminds us that the magnetic moment points opposite to the spin. It is customary to refer to μ, rather than μ itself, as "the magnetic moment" of the particle. For the quarks, we have

$$\mu_u = \frac{2}{3}\frac{e\hbar}{2m_u c}, \qquad \mu_d = -\frac{1}{3}\frac{e\hbar}{2m_d c}, \qquad \mu_s = -\frac{1}{3}\frac{e\hbar}{2m_s c} \tag{5.116}$$

The magnetic moment of baryon B, then, is

$$\mu_B = \langle B\uparrow|(\mu_1 + \mu_2 + \mu_3)_z|B\uparrow\rangle = \frac{2}{\hbar}\sum_{i=1}^{3}\langle B\uparrow|(\mu_i S_{i_z})|B\uparrow\rangle \tag{5.117}$$

EXAMPLE 5.3

Calculate the magnetic moment of the proton.

Solution. The wave function was found in Example 5.2. The first term is

$$\frac{2}{3\sqrt{2}}[u(\uparrow)u(\uparrow)d(\downarrow)]$$

Now
$$(\mu_1 S_{1_z} + \mu_2 S_{2_z} + \mu_3 S_{3_z})|u(\uparrow)u(\uparrow)d(\downarrow)\rangle = \left[\mu_u\frac{\hbar}{2} + \mu_u\frac{\hbar}{2} + \mu_d\left(-\frac{\hbar}{2}\right)\right]|u(\uparrow)u(\uparrow)d(\downarrow)\rangle$$

so this term contributes an amount

$$\left(\frac{2}{3\sqrt{2}}\right)^2\frac{\hbar}{2}\sum_{i=1}^{3}\langle u(\uparrow)u(\uparrow)d(\downarrow)|(\mu_i S_{i_z})|u(\uparrow)u(\uparrow)d(\downarrow)\rangle = \tfrac{2}{9}(2\mu_u - \mu_d)$$

Similarly, the second term $(u(\uparrow)u(\downarrow)d(\uparrow))$ gives $\frac{1}{18}\mu_d$, as does the third.* We could continue in this way to evaluate all nine terms, but the rest are simply permutations, in which d occupies position 2 or position 1. The result, then, is

$$\mu_p = 3[\tfrac{2}{9}(2\mu_u - \mu_d) + \tfrac{1}{18}\mu_d + \tfrac{1}{18}\mu_d] = \tfrac{1}{3}(4\mu_u - \mu_d)$$

In this way we can calculate all the octet magnetic moments in terms of μ_u, μ_d, and μ_s (Problem 5.29). The results are listed in the second column of Table 5.5. To get *numbers*, we need to know the quark magnetic moments (5.116). Using the baryon constituent quark masses in Table 4.4, we obtain the

* Note that everything is normalized, so that for instance $\langle u(\uparrow)|u(\uparrow)\rangle = 1$, and the states are orthogonal $\langle u(\uparrow)|u(\downarrow)\rangle = 0$.

TABLE 5.5 MAGNETIC DIPOLE MOMENTS OF OCTET BARYONS

Baryon	Moment	Prediction	Experiment
p	$(\frac{4}{3})\mu_u - (\frac{1}{3})\mu_d$	2.79	2.793
n	$(\frac{4}{3})\mu_d - (\frac{1}{3})\mu_u$	−1.86	−1.913
Λ	μ_s	−0.58	−0.61
Σ^+	$(\frac{4}{3})\mu_u - (\frac{1}{3})\mu_s$	2.68	2.33 ± 0.13
Σ^0	$(\frac{2}{3})(\mu_u + \mu_d) - (\frac{1}{3})\mu_s$	0.82	
Σ^-	$(\frac{4}{3})\mu_d - (\frac{1}{3})\mu_s$	−1.05	−1.41 ± 0.25
Ξ^0	$(\frac{4}{3})\mu_s - (\frac{1}{3})\mu_u$	−1.40	−1.253 ± 0.014
Ξ^-	$(\frac{4}{3})\mu_s - (\frac{1}{3})\mu_d$	−0.47	−0.69 ± 0.04

The numerical values are given as multiples of the nuclear magneton, $e\hbar/2m_pc = 3.152 \times 10^{-18}$ MeV/gauss.

Source: S. Gasiorowicz and J. L. Rosner, *Am. J. Phys.* **49**, 954 (1981).

figures in the third column of Table 5.5. The comparison with experiment is reasonably good, considering the uncertainties in the quark masses and anomalous magnetic moments. Somewhat better predictions are obtained if we take *ratios*. In particular, to the extent that $m_u = m_d$, we have

$$\frac{\mu_n}{\mu_p} = -\frac{2}{3} \tag{5.118}$$

which compares well with the experimental value, $0.68497945 \pm 0.00000058$.

Finally, we come to the problem of baryon masses. The situation is the same as for the mesons: If flavor $SU(3)$ were a perfect symmetry, all the octet baryons would weigh the same. But they don't. We attribute this in the first instance to the fact that the s quark is more massive than u and d. But that can't be the *whole* story, or the Λ would carry the same mass as the Σ's and the Δ's would match the proton. Evidently there is a significant spin-spin ("hyperfine") contribution, which, as before, we take to be proportional to the dot product of the spins and inversely proportional to the product of the masses. The only difference is that this time there are *three* pairs of spins to contend with:

$$M(\text{baryon}) = m_1 + m_2 + m_3 + A'\left[\frac{\mathbf{S}_1 \cdot \mathbf{S}_2}{m_1 m_2} + \frac{\mathbf{S}_1 \cdot \mathbf{S}_3}{m_1 m_3} + \frac{\mathbf{S}_2 \cdot \mathbf{S}_3}{m_2 m_3}\right] \tag{5.119}$$

Here A' [like A in equation (5.99)] is a constant, which we adjust to obtain the optimal fit to the data.

The spin products are easiest when the three quark masses are equal, for

$$J^2 = (\mathbf{S}_1 + \mathbf{S}_2 + \mathbf{S}_3)^2 = S_1^2 + S_2^2 + S_3^2 + 2(\mathbf{S}_1 \cdot \mathbf{S}_2 + \mathbf{S}_1 \cdot \mathbf{S}_3 + \mathbf{S}_2 \cdot \mathbf{S}_3) \tag{5.120}$$

and hence

$$\begin{aligned} \mathbf{S}_1 \cdot \mathbf{S}_2 + \mathbf{S}_1 \cdot \mathbf{S}_3 + \mathbf{S}_2 \cdot \mathbf{S}_3 &= \frac{\hbar^2}{2}[j(j+1) - \tfrac{9}{4}] \\ &= \begin{cases} \frac{3}{4}\hbar^2, & \text{for } j = \frac{3}{2}\text{(decuplet)} \\ -\frac{3}{4}\hbar^2, & \text{for } j = \frac{1}{2}\text{(octet)} \end{cases} \end{aligned} \tag{5.121}$$

Thus the nucleon (neutron or proton) mass is

$$M_N = 3m_u - \frac{3}{4}\frac{\hbar^2}{m_u^2}A' \tag{5.122}$$

the Δ is

$$M_\Delta = 3m_u + \frac{3}{4}\frac{\hbar^2}{m_u^2}A' \tag{5.123}$$

and the Ω^- is

$$M_\Omega = 3m_s + \frac{3}{4}\frac{\hbar^2}{m_s^2}A' \tag{5.124}$$

Indeed, in the case of the decuplet the spins are all "parallel" (every pair combines to make spin 1) so

$$(\mathbf{S}_1 + \mathbf{S}_2)^2 = S_1^2 + S_2^2 + 2\mathbf{S}_1 \cdot \mathbf{S}_2 = 2\hbar^2 \tag{5.125}$$

(and the same for 1 and 3, or 2 and 3). Hence for the decuplet

$$\mathbf{S}_1 \cdot \mathbf{S}_2 = \mathbf{S}_1 \cdot \mathbf{S}_3 = \mathbf{S}_2 \cdot \mathbf{S}_3 = \frac{\hbar^2}{4} \tag{5.126}$$

[which is consistent, notice, with eq. (5.121)], and therefore

$$M_{\Sigma^*} = 2m_u + m_s + \frac{\hbar^2}{4}A'\left(\frac{1}{m_u^2} + \frac{2}{m_u m_s}\right) \tag{5.127}$$

while

$$M_{\Xi^*} = m_u + 2m_s + \frac{\hbar^2}{4}A'\left(\frac{2}{m_u m_s} + \frac{1}{m_s^2}\right) \tag{5.128}$$

The Σ and Λ can be done by noting that the up and down quarks combine to make isospin 1 and 0, respectively, and in order for the spin/flavor wave function to be symmetric, under the interchange of u and d, the *spins* must therefore combine to a total of 1 and 0, respectively. For the Σ's, then

$$(\mathbf{S}_u + \mathbf{S}_d)^2 = S_u^2 + S_d^2 + 2\mathbf{S}_u \cdot \mathbf{S}_d = 2\hbar^2, \qquad \text{so that } \mathbf{S}_u \cdot \mathbf{S}_d = \frac{\hbar^2}{4} \tag{5.129}$$

whereas for the Λ

$$(\mathbf{S}_u + \mathbf{S}_d)^2 = 0, \qquad \text{so that } \mathbf{S}_u \cdot \mathbf{S}_d = -\tfrac{3}{4}\hbar^2 \tag{5.130}$$

Using these results together with equation (5.121), we find

$$M_\Sigma = 2m_u + m_s + A'\left[\frac{\mathbf{S}_u \cdot \mathbf{S}_d}{m_u m_d} + \frac{(\mathbf{S}_1 \cdot \mathbf{S}_2 + \mathbf{S}_1 \cdot \mathbf{S}_3 + \mathbf{S}_2 \cdot \mathbf{S}_3 - \mathbf{S}_u \cdot \mathbf{S}_d)}{m_u m_s}\right]$$

$$= 2m_u + m_s + \frac{\hbar^2}{4}A'\left(\frac{1}{m_u^2} - \frac{4}{m_u m_s}\right) \tag{5.131}$$

TABLE 5.6 BARYON OCTET AND DECUPLET MASSES (MeV/c^2)

Baryon	Calculated	Observed
N	939	939
Λ	1116	1114
Σ	1179	1193
Ξ	1327	1318
Δ	1239	1232
Σ^*	1381	1384
Ξ^*	1529	1533
Ω	1682	1672

and

$$M_\Lambda = 2m_u + m_s - \frac{3}{4}\frac{\hbar^2}{m_u^2}A' \tag{5.132}$$

I'll let you figure out the mass of the Ξ's (Problem 5.32):

$$M_\Xi = 2m_s + m_u + \frac{\hbar^2}{4}A'\left(\frac{1}{m_s^2} - \frac{4}{m_u m_s}\right) \tag{5.133}$$

Using the constituent quark masses in Table 4.4, and picking $A' = (2m_u/\hbar)^2 \cdot 50\ \text{MeV}/c^2$, we obtain an excellent fit to the experimental data (Table 5.6).

REFERENCES AND NOTES

1. The material that follows may be very familiar to some readers, and much too brief for others. Those in the first category are encouraged to skim through it quickly: those in the second should refer to any of the standard quantum texts, such as D. Park, *Introduction to the Quantum Theory,* 2d Ed. (New York: McGraw-Hill, 1974), or E. Merzbacher, *Quantum Mechanics,* 2d Ed. (New York: Wiley, 1970).
2. A derivation will be found in any intermediate-level electromagnetism text, for example, D. Griffiths, *Introduction to Electrodynamics* (Englewood Cliffs, NJ: Prentice-Hall, 1981), App. A.
3. Merzbacher (ref. 1), p. 184.
4. Merzbacher (ref. 1), p. 209.
5. For a fascinating account of the experimental study of the hydrogen spectrum, from its beginnings in the mid-nineteenth century up to the present day, see the article by T. W. Hänsch, A. L. Schawlow, and G. W. Series in *Scientific American* (March 1979).
6. See, for example, H. Margenau and G. M. Murphy, *The Mathematics of Physics and Chemistry,* 2d Ed. (Princeton, NJ: Van Nostrand, 1956), p. 388, or M. Mizushima, *Quantum Mechanics of Atomic Spectra and Atomic Structure* (New York: Benjamin, 1970), Sect. 5.2. Equation (5.40) can be obtained very simply from the Virial Theorem—see Merzbacher (ref. 1) p. 168. Expectation values of powers of r are connected by a beautiful formula due to Kramers—see A. Messiah, *Quantum Mechanics, Vol. I* (New York: Wiley, 1958), p. 431.

7. If the electrodynamics involved in equations (5.43), (5.45), and (5.46) is not familiar to you, refer to any electricity and magnetism text, for instance, E. M. Purcell, *Electricity and Magnetism,* 2d Ed. (New York: McGraw-Hill, 1985), or J. D. Jackson, *Classical Electrodynamics,* 2d Ed. (New York: Wiley, 1975).
8. See, for example, R. B. Leighton, *Principles of Modern Physics* (New York: McGraw-Hill, 1959).
9. In quantum-mechanics this is confirmed by showing that **J**, and also L^2 and S^2, commute with ΔH_{so}. See Park (ref. 1), p. 469. My treatment of fine structure closely follows Park's.
10. W. E. Lamb, Jr., and R. C. Retherford, *Phys. Rev.* **72,** 241 (1947).
11. Each normal mode of the electromagnetic field functions as an oscillator; in quantum-mechanics the ground-state energy of a harmonic oscillator is not zero, but rather, $\frac{1}{2}\hbar\omega$. See J. D. Bjorken and S. D. Drell, *Relativistic Quantum Mechanics* (New York: McGraw-Hill, 1964), p. 58.
12. See H. A. Bethe and E. E. Salpeter, *Quantum Mechanics of One- and Two-Electron Atoms* (Berlin: Springer-Verlag, 1957), Sect. 21, for details.
13. See, for instance, reference 2, p. 213.
14. See, for example, D. J. Griffiths, *Am. J. Phys.* **50,** 698 (1982).
15. H. I. Ewen and E. M. Purcell, *Nature* **168,** 356 (1951).
16. See reference 12, p. 109.
17. The spectrum of positronium is discussed in Bethe and Salpeter (ref. 12), Sect. 23. Recent work is reviewed by S. Berko and H. N. Pendleton, *Ann. Rev. Nucl. Part. Sci.* **30,** 543 (1980) and A. Rich, *Rev. Mod. Phys.* **53,** 127 (1981).
18. See reference 17.
19. T. Appelquist and H. D. Politzer, *Phys. Rev. Lett.* **34,** 43 (1975); *Phys. Rev.* **D12,** 1404 (1975).
20. For an interesting account of these discoveries, see the article "Quarkonium" by E. D. Bloom and G. J. Feldman in *Scientific American* (May 1982). The details of the fine structure are not quite the same in quarkonium as in positronium; see E. Eichten and F. Feinberg, *Phys. Rev. Lett.* **43,** 1205 (1979) and *Phys. Rev.* **D23,** 2724 (1981).
21. There are other ways to estimate F_0, which give roughly the same answer. See D. H. Perkins, *Introduction to High Energy Physics,* 2d Ed. (Reading, MA: Addison-Wesley, 1982), p. 209.
22. E. Eichten and K. Gottfried, *Phys. Lett.* **66B,** 286 (1977).
23. The "MIT Bag Model" offers a possible approach to relativistic light-quark systems, but at the cost of vastly oversimplified dynamics. The quarks are treated as free particles confined within a spherical "bag", which is stabilized by an *ad hoc* external pressure. Many interesting calculations have been carried out using the bag model, but no one would pretend that it is a realistic picture of hadron structure. See F. E. Close, *An Introduction to Quarks and Partons* (London: Academic, 1979), Chapter 18.
24. See, for example, F. Halzen and A. D. Martin, *Quarks and Leptons* (New York: Wiley, 1984), p. 42.
25. The construction of mesonic and baryonic wave functions is accomplished by somewhat different means in S. Gasiorowicz and J. L. Rosner, *Am. J. Phys.* **49,** 954 (1981), an article that contains a wealth of useful and accessible information about the quark model.

PROBLEMS

5.1. **(a)** The deuteron's mass is 1875.6 MeV/c^2. What is its binding energy? Is this a relativistic system?

(b) If you take the up- and down-quark masses to be those given in Table 4.4, ("effective mass in mesons"), what is the binding energy of a pion? Is this a relativistic system?

5.2. Derive equations (5.16), (5.17), and (5.18) from equation (5.8), using equations (5.14) and (5.15).

5.3. Show that the spherical harmonics are either even or odd under the parity transformation, depending on whether l is even or odd. That is

$$P\text{: } Y_l^m(\theta, \phi) \rightarrow (-1)^l Y_l^m(\theta, \phi)$$

5.4. Using equations (5.21) and (5.22), find Y_0^0, Y_3^0, and Y_3^1. Check your results against Table 5.1.

5.5. Show that putting equation (5.25) into equation (5.24) leads to the Bohr energies, equation (5.26). [*Hint:* The basic idea is to solve equation (5.24) by a power series, and demand that the result be normalizable. But if you have never been through this derivation before, you had better refer to a quantum text before proceeding. See, for example, Merzbacher (ref. 1), Chap. 10, Sects. 5 and 6.]

5.6. Use equation (5.28) to obtain the ground-state wave function ψ_{100}. Check that it satisfies the Schrödinger equation (5.8), with the appropriate energy, and that it is properly normalized. [*Answer:* $\psi_{100} = (1/\sqrt{\pi a^3})e^{-r/a}$]

5.7. Work out all of the hydrogen wave functions for $n = 2$, using equation (5.28). (How many are there?)

5.8. Suppose you are interested in some dynamical quantity, $Q(\mathbf{r}, \mathbf{p})$—for instance, kinetic energy ($\mathbf{p}^2/2m$), potential energy ($V(\mathbf{r})$), or angular momentum ($\mathbf{r} \times \mathbf{p}$). If you measure Q on an ensemble of particles all in the same state ψ, you will *not* in general get reproducible results; quantum mechanics can only tell you the *probability* of obtaining a given answer. In particular, the *average* (or "expectation") value of Q is given by the formula

$$\langle Q \rangle = \int \psi^* Q\left(\mathbf{r}, \frac{\hbar}{i}\nabla\right)\psi \, d^3x$$

(a) Compute $\langle r \rangle$, $\langle r^2 \rangle$, $\langle r^{-1} \rangle$, $\langle r^{-2} \rangle$ in the ground state of hydrogen. [Use the wave function from Problem 5.6.]

(b) Find the expectation values of the kinetic and potential energies in the ground state of hydrogen. Is their *sum* what you would expect?

5.9. Derive the fine structure formula (5.52), starting with the relativistic correction (5.42) and the spin-orbit coupling (5.51).

5.10. Find the energy splitting between the $j = \frac{3}{2}$ and $j = \frac{1}{2}$ levels for $n = 2$ (see Fig. 5.2), in electron volts. How does this compare with the spacing between the $n = 2$ and $n = 1$ Bohr energies?

5.11. Estimate the Lamb shift energy gap between the $2S_{1/2}$ and $2P_{1/2}$ levels in hydrogen, using equations (5.53) and (5.54). What is the frequency of the photon emitted in such a transition? [The experimental value is 1057 MHz.]

5.12. Suppose **A** and **B** are two fixed vectors. Show that the expectation value of

$$Q \equiv [3(\mathbf{A}\cdot\hat{r})(\mathbf{B}\cdot\hat{r}) - (\mathbf{A}\cdot\mathbf{B})]$$

is zero in any spherically symmetrical state. [*Hint:* First convince yourself that $\langle Q\rangle$ must be proportional to $(\mathbf{A}\cdot\mathbf{B})$:

$$\langle Q\rangle = \alpha(\mathbf{A}\cdot\mathbf{B})$$

where α, the proportionality constant, is independent of **A** and **B**. To evaluate α, take the case where **A** and **B** are parallel, and choose the z axis to lie along this direction. Then

$$\alpha = \langle(3\cos^2\theta - 1)\rangle$$

Now perform the θ integral. [Notice that the argument is not affected if an arbitrary function of r is thrown into the expectation value. In particular, the expectation value of the first term in equation (5.59) must vanish for an electron in the S state.]

5.13. Derive equation (5.61). [*Hint:* First show, using (5.30), that $L_{n-1}^1(0) = n(n!)$.]

5.14. If you include the fine structure, Lamb shift, and hyperfine splitting, how many different $n = 2$ energy levels are there altogether in hydrogen? Find the hyperfine splitting between the $2S_{1/2}$ and $2P_{1/2}$ levels, and compare the Lamb shift (Problem 5.11).

5.15. Analyze the splitting of the $n = 3$ Bohr level in positronium. How many different levels are there, and what are their relative energies? Construct the level diagram, analogous to Figure 5.7.

5.16. The cross section for $e^+ + e^- \to 3\gamma$, in the triplet spin configuration, is

$$\sigma = \frac{16}{9}(\pi^2 - 9)\alpha^3\left(\frac{\hbar^2}{m^2cv}\right)$$

Find the lifetime of positronium in the 1^3S_1 state. [The experimental answer is 1.45×10^{-7} sec.]

5.17. Suppose particle A has charge e_a, and particle B has charge e_b. Assume A is much heavier than B. As we shall see in Chapter 7, the cross section for the electromagnetic process $A + \bar{A} \to B + \bar{B}$ is

$$\sigma = \frac{16\pi}{3}\left(\frac{e_a e_b}{m_a c^2}\right)^2 \frac{c}{v}$$

Calculate the decay rate for "muonic muonium" $\mu^+\mu^-$ in its ground state (a) to $\gamma + \gamma$ and (b) to $e^+ + e^-$.

5.18. Just as positronium, in the triplet configuration, decays to three photons (Problem 5.16), the ψ meson (charmonium in the triplet configuration) decays to three gluons. (The gluons subsequently turn into various combinations of hadrons.) Indeed, the cross section for $c + \bar{c} \to 3g$ is the same as for $e^+ + e^- \to 3\gamma$, only with $\alpha^3 \to \frac{5}{18}\alpha_s^3$ (the $\frac{5}{18}$ is a so-called "color factor," which you'll learn how to calculate in Chapter 9). Use this, together with the formulas in Problems 5.16 and 5.17, and equation (5.82), to determine the "branching ratios"

$$\Gamma(\psi \to \text{hadrons}):\Gamma(\psi \to e^+e^-):\Gamma(\psi \to \mu^+\mu^-)$$

Compare the experimental results.

5.19. Would you consider the $\phi(s\bar{s})$ meson *bound* or *quasi*-bound?

5.20. On dimensional grounds, show that the energy levels of a purely linear potential, $V(r) = F_0 r$, must be of the form

$$E_n = \left(\frac{(F_0\hbar)^2}{m}\right)^{1/3} a_n$$

where a_n is a dimensionless numerical factor.

5.21. Use the numerical results on Table 5.2 to "predict" the masses of the four lightest ψ's and Υ's; compare the experimental results (Fig. 5.10). What value of F_0 gives the best fit to the level spacings? Why aren't the calculated masses in better agreement with the experiments?

5.22. On the basis of equation (5.99), using the values of m_u, m_d, m_s, and A given in the text, calculate all the meson masses in Table 5.3. [*Hint:* For the η, ϕ, and ω, first find the mass for pure $u\bar{u}$, pure $d\bar{d}$, and pure $s\bar{s}$. Think of the η, for instance, as being $\frac{1}{6}u\bar{u}$, $\frac{1}{6}d\bar{d}$, and $\frac{1}{3}s\bar{s}$.] Also apply the formula to the η', and note the disastrous result. [For commentary on the η' mass problem, see C. Quigg, *Gauge Theories of the Strong, Weak, and Electromagnetic Interactions* (New York: Benjamin, 1983), p. 252.]

5.23. In the text we used equation (5.99) to calculate the masses of light quark pseudoscalar and vector mesons. But the same formula can be applied to heavy quark systems involving charm and beauty quarks.

(a) Calculate the masses of the pseudoscalar mesons $\eta_c(c\bar{c})$, $D(c\bar{u})$, $F(c\bar{s})$, and the corresponding vector mesons $\psi(c\bar{c})$, $D^*(c\bar{u})$, and $F^*(c\bar{s})$. Compare the experimental values, from the Particle Data Booklet.

(b) Do the same for the "beautiful" mesons $b\bar{u}$, $b\bar{s}$, $b\bar{c}$, and $b\bar{b}$. [At present only the pseudoscalar $B(b\bar{u})$ and the vector $\Upsilon(b\bar{b})$ have been detected experimentally.]

5.24. Construct the eight states ψ_{12} in Section 5.9. [*Hint:* The six outer ones are easy—the quark content is determined by Q and S, and all you have to do is antisymmetrize in 1 and 2. To get the two states in the center, remember that the one in the "Σ^0" position forms an isotriplet with the "Σ^+" and "Σ^-"; the "Λ" may then be constructed by orthogonalizing with respect to "Σ^0" and ψ_A.]

5.25. Find the color wave function for mesons, analogous to equation (5.111).

5.26. Check that the baryon octet spin/flavor wave function (5.113) is correctly normalized. Remember that ψ_{13} is not independent of ψ_{12} and ψ_{23}.

5.27. Construct the spin-flavor wave functions, as in Example 5.2, for Σ^+ with spin up and Λ with spin down.

5.28. Construct a totally *anti*symmetric spin/flavor baryon octet. [In this configuration we do not need color to antisymmetrize the wave function. However, an antisymmetric *decuplet cannot* be constructed.] (See Halzen and Martin, ref. 24, Exercise 2.18.)

5.29. **(a)** Derive the expressions in the second column of Table 5.5.

(b) From these results, calculate the numbers in the third column of Table 5.5.

5.30. Calculate the ratio μ_n/μ_p in the configuration you found for Problem 5.28. Notice that μ_p is *negative* in this case (!). Is your result consistent with experiment? (Here, then, is a second strike against the quark model without color, the first strike being its failure to account for the decuplet.)

5.31. Show that $\mu_{p+} = -\mu_{p-} = \mu_p$. (See Halzen and Martin, ref. 24, Exercise 2.19).

5.32. Use equation (5.119) to determine the mass of the Ξ.

Chapter 6

The Feynman Calculus

In this chapter we begin the quantitative formulation of elementary particle dynamics, which amounts, in practice, to the calculation of decay rates (Γ) and scattering cross sections (σ). The procedure involves two distinct parts: (1) evaluation of the relevant Feynman diagrams to determine the "amplitude" ($\mathcal{M}$) for the process in question, and (2) insertion of $\mathcal{M}$ into Fermi's "Golden Rule" to compute Γ or σ, as the case may be. To avoid distracting algebraic complications, I introduce here a simplified model. Realistic theories—QED, QCD, and GWS—are developed in succeeding chapters. If you like, Chapter 6 can be read immediately after Chapter 3. Study it with scrupulous care, or what follows will be unintelligible.

6.1 LIFETIMES AND CROSS SECTIONS

As I mentioned in the Introduction, we have three experimental probes of elementary particle interactions: bound states, decays, and scattering. Nonrelativistic quantum mechanics (in Schrödinger's formulation) is particularly well adapted to handle bound states, which is why we used it, as far as possible, in Chapter 5. By contrast, the relativistic theory (in Feynman's formulation) is especially well suited to describe decays and scattering. In this chapter I'll introduce the basic ideas and strategies of the Feynman "calculus"; in subsequent chapters we will use it to develop the theories of strong, electromagnetic, and weak interactions.

To begin with, we must decide what physical quantities we would like to calculate. In the case of decays, the item of greatest interest is the *lifetime* of the particle in question. What precisely do we mean by the lifetime of, say, the muon? We have in mind, of course, a muon at *rest;* a moving muon lasts longer

(from our perspective) because of time dilation. But even stationary muons don't all last the same amount of time, for there is an intrinsically random element in the decay process. We cannot hope to calculate the lifetime of any *particular* muon; rather, what we are after is the *average* (or "mean") lifetime, τ, of the muons in any large sample. Now, elementary particles have no memories, so the probability of a given muon decaying in the next microsecond is independent of how long ago that muon was created. (It's quite different in biological systems: An 80-year-old man is much more likely to die in the next year than is a 20-year-old, and his body shows the signs of eight decades of wear and tear. But all muons are identical, regardless of when they were produced; from an actuarial point of view they're all on an equal footing.) The critical parameter, then, is the *decay rate,* Γ, the *probability per unit time* that any given muon will disintegrate. If we had a large collection of muons, say, $N(t)$, at time t, then $N\Gamma\, dt$ of them would decay in the next instant dt. This would, of course, *decrease* the number remaining:

$$dN = -\Gamma N\, dt \tag{6.1}$$

It follows that

$$N(t) = N(0)e^{-\Gamma t} \tag{6.2}$$

Evidently, the number of particles left decreases exponentially with time. As you can check for yourself (Problem 6.1), the mean lifetime is simply the reciprocal of the decay rate:

$$\tau = \frac{1}{\Gamma} \tag{6.3}$$

Actually, most particles can decay by several different routes. The π^+, for instance, usually decays to $\mu^+ + \nu_\mu$, but sometimes one goes to $e^+ + \nu_e$; occasionally, a π^+ decays to $\mu^+ + \nu_\mu + \gamma$, and they have even been known to go to $e^+ + \nu_e + \pi^0$. In such circumstances the *total* decay rate is the sum of the individual decay rates:

$$\Gamma_{\text{tot}} = \sum_{i=1}^{n} \Gamma_i \tag{6.4}$$

and the lifetime of the particle is the reciprocal of Γ_{tot}:

$$\tau = 1/\Gamma_{\text{tot}} \tag{6.5}$$

In addition to τ, we want to calculate the various *branching ratios,* that is, the fraction of all particles of the given type that decay by each mode. Branching ratios are determined by the decay rates:

$$\text{Branching ratio for } i\text{th decay mode} = \Gamma_i/\Gamma_{\text{tot}} \tag{6.6}$$

For decays, then, the essential problem is to calculate the decay rate Γ_i for each mode; from there it is an easy matter to obtain the lifetime and branching ratios.

How about scattering? What quantity should the experimentalist measure and the theorist calculate? If we were talking about an *archer* aiming at a "bull's-eye," the parameter of interest would be the *size of the target,* or more precisely

the cross-sectional area it presents to a stream of incoming arrows. In a crude sense, the same goes for elementary particle scattering: If you fire a stream of electrons into a tank of hydrogen (which is essentially a collection of protons) the parameter of interest is the size of the proton—the cross-sectional area σ it presents to the incident beam. The situation is more complicated than in archery, however, for several reasons. First of all the target is "soft"; it's not a simple case of "hit-or-miss," but rather "the closer you come the greater the deflection." Nevertheless, it is still possible to define an "effective" cross section; I'll show you how in the next paragraph. Second, the cross section depends on the nature of the "arrow" as well as the structure of the "target." Electrons scatter off hydrogen more sharply than neutrinos and less so than pions, because different interactions are involved. It depends, too, on the *outgoing* particles; if the energy is high enough we can have not only *elastic* scattering ($e + p \rightarrow e + p$), but a variety of *inelastic* processes, such as $e + p \rightarrow e + p + \gamma$, or $e + p + \pi^0$, or even, in principle, $\nu_e + \Lambda$. Each one of these has its own ("exclusive") scattering cross section, σ_i (for process i). In some experiments, however, the final products are not examined, and we are interested only in the *total* ("inclusive") cross section.

$$\sigma_{\text{tot}} = \sum_{i=1}^{n} \sigma_i \tag{6.7}$$

Finally, each cross section typically depends on the *velocity* of the incident particle. At the most naïve level we might expect the cross section to be proportional to the amount of time the incident particle spends in the vicinity of the target, which is to say that σ should be inversely proportional to v. But this behavior is dramatically altered in the neighborhood of a "resonance"—a special energy at which the particles involved "like" to interact, forming a short-lived semibound state before breaking apart. Such "bumps" in the graph of σ versus v (or, as it is more commonly plotted, σ versus E) are in fact the principal means by which short-lived particles are discovered (see Fig. 4.6). So, unlike the archer's target, there's a lot of physics in an elementary particle cross section.

Let's go back, now, to the question of what we mean by a "cross section" when the target is "soft." Suppose a particle (maybe an electron) comes along, encounters some kind of potential (perhaps the Coulomb potential of a stationary proton), and scatters off at an angle θ. This *scattering angle* is a function of the *impact parameter* b, the distance by which the incident particle would have missed the scattering center, had it continued on its original trajectory (Fig. 6.1). Ordinarily, the smaller the impact parameter, the larger the deflection, but the actual functional form of $\theta(b)$ depends on the particular potential involved.

EXAMPLE 6.1 Hard-Sphere Scattering

Suppose the particle bounces elastically off a sphere of radius R. From Figure 6.2, we have

$$b = R \sin \alpha, \qquad 2\alpha + \theta = \pi$$

Thus $$\sin \alpha = \sin(\pi/2 - \theta/2) = \cos(\theta/2)$$

and hence $$b = R \cos(\theta/2) \quad \text{or} \quad \theta = 2 \cos^{-1}(b/R)$$

This is the relation between θ and b for classical hard-sphere scattering.

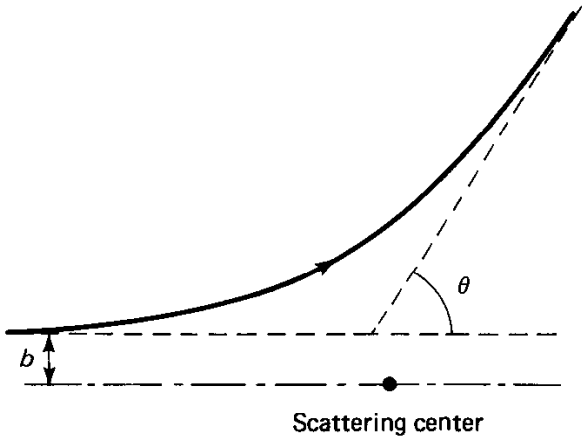


Figure 6.1 Scattering from a fixed potential: θ is the scattering angle, b is the impact parameter.

If the particle comes in with an impact parameter between b and $b + db$, it will emerge with a scattering angle between θ and $\theta + d\theta$. More generally, if it passes through an infinitesimal *area* $d\sigma$, it will scatter into a corresponding *solid angle* $d\Omega$ (Fig. 6.3). Naturally, the larger we make $d\sigma$, the larger $d\Omega$ will be. The proportionality factor is called the *differential scattering cross section, D:*

$$d\sigma = D(\theta)d\Omega \tag{6.8}$$

In principle, D might depend on the azimuthal angle ϕ; however, most potentials of interest are spherically symmetrical, in which case the differential cross section depends only on θ (or, if you prefer, on b). By the way, the notation, D, is my own; most people call it simply $d\sigma/d\Omega$, and in the rest of the book I'll revert to the standard terminology. The name "differential cross section" is poorly chosen; it's not a differential at all, in the mathematical sense (the words would apply more naturally to $d\sigma$ than to $d\sigma/d\Omega$).

Now, from Figure 6.3 we see that

$$d\sigma = |b\, db\, d\phi|, \qquad d\Omega = |\sin\theta\, d\theta\, d\phi| \tag{6.9}$$

(Areas and solid angles are intrinsically positive, hence the absolute value signs.) Accordingly,

$$D(\theta) = \frac{d\sigma}{d\Omega} = \left|\frac{b}{\sin\theta}\left(\frac{db}{d\theta}\right)\right| \tag{6.10}$$

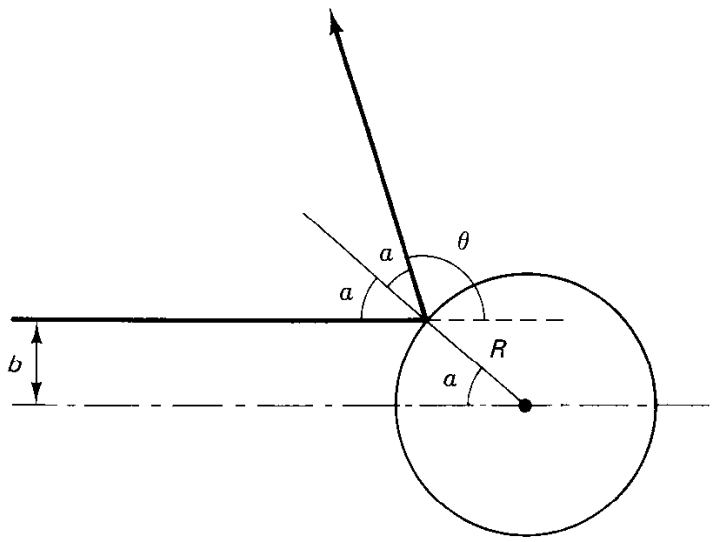


Figure 6.2 Hard-sphere scattering.

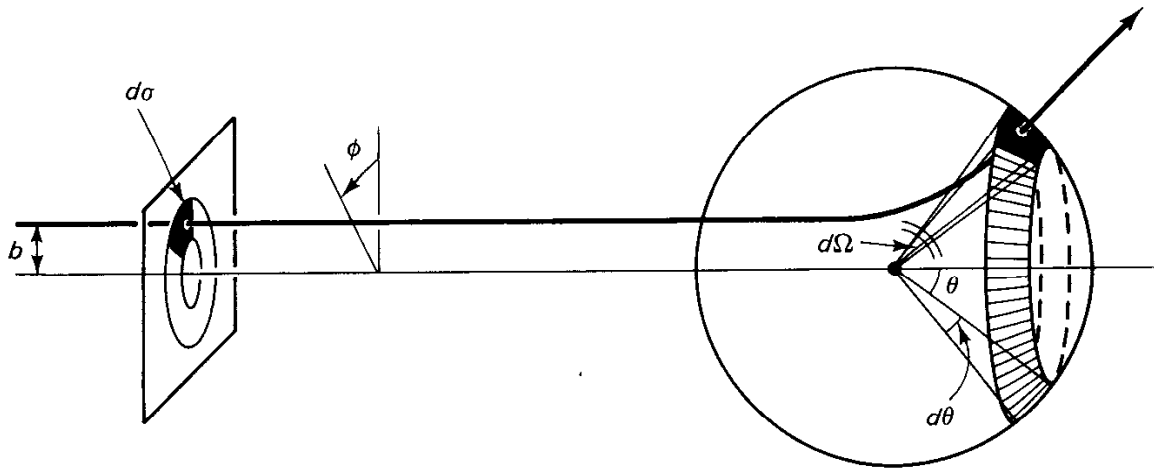


Figure 6.3 Particle incident in area $d\sigma$ scatters into solid angle $d\Omega$.

EXAMPLE 6.2

In the case of hard-sphere scattering, Example 6.1, we find

$$\frac{db}{d\theta} = -\frac{R}{2}\sin\left(\frac{\theta}{2}\right)$$

and hence

$$D(\theta) = \frac{Rb\,\sin(\theta/2)}{2\sin\theta} = \frac{R^2}{2}\frac{\cos(\theta/2)\,\sin(\theta/2)}{\sin\theta} = \frac{R^2}{4}$$

Finally, the *total* cross section is the integral of $d\sigma$ over all solid angles:

$$\sigma = \int d\sigma = \int D(\theta)d\Omega \tag{6.11}$$

EXAMPLE 6.3

For hard-sphere scattering

$$\sigma = \int \frac{R^2}{4}\,d\Omega = \pi R^2$$

which is, of course, the total cross section the sphere presents to an incoming beam: Any particles *within* this area will scatter, any *outside* will pass by unaffected.

As Example 6.3 indicates, the formalism developed here is consistent with our naïve sense of the term "cross section," in the case of a "hard" target; its virtue is that it applies as well to "soft" targets, which do not have sharp edges.

EXAMPLE 6.4 Rutherford Scattering

A particle of charge q_1 scatters off a stationary particle of charge q_2. In classical mechanics the formula relating the impact parameter to the scattering angle is[1]

$$b = \frac{q_1 q_2}{2E}\cot(\theta/2)$$

where E is the initial kinetic energy of the incident charge. The differential cross section is therefore

$$D(\theta) = \left(\frac{q_1 q_2}{4E \sin^2(\theta/2)}\right)^2$$

In this case the total cross section is actually *infinite*:*

$$\sigma = 2\pi\left(\frac{q_1 q_2}{4E}\right)^2 \int_0^\pi \frac{1}{\sin^4(\theta/2)} \sin\theta \, d\theta = \infty$$

Suppose now, that we have a *beam* of incoming particles, with uniform *luminosity* $\mathcal{L}$ ($\mathcal{L}$ is the number of particles passing down the line per unit time, per unit area). Then $dN = \mathcal{L}\, d\sigma$ is the number of particles per unit time passing through area $d\sigma$, and hence also the number per unit time scattered into solid angle $d\Omega$:

$$dN = \mathcal{L}\, d\sigma = \mathcal{L} D(\theta) d\Omega$$

It follows that

$$\frac{d\sigma}{d\Omega} = D(\theta) = \frac{1}{\mathcal{L}} \frac{dN}{d\Omega} \tag{6.12}$$

This is frequently a more convenient way to think of the differential cross section: It is the number of particles per unit time scattered into solid angle $d\Omega$, divided by $d\Omega$ and by the luminosity. (Or, as accelerator physicists like to put it, "the event rate is the cross section times the luminosity".)†

6.2 THE GOLDEN RULE

In Section 6.1 I introduced the basic physical quantities we need to calculate: decay rates and scattering cross sections. In either case there are two ingredients in the recipe: (1) the *amplitude* ($\mathcal{M}$) for the process and (2) the *phase space* available.‡ The amplitude contains all the *dynamical* information; we calculate it by evaluating the relevant Feynman diagrams, using the "Feynman rules" appropriate to the interaction in question. The phase space factor contains only *kinematical* information; it depends on the masses, energies, and momenta of the participants, and reflects the fact that a given process is more likely to occur

* This is related to the fact that the Coulomb potential has infinite range (see footnote on p. 17).

† In this discussion I have assumed that the target itself is *stationary*, and that the incident particle is simply *deflected* as it passes through the scattering potential. My purpose was to introduce the essential ideas in the simplest possible context. But in Section 6.2 the formalism is completely general; it includes the recoil of the target, and allows for a change in the identity of the participants during the scattering process (in the reaction $\pi^- + p^+ \to K^+ + \Sigma^-$, for example, $d\Omega$ might represent the solid angle into which the K^+ scatters).

‡ The *amplitude* is also called the *matrix element;* the *phase space* is sometimes called the *density of final states.*

the more "room to maneuver" there is in the final state. For example, the decay of a heavy particle into many light secondaries involves a large phase space factor, for there are many different ways to apportion the available energy. By contrast, the decay of the neutron ($n \to p + e + \bar{\nu}_e$), in which there is almost no extra mass to spare, is tightly constrained, and the phase space factor is very small.*

The transition rate for a given process is determined by the amplitude and the phase space according to Fermi's "Golden Rule":

$$\text{transition rate} = \frac{2\pi}{\hbar} |\mathcal{M}|^2 \times (\text{phase space}) \tag{6.13}$$

A derivation of the Golden Rule in the nonrelativistic context will be found in any quantum mechanics text[2]; for the relativistic version one must consult a book on quantum field theory.[3] We shall not go into that here; for our purposes it will suffice to quote the quantitative formulation of the Golden Rule in the two cases of interest:

Golden Rule for Decays. Suppose particle 1 decays into several other particles 2, 3, 4, . . . , n:

$$1 \to 2 + 3 + 4 + \cdots + n \tag{6.14}$$

The decay rate is given by the formula†

$$d\Gamma = |\mathcal{M}|^2 \frac{S}{2\hbar m_1} \left[\left(\frac{c\, d^3\mathbf{p}_2}{(2\pi)^3 2E_2} \right) \left(\frac{c\, d^3\mathbf{p}_3}{(2\pi)^3 2E_3} \right) \cdots \left(\frac{c\, d^3\mathbf{p}_n}{(2\pi)^3 2E_n} \right) \right] \times (2\pi)^4\, \delta^4(p_1 - p_2 - p_3 \cdots - p_n) \tag{6.15}$$

where $p_i = (E_i/c,\ \mathbf{p}_i)$ is the four-momentum of the ith particle (which carries mass m_i, so that $E_i^2 - \mathbf{p}_i^2 c^2 = m_i^2 c^4$).‡ The delta function§ enforces conservation of energy and momentum; it is zero unless $p_1 = p_2 + p_3 + \cdots + p_n$. The decaying particle is presumed to be at rest: $p_1 = (m_1 c,\ \mathbf{0})$. S is a product of statistical factors: $1/j!$ for each group of j identical particles in the final state.

Equation (6.15) determines the *differential* rate for a decay in which the three-momentum of particle 2 lies in the range $d^3\mathbf{p}_2$ about the value $\mathbf{p}_2$, that of particle 3 lies in the range $d^3\mathbf{p}_3$ about $\mathbf{p}_3$, and so on. Ordinarily, we are not interested in the individual momenta of the decay products, and so we integrate over all

* For a more extreme case, consider the (kinematically forbidden) decay $\Omega^- \to \Xi^- + \bar{K}^0$. Since the final products weigh more than the Ω, there is no phase space available at all, and the decay rate is zero.

† The formula *looks* simpler if you collect together the various factors of c, 2π, and so on, but its *structure* is clearer when they are grouped as shown.

‡ Notice that this makes E_i a function of $\mathbf{p}_i$: $E_i = c\sqrt{m_i^2 c^2 + \mathbf{p}_i^2}$; in fact, as far as equation (6.15) is concerned, E_i should be regarded as shorthand for this expression.

§ Those unfamiliar with the Dirac delta function should study Appendix A before proceeding.

outgoing momenta to get the total decay rate Γ for mode in question (6.14). In particular, if there are only two particles in the final state

$$\Gamma = \frac{S}{\hbar m_1}\left(\frac{c}{4\pi}\right)^2 \frac{1}{2}\int \frac{|\mathcal{M}|^2}{E_2 E_3}\,\delta^4(p_1 - p_2 - p_3)d^3\mathbf{p}_2\, d^3\mathbf{p}_3 \tag{6.16}$$

In general, the amplitude $\mathcal{M}$ is a function of $\mathbf{p}_2$ and $\mathbf{p}_3$, and it cannot be taken outside the integral. Nevertheless, for a two-body decay we can carry out the integration explicitly, without knowing the functional form of $\mathcal{M}$, as the following two examples show. Because the general case (Example 6.6) involves some cumbersome algebra, we shall first consider the special case in which the decay products are both massless (Example 6.5).

EXAMPLE 6.5

A particle of mass m decays into two massless secondaries (for instance, $\pi^0 \longrightarrow \gamma + \gamma$). If the amplitude for the process is $\mathcal{M}(\mathbf{p}_2, \mathbf{p}_3)$, find the decay rate.

Solution. First, rewrite the delta function, using the fact that $E_1 = mc^2$ and $\mathbf{p}_1 = \mathbf{0}$:

$$\delta^4(p_1 - p_2 - p_3) = \delta\left(mc - \frac{E_2}{c} - \frac{E_3}{c}\right)\delta^3(-\mathbf{p}_2 - \mathbf{p}_3) \tag{6.17}$$

Since $m_2 = m_3 = 0$, we have $E_2 = |\mathbf{p}_2|c$, $E_3 = |\mathbf{p}_3|c$. Thus

$$\Gamma = \frac{S}{\hbar m}\left(\frac{1}{4\pi}\right)^2 \frac{1}{2}\int \frac{|\mathcal{M}|^2}{|\mathbf{p}_2||\mathbf{p}_3|} \times \delta(mc - |\mathbf{p}_2| - |\mathbf{p}_3|)\delta^3(-\mathbf{p}_2 - \mathbf{p}_3)d^3\mathbf{p}_2\, d^3\mathbf{p}_3 \tag{6.18}$$

Next, use $\delta^3(-\mathbf{p}_2 - \mathbf{p}_3)$ to do the $\mathbf{p}_3$ integral. This simply replaces every $\mathbf{p}_3$ by $-\mathbf{p}_2$, reflecting the conservation of momentum:

$$\Gamma = \frac{S}{2(4\pi)^2\hbar m}\int \frac{|\mathcal{M}|^2}{|\mathbf{p}_2|^2}\,\delta(mc - 2|\mathbf{p}_2|)d^3\mathbf{p}_2 \tag{6.19}$$

At this stage $|\mathcal{M}|^2$ is a function of $\mathbf{p}_2$ alone; indeed, since it has to be a *scalar,* it can only depend on $|\mathbf{p}_2|$.* Going to spherical coordinates,

$$d^3\mathbf{p}_2 = |\mathbf{p}_2|^2\, d|\mathbf{p}_2|\sin\theta\, d\theta\, d\phi \tag{6.20}$$

and performing the angular integration, $\int \sin\theta\, d\theta\, d\phi = 4\pi$, we have

$$\Gamma = \frac{S}{8\pi\hbar m}\int_0^\infty |\mathcal{M}|^2\,\delta(mc - 2|\mathbf{p}_2|)d|\mathbf{p}_2| \tag{6.21}$$

* If the particles carry spin, then $|\mathcal{M}|^2$ might depend also on $(\mathbf{p}_i \cdot \mathbf{S}_j)$ and $(\mathbf{S}_i \cdot \mathbf{S}_j)$. However, since experiments rarely measure the spin orientation, we almost always work with the spin-averaged amplitude. In that case, as in the case of spin 0, the only vector in sight is $\mathbf{p}_2$, and the only scalar variable is $\mathbf{p}_2^2$.

Now [see eq. (A.9)]

$$\delta(mc - 2|\mathbf{p}_2|) = \frac{1}{2}\,\delta\left(|\mathbf{p}_2| - \frac{mc}{2}\right) \tag{6.22}$$

and we conclude that

$$\Gamma = \frac{S}{16\pi\hbar m}\,|\mathcal{M}|^2 \tag{6.23}$$

where $\mathcal{M}$ is evaluated at the momenta dictated by the conservation laws, to wit: $\mathbf{p}_3 = -\mathbf{p}_2$ and $|\mathbf{p}_2| = mc/2$.*

EXAMPLE 6.6 Two-Body Decay

Consider now the *general* case of a two-body decay, in which the outgoing particles carry masses m_2 and m_3. Find the decay rate, assuming $\mathcal{M}$ is given.

Solution. Again, we begin by rewriting the delta function, as in equation (6.17), and performing the $\mathbf{p}_3$ integral; but this time $E_2 = c\sqrt{m_2^2c^2 + \mathbf{p}_2^2}$, $E_3 = c\sqrt{m_3^2c^2 + \mathbf{p}_2^2}$, so in place of equation (6.19) we have

$$\Gamma = \frac{S}{2(4\pi)^2\hbar m_1}\int \frac{|\mathcal{M}|^2\,\delta(m_1c - \sqrt{m_2^2c^2 + \mathbf{p}_2^2} - \sqrt{m_3^2c^2 + \mathbf{p}_2^2})}{\sqrt{m_2^2c^2 + \mathbf{p}_2^2}\,\sqrt{m_3^2c^2 + \mathbf{p}_2^2}}\,d^3\mathbf{p}_2 \tag{6.24}$$

As before, $|\mathcal{M}|^2$ is now a function only of $|\mathbf{p}_2|$, so we can introduce spherical coordinates and do the angular integration:

$$\Gamma = \frac{S}{8\pi\hbar m_1}\int_0^\infty \frac{|\mathcal{M}|^2\,\delta(m_1c - \sqrt{m_2^2c^2 + \rho^2} - \sqrt{m_3^2c^2 + \rho^2})}{\sqrt{m_2^2c^2 + \rho^2}\,\sqrt{m_3^2c^2 + \rho^2}}\,\rho^2\,d\rho \tag{6.25}$$

where ρ is shorthand for $|\mathbf{p}_2|$. We could use the general formula (A.13) to reduce the remaining delta function, but it is simpler, and more illuminating, to make a change of variables: Let

$$E \equiv c(\sqrt{m_2^2c^2 + \rho^2} + \sqrt{m_3^2c^2 + \rho^2}) \tag{6.26}$$

(Physically, E represents the total energy of the outgoing particles—hence the choice of letter.) It follows that

$$dE = \frac{E\rho}{\sqrt{m_2^2c^2 + \rho^2}\,\sqrt{m_3^2c^2 + \rho^2}}\,d\rho \tag{6.27}$$

and therefore

$$\Gamma = \frac{S}{8\pi\hbar m_1}\int_{(m_2+m_3)c^2}^\infty |\mathcal{M}|^2\,\frac{\rho}{E}\,\delta\left(m_1c - \frac{E}{c}\right)dE \tag{6.28}$$

* In the specific case $\pi^0 \to \gamma + \gamma$ there are two identical particles in the final state, so $S = \frac{1}{2}$. For $\pi^0 \to \nu + \bar{\nu}$, $S = 1$.

But [eq. (A.9)]

$$\delta(m_1c - E/c) = c\delta(E - m_1c^2) \tag{6.29}$$

and we conclude that

$$\Gamma = \frac{S|\mathcal{M}|^2\rho_0}{8\pi\hbar m_1^2 c} \tag{6.30}$$

provided $m_1 > (m_2 + m_3)$; otherwise the delta function spike is outside the domain of integration and we get $\Gamma = 0$, recording the fact that a particle cannot decay into heavier secondaries. Here ρ_0 is the value of ρ for which $E = m_1c^2$. Solving equation (6.26), we find (Problem 6.5) that

$$\rho_0 = \frac{c}{2m_1}\sqrt{m_1^4 + m_2^4 + m_3^4 - 2m_1^2m_2^2 - 2m_1^2m_3^2 - 2m_2^2m_3^2} \tag{6.31}$$

Remember that ρ was short for the variable $|\mathbf{p}_2|$; ρ_0 is the *particular* value of $|\mathbf{p}_2|$ that is consistent with conservation of energy, and equation (6.31) simply reproduces the result we obtained back in Chapter 3 (see Problem 3.16). In more comprehensible notation, then,

$$\Gamma = \frac{S|\mathbf{p}|}{8\pi\hbar m_1^2 c}|\mathcal{M}|^2 \tag{6.32}$$

where $|\mathbf{p}|$ is the magnitude of either outgoing momentum, given in terms of the three masses by equation (6.31), and $\mathcal{M}$ is evaluated at the momenta required by the conservation laws. Notice that if $m_2 = m_3 = 0$, then $|\mathbf{p}| = m_1c/2$ and we recover equation (6.23).

The final two-body decay formula (6.32) is surprisingly simple and general. We shall use it frequently in later chapters. Unfortunately, when there are three or more particles in the final state the integrals cannot be done until we know the specific functional form of $\mathcal{M}$ for the process in question. In such cases (of which we shall encounter mercifully few) you have to go back to the Golden Rule, and work it out from scratch.

Golden Rule for Scattering Suppose particles 1 and 2 collide, producing particles 3, 4, . . . , *n*:

$$1 + 2 \longrightarrow 3 + 4 + \cdots + n \tag{6.33}$$

The cross section is given by the formula

$$d\sigma = |\mathcal{M}|^2 \frac{\hbar^2 S}{4\sqrt{(p_1 \cdot p_2)^2 - (m_1m_2c^2)^2}}\left[\left(\frac{c\,d^3\mathbf{p}_3}{(2\pi)^3 2E_3}\right)\left(\frac{c\,d^3\mathbf{p}_4}{(2\pi)^3 2E_4}\right)\cdots\left(\frac{c\,d^3\mathbf{p}_n}{(2\pi)^3 2E_n}\right)\right]$$
$$\times (2\pi)^4\,\delta^4(p_1 + p_2 - p_3 - p_4 \cdots - p_n) \tag{6.34}$$

where, as before, $p_i = (E_i/c, \mathbf{p}_i)$ is the four-momentum of particle i (mass m_i), $E_i = c\sqrt{m_i^2c^2 + \mathbf{p}_i^2}$, and S is a statistical factor ($1/j!$ for each group of j identical

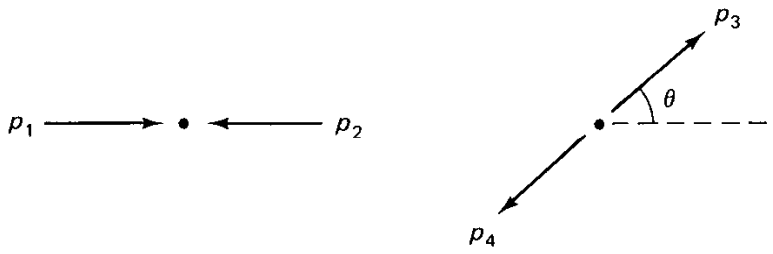


Figure 6.4 Two-body scattering in the *CM* frame.

particles in the final state). Note that the delta function enforces conservation of energy and momentum.

Equation (6.34) determines the cross section for a process in which the three-momentum of particle 3 lies in the range $d^3\mathbf{p}_3$ about the value $\mathbf{p}_3$, that of particle 4 falls in the range $d^3\mathbf{p}_4$ about $\mathbf{p}_4$, and so on. In a typical situation we study only the angle at which particle 3 emerges. In that case we integrate over all the *other* momenta ($\mathbf{p}_4, \mathbf{p}_5, \ldots, \mathbf{p}_n$), and over the magnitude of $\mathbf{p}_3$; what's left gives us $d\sigma/d\Omega$, the differential cross section for the scattering of particle 3 into solid angle $d\Omega$.

EXAMPLE 6.7 Two-Body Scattering in the *CM* Frame

Consider the process

$$1 + 2 \rightarrow 3 + 4 \tag{6.35}$$

in the *CM* frame (Fig. 6.4). If the amplitude is $\mathcal{M}$, calculate the differential cross section.

Solution. In the *CM* frame, $\mathbf{p}_2 = -\mathbf{p}_1$, and hence $p_1 \cdot p_2 = E_1E_2/c^2 + \mathbf{p}_1^2$. It follows, after some simple algebra (see Problem 6.7), that

$$\sqrt{(p_1 \cdot p_2)^2 - (m_1m_2c^2)^2} = (E_1 + E_2)|\mathbf{p}_1|/c \tag{6.36}$$

$$\text{Thus} \quad d\sigma = \left(\frac{\hbar c}{8\pi}\right)^2 \frac{S|\mathcal{M}|^2 c}{(E_1 + E_2)|\mathbf{p}_1|} \frac{d^3\mathbf{p}_3\, d^3\mathbf{p}_4}{E_3E_4} \delta^4(p_1 + p_2 - p_3 - p_4) \tag{6.37}$$

As usual, we begin by rewriting the delta function:*

$$\delta^4(p_1 + p_2 - p_3 - p_4) = \delta\left(\frac{E_1 + E_2 - E_3 - E_4}{c}\right)\delta^3(-\mathbf{p}_3 - \mathbf{p}_4) \tag{6.38}$$

Next we express the outgoing energies in terms of $\mathbf{p}_3$ and $\mathbf{p}_4$ ($E_i = c\sqrt{m_i^2c^2 + \mathbf{p}_i^2}$, and carry out the $\mathbf{p}_4$ integral (which sends $\mathbf{p}_4 \rightarrow -\mathbf{p}_3$):†

* Observe that $\mathbf{p}_1$ and $\mathbf{p}_2$ are *fixed* vectors (related by our choice of reference frame: $\mathbf{p}_2 = -\mathbf{p}_1$), but at this stage $\mathbf{p}_3$ and $\mathbf{p}_4$ are integration variables. It is only *after* the $\mathbf{p}_4$ integration that they are restricted ($\mathbf{p}_4 = -\mathbf{p}_3$), and after the $|\mathbf{p}_3|$ integration that they are *determined* by the scattering angle θ.

† I follow the standard (sloppy) notation, in which we use the same symbol, $d\sigma$, before and after integration over $\mathbf{p}_4$ (and indeed over $|\mathbf{p}_3|$ as well, as you'll see in a moment).

$$d\sigma = \left(\frac{\hbar}{8\pi}\right)^2 \frac{S|\mathcal{M}|^2 c}{(E_1 + E_2)|\mathbf{p}_1|} \times \frac{\delta((E_1 + E_2)/c) - \sqrt{m_3^2c^2 + \mathbf{p}_3^2} - \sqrt{m_4^2c^2 + \mathbf{p}_3^2})}{\sqrt{m_3^2c^2 + \mathbf{p}_3^2}\,\sqrt{m_4^2c^2 + \mathbf{p}_3^2}}\, d^3\mathbf{p}_3 \qquad (6.39)$$

This time, however, $|\mathcal{M}|^2$ depends on the *direction* of $\mathbf{p}_3$ as well as its magnitude,* so we cannot carry out the angular integration. Writing

$$d^3\mathbf{p}_3 = \rho^2\, d\rho\, d\Omega \qquad (6.40)$$

(where ρ is shorthand for $|\mathbf{p}_3|$ and $d\Omega = \sin\theta\, d\theta\, d\phi$), we obtain

$$\frac{d\sigma}{d\Omega} = \left(\frac{\hbar}{8\pi}\right)^2 \frac{Sc}{(E_1 + E_2)|\mathbf{p}_1|} \int_0^\infty |\mathcal{M}|^2 \times \frac{\delta((E_1 + E_2)/c) - \sqrt{m_3^2c^2 + \rho^2} - \sqrt{m_4^2c^2 + \rho^2})}{\sqrt{m_3^2c^2 + \rho^2}\,\sqrt{m_4^2c^2 + \rho^2}}\, \rho^2\, d\rho \qquad (6.41)$$

The integral over ρ is the same as in equation (6.25), with $m_2 \to m_4$ and $m_1 \to ((E_1 + E_2)/c^2)$. Quoting our previous result (6.32), I conclude that

$$\frac{d\sigma}{d\Omega} = \left(\frac{\hbar c}{8\pi}\right)^2 \frac{S|\mathcal{M}|^2}{(E_1 + E_2)^2} \frac{|\mathbf{p}_f|}{|\mathbf{p}_i|} \qquad (6.42)$$

where $|\mathbf{p}_f|$ is the magnitude of either outgoing momentum and $|\mathbf{p}_i|$ is the magnitude of either incoming momentum.

As in the case of decays, the two-body final state is peculiarly simple, in the sense that we are able to carry the calculation through to the end *without knowing the explicit functional form of* $\mathcal{M}$. We will be using equation (6.42) frequently in later chapters.

By the way, lifetimes obviously carry the dimensions of *time* (seconds); decay rates ($\Gamma = 1/\tau$), therefore, are measured in inverse seconds. Cross sections, have dimensions of *area*—cm^2, or, more conveniently, "barns":

$$1 \text{ barn} = 10^{-24} \text{ cm}^2 \qquad (6.43)$$

Differential cross sections, $d\sigma/d\Omega$, are given in barns per steradian, or simply barns (steradians, like radians, being dimensionless). The amplitude, $\mathcal{M}$, has units which depend on the number of particles involved: If there are n external lines (incoming plus outgoing), the dimensions of $\mathcal{M}$ are those of momentum raised to the power $4 - n$:

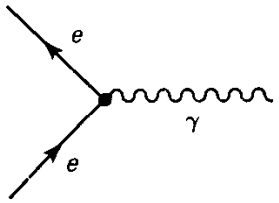
$$\text{Dimensions of } \mathcal{M} = (mc)^{4-n} \qquad (6.44)$$

* In general, $|\mathcal{M}|^2$ depends on all four momenta. However, in this case $\mathbf{p}_2 = -\mathbf{p}_1$ and $\mathbf{p}_4 = -\mathbf{p}_3$, so it remains a function only of $\mathbf{p}_1$ and $\mathbf{p}_3$, (assuming again that spin does not come into it). From these vectors we can construct three scalars: $\mathbf{p}_1 \cdot \mathbf{p}_1 = |\mathbf{p}_1|^2$, $\mathbf{p}_3 \cdot \mathbf{p}_3 = |\mathbf{p}_3|^2$, and $\mathbf{p}_1 \cdot \mathbf{p}_3 = |\mathbf{p}_1||\mathbf{p}_3| \cos\theta$. But $\mathbf{p}_1$ is fixed, so the only *integration* variables on which $|\mathcal{M}|^2$ can depend are $|\mathbf{p}_3|$ and θ.

For example, in a three-body process ($A \rightarrow B + C$), $\mathcal{M}$ has dimensions of momentum; in a four-body process ($A \rightarrow B + C + D$, or $A + B \rightarrow C + D$), $\mathcal{M}$ is dimensionless. You can check for yourself that the two Golden Rules then yield the correct units for Γ and σ.

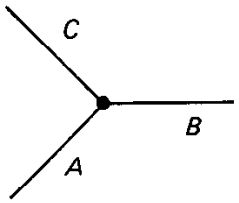
6.3 THE FEYNMAN RULES FOR A TOY THEORY

In Section 6.2 we learned how to calculate decay rates and scattering cross sections, in terms of the amplitude $\mathcal{M}$ for the process in question. Now I'll show you how to determine $\mathcal{M}$ itself, using the "Feynman rules" to evaluate the relevant diagrams. We could go straight to a "real-life" system, such as quantum electrodynamics, with electrons and photons interacting via the primitive vertex:

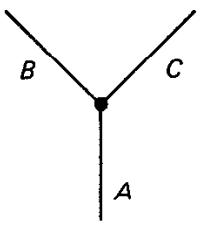


This is the original, the most important, and the best understood application of Feynman's technique. Unfortunately, it involves diverting complications (due to the fact that the electron carries spin $\frac{1}{2}$ and the photon carries spin 1) which have nothing to do with the Feynman calculus as such. In Chapter 7 I'll show you how to handle particles with spin, but for the moment I don't want to confuse the issue, so I'm going to introduce a "toy" theory, which does not pretend to represent the real world, but will serve to illustrate the *method*, with a minimum of extraneous baggage.[4]

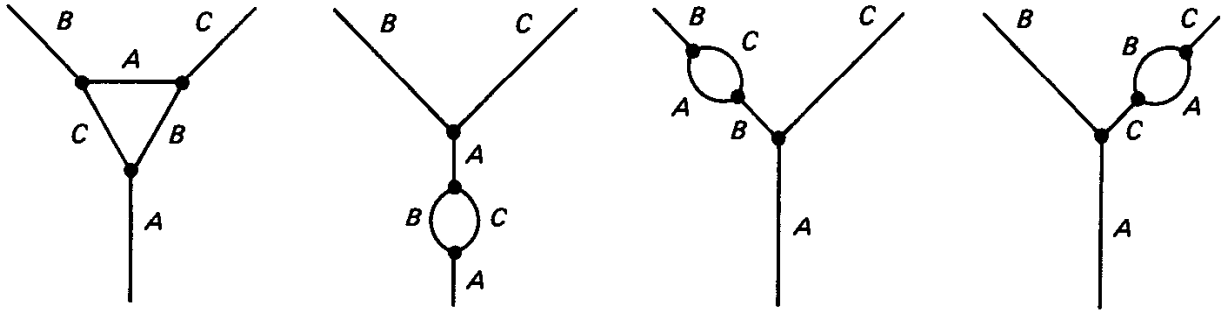
Imagine a world in which there are just three kinds of particles—call them A, B, and C—with masses m_A, m_B, and m_C. They all have spin 0, and each is its own antiparticle. There is one primitive vertex, by which the three particles interact:



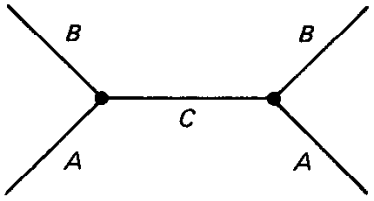
I shall assume that A is the heaviest of the three, and in fact weighs more than B and C combined, so that it can decay into $B + C$. The lowest-order diagram describing this disintegration is:



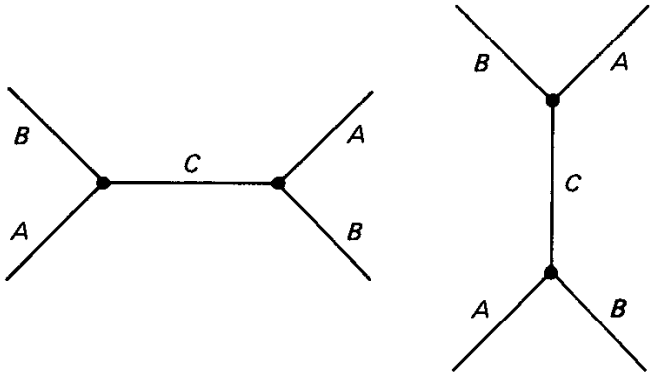
to which there are (small) third-order corrections:



and even smaller ones of higher order. Our first project will be to calculate the lifetime of the A, to lowest order. After that, we'll look at various scattering processes, such as $A + A \rightarrow B + B$:



$A + B \rightarrow A + B$:



and so on. We shall determine the cross sections for such events. Our problem is to find the amplitude $\mathcal{M}$ associated with a given Feynman diagram. The ritual is as follows:[5]

1. *Notation.* Label the incoming and outgoing four-momenta $p_1, p_2, \ldots, p_n$ (Fig. 6.5). Label the internal momenta $q_1, q_2, \ldots$. Put an arrow on each line, to keep track of the "positive" direction (arbitrarily assigned, for the internal lines).*
2. *Coupling Constant.* For each vertex, write down a factor of

$$-ig$$

g is called the *coupling constant;* it specifies the strength of the interaction between A, B, and C. In this toy theory g has the dimensions of momentum; in the "real-world" theories we shall encounter later on the coupling constant is always dimensionless.

* Since these particles are their own antiparticles, we do not need the arrows here to keep track of *that* distinction. In later chapters we shall have to be more careful.

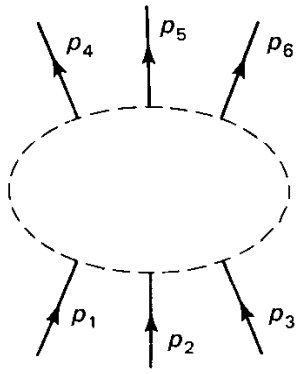


Figure 6.5 A typical Feynman diagram, with external lines labeled (internal lines not shown).

3. *Propagator.* For each internal line, write a factor

$$\frac{i}{q_j^2 - m_j^2 c^2}$$

where q_j is the four-momentum of the line ($q_j^2 \equiv q_j^\mu q_{j_\mu}$) and m_j is the mass of the particle the line describes. (Note that $q_j^2 \neq m_j^2 c^2$, because a virtual particle does not lie on its mass shell.)

4. *Conservation of Energy and Momentum.* For each vertex, write a delta function of the form

$$(2\pi)^4 \, \delta^4(k_1 + k_2 + k_3)$$

where the k's are the three four-momenta coming *into* the vertex (if the arrow leads outward, then k is *minus* the four-momentum of that line). This factor imposes conservation of energy and momentum at each vertex, since the delta function is zero unless the sum of the incoming momenta equals the sum of the outgoing momenta.

5. *Integration over Internal Momenta.* For each internal line, write down a factor

$$\frac{1}{(2\pi)^4} \, d^4 q_j$$

and integrate over all internal momenta.

6. *Cancel the Delta Function.* The result will include a delta function

$$(2\pi)^4 \, \delta^4(p_1 + p_2 + \cdots - p_n)$$

enforcing overall conservation of energy and momentum. Erase this factor, and what remains is $-i\mathcal{M}$.*

* Once you get used to it, steps 4, 5, and 6 can be collapsed into a single rule: "Integrate over all *undetermined* internal momenta." This is how most books do it, right from the start, but I think the method presented here is clearer, even if it does take a little extra time. By the way, you'll notice that every delta function carries a factor of $(2\pi)^4$, and every four-dimensional volume element carries a factor of $(2\pi)^{-4}$. *Most* of these factors eventually cancel out, and you might wonder if they are really necessary (similar remarks apply to the i's in the propagators and coupling constants). They *are* necessary, and the prescription given here is the most systematic way to keep track of them. Feynman is supposed to have shouted once in exasperation (at a graduate student who "couldn't be bothered with such trivial matters"), "If you can't get the 4π's in the right place, you don't know *nothing*!"

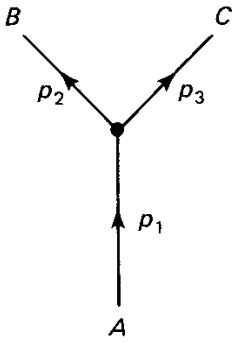


Figure 6.6 Lowest-order contribution to $A \rightarrow B + C$.

In the following sections we'll see how these rules are used to evaluate some elementary Feynman diagrams in the "ABC theory."

6.4 LIFETIME OF THE *A*

The simplest possible diagram, representing the lowest-order contribution to $A \rightarrow B + C$, has no internal lines at all (Fig. 6.6). There is one vertex, at which we pick up a factor of $-ig$ (rule 2) and a delta function

$$(2\pi)^4 \, \delta^4(p_1 - p_2 - p_3)$$

(rule 4), which we promptly discard (rule 6), obtaining $-i\mathcal{M} = -ig$, or

$$\mathcal{M} = g \tag{6.45}$$

This is the *amplitude* (to lowest order); the decay rate is found by plugging $\mathcal{M}$ into equation (6.32):

$$\Gamma = \frac{g^2 |\mathbf{p}|}{8\pi \hbar m_A^2 c} \tag{6.46}$$

where $|\mathbf{p}|$ (the magnitude of either outgoing momentum) is

$$|\mathbf{p}| = \frac{c}{2m_A} \sqrt{m_A^4 + m_B^4 + m_C^4 - 2m_A^2 m_B^2 - 2m_A^2 m_C^2 - 2m_B^2 m_C^2} \tag{6.47}$$

The lifetime of the A, then, is

$$\tau = \frac{1}{\Gamma} = \frac{8\pi \hbar m_A^2 c}{g^2 |\mathbf{p}|} \tag{6.48}$$

You should check for yourself that τ comes out with the correct units.

6.5 SCATTERING

The lowest-order contribution to the process $A + A \rightarrow B + B$ is shown in Figure 6.7. In this case there are two vertices (hence two factors of $-ig$), one internal line, with the propagator

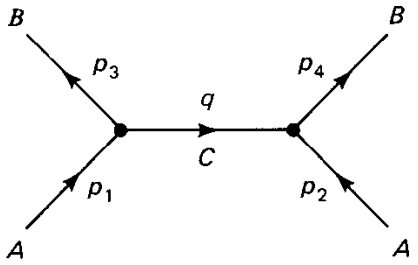


Figure 6.7 Lowest-order contribution to $A + A \rightarrow B + B$.

$$\frac{i}{q^2 - m_C^2 c^2}$$

two delta functions:

$$(2\pi)^4\, \delta^4(p_1 - p_3 - q) \quad \text{and} \quad (2\pi)^4\, \delta^4(p_2 + q - p_4)$$

and one integration:

$$\frac{1}{(2\pi)^4}\, d^4q$$

Rules 1 through 5, then, yield

$$-i(2\pi)^4 g^2 \int \frac{1}{q^2 - m_C^2 c^2}\, \delta^4(p_1 - p_3 - q)\delta^4(p_2 + q - p_4) d^4q$$

The second delta function serves to pick out the value of everything else at the point $q = p_4 - p_2$, so we have

$$-ig^2 \frac{1}{(p_4 - p_2)^2 - m_C^2 c^2}\, (2\pi)^4\, \delta^4(p_1 + p_2 - p_3 - p_4)$$

As promised, there is one remaining delta function, reflecting overall conservation of energy and momentum. Erasing it (rule 6), we are left with

$$\mathcal{M} = \frac{g^2}{(p_4 - p_2) - m_C^2 c^2} \tag{6.49}$$

But that's not the whole story, for there is another diagram of order g^2, obtained by "twisting" the B lines (Fig. 6.8). (You don't get yet another new diagram by twisting the A lines; the only choice here is whether p_3 connects to p_1 or to p_2). Since this differs from Figure 6.7 only by the interchange $p_3 \leftrightarrow p_4$, there is no need to compute it from scratch; quoting equation (6.49), we

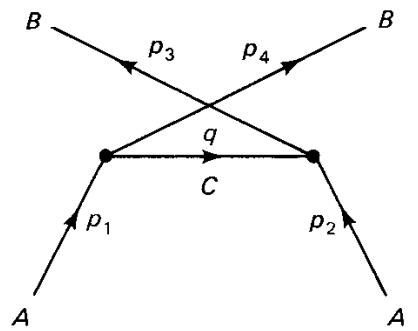


Figure 6.8 Second diagram contributing in lowest order to $A + A \rightarrow B + B$.

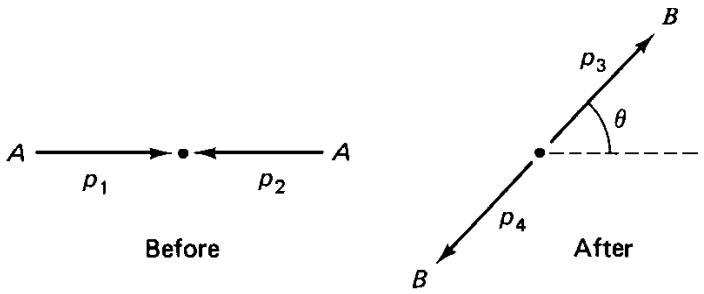


Figure 6.9 $A + A \rightarrow B + B$ in the *CM* frame.

can write down immediately the total amplitude (to order g^2) for the process $A + A \rightarrow B + B$:

$$\mathcal{M} = \frac{g^2}{(p_4 - p_2)^2 - m_C^2c^2} + \frac{g^2}{(p_3 - p_2)^2 - m_C^2c^2} \tag{6.50}$$

Notice, incidentally, that $\mathcal{M}$ is a Lorentz-invariant quantity. This is *always* the case; it is built into the Feynman rules.

Suppose we are interested in the differential cross section $(d\sigma/d\Omega)$ for this process, in the *CM* system (Fig. 6.9). Say, for simplicity, that $m_A = m_B = m$ and $m_C = 0$. Then

$$(p_4 - p_2)^2 - m_C^2c^2 = p_4^2 + p_2^2 - 2p_2 \cdot p_4 = -2\mathbf{p}^2(1 - \cos\theta) \tag{6.51}$$

$$(p_3 - p_2)^2 - m_C^2c^2 = p_3^2 + p_2^2 - 2p_3 \cdot p_2 = -2\mathbf{p}^2(1 + \cos\theta) \tag{6.52}$$

(where $\mathbf{p}$ is the incident momentum of particle 1), and hence

$$\mathcal{M} = -\frac{g^2}{\mathbf{p}^2 \sin^2\theta} \tag{6.53}$$

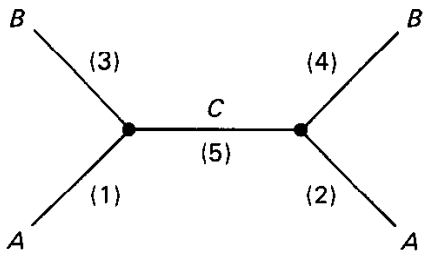
According to equation (6.42), then,

$$\frac{d\sigma}{d\Omega} = \frac{1}{2}\left(\frac{\hbar c g^2}{16\pi E \mathbf{p}^2 \sin^2\theta}\right)^2 \tag{6.54}$$

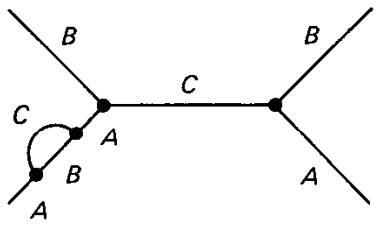
In this case, as for Rutherford scattering (Example 6.4), the *total* cross section is infinite.

6.6 HIGHER-ORDER DIAGRAMS

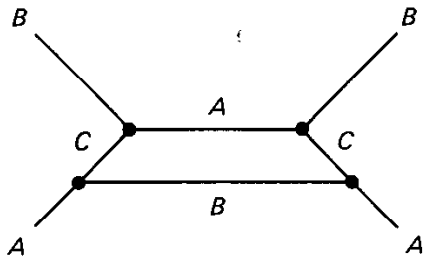
So far we have looked only at lowest-order ("tree level") Feynman diagrams; in the case of $A + A \rightarrow B + B$, for instance, we considered the graph:



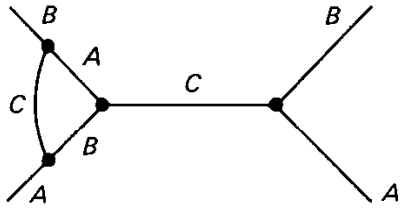
This diagram has two vertices, so that $\mathcal{M}$ is proportional to g^2. But there are a number of diagrams with *four* vertices which contribute to this process. If the added line starts on line (1), it could terminate also on line (1):



or it could terminate on line (2):



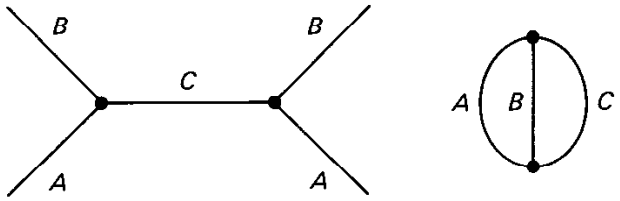
or on line (3):



or on line (4) or line (5). So there are five diagrams in which the added line attaches to line (1). There are five also, of course, in which it attaches to line (2), but we have already counted *one* of these (the one joining lines (1) and (2)), so there are four *new* diagrams. Likewise, three for line (3), two for line (4), and one for line (5). All told, then, there are

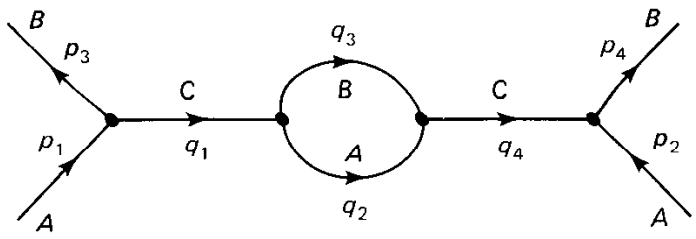
$$5 + 4 + 3 + 2 + 1 = 15$$

fourth-order diagrams for this process, and another 15 for the "twisted" version. (Disconnected diagrams, such as



don't count.)

I am certainly not going to evaluate all these "one-loop" diagrams (or even *think* about *two*-loop diagrams), but I would like to take a closer look at *one* of them—the one in which line (5) joins onto itself:



Applying Feynman Rules 1 through 5, we obtain

$$g^4 \int \frac{\delta^4(p_1 - q_1 - p_3)\delta^4(q_1 - q_2 - q_3)\delta^4(q_2 + q_3 - q_4)\delta^4(q_4 + p_2 - p_4)}{(q_1^2 - m_C^2c^2)(q_2^2 - m_A^2c^2)(q_3^2 - m_B^2c^2)(q_4^2 - m_C^2c^2)} \times d^4q_1\, d^4q_2\, d^4q_3\, d^4q_4 \qquad (6.55)$$

Integration over q_1, using the first delta function, replaces q_1 by $(p_1 - p_3)$; integration over q_4, using the last delta function, replaces q_4 by $(p_4 - p_2)$:

$$\frac{g^4}{[(p_1 - p_3)^2 - m_C^2c^2][(p_4 - p_2)^2 - m_C^2c^2]} \times \int \frac{\delta^4(p_1 - p_3 - q_2 - q_3)\delta^4(q_2 + q_3 - p_4 + p_2)}{(q_2^2 - m_A^2c^2)(q_3^2 - m_B^2c^2)}\, d^4q_2\, d^4q_3 \qquad (6.56)$$

Here the first delta function replaces q_2 by $p_1 - p_3 - q_3$, so the second delta function becomes

$$\delta^4(p_1 + p_2 - p_3 - p_4)$$

which, by rule 6, we erase, leaving

$$\mathcal{M} = i\left(\frac{g}{2\pi}\right)^4 \frac{1}{[(p_1 - p_3)^2 - m_C^2c^2]^2} \int \frac{1}{[(p_1 - p_3 - q)^2 - m_A^2c^2](q^2 - m_B^2c^2)}\, d^4q \qquad (6.57)$$

(I drop the subscript on q_3 at this point.)

You can try calculating this integral, if you've got the energy, but I'll tell you right now you're going to hit a snag. For the four-dimensional volume element could be written $d^4q = q^3\, dq\, d\Omega'$ (where $d\Omega'$ stands for the angular part) (just as in two-dimensional polar coordinates the element of area is $r\, dr\, d\theta$ and in three-dimensional spherical coordinates the volume element is $r^2\, dr \sin\theta\, d\theta\, d\phi$). At large q the integrand is essentially just $1/q^4$, so the q integral has the form

$$\int^{\infty} \frac{1}{q^4} q^3\, dq = \ln q|^{\infty} = \infty \qquad (6.58)$$

The integral is logarithmically divergent at large q. This disaster, in one form or another, held up the development of quantum electrodynamics for nearly two decades, until, through the combined efforts of many great physicists—from Dirac, Pauli, Kramers, Weisskopf, and Bethe through Tomonaga, Schwinger, and Feynman—systematic methods were developed for "sweeping the infinities under the rug." The first step is to *regularize* the integral, using a suitable cutoff procedure that renders it finite without spoiling other desirable features (such as Lorentz invariance). In the case of equation (6.57) this can be accomplished by introducing a factor

$$\frac{-M^2c^2}{(q^2 - M^2c^2)} \qquad (6.59)$$

under the integral sign. The *cutoff* mass M is assumed to be very large, and will be taken to infinity at the end of the calculation (note that the "fudge factor" (6.59) goes to 1 as $M \to \infty$).* The integral can now be calculated[6] and separated into two parts: a finite term, independent of M, and a term involving (in this case) the logarithm of M, which blows up as $M \to \infty$.

At this point a miraculous thing happens: all the divergent, M-dependent terms appear in the final answer in the form of *additions to the masses and the coupling constant.* If we take this seriously, it means that the *physical* masses and couplings are not the m's and g's that appeared in the original Feynman rules, but rather the "renormalized" ones, containing these extra factors:

$$m_{\text{physical}} = m + \delta m; \qquad g_{\text{physical}} = g + \delta g \tag{6.60}$$

The fact that δm and δg are infinite (in the limit $M \to \infty$) is disturbing, but not catastrophic, for we never measure them anyway; all we ever see in the laboratory are the *physical* values, and these are (obviously) finite (evidently the unmeasurable "bare" masses and couplings, m and g, contain compensating infinities).† As a practical matter, we take account of the infinities by using the *physical* values of m and g in the Feynman rules, and then systematically ignoring the divergent contributions from higher-order diagrams.

Meanwhile, there remain the *finite* (*M-in*dependent) contributions from the loop diagrams. They, too, lead to modifications in m and g (perfectly calculable ones, in this case)—which, however, are functions of the four-momentum of the line in which the loop is inserted ($p_1 - p_3$ in the example). This means that the *effective* masses and coupling constants actually depend on the *energies* of the particles involved; we call them "running" masses and "running" coupling constants. The dependence is typically rather slight, at low energies, and can ordinarily be ignored, but it does have observable consequences, in the form of the Lamb shift (in QED) and asymptotic freedom (in QCD).‡

The procedure I have sketched in the last three paragraphs is called *renor-*

* No one would deny that this procedure is artificial. Still, it can be argued that expression (6.59) merely reflects our ignorance of the high-energy (short distance) behavior of quantum field theory. Perhaps the Feynman propagators are not quite right in this régime, and M is simply a crude way of accounting for the unknown modification. (This would be the case, for example, if the "particles" have substructure that becomes relevant at extremely close range.) Dirac once remarked:

> It's just a stop-gap procedure. There must be some fundamental change in our ideas, probably a change just as fundamental as the passage from Bohr's orbit theory to quantum mechanics. When you get a number turning out to be infinite which ought to be finite, you should admit that there is something wrong with your equations, and not hope that you can get a good theory just by doctoring up that number. [From P. Buckley and F. D. Peat, *A Question of Physics* (Toronto: University of Toronto Press, 1979), page 39.]

† In case it is some comfort, I should point out that essentially the same thing occurs in *classical* electrodynamics: the electrostatic energy of a point charge is infinite, and makes an infinite contribution (via $E = mc^2$) to the particle's mass. Perhaps this means that there are no true point charges, in classical electrodynamics; perhaps that's what it means in quantum field theory, too. In neither case, however, do we know how to avoid the point particle as a theoretical construct.

‡ A *physical* interpretation of the running coupling constant in *QED* and *QCD* was suggested in Chapter 2, Section 2.3. A nice explanation of mass renormalization is given by P. Nelson in *American Scientist,* **73,** 66 (1985): *(footnote continues on p. 210)*

malization. If *all* the infinities arising from higher-order diagrams can be accommodated in this way, we say that the theory is *renormalizable. ABC* theory and quantum electrodynamics are renormalizable. In the early seventies 't Hooft showed that all *gauge* theories, including chromodynamics and the electroweak theory of Glashow, Weinberg, and Salam, are renormalizable. This was a profoundly important discovery, because, beyond lowest-order calculations, a nonrenormalizable theory yields answers that are cutoff-dependent and, therefore, really, quite meaningless.

REFERENCES AND NOTES

1. See, for example, H. Goldstein, *Classical Mechanics,* 2d Ed. (Reading, MA: Addison-Wesley, 1980), Sects. 3–10.
2. See, for example, D. Park, *Introduction to the Quantum Theory,* 2d Ed. (New York: McGraw-Hill, 1974), Sect. 8.7.
3. Convincing and accessible accounts are in fact difficult to find. R. P. Feynman, *Theory of Fundamental Processes* (Reading, MA: Benjamin, 1961), Chapters 15 and 16, is a good place to start. See also J. M. Jauch and F. Rohrlich, *Theory of Photons and Electrons,* 2d Ed. (Berlin: Springer, 1976), Sect. 8.6, or R. Hagedorn, *Relativistic Kinematics* (New York: Benjamin, 1963), Chap. 7.
4. This model was shown to me by Max Dresden.
5. You may well ask where these rules *come* from, and I am not sure Feynman himself could have given you a completely satisfactory answer in 1949, when he first published the rules for QED (R. P. Feynman, *Phys. Rev.* **76,** 749 and 769 (1949)). It remained for Freeman Dyson to show how the Feynman Rules could be obtained from quantum field theory (F. J. Dyson, *Phys. Rev.* **75,** 486 and 1376 (1949) and **82,** 428 (1951)). For a fascinating personal history of these events, see Chapters 5 and 6 of Dyson's book *Disturbing The Universe* (New York: Harper & Row, 1979). We shall return to the question of how Feynman's Rules are derived in Chapter 11; for now I will simply treat them as axioms.
6. The method is explained in J. J. Sakurai, *Advanced Quantum Mechanics* (Reading, MA: Addison-Wesley, 1967); see particularly the useful collection of formulas in Appendix E.

According to renormalization theory, not only the strengths of the various interactions but the masses of the participating particles appear to vary on differing length scales. To get a feel for this seemingly paradoxical statement, imagine firing a cannon underwater. Even neglecting friction, the trajectory will be very different from the corresponding one on land, since the cannonball must now drag with it a considerable amount of water, modifying its apparent, or "effective," mass. We can experimentally measure the cannonball's effective mass by shaking it to and fro at a rate ω, computing the mass from $F = ma$. (This is how astronauts "weigh" themselves in space.) Having found the effective mass, we can now replace the difficult problem of underwater ballistics by a simplified approximation: we ignore the water altogether, but in Newton's equations we simply replace the true cannonball mass by the effective mass. The complicated details of the interaction with the medium are thus reduced to determining one effective parameter.

A key feature of this approach is that the effective mass so computed depends on ω, since as ω approaches zero, for example, the water has no effect whatever. In other words, the presence of a medium can introduce a scale-dependent effective mass. We say that the effective mass is "renormalized" by the medium. In quantum physics, every particle moves through a "medium" consisting of the quantum fluctuations of all particles present in the theory. We again take into account this medium by ignoring it but changing the values of our parameters to scale-dependent "effective" values.

PROBLEMS

6.1. Derive equation (6.3). [*Hint:* What fraction of the original sample decays between t and $t + dt$? What, then, is the probability of any given particle decaying between t and $t + dt$? Take it from there.]

6.2. Nuclear physicists traditionally work with "half-life" ($t_{1/2}$) instead of mean life (τ); $t_{1/2}$ is the time it takes for half the members of a large sample to decay. Show that for exponential decay [eq. (6.2)]

$$t_{1/2} = \tau \ln 2$$

6.3. (a) Suppose you started out with a million muons (at rest); how many would still be around 2.2×10^{-5} sec later? (b) What is the probability of a π^- lasting more than 1 sec?

6.4. A nonrelativistic particle of mass m scatters from a fixed repulsive potential, $V(r) = k/r^2$, where k is a constant.
(a) Find the scattering angle, θ, as a function of the impact parameter, b.
(b) Determine the differential cross section $d\sigma/d\Omega$.
(c) Find the total cross section.
[References: Goldstein (ref. 1), p. 108, eq. (3-97), and Becker, *Introduction to Theoretical Mechanics* (New York: McGraw-Hill, 1954), p. 231, Example 10-3.]

6.5. Derive equation (6.31), using definition (6.26).

6.6. As an application of Example 6.5, consider the decay of $\pi^0 \rightarrow \gamma + \gamma$. Of course, the π^0 is a composite object ($u\bar{u}$ and $d\bar{d}$), and so equation (6.23) does not really apply. But let's pretend that the π^0 is a true elementary particle, and see how close we come. Unfortunately, we don't know the amplitude $\mathcal{M}$; however, it must have the dimensions of mass times velocity [eq. (6.44)], and there is only one mass and one velocity available. Moreover, the emission of each photon introduces a factor of $\sqrt{\alpha}$ (the fine structure constant) into $\mathcal{M}$, as we shall see in Chapter 7, so the amplitude must be proportional to α. On this basis, estimate the lifetime of the π^0. Compare the experimental value. [Evidently, the decay of the π^0 is a much more complicated process than this crude model suggests. See C. Quigg, *Gauge Theories of the Strong, Weak, and Electromagnetic Interactions,* Reading, MA: Benjamin/Cummings, 1983, eq. (1.2.25).]

6.7. (a) Derive equation (6.36) for scattering of particles 1 and 2 in the *CM*.
(b) Obtain the corresponding formula for the lab frame (particle 2 at rest).

$$[\textit{Answer: } \sqrt{(p_1 \cdot p_2)^2 - (m_1 m_2 c^2)^2} = m_2 |\mathbf{p}_1| c]$$

6.8. Consider the case of elastic scattering, $A + B \rightarrow A + B$, in the lab frame (B initially at rest) assuming the target is so heavy ($m_B c^2 \gg E_A$) that its recoil is negligible. Use equation (6.34) to determine the differential scattering cross section.

$$[\textit{Answer: } (d\sigma/d\Omega) = (\hbar/8\pi m_B c)^2 |\mathcal{M}|^2]$$

6.9. Consider the collision $1 + 2 \rightarrow 3 + 4$ in the lab frame (2 at rest), with particles 3 and 4 massless. Obtain the formula for the differential cross section.

$$\left[\textit{Answer: } \frac{d\sigma}{d\Omega} = \left(\frac{\hbar}{8\pi}\right)^2 \frac{S|\mathcal{M}|^2 |\mathbf{p}_3|}{m_2 |\mathbf{p}_1| (E_1 + m_2 c^2 - |\mathbf{p}_1| c \cos\theta)}\right]$$

6.10. (a) Analyze the problem of elastic scattering ($m_3 = m_1$, $m_4 = m_2$) in the lab frame (particle 2 at rest). Derive the formula for the differential cross section.

$$\left[\textit{Answer: } \frac{d\sigma}{d\Omega} = \left(\frac{\hbar}{8\pi}\right)^2 \frac{\mathbf{p}_3^2 S|\mathcal{M}|^2}{m_2|\mathbf{p}_1||\mathbf{p}_3|(E_1 + m_2c^2) - |\mathbf{p}_1|E_3\cos\theta|}\right]$$

(b) If the incident particle is massless ($m_1 = 0$), show that the result in part (a) simplifies to

$$\left(\frac{d\sigma}{d\Omega}\right) = S\left(\frac{\hbar E_3}{8\pi m_2 c E_1}\right)^2 |\mathcal{M}|^2$$

6.11. **(a)** Is $A \to B + B$ a possible process in the ABC theory?

(b) Suppose a diagram has n_A external A lines, n_B external B lines, and n_C external C lines. Develop a simple criterion for determining whether it is an allowed reaction.

(c) Assuming A is heavy enough, what are the next most likely decay modes, after $A \to B + C$? Draw a Feynman diagram for each decay.

6.12. **(a)** Draw all the lowest-order diagrams for $A + A \to A + A$. (There are six of them.)

(b) Find the amplitude for this process, in lowest order, assuming $m_B = m_C = 0$. Leave your answer in the form of an integral over one remaining four-momentum, q.

6.13. Calculate $d\sigma/d\Omega$ for $A + A \to B + B$, in the CM frame, assuming $m_B = m_C = 0$. Find the total cross section, σ.

6.14. Find $d\sigma/d\Omega$ and σ for $A + A \to B + B$ in the lab frame. (Let E be the energy, and $\mathbf{p}$ the momentum, of the incident A. Assume $m_B = m_C = 0$.) Determine the non-relativistic and ultrarelativistic limits of your formula.

6.15. **(a)** Determine the lowest-order amplitude for $A + B \to A + B$. [*Note:* There are two diagrams.]

(b) Find the differential cross section for this process in the CM frame, assuming $m_A = m_B = m$, $m_C = 0$. Express your answer in terms of the incident energy, E, and the scattering angle, θ.

(c) Find $d\sigma/d\Omega$ for this process in the lab frame, assuming B is much heavier than A, and remains stationary. A is incident with energy E. [*Hint:* See Problem (6.8). Assume $m_B \gg m_A$, m_C, and E/c^2.]

(d) In case (c), find the total cross section, σ.

Chapter 7

Quantum Electrodynamics

In this chapter I introduce the Dirac equation, state the Feynman rules for quantum electrodynamics, develop some useful calculational tools, and derive some of the classic QED results. The treatment leans heavily on material from Chapters 2, 3, and 6, as well as on the spin-$\frac{1}{2}$ formalism in Chapter 4. In turn, Chapter 7 is the indispensable foundation for everything that follows (however, you may want to skip Example 7.8 and Section 7.9, together with the related passages in Chapters 8 and 9).

7.1 THE DIRAC EQUATION

Although the "ABC" model in Chapter 6 is a perfectly legitimate quantum field theory, it does not describe the real world, because the particles A, B, and C have spin 0, whereas quarks and leptons carry spin $\frac{1}{2}$, and mediators carry spin 1. The inclusion of spin can be algebraically cumbersome; that's why I decided to introduce the Feynman calculus in the context of a "toy" theory free of such distractions. In *nonrelativistic* quantum mechanics particles are described by the *Schrödinger* equation; in *relativistic* quantum mechanics particles of spin 0 are described by the *Klein–Gordon* equation, particles of spin $\frac{1}{2}$ by the *Dirac* equation, and particles of spin 1 by the *Proca* equation. Once the Feynman rules have been established, however, the underlying field equation fades into the background—that's how we got through Chapter 6 without ever mentioning the Klein–Gordon equation. But for spin $\frac{1}{2}$ the very *notation* of the Feynman rules presupposes some familiarity with the Dirac equation. So for the next three sections we shall study the Dirac theory in its own right.

In Chapter 5 I "derived" the Schrödinger equation by starting with the *classical* energy-momentum relation:

$$\frac{\mathbf{p}^2}{2m} + V = E \tag{7.1}$$

applying the quantum prescription

$$\mathbf{p} \rightarrow \frac{\hbar}{i}\nabla, \qquad E \rightarrow i\hbar\frac{\partial}{\partial t} \tag{7.2}$$

and letting the resulting operator act on the "wave function," ψ:

$$-\frac{\hbar^2}{2m}\nabla^2\Psi + V\Psi = i\hbar\frac{\partial\Psi}{\partial t} \qquad \textbf{(Schrödinger equation)} \tag{7.3}$$

The Klein–Gordon equation can be obtained in exactly the same way, beginning with the *relativistic* energy-momentum relation, $E^2 - \mathbf{p}^2c^2 = m^2c^4$, or (better)

$$p^\mu p_\mu - m^2c^2 = 0 \tag{7.4}$$

(I'll leave out the potential energy, from now on; we'll stick to *free* particles). Surprisingly, the quantum prescription (7.2) requires no relativistic modification; in four-vector notation, it reads

$$p_\mu \rightarrow i\hbar\,\partial_\mu \tag{7.5}$$

Here*

$$\partial_\mu \equiv \frac{\partial}{\partial x^\mu} \tag{7.6}$$

which is to say

$$\partial_0 = \frac{1}{c}\frac{\partial}{\partial t}, \qquad \partial_1 = \frac{\partial}{\partial x}, \qquad \partial_2 = \frac{\partial}{\partial y}, \qquad \partial_3 = \frac{\partial}{\partial z} \tag{7.7}$$

Putting (7.5) into (7.4), and letting the derivatives act on a wave function ψ, we obtain

$$-\hbar^2\,\partial^\mu\partial_\mu\psi - m^2c^2\psi = 0 \tag{7.8}$$

or

$$-\frac{1}{c^2}\frac{\partial^2\psi}{\partial t^2} + \nabla^2\psi = \left(\frac{mc}{\hbar}\right)^2\psi \qquad \textbf{(Klein–Gordon equation)} \tag{7.9}$$

Schrödinger apparently discovered this equation even before the nonrelativistic one that bears his name; it was eventually rejected on the ground that (for reasons we need not go into) it was incompatible with the statistical interpretation of ψ [which says that $|\psi|^2$ is the probability of finding the particle at the point (x, y, z)]. The source of the difficulty was traced to the fact that the Klein–Gordon equation is second order in t.† So Dirac set out to find an equation consistent with the relativistic energy-momentum formula, and yet first order in time. Ironically, in 1934 Pauli and Weisskopf showed that the statistical in-

* The gradient with respect to a *contravariant* position-time four-vector x^μ is itself a *co*variant four-vector, hence the placement of the index. Written out in full, equation (7.5) says $(E/c, -\mathbf{p}) \rightarrow i\hbar\left(\frac{1}{c}\frac{\partial}{\partial t}, \nabla\right)$. Of course, $\partial^\mu \equiv \partial/\partial x_\mu$ (See Problem 7.1.)

† Notice that the Schrödinger equation (7.3) is *first* order in t.

terpretation itself is flawed in relativistic quantum theory,* and restored the Klein–Gordon equation to its rightful place, while keeping the Dirac equation for particles of spin $\frac{1}{2}$.

Dirac's basic strategy was to "factor" the energy-momentum relation (7.4). This would be easy if we had only p^0 (that is, if $\mathbf{p}$ were zero):

$$(p^0)^2 - m^2c^2 = (p^0 + mc)(p^0 - mc) = 0 \tag{7.10}$$

We then obtain two first-order equations:

$$(p^0 - mc) = 0 \quad \text{or} \quad (p^0 + mc) = 0 \tag{7.11}$$

either one of which guarantees that $p^\mu p_\mu - m^2c^2 = 0$. But it's a different matter when the other three components of p^μ are included; in that case we are looking for something of the form

$$(p^\mu p_\mu - m^2c^2) = (\beta^\kappa p_\kappa + mc)(\gamma^\lambda p_\lambda - mc) \tag{7.12}$$

where β^κ and γ^λ are eight coefficients to be determined.† Multiplying out the right-hand side, we have

$$\beta^\kappa\gamma^\lambda p_\kappa p_\lambda - mc(\beta^\kappa - \gamma^\kappa)p_\kappa - m^2c^2$$

We don't want any terms linear in p_κ, so we must choose $\beta^\kappa = \gamma^\kappa$; to finish the job, we need to find coefficients γ^κ such that

$$p^\mu p_\mu = \gamma^\kappa\gamma^\lambda p_\kappa p_\lambda$$

which is to say

$$\begin{aligned}(p^0)^2 - (p^1)^2 - (p^2)^2 - (p^3)^2 = {} & (\gamma^0)^2(p^0)^2 + (\gamma^1)^2(p^1)^2 + (\gamma^2)^2(p^2)^2 \\ & + (\gamma^3)^2(p^3)^2 + (\gamma^0\gamma^1 + \gamma^1\gamma^0)p_0p_1 \\ & + (\gamma^0\gamma^2 + \gamma^2\gamma^0)p_0p_2 + (\gamma^0\gamma^3 + \gamma^3\gamma^0)p_0p_3 \\ & + (\gamma^1\gamma^2 + \gamma^2\gamma^1)p_1p_2 + (\gamma^1\gamma^3 + \gamma^3\gamma^1)p_1p_3 \\ & + (\gamma^2\gamma^3 + \gamma^3\gamma^2)p_2p_3 \end{aligned} \tag{7.13}$$

You see the problem: we could pick $\gamma^0 = 1$ and $\gamma^1 = \gamma^2 = \gamma^3 = i$, but there doesn't seem to be any way to get rid of the "cross-terms". At this point Dirac had a brilliant inspiration: what if the γ's are *matrices,* instead of *numbers?* Since matrices don't commute, we just might be able to find a set such that

$$\begin{aligned}&(\gamma^0)^2 = 1, \qquad (\gamma^1)^2 = (\gamma^2)^2 = (\gamma^3)^2 = -1, \\ &\gamma^\mu\gamma^\nu + \gamma^\nu\gamma^\mu = 0, \qquad \text{for } \mu \neq \nu\end{aligned} \tag{7.14}$$

Or, more succinctly,

$$\{\gamma^\mu, \gamma^\nu\} = 2g^{\mu\nu} \tag{7.15}$$

* The essential point is that a relativistic theory must account for pair production and annihilation, and hence the number of particles is not a conserved quantity.

† In case the notation confuses you, I'll write equation (7.12) out "long-hand":

$$\begin{aligned}&(p^0)^2 - (p^1)^2 - (p^2)^2 - (p^3)^2 - m^2c^2 \\ &\qquad = (\beta^0p^0 - \beta^1p^1 - \beta^2p^2 - \beta^3p^3 + mc)(\gamma^0p^0 - \gamma^1p^1 - \gamma^2p^2 - \gamma^3p^3 - mc)\end{aligned}$$

where $g^{\mu\nu}$ is the Minkowski metric [eq. (3.13)], and curly brackets denote the *anticommutator:*

$$\{A, B\} \equiv AB + BA \tag{7.16}$$

You might try fiddling with this problem for yourself. It turns out that it *can* be done, although the smallest matrices that work are 4×4. There are a number of essentially equivalent sets of "gamma matrices"; we'll use the standard "Bjorken and Drell" convention:[1]

$$\gamma^0 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}, \qquad \gamma^i = \begin{pmatrix} 0 & \sigma^i \\ -\sigma^i & 0 \end{pmatrix} \tag{7.17}$$

where σ^i ($i = 1, 2, 3$) is the indicated Pauli matrix [eq. (4.26)], 1 denotes the 2×2 unit matrix, and 0 is the 2×2 matrix of zeroes.*

As a 4×4 matrix equation, then, the relativistic energy-momentum relation *does* factor:

$$(p^\mu p_\mu - m^2c^2) = (\gamma^\kappa p_\kappa + mc)(\gamma^\lambda p_\lambda - mc) = 0 \tag{7.18}$$

We obtain the Dirac equation, now, by peeling off one term (it doesn't really matter *which* one, but this is the conventional choice—see Problem 7.10):

$$\gamma^\mu p_\mu - mc = 0 \tag{7.19}$$

Now we make the usual substitution $p_\mu \rightarrow i\hbar\,\partial_\mu$ [eq. (7.5)] and let the result act on the wave function ψ:

$$i\hbar\gamma^\mu\partial_\mu\psi - mc\psi = 0 \qquad \textbf{(Dirac equation)} \tag{7.20}$$

Note that ψ is now a four-element column matrix:

$$\psi = \begin{pmatrix} \psi_1 \\ \psi_2 \\ \psi_3 \\ \psi_4 \end{pmatrix} \tag{7.21}$$

We call it a "bi-spinor," or "Dirac spinor." (Although it carries four components, this object is *not* a four-vector. In Section 7.3 I'll show you how it *does* transform when you change inertial systems; it's *not* going to be an ordinary Lorentz transformation.)

7.2 SOLUTIONS TO THE DIRAC EQUATION

Let's now look for simple solutions to the Dirac equation. Suppose first that ψ is *independent of position:*

* When the context allows no room for ambiguity, I'll use 1 and 0 this way for 2×2 or 4×4 matrices; also, a unit matrix of the appropriate dimension is *implied,* when necessary, as on the right-hand side of equation (7.15). Incidentally, since σ is not the spatial part of a four-vector, we do not distinguish upper and lower indices: $\sigma^i \equiv \sigma_i$.

$$\frac{\partial\psi}{\partial x} = \frac{\partial\psi}{\partial y} = \frac{\partial\psi}{\partial z} = 0 \tag{7.22}$$

In view of equation (7.5), this describes a state with zero momentum ($\mathbf{p} = 0$). The Dirac equation (7.20) reduces to

$$\frac{i\hbar}{c}\gamma^0\frac{\partial\psi}{\partial t} - mc\psi = 0 \tag{7.23}$$

or

$$\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}\begin{pmatrix} \partial\psi_A/\partial t \\ \partial\psi_B/\partial t \end{pmatrix} = -i\frac{mc^2}{\hbar}\begin{pmatrix} \psi_A \\ \psi_B \end{pmatrix} \tag{7.24}$$

where

$$\psi_A = \begin{pmatrix} \psi_1 \\ \psi_2 \end{pmatrix} \tag{7.25}$$

carries the upper two components, and

$$\psi_B = \begin{pmatrix} \psi_3 \\ \psi_4 \end{pmatrix} \tag{7.26}$$

carries the lower two. Thus

$$\frac{\partial\psi_A}{\partial t} = -i\left(\frac{mc^2}{\hbar}\right)\psi_A, \qquad -\frac{\partial\psi_B}{\partial t} = -i\left(\frac{mc^2}{\hbar}\right)\psi_B \tag{7.27}$$

and the solutions are

$$\psi_A(t) = e^{-i(mc^2/\hbar)t}\psi_A(0), \qquad \psi_B(t) = e^{+i(mc^2/\hbar)t}\psi_B(0) \tag{7.28}$$

Referring to equation (5.10) we recognize the factor

$$e^{-iEt/\hbar} \tag{7.29}$$

as the characteristic time dependence of a quantum state with energy E. For a particle at rest, $E = mc^2$, so ψ_A is exactly what we should have expected, in the case $\mathbf{p} = 0$. But what about ψ_B? It ostensibly represents a state with *negative energy* ($E = -mc^2$). This is the famous disaster I mentioned back in Chapter 1, which Dirac at first tried to avoid by postulating an unseen infinite "sea" of negative energy particles, which fill up all those unwanted states.* Instead, we now interpret the "negative energy" solutions as representing *anti*particles with *positive* energy. Thus ψ_A describes electrons (for example), whereas ψ_B describes *positrons*. Each is a two-component spinor, just right for a system of spin $\frac{1}{2}$. In conclusion then, the Dirac equation with $\mathbf{p} = 0$ admits four independent solutions (ignoring normalization factors, for the moment):

* You might ask why we don't simply assume that ψ_B is always zero; call the negative energy solutions "physically unacceptable" and forget about them. Unfortunately, this can't be done. In a quantum system we need a *complete* set of states, and the positive energy states by themselves are *not* complete. In the Schrödinger equation the sign of i is purely conventional. Had we made the other choice, then $e^{iEt/\hbar}$ would replace (7.29) as the characteristic time dependence for a stationary state of energy E. In relativistic quantum theory *both* signs are forced on us, and this, when properly interpreted, implies the existence of antiparticles.

$$\psi^{(1)} = e^{-i(mc^2/\hbar)t}\begin{pmatrix}1\\0\\0\\0\end{pmatrix}, \qquad \psi^{(2)} = e^{-i(mc^2/\hbar)t}\begin{pmatrix}0\\1\\0\\0\end{pmatrix}$$

$$\psi^{(3)} = e^{+i(mc^2/\hbar)t}\begin{pmatrix}0\\0\\1\\0\end{pmatrix}, \qquad \psi^{(4)} = e^{+i(mc^2/\hbar)t}\begin{pmatrix}0\\0\\0\\1\end{pmatrix} \tag{7.30}$$

describing, respectively, an electron with spin up, an electron with spin down, a positron with spin up, and a positron with spin down.

We look next for *plane-wave* solutions, of the form

$$\psi(\mathbf{r}, t) = ae^{-i/\hbar(Et-\mathbf{p}\cdot\mathbf{r})}u(E, \mathbf{p}) \tag{7.31}$$

or, in tidier notation

$$\psi(x) = ae^{-(i/\hbar)x\cdot p}u(p) \tag{7.32}$$

(Here a is a normalization constant, irrelevant to our present purpose, but necessary later to keep the units consistent.) We're hoping to find a bispinor $u(p)$ such that $\psi(x)$ satisfies the Dirac equation. (At this stage $p \equiv (E/c, \mathbf{p})$ is simply a set of four arbitrary parameters, but since they turn out to represent energy and momentum it seems simplest to assign them the appropriate letters right from the start.) Because the x dependence is confined to the exponent

$$\partial_\mu\psi = -\frac{i}{\hbar}p_\mu ae^{-(i/\hbar)x^\mu p_\mu}u \tag{7.33}$$

Putting this into the Dirac equation (7.20), we get

$$\gamma^\mu p_\mu ae^{-(i/\hbar)x\cdot p}u - mcae^{-(i/\hbar)x\cdot p}u = 0$$

or

$$(\gamma^\mu p_\mu - mc)u = 0 \tag{7.34}$$

This is known as the "momentum space Dirac equation". Notice that it is purely algebraic, with no derivatives. If u satisfies equation (7.34), then ψ (7.32) satisfies the Dirac equation (7.20).

Now

$$\gamma^\mu p_\mu = \gamma^0 p^0 - \boldsymbol{\gamma}\cdot\mathbf{p} = \frac{E}{c}\begin{pmatrix}1 & 0\\0 & -1\end{pmatrix} - \mathbf{p}\cdot\begin{pmatrix}0 & \sigma\\-\sigma & 0\end{pmatrix} = \begin{pmatrix}E/c & -\mathbf{p}\cdot\sigma\\\mathbf{p}\cdot\sigma & -E/c\end{pmatrix} \tag{7.35}$$

Thus

$$(\gamma^\mu p_\mu - mc)u = \begin{pmatrix}\left(\frac{E}{c} - mc\right) & -\mathbf{p}\cdot\sigma\\ \mathbf{p}\cdot\sigma & \left(-\frac{E}{c} - mc\right)\end{pmatrix}\begin{pmatrix}u_A\\u_B\end{pmatrix}$$

$$= \begin{pmatrix}\left(\frac{E}{c} - mc\right)u_A - \mathbf{p}\cdot\sigma u_B\\ \mathbf{p}\cdot\sigma u_A - \left(\frac{E}{c} + mc\right)u_B\end{pmatrix}$$

where, as before, the subscript A denotes the upper two components, and B stands for the lower two. In order to satisfy equation (7.34), then, we must have

$$u_A = \frac{c}{E - mc^2}(\mathbf{p}\cdot\sigma)u_B, \qquad u_B = \frac{c}{E + mc^2}(\mathbf{p}\cdot\sigma)u_A \tag{7.36}$$

Substituting the second of these into the first gives

$$u_A = \frac{c^2}{E^2 - m^2c^4}(\mathbf{p}\cdot\sigma)^2 u_A \tag{7.37}$$

But

$$\begin{aligned}\mathbf{p}\cdot\sigma &= p_x\begin{pmatrix}0 & 1\\ 1 & 0\end{pmatrix} + p_y\begin{pmatrix}0 & -i\\ i & 0\end{pmatrix} + p_z\begin{pmatrix}1 & 0\\ 0 & -1\end{pmatrix}\\ &= \begin{pmatrix}p_z & (p_x - ip_y)\\ (p_x + ip_y) & -p_z\end{pmatrix}\end{aligned} \tag{7.38}$$

so

$$(\mathbf{p}\cdot\sigma)^2 = \begin{pmatrix}p_z^2 + (p_x - ip_y)(p_x + ip_y) & p_z(p_x - ip_y) - p_z(p_x - ip_y)\\ p_z(p_x + ip_y) - p_z(p_x + ip_y) & (p_x + ip_y)(p_x - ip_y) + p_z^2\end{pmatrix} = \mathbf{p}^2 \tag{7.39}$$

where 1 is the 2×2 unit matrix (written in explicitly, just this once).

Thus

$$u_A = \frac{\mathbf{p}^2c^2}{E^2 - m^2c^4}u_A \tag{7.40}$$

and hence*

$$E^2 - m^2c^4 = \mathbf{p}^2c^2 \tag{7.41}$$

That is, in order to satisfy the Dirac equation, E and $\mathbf{p}$ [in eq. (7.31)] must obey the usual relativistic energy-momentum relation. That's hardly *surprising,* but it is interesting to see how the Dirac equation enforces this requirement. As an equation for E, (7.41) admits two solutions:

$$E = \pm\sqrt{m^2c^4 + \mathbf{p}^2c^2} \tag{7.42}$$

The positive root is associated with particle states, and the negative root with antiparticle states.

Returning to equation (7.36), and using (7.38), it is a simple matter to construct four independent solutions to the Dirac equation (ignoring normalization factors for a moment):

* Equation (7.40) would also allow $u_A = 0$ as a solution; however, the same argument, starting with equation (7.36) but inserting the first into the second, yields equation (7.40) with u_B in place of u_A. Thus unless u_A and u_B are *both* zero (in which case we have no solution at all) equation (7.41) must hold.

$$
\begin{aligned}
&(1)\quad \text{Pick } u_A = \begin{pmatrix}1\\0\end{pmatrix}, \qquad \text{then } u_B = \frac{c}{E+mc^2}(\mathbf{p}\cdot\sigma)\begin{pmatrix}1\\0\end{pmatrix} = \frac{c}{E+mc^2}\begin{pmatrix}p_z\\p_x+ip_y\end{pmatrix}\\
&(2)\quad \text{Pick } u_A = \begin{pmatrix}0\\1\end{pmatrix}, \qquad \text{then } u_B = \frac{c}{E+mc^2}(\mathbf{p}\cdot\sigma)\begin{pmatrix}0\\1\end{pmatrix} = \frac{c}{E+mc^2}\begin{pmatrix}p_x-ip_y\\-p_z\end{pmatrix}\\
&(3)\quad \text{Pick } u_B = \begin{pmatrix}0\\1\end{pmatrix}, \qquad \text{then } u_A = \frac{c}{E-mc^2}(\mathbf{p}\cdot\sigma)\begin{pmatrix}1\\0\end{pmatrix} = \frac{c}{E-mc^2}\begin{pmatrix}p_z\\p_x+ip_y\end{pmatrix}\\
&(4)\quad \text{Pick } u_B = \begin{pmatrix}0\\1\end{pmatrix}, \qquad \text{then } u_A = \frac{c}{E-mc^2}(\mathbf{p}\cdot\sigma)\begin{pmatrix}0\\1\end{pmatrix} = \frac{c}{E-mc^2}\begin{pmatrix}p_x-ip_y\\-p_z\end{pmatrix}
\end{aligned}
\tag{7.43}
$$

For (1) and (2) we must use the plus sign in equation (7.42), otherwise u_B blows up as $\mathbf{p} \to 0$; these are evidently *particle* solutions. For (3) and (4) we are obliged to use the minus sign; these are *anti*particle states. It is convenient to *normalize* these spinors in such a way that*

$$u^\dagger u = 2|E|/c \tag{7.44}$$

where the dagger signifies the transpose conjugate (or "Hermitian conjugate"):

$$u = \begin{pmatrix}\alpha\\\beta\\\gamma\\\delta\end{pmatrix} \Rightarrow u^\dagger = (\alpha^*\beta^*\gamma^*\delta^*)$$

so that

$$u^\dagger u = |\alpha|^2 + |\beta|^2 + |\gamma|^2 + |\delta|^2 \tag{7.45}$$

Thus the four solutions are:

$$u^{(1)} = N\begin{pmatrix}1\\0\\\dfrac{c(p_z)}{E+mc^2}\\\dfrac{c(p_x+ip_y)}{E+mc^2}\end{pmatrix}, \qquad u^{(2)} = N\begin{pmatrix}0\\1\\\dfrac{c(p_x-ip_y)}{E+mc^2}\\\dfrac{c(-p_z)}{E+mc^2}\end{pmatrix},$$

(with $E = \sqrt{m^2c^4 + \mathbf{p}^2c^2}$)

* Actually, there are at least three different conventions in the literature: $u^\dagger u = 2|E|/c$ (Halzen and Martin), $u^\dagger u = |E|/mc^2$ (Bjorken and Drell), $u^\dagger u = 1$ (Bogoliubov and Shirkov). In this one instance I depart from Bjorken and Drell, whose choice introduces spurious difficulties when $m \to 0$.

$$u^{(3)} = N\begin{pmatrix} \dfrac{c(p_z)}{E - mc^2} \\ \dfrac{c(p_x + ip_y)}{E - mc^2} \\ 1 \\ 0 \end{pmatrix}, \qquad u^{(4)} = N\begin{pmatrix} \dfrac{c(p_x - ip_y)}{E - mc^2} \\ \dfrac{c(-p_z)}{E - mc^2} \\ 0 \\ 1 \end{pmatrix},$$

(with $E = -\sqrt{m^2c^4 + \mathbf{p}^2c^2}$) (7.46)

and the normalization constant is (see Problem 7.3)

$$N = \sqrt{(|E| + mc^2)/c} \tag{7.47}$$

You might guess that $u^{(1)}$ describes an electron with spin up, $u^{(2)}$ an electron with spin down, and so on, but this is not quite the case. For Dirac particles the spin matrices [generalizing (4.21)] are

$$\mathbf{S} = \frac{\hbar}{2}\boldsymbol{\Sigma}, \qquad \text{with } \boldsymbol{\Sigma} \equiv \begin{pmatrix} \sigma & 0 \\ 0 & \sigma \end{pmatrix} \tag{7.48}$$

and it's easy to check that $u^{(1)}$, for instance, is *not* an eigenstate of Σ_z. However, if we orient the z axis so that it points along the direction of motion (in which case $p_x = p_y = 0$) then $u^{(1)}$, $u^{(2)}$, $u^{(3)}$, and $u^{(4)}$ are eigenspinors of S_z; $u^{(1)}$ and $u^{(3)}$ are spin up, and $u^{(2)}$ and $u^{(4)}$ are spin down* (Problem 7.6).

I said earlier that E and $\mathbf{p}$ [in eq. (7.31)] are mathematical parameters which correspond physically to energy and momentum, and this is quite true for the electron states, $u^{(1)}$ and $u^{(2)}$. However, the E in $u^{(3)}$ and $u^{(4)}$ *cannot* represent positron energy; all free particles, positrons and electrons alike, carry *positive* energy. The "negative-energy" solutions must be reinterpreted as *positive* energy *anti*particle states. To express these solutions in terms of the *physical* energy and momentum of the positron, we flip the signs of E and $\mathbf{p}$:

$$\psi(\mathbf{r}, t) = ae^{i/\hbar(Et - \mathbf{p}\cdot\mathbf{r})}u(-E, -\mathbf{p}) \qquad \text{[for solutions (3) and (4)]} \tag{7.49}$$

Mind you, these are the same old solutions to the Dirac equation; I have simply adopted a different sign convention for the parameters—one that better conforms to their physical interpretation.† It is customary to use the letter v for positron states, expressed in terms of the physical energy and momentum:‡

* As a matter of fact, it is *impossible* to construct spinors that satisfy equation (7.34) and are, at the same time, eigenstates of S_z (except in the special case $\mathbf{p} = p_z\hat{z}$). The reason is that $\mathbf{S}$ by itself is not a conserved quantity; only the *total* angular momentum, $\mathbf{L} + \mathbf{S}$, is conserved here (see Problem 7.8). It *is* possible to construct eigenstates of *helicity*, $\boldsymbol{\Sigma}\cdot\hat{p}$ (there's no *orbital* angular momentum about the direction of motion), but these are rather cumbersome (see Problem 7.7), and in practice it is easier to work with the spinors (7.46), even though their physical interpretation is not so clean. All that *really* matters is that we have a *complete* set of solutions.

† If it bothers you to change notation in "midstream" like this, go back to equation (7.32) and don't call it p^μ at all—call it k^μ (or something). Then, at the end, identify $k^0 = E/c$, $\mathbf{k} = \mathbf{p}$ for solutions (1) and (2), $k^0 = -E/c$, $\mathbf{k} = -\mathbf{p}$ for solutions (3) and (4).

‡ It is conventional to associate $v^{(1)}$ with $u^{(4)}$, and $v^{(2)}$ with $-u^{(3)}$. In the special case $p_x = p_y = 0$, then, $v^{(1)}$ is spin *down* and $v^{(2)}$ is spin *up*. This seems silly at first, but there is a reason for it:

$$v^{(1)}(E, \mathbf{p}) = u^{(4)}(-E, -\mathbf{p}) = N\begin{pmatrix} \dfrac{c(p_x - ip_y)}{E + mc^2} \\ \dfrac{c(-p_z)}{E + mc^2} \\ 0 \\ 1 \end{pmatrix}$$

$$v^{(2)}(E, \mathbf{p}) = -u^{(3)}(-E, -\mathbf{p}) = -N\begin{pmatrix} \dfrac{c(p_z)}{E + mc^2} \\ \dfrac{c(p_x + ip_y)}{E + mc^2} \\ 1 \\ 0 \end{pmatrix},$$

$$(\text{with } E = \sqrt{m^2c^4 + \mathbf{p}^2c^2}) \tag{7.50}$$

From now on I will never mention $u^{(3)}$ and $u^{(4)}$; the solutions we shall use are $u^{(1)}$, $u^{(2)}$ (representing the two spin states of an electron with energy E and momentum $\mathbf{p}$), and $v^{(1)}$, $v^{(2)}$ (representing the two spin states of a positron with energy E and momentum $\mathbf{p}$). Notice that whereas the u's satisfy the momentum space Dirac equation (7.34) in the form

$$(\gamma^\mu p_\mu - mc)u = 0 \tag{7.34}$$

the v's obey the equation with the sign of p_μ reversed:

$$(\gamma^\mu p_\mu + mc)v = 0 \tag{7.51}$$

Incidentally, plane waves are, of course, rather special solutions to the Dirac equation. They are the ones of interest to us, however, because they describe particles with specified energies and momenta, and in a typical experiment these are the parameters we control and measure.

7.3 BILINEAR COVARIANTS

I mentioned in Section 7.1 that the components of a Dirac spinor do not transform as a four-vector, when you go from one inertial system to another. How, then, *do* they transform? I shall not work it out here, but merely quote the result:[2] If we go to a system moving with speed v in the x direction, the transformation rule is

$$\psi \rightarrow \psi' = S\psi \tag{7.52}$$

The charge conjugation operator takes an electron with spin up into a positron with spin down, so this way $u^{(1)}$, $v^{(1)}$ are "particle-antiparticle pairs," as are $u^{(2)}$, $v^{(2)}$ (see Problem 7.9).

where S is the following 4×4 matrix:

$$S = a_+ + a_-\gamma^0\gamma^1 = \begin{pmatrix} a_+ & a_-\sigma_1 \\ a_-\sigma_1 & a_+ \end{pmatrix} \tag{7.53}$$

with

$$a_\pm = \pm\sqrt{\tfrac{1}{2}(\gamma \pm 1)} \tag{7.54}$$

and $\gamma = 1/\sqrt{1 - v^2/c^2}$, as usual.

Suppose we want to construct a scalar quantity out of a spinor ψ. It would be reasonable to try the expression

$$\psi^\dagger\psi = (\psi_1^*\psi_2^*\psi_3^*\psi_4^*)\begin{pmatrix} \psi_1 \\ \psi_2 \\ \psi_3 \\ \psi_4 \end{pmatrix} = |\psi_1|^2 + |\psi_2|^2 + |\psi_3|^2 + |\psi_4|^2 \tag{7.55}$$

Unfortunately, this is *not* a scalar, as you can check by applying the preceding transformation rule:*

$$(\psi^\dagger\psi)' = (\psi')^\dagger\psi' = \psi^\dagger S^\dagger S\psi \neq (\psi^\dagger\psi) \tag{7.56}$$

In fact (see Problem 7.11):

$$S^\dagger S = S^2 = \gamma\begin{pmatrix} 1 & -\dfrac{v}{c}\sigma_1 \\ -\dfrac{v}{c}\sigma_1 & 1 \end{pmatrix} \neq 1 \tag{7.57}$$

Of course, the sum of the squares of the elements of a four-vector is not invariant either; we need minus signs for the spatial components [eq. (3.12)]. With a little trial-and-error you will discover that in the case of spinors we need minus signs for the third and fourth components. Just as we introduced *covariant* four-vectors to keep track of the signs in Chapter 3, we now introduce the *adjoint* spinor:

$$\bar{\psi} \equiv \psi^\dagger\gamma^0 = (\psi_1^* \ \psi_2^* \ -\psi_3^* \ -\psi_4^*) \tag{7.58}$$

I claim that the quantity

$$\bar{\psi}\psi = \psi^\dagger\gamma^0\psi = |\psi_1|^2 + |\psi_2|^2 - |\psi_3|^2 - |\psi_4|^2 \tag{7.59}$$

is a relativistic invariant. For $S^\dagger\gamma^0 S = \gamma^0$ (Problem 7.11), and hence

$$(\bar{\psi}\psi)' = (\psi')^\dagger\gamma^0\psi' = \psi^\dagger S^\dagger\gamma^0 S\psi = \psi^\dagger\gamma^0\psi = \bar{\psi}\psi \tag{7.60}$$

In Chapter 4 we learned to distinguish scalars and *pseudo*scalars, according to their behavior under the parity transformation, P: $(x, y, z) \to (-x, -y, -z)$. Pseudoscalars change sign; scalars do not. It is natural to ask whether $\bar{\psi}\psi$ is the

* Note that the transpose of a product is the product of the transposes *in reverse order:*

$$(\widetilde{AB})_{ij} = (AB)_{ji} = A_{jk}B_{ki} = \tilde{B}_{ik}\tilde{A}_{kj} = (\tilde{B}\tilde{A})_{ij}$$

The same goes for the Hermitian conjugate:

$$(AB)^\dagger = B^\dagger A^\dagger$$

former type, or the latter. First, we need to know how Dirac spinors transform under P. Again, I won't derive it, but simply quote the result:[3]

$$\psi \rightarrow \psi' = \gamma^0 \psi \tag{7.61}$$

It follows that

$$(\bar{\psi}\psi)' = (\psi')^\dagger \gamma^0 \psi' = \psi^\dagger (\gamma^0)^\dagger \gamma^0 \gamma^0 \psi = \psi^\dagger \gamma^0 \psi = \bar{\psi}\psi \tag{7.62}$$

so $(\bar{\psi}\psi)$ is *invariant* under P; it's a "true" scalar. But we can *also* make a *pseudo*-scalar out of ψ:

$$\bar{\psi}\gamma^5\psi \tag{7.63}$$

where

$$\gamma^5 \equiv i\gamma^0\gamma^1\gamma^2\gamma^3 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \tag{7.64}$$

I'll let you check that it is Lorentz-invariant (Problem 7.12). As for its behavior under parity

$$(\bar{\psi}\gamma^5\psi)' = (\psi')^\dagger \gamma^0 \gamma^5 \psi' = \psi^\dagger \gamma^0 \gamma^0 \gamma^5 \gamma^0 \psi = \psi^\dagger \gamma^5 \gamma^0 \psi \tag{7.65}$$

(I used the fact that $(\gamma^0)^2 = 1$ in the last step.) Now, the γ^0 is on the "wrong side" of the γ^5, but we can "pull it through" by noting that it *anti*commutes with γ^1, γ^2, and γ^3 [eq. (7.15)] and *commutes* (of course) with itself ($\gamma^3\gamma^0 = -\gamma^0\gamma^3$, $\gamma^2\gamma^0 = -\gamma^0\gamma^2$, $\gamma^1\gamma^0 = -\gamma^0\gamma^1$, $\gamma^0\gamma^0 = \gamma^0\gamma^0$), so

$$\gamma^5\gamma^0 = i\gamma^0\gamma^1\gamma^2\gamma^3\gamma^0 = (-1)^3\gamma^0(i\gamma^0\gamma^1\gamma^2\gamma^3) = -\gamma^0\gamma^5$$

By the same token, γ^5 anticommutes with all the other γ matrices:

$$\{\gamma^\mu, \gamma^5\} = 0 \tag{7.66}$$

At any rate

$$(\bar{\psi}\gamma^5\psi)' = -\psi^\dagger \gamma^0 \gamma^5 \psi = -(\bar{\psi}\gamma^5\psi) \tag{7.67}$$

so it's a *pseudo*scalar.

All told, there are 16 products of the form $\psi_i^*\psi_j$ (taking one component from ψ^* and one from ψ), since i and j run from 1 to 4. These 16 products can be added together in various linear combinations to construct quantities with distinct transformation behavior, as follows:

$$\left\{\begin{array}{lll} (1) & \bar{\psi}\psi = \text{scalar} & \text{(one component)} \\ (2) & \bar{\psi}\gamma^5\psi = \text{pseudoscalar} & \text{(one component)} \\ (3) & \bar{\psi}\gamma^\mu\psi = \text{vector} & \text{(four components)} \\ (4) & \bar{\psi}\gamma^\mu\gamma^5\psi = \text{pseudovector} & \text{(four components)} \\ (5) & \bar{\psi}\sigma^{\mu\nu}\psi = \text{antisymmetric tensor} & \text{(six components)} \end{array}\right\} \tag{7.68}$$

where

$$\sigma^{\mu\nu} \equiv \frac{i}{2}(\gamma^\mu\gamma^\nu - \gamma^\nu\gamma^\mu) \tag{7.69}$$

This gives 16 terms, so it's all we can hope to make, in this way. You cannot, for example, construct a *symmetric* tensor bilinear in ψ^* and ψ, and if you're

looking for a vector, $\bar{\psi}\gamma^\mu\psi$ is the only candidate.* (Another way to think of it is this: 1, γ^5, γ^μ, $\gamma^\mu\gamma^5$, and $\sigma^{\mu\nu}$ constitute a "basis" for the space of all 4×4 matrices; *any* 4×4 matrix can be written as a linear combination of these 16. In particular, if you ever encounter a product of five γ matrices, say, you may be sure that it can be reduced down to a product of no more than two.) Pause a moment to admire the notation in (7.68). The tensorial character of the *bilinear covariants,* and even their behavior under parity, is indicated at a glance: $\bar{\psi}\gamma^\mu\psi$ *looks* like a four-vector, and it *is* a four-vector. But γ^μ by itself is certainly *not* a four-vector; it's a collection of four fixed matrices (7.17); *they* don't change when you go to a different inertial system—it's ψ that changes.

7.4 THE PHOTON

In classical electrodynamics the electric and magnetic fields (**E** and **B**) set up by a charge density ρ and a current density **J** are determined by Maxwell's equations:†

$$\left\{ \begin{array}{ll} \text{(i)} \quad \nabla \cdot \mathbf{E} = 4\pi\rho & \text{(iii)} \quad \nabla \cdot \mathbf{B} = 0 \\ \text{(ii)} \quad \nabla \times \mathbf{E} + \dfrac{1}{c}\dfrac{\partial \mathbf{B}}{\partial t} = 0 & \text{(iv)} \quad \nabla \times \mathbf{B} - \dfrac{1}{c}\dfrac{\partial \mathbf{E}}{\partial t} = \dfrac{4\pi}{c}\mathbf{J} \end{array} \right\} \tag{7.70}$$

In relativistic notation, **E** and **B** together form an antisymmetric second-rank tensor, the "field strength tensor," $F^{\mu\nu}$:

$$F^{\mu\nu} = \begin{pmatrix} 0 & -E_x & -E_y & -E_z \\ E_x & 0 & -B_z & B_y \\ E_y & B_z & 0 & -B_x \\ E_z & -B_y & B_x & 0 \end{pmatrix} \tag{7.71}$$

(that is, $F^{01} = E_x$, $F^{12} = -B_z$, etc.), while ρ and **J** constitute a four-vector:

$$J^\mu = (c\rho, \mathbf{J}) \tag{7.72}$$

The inhomogeneous Maxwell equations [(i) and (iv)] can now be written more neatly (Problem 7.18)

$$\partial_\mu F^{\mu\nu} = \frac{4\pi}{c} J^\nu \tag{7.73}$$

* Notice that $\bar{\psi}\gamma^0\psi = \psi^\dagger\gamma^0\gamma^0\psi = \psi^\dagger\psi$, so $\psi^\dagger\psi$ is actually the zeroth component of a four-vector. That's why the normalization convention (7.44), which no doubt looked peculiar at the time, is actually very sensible. By normalizing $u^\dagger u$ to the zeroth component of the four-vector p^μ, we obtain a relativistically "natural" convention (see Problem 7.14). By the way, as in nonrelativistic quantum mechanics, $\psi^*\psi$ has the dimensions of (volume)$^{-1}$, so the constant a in equation (7.31) carries the units $mc(\hbar^{-3/2})$.

† This section presupposes some familiarity with classical electrodynamics; it is designed to make the description of photons in quantum electrodynamics more plausible, but if you don't understand it, skip directly to Section 7.5. As always, I use Gaussian cgs units.

From the antisymmetry of $F^{\mu\nu}$ (that is: $F^{\nu\mu} = -F^{\mu\nu}$) it follows (Problem 7.18) that J^μ is divergenceless:

$$\partial_\mu J^\mu = 0 \tag{7.74}$$

Or, in three-vector notation, $\nabla \cdot \mathbf{J} = -\partial\rho/\partial t$; this is the "continuity equation," expressing local conservation of charge (Problem 7.19).

As for the homogeneous Maxwell equations, (iii) is equivalent to the statement that **B** can be written as the curl of a vector potential, **A**:

$$\mathbf{B} = \nabla \times \mathbf{A} \tag{7.75}$$

With this, (ii) becomes

$$\nabla \times \left(\mathbf{E} + \frac{1}{c}\frac{\partial \mathbf{A}}{\partial t}\right) = 0 \tag{7.76}$$

which is equivalent to the statement that $\mathbf{E} + (1/c)(\partial\mathbf{A}/\partial t)$ can be written as the gradient of a scalar potential, V:

$$\mathbf{E} = -\nabla V - \frac{1}{c}\frac{\partial \mathbf{A}}{\partial \mathbf{t}} \tag{7.77}$$

In relativistic notation, equations (7.75) and (7.77) become

$$F^{\mu\nu} = \partial^\mu A^\nu - \partial^\nu A^\mu \tag{7.78}$$

where

$$A^\mu = (V, \mathbf{A}) \tag{7.79}$$

In terms of this four-vector potential, the inhomogeneous Maxwell equations (7.73) read:

$$\partial_\mu \partial^\mu A^\nu - \partial^\nu(\partial_\mu A^\mu) = \frac{4\pi}{c} J^\nu \tag{7.80}$$

In classical electrodynamics the *fields* are the physical entities; the potentials are simply useful mathematical constructs. The *virtue* of the potential formulation is that it *automatically* takes care of the homogeneous Maxwell equations: given equations (7.75) and (7.77), (ii) and (iii) follow immediately, no matter what V and **A** may be. This leaves us only the inhomogeneous equation (7.80) to worry about. The *defect* of the potential formulation is that V and **A** are not uniquely determined. Indeed, it is clear from equation (7.78) that new potentials

$$A'_\mu = A_\mu + \partial_\mu \lambda \tag{7.81}$$

(where λ is any function of position and time) would do just as well, since $\partial^\mu A'^\nu - \partial^\nu A'^\mu = \partial^\mu A^\nu - \partial^\nu A^\mu$. Such a change of potentials, which has no effect on the fields, is called a *gauge transformation.* We can exploit this gauge freedom to impose an extra constraint on the potential:[4]

$$\partial_\mu A^\mu = 0 \tag{7.82}$$

This is called the *Lorentz condition;* with it Maxwell's equations (7.80) simplify still further:

$$\Box A^\mu = \frac{4\pi}{c} J^\mu \tag{7.83}$$

Here $\Box \equiv \partial^\mu \partial_\mu = \frac{1}{c^2}\frac{\partial^2}{\partial t^2} - \nabla^2$; it is called the *d'Alembertian operator.*

Even the Lorentz condition, however, does not uniquely specify A^μ. Further gauge transformations are possible, without disturbing equation (7.82), provided that the gauge function λ satisfies the wave equation:

$$\Box \lambda = 0 \tag{7.84}$$

Unfortunately, there is no clean way to eliminate the residual ambiguity in A^μ, and one can either (1) live with the indeterminacy, which means carrying along spurious degrees of freedom, or (2) impose an additional constraint, which spoils the manifest Lorentz covariance of the theory. Both approaches have been used in formulating quantum electrodynamics; we shall follow the latter course. In empty space, where $J^\mu = 0$, we pick (see Problem 7.20)

$$A^0 = 0 \tag{7.85}$$

The Lorentz condition then reads

$$\nabla \cdot \mathbf{A} = 0 \tag{7.86}$$

This choice (the *Coulomb gauge*) is attractively simple, but by selecting one component (A^0) for special treatment, it ties us down to a particular inertial system (either that, or it obliges us to perform a gauge transformation in conjunction with every Lorentz transformation, in order to restore the Coulomb gauge condition). In practice, this is very seldom a problem, but it is aesthetically displeasing.

In quantum electrodynamics A^μ becomes the wave function of the photon. The free photon satisfies equation (7.83) with $J^\mu = 0$

$$\Box A^\mu = 0 \tag{7.87}$$

which we recognize in this context as the Klein–Gordon equation (7.9) for a massless particle. As in the case of the Dirac equation, we look for plane-wave solutions with momentum $p = (E/c, \mathbf{p})$:

$$A^\mu(x) = ae^{-(i/\hbar)p\cdot x}\epsilon^\mu(p) \tag{7.88}$$

Here ϵ^μ is the *polarization vector*—it characterizes the spin of the photon—and a is a normalization factor. Substituting equation (7.88) into equation (7.87), we obtain a constraint on p^μ:

$$p^\mu p_\mu = 0, \qquad \text{so that} \quad E = |\mathbf{p}|c \tag{7.89}$$

which is as it should be for a massless particle.

Meanwhile, ϵ^μ has four components, but they are not all independent. The Lorentz condition (7.82) requires that

$$p^\mu \epsilon_\mu = 0 \tag{7.90}$$

Moreover, in the Coulomb gauge we have

$$\epsilon^0 = 0, \qquad \boldsymbol{\epsilon} \cdot \mathbf{p} = 0 \tag{7.91}$$

which is to say that the polarization three-vector ($\boldsymbol{\epsilon}$) is perpendicular to the direction of propagation; we say that a free photon is *transversely polarized.* For this reason the Coulomb gauge is also known as the *transverse* gauge. Now, there are two linearly independent three-vectors that are perpendicular to $\mathbf{p}$; for example, if $\mathbf{p}$ points in the z direction, we might choose

$$\boldsymbol{\epsilon}_{(1)} = (1, 0, 0), \qquad \boldsymbol{\epsilon}_{(2)} = (0, 1, 0) \tag{7.92}$$

Thus instead of *four* independent solutions for a given momentum (too many, for a particle of spin 1), we are left with only *two.* That looks like too *few:* shouldn't the photon have *three* spin states? The answer is *no:* a *massive* particle of spin s admits $2s + 1$ different spin orientations, but a *massless* particle has only two, regardless of its spin (except for $s = 0$, which has only *one*). Along its direction of motion it can only have $m_s = +s$ or $m_s = -s$; its helicity, in other words, can only be $+1$ or -1.*

7.5 THE FEYNMAN RULES FOR QUANTUM ELECTRODYNAMICS

In Section 7.2 we found that free electrons and positrons of momentum $p = (E/c, \mathbf{p})$, with $E = \sqrt{m^2c^4 + \mathbf{p}^2c^2}$, are represented by the wave functions†

Electrons	**Positrons**	
$\psi(x) = ae^{-(i/\hbar)p\cdot x}u^{(s)}(p)$	$\psi(x) = ae^{(i/\hbar)p\cdot x}v^{(s)}(p)$	(7.93)

where $s = 1, 2$ for the two spin states. The spinors $u^{(s)}$ and $v^{(s)}$ satisfy the *momentum space Dirac equations:*

$$(\gamma^\mu p_\mu - mc)u = 0 \qquad (\gamma^\mu p_\mu + mc)v = 0 \tag{7.94}$$

and their *adjoints,* $\bar{u} = u^\dagger\gamma^0$, $\bar{v} = v^\dagger\gamma^0$, satisfy

$$\bar{u}(\gamma^\mu p_\mu - mc) = 0 \qquad \bar{v}(\gamma^\mu p_\mu + mc) = 0$$

They are *orthogonal,*

$$\bar{u}^{(1)}u^{(2)} = 0 \qquad \bar{v}^{(1)}v^{(2)} = 0 \tag{7.95}$$

* Photon states with $m_s = \pm 1$ correspond to right- and left-circular polarization; the respective polarization vectors are $\epsilon_\pm = \mp(\epsilon_1 \pm i\epsilon_2)/\sqrt{2}$. Notice that it is by specifying a particular gauge that we eliminate the nonphysical ($m_s = 0$) solution. If we were to follow a "covariant" approach, in which we avoid imposing the Coulomb gauge condition, longitudinal free photons would be present in the theory. But these "ghosts" decouple from everything else, and they do not affect the final results.

† To make this section as self-contained as possible for easier reference, and also to emphasize the similarities and differences in the theories of electrons, positrons, and photons, I begin with a summary of the essential results from earlier sections. For the sake of argument I speak of "electrons" and "positrons," but they could as well be μ^- and μ^+, or τ^- and τ^+, or (with the appropriate electric charges) quarks and antiquarks—in short, any point charges of spin $\frac{1}{2}$.

normalized,

$$\bar{u}u = 2mc \qquad \bar{v}v = -2mc \tag{7.96}$$

and *complete,* in the sense that

$$\sum_{s=1,2} u^{(s)}\bar{u}^{(s)} = (\gamma^\mu p_\mu + mc) \qquad \sum_{s=1,2} v^{(s)}\bar{v}^{(s)} = (\gamma^\mu p_\mu - mc) \tag{7.97}$$

(see Problem 7.22). A convenient explicit set ($u^{(1)}$, $u^{(2)}$, $v^{(1)}$, $v^{(2)}$) is given in equations (7.46) and (7.50). Ordinarily, we shall be averaging over electron and positron spins, and in that case it doesn't matter that these are not pure spin up and spin down; all we really need is completeness. For the occasional problem in which the spins are specified, we must, of course, use the spinors appropriate to the case at hand.

Meanwhile, a free *photon* of momentum $p = (E/c, \mathbf{p})$, with $E = |\mathbf{p}|c$, is represented by the wave function

Photons

$$A^\mu(x) = ae^{-(i/\hbar)p\cdot x}\epsilon^\mu_{(s)} \tag{7.98}$$

where $s = 1, 2$ for the two spin states (or "polarizations") of the photon. The polarization vectors $\epsilon^\mu_{(s)}$ satisfy the *momentum space Lorentz condition:*

$$\epsilon^\mu p_\mu = 0 \tag{7.99}$$

They are *orthogonal,* in the sense that

$$\epsilon^{\mu *}_{(1)}\epsilon_{\mu(2)} = 0 \tag{7.100}$$

and *normalized*

$$\epsilon^{\mu *}\epsilon_\mu = 1 \tag{7.101}$$

In the Coulomb gauge

$$\epsilon^0 = 0, \qquad \boldsymbol{\epsilon}\cdot\mathbf{p} = 0 \tag{7.102}$$

and the polarization three-vectors obey the completeness relation (Problem 7.23)

$$\sum_{s=1,2} (\boldsymbol{\epsilon}_{(s)})_i(\boldsymbol{\epsilon}^*_{(s)})_j = \delta_{ij} - \hat{p}_i\hat{p}_j \tag{7.103}$$

A convenient explicit pair ($\boldsymbol{\epsilon}_{(1)}$, $\boldsymbol{\epsilon}_{(2)}$) is given in equation (7.92).

To calculate the amplitude, $\mathcal{M}$, associated with a particular Feynman diagram, proceed as follows:

1. *Notation.* Label the incoming and outgoing four-momenta $p_1, p_2, \ldots, p_n$, and the corresponding spins $s_1, s_2, \ldots, s_n$; label the internal four-momenta $q_1, q_2, \ldots$. Assign arrows to the lines as follows: the arrows on *external* fermion lines indicate whether it is an electron or a positron; arrows on *internal* fermion lines are assigned so that the "direction of the flow" through the diagram is preserved (i.e., every vertex must have one arrow entering and one arrow leaving). The arrows on external

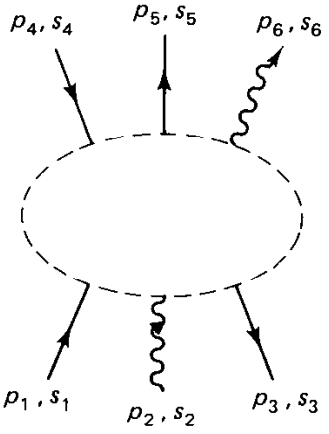


Figure 7.1 A typical QED diagram, with external lines labeled. (Internal lines not shown.)

photon lines point "forward"; for internal photon lines the choice is arbitrary. (See Fig. 7.1.)

2. *External Lines.* External lines contribute factors as follows:

$$\textit{Electrons:} \begin{cases} \text{Incoming } (\nearrow): u \\ \text{Outgoing } (\nearrow): \bar{u} \end{cases}$$

$$\textit{Positrons:} \begin{cases} \text{Incoming } (\swarrow): \bar{v} \\ \text{Outgoing } (\swarrow): v \end{cases}$$

$$\textit{Photons:} \begin{cases} \text{Incoming } (\nearrow): \epsilon^{\mu} \\ \text{Outgoing } (\nearrow): \epsilon^{\mu *} \end{cases}$$

3. *Vertex Factors.* Each vertex contributes a factor

$$ig_e\gamma^{\mu}$$

The dimensionless coupling constant g_e is related to the charge of the positron: $g_e = e\sqrt{4\pi/\hbar c} = \sqrt{4\pi\alpha}$.*

4. *Propagators.* Each internal line contributes a factor as follows:

$$\textit{Electrons and positrons:} \quad \frac{i(\gamma^{\mu}q_{\mu} + mc)}{q^2 - m^2c^2}$$

$$\textit{Photons:} \quad \frac{-ig_{\mu\nu}}{q^2}$$

5. *Conservation of Energy and Momentum.* For each vertex, write a delta function of the form

$$(2\pi)^4\,\delta^4(k_1 + k_2 + k_3)$$

* In Heaviside–Lorentz units, with $\hbar$ and c set equal to 1, g_e *is* the charge of the positron, and hence is written "e" in most texts. In this book I use Gaussian units, and keep all factors of $\hbar$ and c. The easiest way to avoid trouble over units is to express all results in terms of the universal dimensionless quantity α. In writing the Feynman rules for QED I assume we are dealing with electrons and positrons. In *general*, the QED coupling constant is $-q\sqrt{4\pi/\hbar c}$, where q is the charge of the *particle* (as opposed to the *anti*particle). For electrons, $q = -e$, but for "up" quarks, say, $q = \frac{2}{3}e$.

where the k's are the three four-momenta coming *into* the vertex (if an arrow leads *outward,* then k is *minus* the four-momentum of that line, except for external positrons*). This factor enforces conservation of energy and momentum at the vertex.

6. *Integrate Over Internal Momenta.* For each internal momentum q, write a factor

$$\frac{d^4q}{(2\pi)^4}$$

and integrate.

7. *Cancel the Delta Function.* The result will include a factor

$$(2\pi)^4\delta^4(p_1 + p_2 + \cdots - p_n)$$

corresponding to overall energy-momentum conservation. Cancel this factor, and what remains is $-i\mathcal{M}$.

As before, the procedure is to write down all diagrams contributing to the process in question (up to the desired order), calculate the amplitude ($\mathcal{M}$) for each one, and add them up to get the *total* amplitude, which is then inserted into the appropriate formula for the cross section or the lifetime, as the case may be. There's just one new twist, here: the antisymmetrization of fermion wave functions requires that we insert a minus sign in combining amplitudes that differ only in the interchange of two identical external fermions. It doesn't matter *which* diagram you associate the minus sign with, since the total will be squared eventually anyway; but there must be a *relative* minus sign between them.

8. *Antisymmetrization.* Include a minus sign between diagrams that differ only in the interchange of two incoming (or outgoing) electrons (or positrons), or of an incoming electron with an outgoing positron (or vice versa).

The handling of fermion loops will be discussed in the final section of this chapter.

7.6 EXAMPLES

We are now in a position to reproduce many of the classic calculations in quantum electrodynamics. Just so you don't get lost in the details, let me begin by giving you a catalog of the most important processes (see Table 7.1). The simplest

* The problem here is that the arrows are being asked to do double duty: they establish the convention for the sign of momentum, and in the case of external fermion lines, they tell you whether it is a particle or an antiparticle (for internal lines we need not distinguish). The latter role takes precedence, so for external positrons the "positive" direction for momentum is *opposite* to the direction of the arrow.

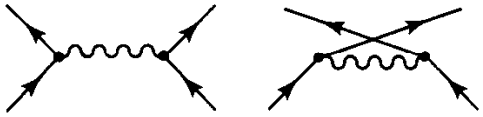
TABLE 7.1 CATALOG OF BASIC QUANTUM ELECTRODYNAMIC PROCESSES

Second-order processes

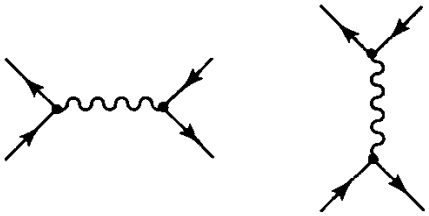
Elastic



Electron–muon scattering ($e + \mu \to e + \mu$)
(Mott scattering ($M \gg m$) ⇒ Rutherford scattering ($v \ll c$))

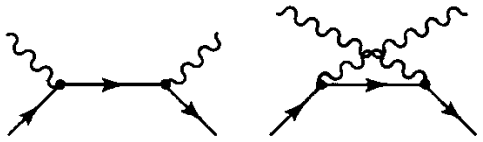


Electron–electron scattering ($e^- + e^- \to e^- + e^-$)
(Møller scattering)

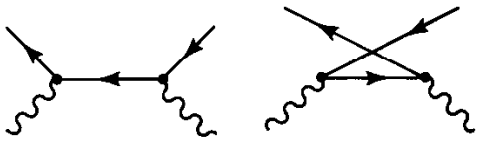


Electron–positron scattering ($e^- + e^+ \to e^- + e^+$)
(Bhabha scattering)

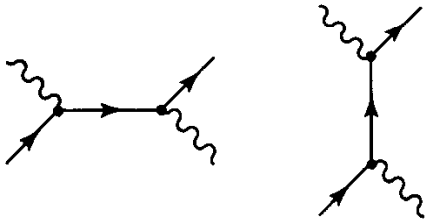
Inelastic



Pair annihilation ($e^- + e^+ \to \gamma + \gamma$)

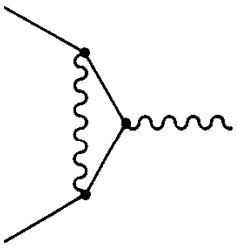


Pair production ($\gamma + \gamma \to e^- + e^+$)



Compton scattering ($\gamma + e^- \to \gamma + e^-$)

Most important third-order process



⇒ Anomalous magnetic moment of electron

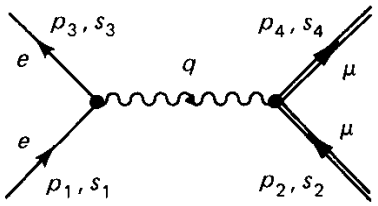


Figure 7.2 Electron-muon scattering.

case is electron-muon scattering, for here only one diagram contributes in second order.*

EXAMPLE 7.1 Electron-Muon Scattering

In applying the Feynman rules, we proceed "backward" along each fermion line (Fig. 7.2):

$$(2\pi)^4 \int [\bar{u}^{(s_3)}(p_3)(ig_e\gamma^\mu)u^{(s_1)}(p_1)] \frac{-ig_{\mu\nu}}{q^2} [\bar{u}^{(s_4)}(p_4)(ig_e\gamma^\nu)u^{(s_2)}(p_2)] \times \delta^4(p_1 - p_3 - q)\delta^4(p_2 + q - p_4)d^4q$$

Notice that the space–time indices on the photon propagator contract with those of the vertex factors at either end of the photon line. Carrying out the (trivial) q integration, and dropping the overall delta function, we find

$$\mathcal{M} = -\frac{g_e^2}{(p_1 - p_3)^2} [\bar{u}^{(s_3)}(p_3)\gamma^\mu u^{(s_1)}(p_1)][\bar{u}^{(s_4)}(p_4)\gamma_\mu u^{(s_2)}(p_2)] \qquad (7.104)$$

In spite of its complicated appearance, with four spinors and eight γ matrices, this is just a *number,* which you can work out once the spins are specified (see Problem 7.24).

EXAMPLE 7.2 Electron-Electron Scattering

In this case there is a second diagram, in which the electron that emerges with momentum p_3 and spin s_3 comes from the p_2, s_2 electron, instead of the p_1, s_1 electron (Fig. 7.3). We can obtain this amplitude from equation (7.104) simply by the replacement $p_3, s_3 \leftrightarrow p_4, s_4$. According to rule 8, the two diagrams are to be *subtracted,* so the total amplitude is

$$\mathcal{M} = -\frac{g_e^2}{(p_1 - p_3)^2} [\bar{u}(3)\gamma^\mu u(1)][\bar{u}(4)\gamma_\mu u(2)] + \frac{g_e^2}{(p_1 - p_4)^2} [\bar{u}(4)\gamma^\mu u(1)][\bar{u}(3)\gamma_\mu u(2)] \qquad (7.105)$$

* It doesn't have to be an e and a μ, of course. Any spin-$\frac{1}{2}$ point charges would do (e and τ, for instance, or μ and τ, or electron and quark, etc.), as long as you put in the correct masses and charges. As a matter of fact, most books use electron-*proton* scattering as the canonical example, but that is actually a rather *inappropriate* choice, since the proton is a composite structure, not a point particle. Still, to the extent that the internal structure of the proton can be ignored, it is not a bad *approximation* (it is rather like treating the sun as a point mass in the theory of the solar system). In the regime where the "muon" is much heavier than the "electron," we have *Mott scattering;* if, moreover, the "electron" is nonrelativistic, it reduces to *Rutherford scattering,* and actually reproduces exactly the same formula for the cross section as Rutherford obtained using classical mechanics.

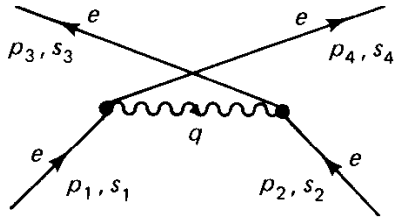


Figure 7.3 "Twisted" diagram for electron-electron scattering.

EXAMPLE 7.3 Electron–Positron Scattering

Again, there are two diagrams.* The first is similar to the electron-muon diagram (Fig. 7.4):

$$(2\pi)^4 \int [\bar{u}(3)(ig_e\gamma^\mu)u(1)] \frac{-ig_{\mu\nu}}{q^2} \times [\bar{v}(2)(ig_e\gamma^\nu)v(4)]\delta^4(p_1 - p_3 - q)\delta^4(p_2 + q - p_4)d^4q$$

Notice that "proceeding backwards" along an *anti*particle line means working *forward* in time; the order is always *adjoint spinor/gamma matrix/spinor*. The amplitude for this diagram is thus

$$\mathcal{M}_1 = -\frac{g_e^2}{(p_1 - p_3)^2}[\bar{u}(3)\gamma^\mu u(1)][\bar{v}(2)\gamma_\mu v(4)] \tag{7.106}$$

The other diagram represents virtual annihilation of the electron and positron, followed by pair production (Fig. 7.5):

$$(2\pi)^4 \int [\bar{u}(3)(ig_e\gamma^\mu)v(4)] \frac{-ig_{\mu\nu}}{q^2} \times [\bar{v}(2)(ig_e\gamma^\nu)u(1)]\delta^4(q - p_3 - p_4)\delta^4(p_1 + p_2 - q)d^4q$$

The amplitude for this diagram is therefore

$$\mathcal{M}_2 = -\frac{g_e^2}{(p_1 + p_2)^2}[\bar{u}(3)\gamma^\mu v(4)][\bar{v}(2)\gamma_\mu u(1)] \tag{7.107}$$

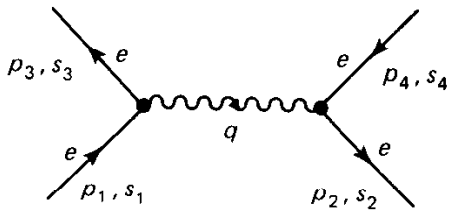


Figure 7.4 Electron-positron scattering.

* The fact that there are *two* diagrams for electron-electron and electron-positron scattering, but only *one* for electron-muon scattering, would appear offhand to be inconsistent with the classical limit. After all, Coulomb's Law says that the force of attraction or repulsion between two particles depends only on their charges, not on whether they happen to be identical (or antiparticles of one another). In the nonrelativistic limit, then, we should get the same answer whether we use the electron-muon formula or the electron-electron formula. The *amplitudes*, to be sure, are *not* the same, but the cross-section formula (6.34) carries a factor of S, which is $\frac{1}{2}$ for electron-electron scattering and 1 for electron-muon scattering. [For electron-positron scattering, $S = 1$, but the second amplitude (7.107) is smaller than the first (7.106) by a factor $(v/c)^2$, so that in the nonrelativistic limit only $\mathcal{M}_1$ contributes.]

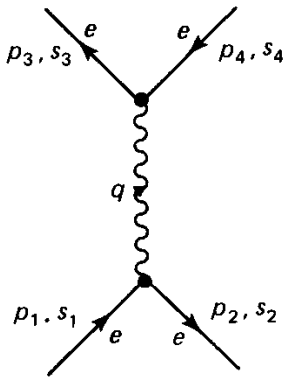
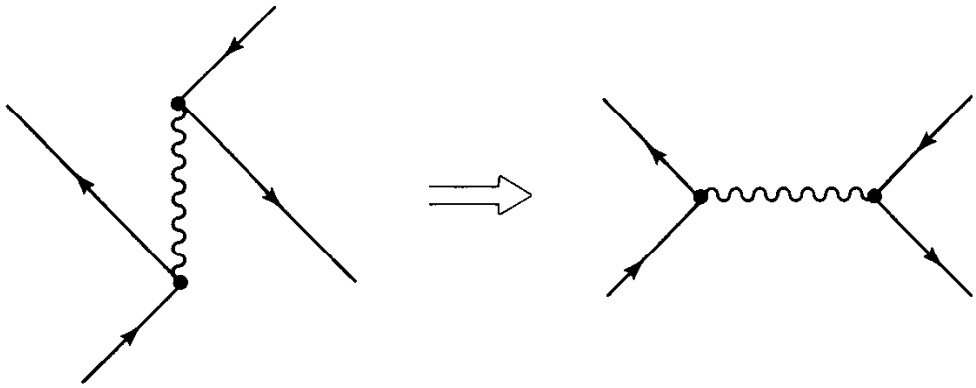


Figure 7.5 Second diagram contributing to electron-positron scattering.

Now, do we *add* these diagrams, or *subtract* them? Interchanging the incoming positron and the outgoing electron in the second diagram (Fig. 7.5), and then redrawing it in a more customary configuration



we recover the first diagram (Fig. 7.4). According to rule 8, then, we need a minus sign:

$$\mathcal{M} = -\frac{g_e^2}{(p_1 - p_3)^2}[\bar{u}(3)\gamma^\mu u(1)][\bar{v}(2)\gamma_\mu v(4)] + \frac{g_e^2}{(p_1 + p_2)^2}[\bar{u}(3)\gamma^\mu v(4)][\bar{v}(2)\gamma_\mu u(1)] \tag{7.108}$$

EXAMPLE 7.4 Compton Scattering

For an example involving the electron propagator and photon polarization, consider the case of Compton scattering, $\gamma + e \rightarrow \gamma + e$. Again there are two diagrams, but they do *not* differ by the interchange of fermions, and the amplitudes *add*. The first diagram (Fig. 7.6) yields

$$(2\pi)^4 \int \epsilon_\mu(2)\left[\bar{u}(4)(ig_e\gamma^\mu)\frac{i(\not{q} + mc)}{(q^2 - m^2c^2)}(ig_e\gamma^\nu)u(1)\right] \times \epsilon_\nu(3)^*\,\delta^4(p_1 - p_3 - q)\delta^4(p_2 + q - p_4)d^4q$$

Notice that the space–time index on each photon polarization vector is contracted with the index of the γ matrix at the vertex where the photon was created or absorbed. Notice also how the electron propagator fits in as we work our way backward along the fermion line. I have introduced here the convenient shorthand "a-slash":

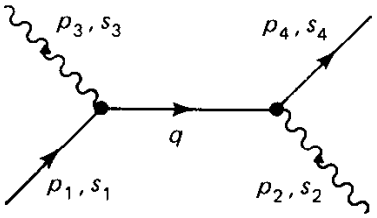


Figure 7.6 Compton scattering.

$$\not{a} \equiv a^\mu \gamma_\mu \tag{7.109}$$

Evidently, the amplitude associated with this diagram is*

$$\mathcal{M}_1 = \frac{g_e^2}{(p_1 - p_3)^2 - m^2c^2} [\bar{u}(4)\not{\epsilon}(2)(\not{p}_1 - \not{p}_3 + mc)\not{\epsilon}(3)^* u(1)] \tag{7.110}$$

Meanwhile, the second diagram (Fig. 7.7) yields

$$\mathcal{M}_2 = \frac{g_e^2}{(p_1 + p_2)^2 - m^2c^2} [\bar{u}(4)\not{\epsilon}(3)^*(\not{p}_1 + \not{p}_2 + mc)\not{\epsilon}(2)u(1)] \tag{7.111}$$

and the total amplitude is $\mathcal{M} = \mathcal{M}_1 + \mathcal{M}_2$.

7.7 CASIMIR'S TRICK AND THE TRACE THEOREMS

In some experiments the incoming and outgoing electron (or positron) spins are specified, and the photon polarizations are given. If so, the next thing to do is insert the appropriate spinors and polarization vectors into the expression for $\mathcal{M}$, and compute $|\mathcal{M}|^2$, the quantity we actually need to determine cross sections and lifetimes. More often, however, we are not interested in the spins. A typical experiment starts out with a beam of particles whose spin orientations are random, and simply counts the *number* of particles scattered in a given direction. In this case the relevant cross section is the *average* over all initial spin configurations, i, and the *sum* over all final spin configurations, f. In principle, we could compute $|\mathcal{M}(i \to f)|^2$ for every possible combination, and *then* do the summing and averaging:

$$\langle |\mathcal{M}|^2 \rangle \equiv \textit{average} \text{ over initial spins, } \textit{sum} \text{ over final spins, of } |\mathcal{M}(i \to f)|^2 \tag{7.112}$$

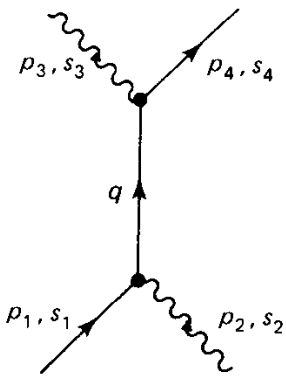


Figure 7.7 Second diagram for Compton scattering.

* Here and below, $\not{\epsilon}^*$ means $\gamma^\mu(\epsilon_\mu^*)$; the γ matrix is *not* conjugated.

In practice, it is much easier to compute $\langle|\mathcal{M}|^2\rangle$ directly, without ever evaluating the individual amplitudes.

Consider, for instance, the electron-muon scattering amplitude (7.104). Squaring, we have

$$|\mathcal{M}|^2 = \frac{g_e^4}{(p_1 - p_3)^4}[\bar{u}(3)\gamma^\mu u(1)][\bar{u}(4)\gamma_\mu u(2)][\bar{u}(3)\gamma^\nu u(1)]^*[\bar{u}(4)\gamma_\nu u(2)]^* \quad (7.113)$$

(I use ν for the second space-time index, to avoid confusion.) A glance at the first and third terms (or the second and fourth) reveals that we must handle quantities of the general form

$$G \equiv [\bar{u}(a)\Gamma_1 u(b)][\bar{u}(a)\Gamma_2 u(b)]^* \quad (7.114)$$

where (a) and (b) stand for the appropriate spins and momenta, and Γ_1, and Γ_2 are two 4×4 matrices. All the other processes described in Section 7.6—Møller, Bhabha, and Compton scattering, as well as pair production and annihilation—lead to expressions with similar structure. To begin with, we evaluate the complex conjugate (which is the same as the Hermitian conjugate, since the quantity in brackets is a 1×1 "matrix"):

$$[\bar{u}(a)\Gamma_2 u(b)]^* = [u(a)^\dagger\gamma^0\Gamma_2 u(b)]^\dagger = u(b)^\dagger\Gamma_2^\dagger\gamma^{0\dagger}u(a) \quad (7.115)$$

Now, $\gamma^{0\dagger} = \gamma^0$, and $(\gamma^0)^2 = 1$, so

$$[\bar{u}(a)\Gamma_2 u(b)]^* = u(b)^\dagger\gamma^0\gamma^0\Gamma_2^\dagger\gamma^0 u(a) = \bar{u}(b)\bar{\Gamma}_2 u(a) \quad (7.116)$$

where* $$\bar{\Gamma}_2 \equiv \gamma^0\Gamma_2^\dagger\gamma^0 \quad (7.117)$$

Thus $$G = [\bar{u}(a)\Gamma_1 u(b)][\bar{u}(b)\bar{\Gamma}_2 u(a)] \quad (7.118)$$

We are ready now to sum over the spin orientations of particle (b). Using the completeness relation (7.97), we have

$$\sum_{b \text{ spins}} G = \bar{u}(a)\Gamma_1\left\{\sum_{s_b=1,2} u^{(s_b)}(p_b)\bar{u}^{(s_b)}(p_b)\right\}\bar{\Gamma}_2 u(a)$$
$$= \bar{u}(a)\Gamma_1(\not{p}_b + m_b c)\bar{\Gamma}_2 u(a) = \bar{u}(a)Qu(a) \quad (7.119)$$

where Q is a temporary shorthand for the 4×4 matrix

$$Q \equiv \Gamma_1(\not{p}_b + m_b c)\bar{\Gamma}_2 \quad (7.120)$$

Next, we do the same for particle (a):

$$\sum_{a \text{ spins}}\sum_{b \text{ spins}} G = \sum_{s_a=1,2} \bar{u}^{(s_a)}(p_a)Qu^{(s_a)}(p_a)$$

Or, writing out the matrix multiplication explicitly (i and j are summed from 1 to 4):

$$\sum_{s_a=1,2} \bar{u}^{(s_a)}(p_a)_i Q_{ij} u^{(s_a)}(p_a)_j = Q_{ij}\left\{\sum_{s_a=1,2} u^{(s_a)}(p_a)\bar{u}^{(s_a)}(p_a)\right\}_{ji}$$
$$= Q_{ij}(\not{p}_a + m_a c)_{ji} = Tr(Q(\not{p}_a + m_a c)) \quad (7.121)$$

* Observe that the overbar now serves two different functions. On a *spinor* it denotes the *adjoint*: $\bar{\psi} \equiv \psi^\dagger\gamma^0$ [eq. (7.58)]; on a 4×4 *matrix* it defines a *new* matrix: $\bar{\Gamma} \equiv \gamma^0\Gamma^\dagger\gamma^0$.

where Tr denotes the trace of the matrix (the sum of its diagonal elements):

$$Tr(A) \equiv \sum_i A_{ii} \tag{7.122}$$

Conclusion:

$$\sum_{\text{all spins}} [\bar{u}(a)\Gamma_1 u(b)][\bar{u}(a)\Gamma_2 u(b)]^* = Tr[\Gamma_1(\not{p}_b + m_b c)\bar{\Gamma}_2(\not{p}_a + m_a c)] \tag{7.123}$$

This may not look like much of a simplification, but notice that there are *no spinors left;* once we do the summation over spins, it all reduces to matrix multiplication and taking the trace. For want of a better name, I call equation (7.123) "Casimir's trick," since Casimir was apparently the first one to use it.[5] If either u (in 7.123) is replaced by a v, the corresponding mass on the right-hand side switches sign (see Problem 7.26).

EXAMPLE 7.5

In the case of electron-muon scattering (7.113), $\Gamma_2 = \gamma^\nu$, and hence $\bar{\Gamma}_2 = \gamma^0\gamma^{\nu\dagger}\gamma^0 = \gamma^\nu$ (Problem 7.27). Applying Casimir's trick twice, we find

$$\langle|\mathcal{M}|^2\rangle = \frac{g_e^4}{4(p_1 - p_3)^4} Tr[\gamma^\mu(\not{p}_1 + mc)\gamma^\nu(\not{p}_3 + mc)] \times Tr[\gamma_\mu(\not{p}_2 + Mc)\gamma_\nu(\not{p}_4 + Mc)] \tag{7.124}$$

where m is the mass of the electron and M is the mass of the muon. The factor of $\frac{1}{4}$ is included because we want the *average* over the initial spins; since there are two particles, each with two allowed spin orientations, the average is a quarter of the sum.

Casimir's trick reduces everything down to a problem of calculating the trace of some complicated product of γ matrices. This algebra is facilitated by a number of theorems, which I collect together below (I'll leave the proofs to you—see Problems 7.29 through 7.32). First of all, I should mention three facts about traces in general: if A and B are any two matrices, and α is any number

1. $Tr(A + B) = Tr(A) + Tr(B)$
2. $Tr(\alpha A) = \alpha Tr(A)$
3. $Tr(AB) = Tr(BA)$

It follows from number 3 that $Tr(ABC) = Tr(CAB) = Tr(BCA)$, but this is *not* equal, in general, to the trace of the matrices taken in the other order: $Tr(ACB) = Tr(BAC) = Tr(CBA)$. In this way you can "peel" matrices off the back end of a product and move them around to the front, but you must preserve the ordering. It is useful to note that

4. $g_{\mu\nu}g^{\mu\nu} = 4$

and to recall the fundamental anticommutation relation for the γ matrices (together with an associated rule for "slash" products):

5. $\gamma^\mu\gamma^\nu + \gamma^\nu\gamma^\mu = 2g^{\mu\nu}$ 5′. $\not{a}\not{b} + \not{b}\not{a} = 2a \cdot b$

From these there follows a sequence of "contraction theorems":

6. $\gamma_\mu\gamma^\mu = 4$
7. $\gamma_\mu\gamma^\nu\gamma^\mu = -2\gamma^\nu$ — 7′. $\gamma_\mu \not{a}\gamma^\mu = -2\not{a}$
8. $\gamma_\mu\gamma^\nu\gamma^\lambda\gamma^\mu = 4g^{\nu\lambda}$ — 8′. $\gamma_\mu \not{a}\not{b}\gamma^\mu = 4a\cdot b$
9. $\gamma_\mu\gamma^\nu\gamma^\lambda\gamma^\sigma\gamma^\mu = -2\gamma^\sigma\gamma^\lambda\gamma^\nu$ — 9′. $\gamma_\mu \not{a}\not{b}\not{c}\gamma^\mu = -2\not{c}\not{b}\not{a}$

And finally, there is a set of "trace theorems":

10. The trace of the product of an *odd* number of gamma matrices is *zero*
11. $Tr(1) = 4$
12. $Tr(\gamma^\mu\gamma^\nu) = 4g^{\mu\nu}$ — 12′. $Tr(\not{a}\not{b}) = 4a\cdot b$
13. $Tr(\gamma^\mu\gamma^\nu\gamma^\lambda\gamma^\sigma) = 4(g^{\mu\nu}g^{\lambda\sigma} - g^{\mu\lambda}g^{\nu\sigma} + g^{\mu\sigma}g^{\nu\lambda})$ — 13′. $Tr(\not{a}\not{b}\not{c}\not{d}) = 4(a\cdot b\, c\cdot d - a\cdot c\, b\cdot d + a\cdot d\, b\cdot c)$

Since $\gamma^5 = i\gamma^0\gamma^1\gamma^2\gamma^3$ is the product of an *even* number of γ matrices, it follows from rule 10 that $Tr(\gamma^5\gamma^\mu) = Tr(\gamma^5\gamma^\mu\gamma^\nu\gamma^\lambda) = 0$. When γ^5 is multiplied by an *even* number of γ's, we find

14. $Tr(\gamma^5) = 0$
15. $Tr(\gamma^5\gamma^\mu\gamma^\nu) = 0$ — 15′. $Tr(\gamma^5\not{a}\not{b}) = 0$
16. $Tr(\gamma^5\gamma^\mu\gamma^\nu\gamma^\lambda\gamma^\sigma) = 4i\epsilon^{\mu\nu\lambda\sigma}$ — 16′. $Tr(\gamma^5\not{a}\not{b}\not{c}\not{d}) = 4i\epsilon^{\mu\nu\lambda\sigma}a_\mu b_\nu c_\lambda d_\sigma$

where

$$\epsilon^{\mu\nu\lambda\sigma} \equiv \begin{cases} -1, \text{ if } \mu\nu\lambda\sigma \text{ is an } \textit{even} \text{ permutation of } 0123, \\ +1, \text{ if } \mu\nu\lambda\sigma \text{ is an } \textit{odd} \text{ permutation}, \\ 0, \text{ if any two indices are the same.}* \end{cases}$$

EXAMPLE 7.6

Evaluate the traces in electron-muon scattering [eq. (7.124)].

$$Tr[\gamma^\mu(\not{p}_1 + mc)\gamma^\nu(\not{p}_3 + mc)]$$
$$= Tr(\gamma^\mu\not{p}_1\gamma^\nu\not{p}_3) + mc[Tr(\gamma^\mu\not{p}_1\gamma^\nu) + Tr(\gamma^\mu\gamma^\nu\not{p}_3)] + (mc)^2Tr(\gamma^\mu\gamma^\nu)$$

Solution. According to rule 10, the terms in square brackets are zero. The last term can be evaluated using rule 12, and the first by rule 13:

$$\begin{aligned} Tr(\gamma^\mu\not{p}_1\gamma^\nu\not{p}_3) &= (p_1)_\lambda(p_3)_\sigma Tr(\gamma^\mu\gamma^\lambda\gamma^\nu\gamma^\sigma) \\ &= (p_1)_\lambda(p_3)_\sigma 4(g^{\mu\lambda}g^{\nu\sigma} - g^{\mu\nu}g^{\lambda\sigma} + g^{\mu\sigma}g^{\lambda\nu}) \\ &= 4(p_1^\mu p_3^\nu - g^{\mu\nu}(p_1\cdot p_3) + p_3^\mu p_1^\nu) \end{aligned}$$

Thus
$$Tr(\gamma^\mu(\not{p}_1 + mc)\gamma^\nu(\not{p}_3 + mc)] = 4[p_1^\mu p_3^\nu + p_3^\mu p_1^\nu + g^{\mu\nu}((mc)^2 - (p_1\cdot p_3))] \qquad (7.125)$$

* By "even permutation" I mean an even number of interchanges of two indices. Thus $\epsilon^{\mu\nu\lambda\sigma} = -\epsilon^{\nu\mu\lambda\sigma} = \epsilon^{\nu\lambda\mu\sigma} = -\epsilon^{\nu\lambda\sigma\mu}$, and so on. Putting it differently, $\epsilon^{\mu\nu\lambda\sigma}$ is antisymmetric in every pair of superscripts. It might seem strange that ϵ^{0123} is *minus* 1; why not make it *plus* 1? It's purely *conventional*, of course. Evidently, whoever established the definition wanted ϵ_{0123} to be plus 1, and from that it follows that $\epsilon^{0123} = -1$, since three spatial indices are raised. By the way, if you are used to working with the three-dimensional Levi-Civita symbol ϵ_{ijk} (see Problem 4.19), be warned that although an even permutation on *three* indices corresponds to preservation of cyclic order ($\epsilon_{ijk} = \epsilon_{jki} = \epsilon_{kij}$), this is *not* the case for *four* indices: $\epsilon^{\mu\nu\lambda\sigma} = -\epsilon^{\nu\lambda\sigma\mu} = \epsilon^{\lambda\sigma\mu\nu} = -\epsilon^{\sigma\mu\nu\lambda}$.

The second trace (in 7.124) is the same, with $m \to M$, $1 \to 2$, $3 \to 4$, and the Greek indices lowered. So

$$\langle|\mathcal{M}|^2\rangle = \frac{4g_e^4}{(p_1 - p_3)^4}\,[p_1^\mu p_3^\nu + p_3^\mu p_1^\nu + g^{\mu\nu}((mc)^2 - (p_1 \cdot p_3))]$$

$$\times\,[p_{2_\mu} p_{4_\nu} + p_{4_\mu} p_{2_\nu} + g_{\mu\nu}((Mc)^2 - (p_2 \cdot p_4))]$$

$$= \frac{8g_e^4}{(p_1 - p_3)^4}\,[(p_1 \cdot p_2)(p_3 \cdot p_4) + (p_1 \cdot p_4)(p_2 \cdot p_3)$$

$$- (p_1 \cdot p_3)(Mc)^2 - (p_2 \cdot p_4)(mc)^2 + 2(mMc^2)^2] \qquad (7.126)$$

7.8 CROSS SECTIONS AND LIFETIMES

We are now back on familiar "turf." Having calculated $|\mathcal{M}|^2$ (or, where appropriate, $\langle|\mathcal{M}|^2\rangle$), we simply plug it into the relevant cross-section formula from Chapter 6: equation (6.34), in the general case; equation (6.42), for two-body scattering in the *CM;* or one of the equations from Problems (6.8), (6.9), or (6.10), in the lab frame.

EXAMPLE 7.7 Mott and Rutherford Scattering

An electron (mass m) scatters off a much heavier "muon" (mass $M \gg m$). Assuming that the recoil of M can be neglected, find the differential scattering cross section in the lab frame (M at rest).

Solution. According to Problem (6.8), the cross section is given by

$$\frac{d\sigma}{d\Omega} = \left(\frac{\hbar}{8\pi Mc}\right)^2 \langle|\mathcal{M}|^2\rangle$$

Because the target is stationary, we have (see Fig. 7.8):

$$p_1 = \left(\frac{E}{c}, \mathbf{p}_1\right), \qquad p_2 = (Mc, \mathbf{0}), \qquad p_3 = \left(\frac{E}{c}, \mathbf{p}_3\right), \qquad p_4 = (Mc, \mathbf{0})$$

where E is the incident (and scattered) electron energy, $\mathbf{p}_1$ is the incident momentum, and $\mathbf{p}_3$ is the scattered momentum (their magnitudes are equal, $|\mathbf{p}_1| = |\mathbf{p}_3| \equiv |\mathbf{p}|$, and the angle between them is θ: $\mathbf{p}_1 \cdot \mathbf{p}_3 = \mathbf{p}^2 \cos\theta$). So

$$(p_1 - p_3)^2 = -(\mathbf{p}_1 - \mathbf{p}_3)^2 = -\mathbf{p}_1^2 - \mathbf{p}_3^2 + 2\mathbf{p}_1 \cdot \mathbf{p}_3$$

$$= -2\mathbf{p}^2(1 - \cos\theta) = -4\mathbf{p}^2 \sin^2\frac{\theta}{2}$$

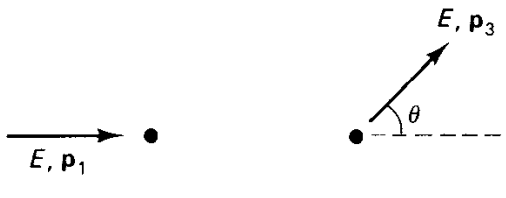


Figure 7.8 Electron scattering from a heavy target.

$$(p_1 \cdot p_3) = \frac{E^2}{c^2} - \mathbf{p}_1 \cdot \mathbf{p}_3 = \mathbf{p}^2 + m^2c^2 - \mathbf{p}^2 \cos\theta = m^2c^2 + 2\mathbf{p}^2 \sin^2\frac{\theta}{2}$$

$$(p_1 \cdot p_2)(p_3 \cdot p_4) = (p_1 \cdot p_4)(p_2 \cdot p_3) = (ME)^2$$

$$(p_2 \cdot p_4) = (Mc)^2$$

Putting this into equation (7.126), we have

$$\langle|\mathcal{M}|^2\rangle = \left(\frac{g_e^2 Mc}{\mathbf{p}^2 \sin^2(\theta/2)}\right)^2 \left((mc)^2 + \mathbf{p}^2 \cos^2\frac{\theta}{2}\right) \tag{7.127}$$

and therefore (recalling that $g_e = \sqrt{4\pi\alpha}$)

$$\frac{d\sigma}{d\Omega} = \left(\frac{\alpha\hbar}{2\mathbf{p}^2 \sin^2(\theta/2)}\right)^2 \left[(mc)^2 + \mathbf{p}^2 \cos^2\frac{\theta}{2}\right] \tag{7.128}$$

This is the Mott formula. It gives, to good approximation, the differential cross section for electron-proton scattering. If the incident electron is nonrelativistic, so that $\mathbf{p}^2 \ll (mc)^2$, equation (7.128) reduces to the Rutherford formula (compare Example 6.4):

$$\frac{d\sigma}{d\Omega} = \left(\frac{e^2}{2mv^2 \sin^2(\theta/2)}\right)^2 \tag{7.129}$$

What about decays? Actually, there is no such thing, in pure QED, for if a single fermion goes *in,* that same fermion must eventually come *out;* a fermion line cannot simply terminate within a diagram, nor is there any mechanism in QED for converting one fermion (say, a muon) into another (such as an electron). To be sure, there exist electromagnetic decays of *composite* particles, for example, $\pi^0 \to \gamma + \gamma$; but the *electromagnetic* component in this process is nothing but quark-antiquark pair annihilation, $q + \bar{q} \to \gamma + \gamma$. It is really a *scattering* event, in which the two colliding particles happen to be in a bound state. The cleanest example of such a process is the decay of positronium: $e^+ + e^- \to \gamma + \gamma$, which we consider in the following example. We'll do the analysis in the positronium rest frame (which is to say, in the *CM* frame of the electron-positron pair). They are typically moving rather slowly; indeed, for purposes of calculating the amplitude we shall assume they are at rest. On the other hand, this is one of those cases in which we cannot average over initial spins, because the composite system is either in the singlet configuration—spins antiparallel—or in the triplet configuration—spins parallel—and the formula for the cross section (and hence the lifetime) is quite different in the two cases.*

EXAMPLE 7.8 Pair Annihilation

Compute the amplitude, $\mathcal{M}$, for $e^+ + e^- \to \gamma + \gamma$, assuming the electron and positron are at rest, and in the singlet spin configuration.

* As a matter of fact, you *can* do this particular problem by Casimir's trick, because of a rather special circumstance: the singlet state can only decay to an *even* number of photons (predominantly *two*) and the triplet to an *odd* number (usually *three*). So in calculating the matrix element for $e^+ + e^- \to \gamma + \gamma$, we are *automatically* selecting out the singlet configuration even if the triplet was included in the sum over spins.

Solution. Two diagrams contribute, as shown in Figure 7.9. The amplitudes are (for simplicity I'll suppress the complex conjugate signs on the ϵ's):

$$\mathcal{M}_1 = \frac{g_e^2}{(p_1 - p_3)^2 - m^2c^2}\,\bar{v}(2)\not{\epsilon}_4(\not{p}_1 - \not{p}_3 + mc)\not{\epsilon}_3 u(1) \tag{7.130}$$

$$\mathcal{M}_2 = \frac{g_e^2}{(p_1 - p_4)^2 - m^2c^2}\,\bar{v}(2)\not{\epsilon}_3(\not{p}_1 - \not{p}_4 + mc)\not{\epsilon}_4 u(1) \tag{7.131}$$

and they *add*

$$\mathcal{M} = \mathcal{M}_1 + \mathcal{M}_2 \tag{7.132}$$

With the initial particles at rest, the photons come out "back to back," and we may as well choose the z axis to coincide with the photon line; then

$$\begin{aligned} p_1 &= mc(1, 0, 0, 0), \qquad p_2 = mc(1, 0, 0, 0), \\ p_3 &= mc(1, 0, 0, 1), \qquad p_4 = mc(1, 0, 0, -1) \end{aligned} \tag{7.133}$$

and hence

$$(p_1 - p_3)^2 - m^2c^2 = (p_1 - p_4)^2 - m^2c^2 = -2(mc)^2 \tag{7.134}$$

The amplitudes simplify somewhat if we exploit rule 5′ from Section 7.7:

$$\not{p}_1\not{\epsilon}_3 = -\not{\epsilon}_3\not{p}_1 + 2(p_1 \cdot \epsilon_3)$$

But ϵ_3 has only spatial components (in the Coulomb gauge), whereas p_1 is purely temporal, so $p_1 \cdot \epsilon_3 = 0$, and hence

$$\not{p}_1\not{\epsilon}_3 = -\not{\epsilon}_3\not{p}_1 \tag{7.135}$$

Similarly $$\not{p}_3\not{\epsilon}_3 = -\not{\epsilon}_3\not{p}_3 + 2(p_3 \cdot \epsilon_3)$$

but $(p_3 \cdot \epsilon_3) = 0$ by virtue of the Lorentz condition (7.90), so

$$\not{p}_3\not{\epsilon}_3 = -\not{\epsilon}_3\not{p}_3 \tag{7.136}$$

Therefore $$(\not{p}_1 - \not{p}_3 + mc)\not{\epsilon}_3 u(1) = \not{\epsilon}_3(-\not{p}_1 + \not{p}_3 + mc)u(1)$$

But $(p_1 - mc)u(1) = 0$, because $u(1)$ satisfies the Dirac equation (7.34), so

$$(\not{p}_1 - \not{p}_3 + mc)\not{\epsilon}_3 u(1) = \not{\epsilon}_3\not{p}_3 u(1) \tag{7.137}$$

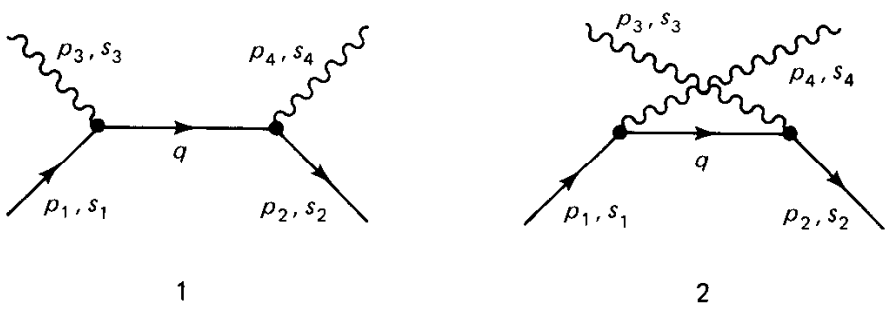


Figure 7.9 Two contributions to pair annihilation.

By the same token

$$(\not{p}_1 - \not{p}_4 + mc)\not{\epsilon}_4 u(1) = \not{\epsilon}_4 \not{p}_4 u(1) \tag{7.138}$$

Putting all this together, we find

$$\mathcal{M} = -\frac{g_e^2}{2(mc)^2}\bar{v}(2)[\not{\epsilon}_4 \not{\epsilon}_3 \not{p}_3 + \not{\epsilon}_3 \not{\epsilon}_4 \not{p}_4]u(1) \tag{7.139}$$

Now $\quad \not{p}_3 = mc(\gamma^0 - \gamma^3), \quad \not{p}_4 = mc(\gamma^0 + \gamma^3)$

so the expression in square brackets can be written as

$$mc[(\not{\epsilon}_4 \not{\epsilon}_3 + \not{\epsilon}_3 \not{\epsilon}_4)\gamma^0 - (\not{\epsilon}_4 \not{\epsilon}_3 - \not{\epsilon}_3 \not{\epsilon}_4)\gamma^3] \tag{7.140}$$

But

$$\not{\epsilon} = -\epsilon \cdot \gamma = -\begin{pmatrix} 0 & \sigma \cdot \epsilon \\ -\sigma \cdot \epsilon & 0 \end{pmatrix} \tag{7.141}$$

and therefore

$$\begin{aligned} \not{\epsilon}_3 \not{\epsilon}_4 &= \begin{pmatrix} 0 & \sigma \cdot \epsilon_3 \\ -\sigma \cdot \epsilon_3 & 0 \end{pmatrix}\begin{pmatrix} 0 & \sigma \cdot \epsilon_4 \\ -\sigma \cdot \epsilon_4 & 0 \end{pmatrix} \\ &= -\begin{pmatrix} (\sigma \cdot \epsilon_3)(\sigma \cdot \epsilon_4) & 0 \\ 0 & (\sigma \cdot \epsilon_3)(\sigma \cdot \epsilon_4) \end{pmatrix} \end{aligned} \tag{7.142}$$

In Chapter 4 (Problem 4.20) we encountered the useful theorem

$$(\sigma \cdot \mathbf{a})(\sigma \cdot \mathbf{b}) = \mathbf{a} \cdot \mathbf{b} + i\sigma \cdot (\mathbf{a} \times \mathbf{b}) \tag{7.143}$$

It follows that

$$(\not{\epsilon}_4 \not{\epsilon}_3 + \not{\epsilon}_3 \not{\epsilon}_4) = -2\epsilon_3 \cdot \epsilon_4 \tag{7.144}$$

(which we could have obtained directly from rule 5′), and

$$(\not{\epsilon}_4 \not{\epsilon}_3 - \not{\epsilon}_3 \not{\epsilon}_4) = 2i(\epsilon_3 \times \epsilon_4) \cdot \Sigma \tag{7.145}$$

where $\Sigma = \begin{pmatrix} \sigma & 0 \\ 0 & \sigma \end{pmatrix}$, as before. Accordingly

$$\mathcal{M} = \frac{g_e^2}{(mc)}\bar{v}(2)[\epsilon_3 \cdot \epsilon_4 \gamma^0 + i(\epsilon_3 \times \epsilon_4) \cdot \Sigma \gamma^3]u(1) \tag{7.146}$$

So far, I have said nothing about the spins of the electron and positron. Remember that we are interested in the *singlet* state:

$$(\uparrow\downarrow - \downarrow\uparrow)/\sqrt{2}$$

Symbolically

$$\mathcal{M}_{\text{singlet}} = (\mathcal{M}_{\uparrow\downarrow} - \mathcal{M}_{\downarrow\uparrow})/\sqrt{2} \tag{7.147}$$

$\mathcal{M}_{\uparrow\downarrow}$ is obtained from equation (7.146) with "spin up" for the electron

$$u(1) = \sqrt{2mc}\begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix} \tag{7.148}$$

and "spin down" for the positron

$$\bar{v}(2) = \sqrt{2mc}(0\,0\,1\,0) \tag{7.149}$$

Using these spinors, we find

$$\bar{v}(2)\gamma^0 u(1) = 0 \tag{7.150}$$

$$\bar{v}(2)\Sigma\gamma^3 u(1) = -2mc\hat{z} \tag{7.151}$$

So

$$\mathcal{M}_{\uparrow\downarrow} = -2ig_e^2(\epsilon_3 \times \epsilon_4)_z \tag{7.152}$$

Meanwhile, for $\mathcal{M}_{\downarrow\uparrow}$ we have

$$u(1) = \sqrt{2mc}\begin{pmatrix}0\\1\\0\\0\end{pmatrix}, \qquad \bar{v}(2) = -\sqrt{2mc}(0\,0\,0\,1) \tag{7.153}$$

from which it follows that

$$\mathcal{M}_{\downarrow\uparrow} = 2ig_e^2(\epsilon_3 \times \epsilon_4)_z = -\mathcal{M}_{\uparrow\downarrow} \tag{7.154}$$

Thus the amplitude for annihilation of a stationary e^+e^- pair into two photons, which emerge in the directions $\pm\hat{z}$, is

$$\mathcal{M}_{\text{singlet}} = -2\sqrt{2}\, ig_e^2(\epsilon_3 \times \epsilon_4)_z \tag{7.155}$$

(I note in passing that since $\mathcal{M}_{\uparrow\downarrow} = -\mathcal{M}_{\downarrow\uparrow}$, the *triplet* configuration $(\uparrow\downarrow + \downarrow\uparrow)/\sqrt{2}$ gives *zero*, confirming our earlier observation that the two-photon decay is forbidden in that case.)

Finally, we must put in the appropriate photon polarization vectors. Recall that for "spin up" ($m_s = +1$) we have

$$\epsilon_\uparrow = -(1/\sqrt{2})(1, i, 0) \tag{7.156}$$

whereas for "spin down" ($m_s = -1$)

$$\epsilon_\downarrow = (1/\sqrt{2})(1, -i, 0) \tag{7.157}$$

If the photon is traveling in the $+z$ direction, these correspond to right- and left-circular polarization, respectively. Since the z component of the total angular momentum must be zero, the photon spins must be oppositely aligned: $\uparrow\downarrow$ or $\downarrow\uparrow$. In the first case we have

$$(\uparrow\downarrow)\text{: } \epsilon_3 = -(1/\sqrt{2})(1, i, 0),\ \epsilon_4 = (1/\sqrt{2})(1, -i, 0),$$

so that

$$\epsilon_3 \times \epsilon_4 = i\hat{k} \tag{7.158}$$

In the second case 3 and 4 are interchanged, so that

$$(\downarrow\uparrow)\text{: } \epsilon_3 \times \epsilon_4 = -i\hat{k} \tag{7.159}$$

Evidently, we need the antisymmetric combination, $(\uparrow\downarrow - \downarrow\uparrow)/\sqrt{2}$, which should come as no surprise: this corresponds to a total spin of zero, just as it did when we combined two particles of spin $\frac{1}{2}$. Again, the amplitude is $(\mathcal{M}_{\uparrow\downarrow} - \mathcal{M}_{\downarrow\uparrow})/\sqrt{2}$, only this time the arrows refer to *photon* polarization. Finally, then:

$$\mathcal{M}_{\text{singlet}} = -4g_e^2 \tag{7.160}$$

(I have restored the complex conjugation of the polarization vectors, suppressed until now; this simply reverses the signs in (7.158) and (7.159).)

That was a lot of work, for a rather modest looking answer. What can we *do* with it? In the first place, we can calculate the total cross section for electron-positron annihilation. In the *CM* frame, the differential cross section is [see eq. (6.42)]

$$\frac{d\sigma}{d\Omega} = \left(\frac{\hbar c}{8\pi(E_1 + E_2)}\right)^2 \frac{|\mathbf{p}_f|}{|\mathbf{p}_i|} |\mathcal{M}|^2 \tag{7.161}$$

Here

$$E_1 = E_2 = mc^2, \qquad |\mathbf{p}_f| = mc \tag{7.162}$$

and, since the collision is nonrelativistic

$$|\mathbf{p}_i| = mv \tag{7.163}$$

where v is the incident electron (or positron) speed. (We used $v = 0$ in calculating $\mathcal{M}$, but obviously we cannot do so here. Is there an inconsistency in this? Not really. Think of it this way: $\mathcal{M}$ (and also E_1, E_2, $|\mathbf{p}_f|$, and $|\mathbf{p}_i|$) could be expanded in powers of v/c. What we have done is to calculate the leading term in each expansion.) Putting all this together, we find

$$\frac{d\sigma}{d\Omega} = \frac{1}{cv}\left(\frac{\hbar\alpha}{m}\right)^2 \tag{7.164}$$

Since there is no angular dependence, the total cross section is 4π times this:[6]

$$\sigma = \frac{4\pi}{cv}\left(\frac{\hbar\alpha}{m}\right)^2 \tag{7.165}$$

Finally, we would like to determine the lifetime of positronium, in the singlet state. This is clearly *related* to the cross section for pair annihilation (7.165), but what is the precise connection? Well, going back to equation (6.12)

$$\frac{d\sigma}{d\Omega} = \frac{1}{\mathcal{L}}\frac{dN}{d\Omega} \tag{6.12}$$

we see that the total number of scattering events per unit time is equal to the luminosity times the total cross section:

$$N = \mathcal{L}\sigma \tag{7.166}$$

If ρ is the number of incident particles per unit volume, and if they are traveling at speed v, then the luminosity (Fig. 7.10) is

$$\mathcal{L} = \rho v \tag{7.167}$$

For a single "atom," the electron density is $|\psi(0)|^2$, and N represents the probability of a disintegration, per unit time—which is to say, the decay rate. Thus

$$\Gamma = v\sigma|\psi(0)|^2 \tag{7.168}$$

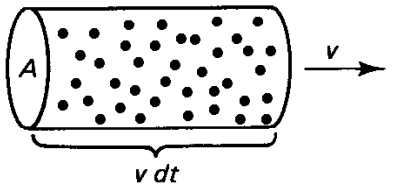


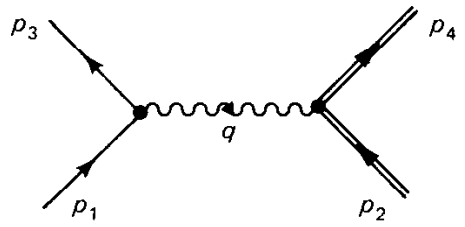
Figure 7.10 The number of particles in the cylinder is $\rho A v\, dt$, so the *luminosity* (number per unit area per unit time) is ρv.

Equations (7.165) and (7.168) are the formulas we used in Chapter 5 to determine the lifetime of positronium, $\tau = 1/\Gamma$ [see eq. (5.84)]:

$$\Gamma = \frac{4\pi}{c}\left(\frac{\hbar\alpha}{m}\right)^2 |\psi(0)|^2 \tag{7.169}$$

7.9 RENORMALIZATION

In Section 7.6 we considered "electron-muon" scattering, described in lowest order by the diagram



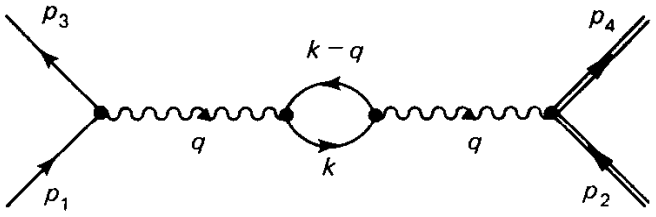
and by the corresponding amplitude

$$\mathcal{M} = -g_e^2[\bar{u}(p_3)\gamma^\mu u(p_1)]\frac{g_{\mu\nu}}{q^2}[\bar{u}(p_4)\gamma^\nu u(p_2)] \tag{7.170}$$

with

$$q = p_1 - p_3 \tag{7.171}$$

There are a number of fourth-order corrections, of which perhaps the most interesting is the "vacuum polarization" diagram



Here the virtual photon momentarily splits into an electron-positron pair, leading (as we saw qualitatively in Chapter 2) to a modification in the effective charge of the electron. My purpose now is to indicate how this works out quantitatively.

The amplitude for this diagram is (Problem 7.38)

$$\mathcal{M} = \frac{-ig_e^4}{q^4}[\bar{u}(p_3)\gamma^\mu u(p_1)]\left\{\int \frac{d^4k}{(2\pi)^4}\frac{Tr[\gamma_\mu(\not{k}+mc)\gamma_\nu(\not{q}-\not{k}+mc)]}{(k^2-m^2c^2)((q-k)^2-m^2c^2)}\right\}[\bar{u}(p_4)\gamma^\nu u(p_2)] \tag{7.172}$$

Its inclusion amounts to a modification of the photon propagator:

$$\frac{g_{\mu\nu}}{q^2} \to \frac{g_{\mu\nu}}{q^2} - \frac{i}{q^4} I_{\mu\nu} \tag{7.173}$$

where [comparing (7.170) and (7.172)]:

$$I_{\mu\nu} = -g_e^2 \int \frac{d^4k}{(2\pi)^4} \frac{Tr[\gamma_\mu(\not{k} + mc)\gamma_\nu(\not{q} - \not{k} + mc)]}{(k^2 - m^2c^2)((q-k)^2 - m^2c^2)} \tag{7.174}$$

Unfortunately, this integral is divergent. Naïvely, it should go like

$$\int |k|^3 \, d|k| \, \frac{|k|^2}{|k|^4} = \int |k| dk = |k|^2, \qquad \text{as } |k| \to \infty \tag{7.175}$$

(That is, it should be "quadratically divergent.") In actual fact, because of cancellations in the algebra, it only goes like $\ln|k|$ (it is "logarithmically divergent"). But never mind—either way, it blows up. We encountered a similar problem in Chapter 6; it seems to be characteristic of closed-loop diagrams in the Feynman calculus. Once again, the strategy will be to absorb the infinities into "renormalized" masses and coupling constants.

The integral (7.174) carries two space–time indices; once we have integrated over k, the only four-vector left is q^μ, so $I_{\mu\nu}$ must have the general form $g_{\mu\nu}(\ \ ) + q_\mu q_\nu(\ \ )$, where the parentheses contain some functions of q^2. We write it thus:[7]

$$I_{\mu\nu} = -ig_{\mu\nu}q^2 I(q^2) + q_\mu q_\nu J(q^2) \tag{7.176}$$

The second term contributes nothing to $\mathcal{M}$, since the q_μ contracts with γ^μ in equation (7.172), giving

$$[\bar{u}(p_3)\not{q}u(p_1)] = \bar{u}(p_3)(\not{p}_1 - \not{p}_3)u(p_1)$$

while, from equation (7.94)

$$\not{p}_1 u(p_1) = mc, \qquad \bar{u}(p_3)\not{p}_3 = \bar{u}(p_3)mc$$

and hence

$$[\bar{u}(p_3)\not{q}u(p_1)] = 0 \tag{7.177}$$

So we'll forget about the second term in equation (7.176). As for the first term, appropriate massaging of the integral (7.174) reduces it to the form (Problem 7.39)

$$I(q^2) = \frac{g_e^2}{12\pi^2}\left\{\int_{m^2}^{\infty} \frac{dx}{x} - 6\int_0^1 z(1-z)\ln\left(1 - \frac{q^2}{m^2c^2}z(1-z)\right)dz\right\} \tag{7.178}$$

The first integral clearly isolates the logarithmic divergence. To handle it, we temporarily impose a cutoff M (not to be confused with the mass of the muon), which we shall send to infinity at the end of the calculation:

$$\int_{m^2}^{\infty} \frac{dx}{x} \to \int_{m^2}^{M^2} \frac{dx}{x} = \ln\frac{M^2}{m^2} \tag{7.179}$$

The second integral

$$f(x) \equiv 6 \int_0^1 z(1 - z) \ln(1 + xz(1 - z))dz \tag{7.180}$$

is perfectly finite. It's irritating that it has to be left in this form; the integral just can't be reduced to elementary functions. It is easy enough, however, to evaluate numerically (Fig. 7.11), and the limiting expressions for large and small x are simple:

$$f(x) \cong \begin{cases} x/5 & (x \ll 1) \\ \ln x & (x \gg 1) \end{cases} \tag{7.181}$$

At any rate

$$I(q^2) = \frac{g_e^2}{12\pi^2}\left\{\ln\left(\frac{M^2}{m^2}\right) - f\left(\frac{-q^2}{m^2c^2}\right)\right\} \tag{7.182}$$

Notice that q^2 is *negative,* here. If the incident electron's three-momentum in the *CM* is $\mathbf{p}$, and the scattering angle is θ, then (Problem 7.40)

$$q^2 = -4\mathbf{p}^2 \sin^2\frac{\theta}{2} \tag{7.183}$$

Thus $-q^2/m^2c^2 \sim v^2/c^2$, and the limiting cases in equation (7.181) correspond to nonrelativistic and ultrarelativistic scattering, respectively.

The amplitude for electron-muon scattering, including vacuum polarization, is therefore

$$\mathcal{M} = -g_e^2[\bar{u}(p_3)\gamma^\mu u(p_1)]\frac{g_{\mu\nu}}{q^2}\left\{1 - \frac{g_e^2}{12\pi^2}\left[\ln\left(\frac{M^2}{m^2}\right) - f\left(\frac{-q^2}{m^2c^2}\right)\right]\right\}$$
$$\times [\bar{u}(p_4)\gamma^\nu u(p_2)] \tag{7.184}$$

Now comes the critical step, in which we "sop up" the infinity (contained for the moment in the cutoff M) by introducing the "renormalized" coupling constant

$$g_R \equiv g_e \sqrt{1 - \frac{g_e^2}{12\pi^2}\ln\left(\frac{M^2}{m^2}\right)} \tag{7.185}$$

Rewriting (7.184) in terms of g_R, we have

$$\mathcal{M} = -g_R^2[\bar{u}(p_3)\gamma^\mu u(p_1)]\frac{g_{\mu\nu}}{q^2}\left\{1 + \frac{g_R^2}{12\pi^2}f\left(\frac{-q^2}{m^2c^2}\right)\right\}[\bar{u}(p_4)\gamma^\nu u(p_2)] \tag{7.186}$$

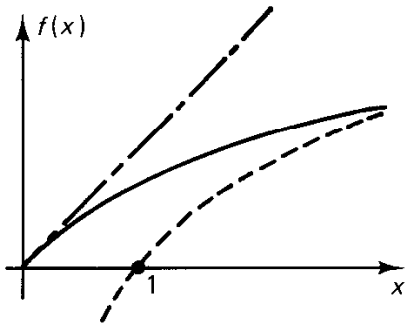


Figure 7.11 Graph of $f(x)$ [eq. (7.180)]. Solid line is the numerical result; dashed line below is $\ln x$ [which approximates $f(x)$ at large x]; straight line above is $x/5$ [which approximates $f(x)$ at small x].

[Equation (7.184) is only valid to order g_e^4 anyway, so it doesn't matter whether we use g_e or g_R within the curly brackets.] There are two important things to notice about this result:

1. The infinities are gone: there is no M in equation (7.186). All reference to the cutoff has been absorbed into the coupling constant. To be sure, everything is now written in terms of g_R, instead of g_e. But that's all to the good: g_R, not g_e, is what we actually measure in the laboratory (in Heaviside–Lorentz units it is the charge of the electron—or muon—and we determine it experimentally as the coefficient of attraction or repulsion between two such particles). If, in our theoretical analysis, we look only at "tree level" (lowest-order) diagrams, we are led to suppose that the physical charge is the same as the "bare" coupling constant, g_e. But as soon as we include higher-order effects we find that it is really g_R, not g_e, that corresponds to the measured electric charge. Does this mean that our earlier results are all wrong? No. What it means is that by naïvely interpreting g_e as the physical electric charge we were unwittingly taking into account the divergent part of the higher-order diagrams.

2. There remains the *finite* correction term, and here the important thing to notice is that it *depends on* q^2. We can absorb this, too, into the coupling constant, but the "constant" is now a function of q^2; we call it a "running" coupling constant:

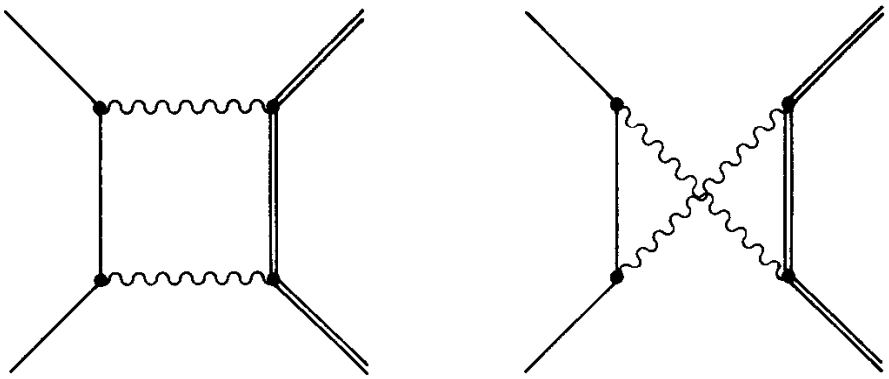
$$g_R(q^2) = g_R(0)\sqrt{1 + \frac{g_R(0)^2}{12\pi^2} f\left(\frac{-q^2}{m^2c^2}\right)} \tag{7.187}$$

or, in terms of the fine structure "constant" ($g_e = \sqrt{4\pi\alpha}$):

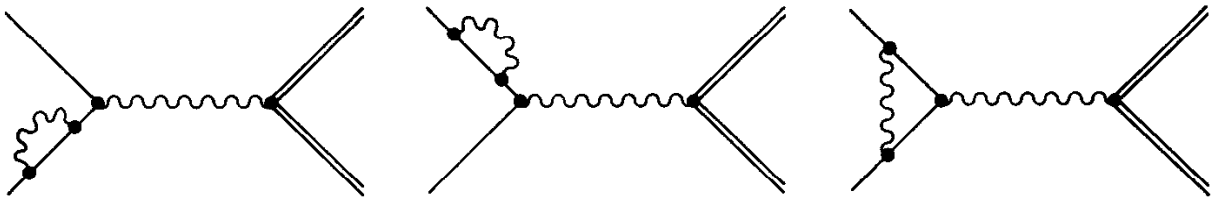
$$\alpha(q^2) = \alpha(0)\left\{1 + \frac{\alpha(0)}{3\pi} f\left(\frac{-q^2}{m^2c^2}\right)\right\} \tag{7.188}$$

The effective charge of the electron (and the muon), then, depends on the momentum transferred in the collision. Higher momentum transfer means closer approach, so another way of saying it is that the effective charge of each particle depends on how far apart they are. This is a consequence of vacuum polarization, which "screens" each charge. We now have an explicit formula for what was, in Chapter 2, a purely qualitative description. How come Millikan and Rutherford, or even Coulomb, never noticed this effect? If the electron's charge is not a constant, why doesn't this foul up everything from electronics to chemistry? The answer is that the variation is extremely slight, in nonrelativistic situations. Even in a head-on collision at $\frac{1}{10}c$, the correction term in equation (7.188) is only about 6×10^{-6} (Problem 7.41). For most purposes, therefore, $\alpha(0) = \frac{1}{137}$ will do just fine. Nevertheless, the second term in (7.188) makes a detectable contribution to the Lamb shift, for example.[8] Moreover, we shall encounter the same problem in quantum chromodynamics, where (because of quark confinement) the short-distance, relativistic régime is the case of interest.

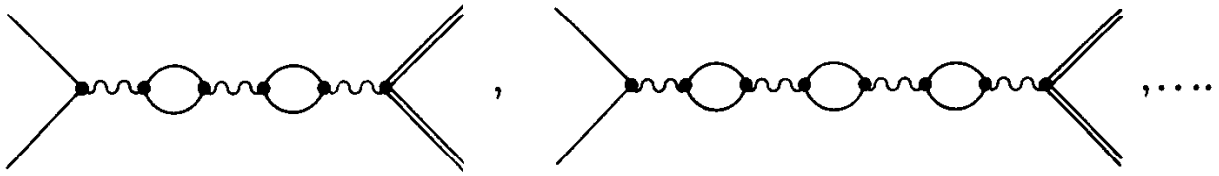
We have concentrated on one particular fourth-order process (vacuum polarization), but there are, of course, several others. There are the "ladder-diagrams":



These are finite and present no particular problems. But there are also three divergent graphs:



(and of course three more in which the extra virtual photon couples to the muon). The first two renormalize the electron's mass; the third modifies its magnetic moment. In addition, all three, considered separately, contribute to the renormalization of the electron's charge. Luckily, the latter contributions cancel one another, so that equation (7.185) remains valid. (I say "luckily," for these corrections depend on the mass of the particle to which the virtual photon line attaches, and if they did *not* cancel we would have a different renormalization for the muon than for the electron. The Ward identity (the official name for this cancellation) guarantees that renormalization preserves the equality of electric charges, irrespective of the mass of the carrier).* And then, there are even higher-order diagrams, such as



These introduce further terms in equation (7.188), of order α^2, α^3, and so on, but we shall not pursue the matter here, for the essential *ideas* are now on the table.

REFERENCES AND NOTES

1. J. D. Bjorken and S. D. Drell, *Relativistic Quantum Mechanics* and *Relativistic Quantum Fields* (New York: McGraw-Hill, 1964).
2. See for instance, Bjorken and Drell (ref. 1) chapter 2.

* Of course, we can put a *muon* bubble onto a photon line, just as we did with the *electron* in equation (7.172). But this will modify the electron and muon charge (and for that matter the tau and the quarks) by the same amount. The electron insertion is the dominant correction, simply because it is the lightest charged particle.

3. Bjorken and Drell (ref. 1), Sect. 2.3. Actually, the *sign* in equation (7.61) is pure convention; that is, $\psi' = -\gamma^0\psi$ would do just as well.
4. See J. D. Jackson, *Classical Electrodynamics,* 2d Ed. (New York: Wiley, 1975), Sect. 6.5.
5. A. Pais, *Inward Bound* (New York: Oxford, 1986) p. 375.
6. For more elegant derivations of (7.165), see J. M. Jauch and F. Rohrlich, *The Theory of Photons and Electrons,* 2d Ed. (New York: Springer-Verlag, 1975), Sect. 12-6, or J. J. Sakurai, *Advanced Quantum Mechanics* (Reading, MA: Addison-Wesley, 1967), pp. 216 ff.
7. My notation follows that of F. Halzen and A. D. Martin, *Quarks and Leptons,* (New York: Wiley, 1984), Chap. 7, and Bjorken and Drell (ref. 1), Chap. 8. I refer the reader to these texts, or to Sakurai (ref. 6), for further discussion.
8. See, for example, Halzen and Martin (ref. 7) Sect. 7.3.

PROBLEMS

7.1. Show that $\partial\phi/\partial x^\mu$ is a *covariant* four-vector (ϕ is a scalar function of x, y, z, and t). [*Hint:* First determine [from (3.7)] how covariant four-vectors transform; then use $\partial\phi/\partial x^{\mu'} = (\partial\phi/\partial x^\nu)(\partial x^\nu/\partial x^{\mu'})$ to find out how $\partial\phi/\partial x^\mu$ transforms.]

7.2. Show that equation (7.17) satisfies equation (7.15).

7.3. Derive equation (7.47), using equations (7.44) and (7.46).

7.4. Show that $u^{(1)}$ and $u^{(2)}$ [eq. (7.46)] are *orthogonal,* in the sense that $u^{(1)\dagger}u^{(2)} = 0$. Likewise, show that $u^{(3)}$ and $u^{(4)}$ are orthogonal. Are $u^{(1)}$ and $u^{(3)}$ orthogonal?

7.5. Show that for $u^{(1)}$ and $u^{(2)}$ [eq. (7.46)] the lower components (u_B) are smaller than the upper ones (u_A), in the nonrelativistic limit, by a factor v/c. [This observation simplifies matters, when we are doing nonrelativistic approximations; we think of u_A as the "big" components and u_B as the "little" components. (For $u^{(3)}$ and $u^{(4)}$ the roles are reversed.) In the relativistic limit, by contrast, u_A and u_B are comparable in size.]

7.6. If the z axis points along the direction of motion, show that equation (7.46) reduces to

$$u^{(1)} = \begin{pmatrix} \sqrt{(E + mc^2)/c} \\ 0 \\ \sqrt{(E - mc^2)/c} \\ 0 \end{pmatrix}, \text{ and so on}$$

Confirm that these are eigenspinors of S_z, and find the eigenvalues.

7.7. Construct the normalized spinors $u^{(+)}$ and $u^{(-)}$ representing an electron of momentum $\mathbf{p}$ with helicity $\pm$ 1. [That is, find the u's that satisfy equation (7.34), with positive E, and are eigenspinors of the helicity operator $(\hat{p}\cdot\boldsymbol{\Sigma})$ with eigenvalues $\pm$ 1.]

[*Solution:* $u^{(\pm)} = N \begin{pmatrix} u_A \\ \dfrac{\pm c|\mathbf{p}|}{(E + mc^2)} u_A \end{pmatrix}$,

where $u_A = \begin{pmatrix} p_z \pm |\mathbf{p}| \\ p_x + ip_y \end{pmatrix}$ and $N^2 = \dfrac{(E + mc^2)}{2|\mathbf{p}|c(|\mathbf{p}| \pm p_z)}$]

7.8. [The purpose of this problem is to demonstrate that particles described by the Dirac equation carry "intrinsic" angular momentum (**S**) in addition to their orbital angular momentum (**L**), neither of which is separately conserved, although their *sum is*. It should be attempted only if you are reasonably familiar with quantum mechanics.]

(a) Construct the Hamiltonian, H, for the Dirac equation. [*Hint:* Solve equation (7.19) for p^0/c. *Solution:* $H = c\gamma^0(\boldsymbol{\gamma}\cdot\mathbf{p} + mc)$, where $\mathbf{p} \equiv (\hbar/i)\nabla$ is the momentum operator.]

(b) Find the commutator of H with the orbital angular momentum $\mathbf{L} \equiv \mathbf{r} \times \mathbf{p}$. [*Solution:* $[H, \mathbf{L}] = -i\hbar c\gamma^0(\boldsymbol{\gamma} \times \mathbf{p})$]
Since $[H, \mathbf{L}]$ is not zero, **L** by itself is not conserved. Evidently there is some *other* form of angular momentum lurking here. Introduce the "spin angular momentum," **S**, defined by equation (7.48).

(c) Find the commutator of H with the spin angular momentum, $\mathbf{S} \equiv (\hbar/2)\boldsymbol{\Sigma}$. [*Solution:* $[H, \mathbf{S}] = i\hbar c\gamma^0(\boldsymbol{\gamma} \times \mathbf{p})$]
It follows that the *total* angular momentum, $\mathbf{J} = \mathbf{L} + \mathbf{S}$, *is* conserved.

(d) Show that every bispinor is an eigenstate of $\mathbf{S}^2$, with eigenvalue $\hbar^2 s(s+1)$, and find s. What, then, is the spin of a particle described by the Dirac equation?

7.9. The charge conjugation operator (C) takes a Dirac spinor ψ into the "charge-conjugate" spinor ψ_c, given by

$$\psi_c = i\gamma^2\psi^*$$

[See Halzen and Martin, ref. 7, Sect. 5.4.] Find the charge-conjugates of $u^{(1)}$ and $u^{(2)}$, and compare them with $v^{(1)}$ and $v^{(2)}$.

7.10. In going from equation (7.18) to (7.19), we (arbitrarily) chose to work with the factor containing the minus sign. How would Section 7.2 be changed if we were to replace (7.19) by $\gamma^\mu p_\mu + mc = 0$?

7.11. **(a)** Starting from equation (7.53), calculate $S^\dagger S$, and confirm equation (7.57).

(b) Show that $S^\dagger\gamma^0 S = \gamma^0$.

7.12. Show that $\bar{\psi}\gamma^5\psi$ [eq. (7.63)] is invariant under equation (7.52).

7.13. Show that the adjoint spinors $\bar{u}^{(1,2)}$ and $\bar{v}^{(1,2)}$ satisfy the equations

$$\bar{u}(\gamma^\mu p_\mu - mc) = 0, \qquad \bar{v}(\gamma^\mu p_\mu + mc) = 0$$

[*Hint:* Take the transpose conjugate of equations (7.34) and (7.51); multiply from the right by γ^0, and show that $(\gamma^\mu)^\dagger\gamma^0 = \gamma^0\gamma^\mu$.]

7.14. Show that the normalization condition (7.44), expressed in terms of the adjoint spinors, becomes

$$\bar{u}u = -\bar{v}v = 2mc$$

7.15. Show that $\bar{\psi}\gamma^\mu\psi$ is a four-vector, by confirming that its components transform according to the Lorentz transformation rule (3.7). Check that it transforms as a (polar) vector under parity (that is, the "time" component is invariant, whereas the "spatial" components change sign).

7.16. Show that the spinor representing an electron at rest [eq. (7.30)] is an eigenstate of the parity operator, P. What is its intrinsic parity? How about the positron? What if you changed the sign convention in equation (7.61) (see ref. 3)? [Notice that whereas the *absolute* parity of a spin-$\frac{1}{2}$ particle is in a sense arbitrary, the fact that particles and antiparticles carry *opposite* parity is *not* arbitrary.]

7.17. **(a)** Express $\gamma^\mu\gamma^\nu$ as a linear combination of 1, γ^5, γ^μ, $\gamma^\mu\gamma^5$, and $\sigma^{\mu\nu}$.
(b) Construct the matrices σ^{12}, σ^{13}, and σ^{23} [eq. (7.69)], and relate them to Σ^1, Σ^2, and Σ^3 [eq. (7.48)].

7.18. **(a)** Derive equations (7.70, i and iv) from equation (7.73).
(b) Prove equation (7.74), from equation (7.73).

7.19. Show that the continuity equation (7.74) enforces conservation of charge. [If you don't see how to do this, look in any electrodynamics textbook.]

7.20. Show that we are always free to pick $A^0 = 0$, in free space [eq. (7.85)]. [That is, given a potential A^μ which does *not* satisfy this constraint, find a gauge function λ, consistent with equation (7.84), such that A'_0, in equation (7.81), is zero.]

7.21. Suppose we apply a gauge transformation (7.81) to the plane-wave potential (7.88), using as the gauge function

$$\lambda = i\hbar\kappa a e^{-(i/\hbar)p\cdot x}$$

where κ is an arbitrary constant and p is the photon four-momentum.
(a) Show that this λ satisfies equation (7.84).
(b) Show that this gauge transformation has the effect of modifying ϵ^μ: $\epsilon^\mu \rightarrow \epsilon^\mu + \kappa p^\mu$. [In particular, if we choose $\kappa = -\epsilon^0/p^0$ we obtain the Coulomb gauge polarization vector (7.91).]
This observation leads to a beautifully simple test for the gauge invariance of QED results: the answer must be unchanged if you replace ϵ^μ by $\epsilon^\mu + \kappa p^\mu$.

7.22. Using $u^{(1)}$, $u^{(2)}$ (7.46) and $v^{(1)}$, $v^{(2)}$ (7.50), prove the completeness relations (7.97). [Note that the matrix multiplication is "backwards": $u\bar{u}$ is a 4×4 matrix, defined by $(u\bar{u})_{ij} \equiv u_i\bar{u}_j$.]

7.23. Using $\epsilon_{(1)}$ and $\epsilon_{(2)}$ (7.92), confirm the completeness relation (7.103).

7.24. Evaluate the amplitude for electron-muon scattering (7.104) in the *CM* system, assuming the e and μ approach one another along the z axis, repel, and return back along the z axis. Assume the initial and final particles all have helicity $+1$. [*Answer:* $\mathcal{M} = -2g_e^2$]

7.25. Determine the total amplitude [the analog to equation (7.104), or (7.105), or (7.108), or (7.110) plus (7.111)] for pair annihilation, $e^+ + e^- \rightarrow \gamma + \gamma$.

7.26. Work out the analog to Casimir's trick (7.123) for *antiparticles*

$$\sum_{\text{all spins}} [\bar{v}(a)\Gamma_1 v(b)][\bar{v}(a)\Gamma_2 v(b)]^*$$

and for the "mixed" cases

$$\sum_{\text{all spins}} [\bar{u}(a)\Gamma_1 v(b)][\bar{u}(a)\Gamma_2 v(b)]^*, \qquad \sum_{\text{all spins}} [\bar{v}(a)\Gamma_1 u(b)][\bar{v}(a)\Gamma_2 u(b)]^*$$

7.27. **(a)** Show that $\gamma^0\gamma^{\nu\dagger}\gamma^0 = \gamma^\nu$, for $\nu = 0$, 1, 2, or 3.
(b) If Γ is any product of γ matrices ($\Gamma = \gamma_a\gamma_b\cdots\gamma_c$) show that $\bar{\Gamma}$ [eq. (7.117)] is the same product in reverse order, $\bar{\Gamma} = \gamma_c\cdots\gamma_b\gamma_a$.

7.28. Apply Casimir's trick to obtain an expression analogous to equation (7.124) for Compton scattering. Note that there are four terms here:

$$|\mathcal{M}|^2 = |\mathcal{M}_1|^2 + |\mathcal{M}_2|^2 + \mathcal{M}_1\mathcal{M}_2^* + \mathcal{M}_1^*\mathcal{M}_2$$

7.29. **(a)** Prove trace theorems 1, 2, and 3, in Section 7.7.
(b) Prove equation 4.
(c) Using the anticommutation relation 5, prove 5′.

7.30. **(a)** Use the anticommutation relation 5 to prove the contraction theorems 6, 7, 8, and 9.
(b) From 7, prove 7′; from 8, prove 8′; from 9, prove 9′.

7.31. **(a)** Confirm the trace theorems 10, 11, 12, and 13.
(b) From 12, prove 12′; from 13, prove 13′.

7.32. **(a)** Prove theorems 14, 15, and 16.
(b) From 15, prove 15′; from 16, prove 16′.

7.33. **(a)** Show that $\epsilon^{\mu\nu\lambda\sigma}\epsilon_{\mu\nu\lambda\tau} = -6\delta^{\sigma}_{\tau}$. (Summation over μ, ν, λ implied.)
(b) Show that $\epsilon^{\mu\nu\lambda\sigma}\epsilon_{\mu\nu\theta\tau} = -2(\delta^{\lambda}_{\theta}\,\delta^{\sigma}_{\tau} - \delta^{\lambda}_{\tau}\,\delta^{\sigma}_{\theta})$.
(c) Find the analogous formula for $\epsilon^{\mu\nu\lambda\sigma}\epsilon_{\mu\phi\theta\tau}$.
(d) Find the analogous formula for $\epsilon^{\mu\nu\lambda\sigma}\epsilon_{\omega\phi\theta\tau}$.
[Here δ^{μ}_{ν} is the Kronecker delta: 1 if $\mu = \nu$, 0 otherwise. It could be written in terms of the mixed (co/contravariant) metric tensor: $\delta^{\mu}_{\nu} = g^{\mu}{}_{\nu} = g_{\nu}{}^{\mu}$.]

7.34. Starting with equation (7.105), determine the spin-averaged amplitude, [analogous to equation (7.126)] for electron-electron scattering. Assume we're working at high energies, so that the mass of the electron can be ignored (i.e., set $m = 0$). [*Hint:* You can read $\langle|\mathcal{M}_1|^2\rangle$ and $\langle|\mathcal{M}_2|^2\rangle$ from equation (7.126). For $\langle\mathcal{M}_1\mathcal{M}_2^*\rangle$ use the same strategy as Casimir's trick to get

$$\langle\mathcal{M}_1\mathcal{M}_2^*\rangle = \frac{-g_e^4}{4(p_1 - p_3)^2(p_2 - p_4)^2}\,Tr(\gamma^{\mu}\not{p}_1\gamma^{\nu}\not{p}_4\gamma_{\mu}\not{p}_2\gamma_{\nu}\not{p}_3)$$

Then exploit the contraction theorems to evaluate the trace. Notice that for *massless* particles the conservation of momentum ($p_1 + p_2 = p_3 + p_4$) implies that $p_1 \cdot p_2 = p_3 \cdot p_4$, $p_1 \cdot p_3 = p_2 \cdot p_4$, and $p_1 \cdot p_4 = p_2 \cdot p_3$.]

$$\left[\textit{Answer: } \langle|\mathcal{M}|^2\rangle = \frac{2g_e^4}{(p_1 \cdot p_3)^2(p_1 \cdot p_4)^2}\left[(p_1 \cdot p_2)^4 + (p_1 \cdot p_3)^4 + (p_1 \cdot p_4)^4\right]\right]$$

7.35. **(a)** Starting with equation (7.126), find the spin-averaged amplitude for electron-muon scattering in the *CM* frame, in the high-energy regime (m, $M \to 0$).
(b) Find the *CM* differential cross section for electron-muon scattering at high energy. Let E be the electron energy and θ the scattering angle.

$$\left[\textit{Answer: } \frac{d\sigma}{d\Omega} = \left(\frac{\hbar c}{8\pi}\right)^2 \frac{g_e^4}{2E^2}\left(\frac{1 + \cos^4\theta/2}{\sin^4\theta/2}\right)\right]$$

7.36. **(a)** Using the result of Problem (7.34), determine the spin-averaged amplitude for electron-electron scattering in the *CM* in the high-energy regime ($m \to 0$).
(b) Find the *CM* differential cross section for electron-electron scattering at high energy.

$$\left[\textit{Answer: } \frac{d\sigma}{d\Omega} = \left(\frac{\hbar c}{8\pi}\right)^2 \frac{g_e^4}{2E^2}\left(1 - \frac{4}{\sin^2\theta}\right)^2\right]$$

Compare the answers to Problems 7.35 and 7.36. (See footnote page 234).

7.37. Starting with equation (7.155), calculate $|\mathcal{M}|^2$, and use equation (7.103) to sum over photon polarizations. Check that the answer is consistent with equation (7.160),

and explain why this method gives the correct answer. [Note that we are now summing over *all* photon polarizations, whereas in *fact* the photons must be in the singlet configuration.]

7.38. Derive equation (7.172). For this we need one last Feynman rule: For a closed fermion loop include a factor -1 and take the trace.

7.39. Derive equation (7.178). [*Hint:* Use the integral theorems in Appendix E of Sakurai (ref. 6).]

7.40. Derive equation (7.183).

7.41. Evaluate the correction term in eq. (7.188) for the case of a head-on collision in the *CM;* assume the electron is traveling at $\frac{1}{10}c$.

Problems 42–44 pertain to the following model:

What if the photon, instead of being a ***massless vector*** *(spin 1) particle, were a* ***massive scalar*** *(spin 0) particle? Specifically, suppose the QED vertex factor were*

$$ig_e 1$$

(where 1 is the 4 × 4 unit matrix), and the "photon" propagator were

$$\frac{-i}{q^2 - (m_\gamma c)^2}.$$

There is no photon polarization vector now, and hence no factor for external photon lines. Apart from this, the Feynman rules for QED are unchanged.

7.42. Assuming it is heavy enough, this "photon" can decay.

(a) Calculate the decay rate for $\gamma \rightarrow e^+ + e^-$.

(b) If $m_\gamma = 300\ \text{MeV}/c^2$, find the lifetime of the "photon," in seconds.

7.43. **(a)** Find the amplitude, $\mathcal{M}$, for electron-muon scattering, in this theory.

(b) Calculate the spin-averaged quantity, $\langle|\mathcal{M}|^2\rangle$.

(c) Determine the differential cross section for electron-muon scattering in the *CM* frame. Assume the energy is high enough so that the electron and muon masses can be neglected: $m_e,\ m_\mu \rightarrow 0$. Express your answer in terms of the incident electron energy, E and the scattering angle, θ.

(d) From your result in (c), calculate the total cross section, assuming the "photon" is extremely heavy, $m_\gamma c^2 \gg E$.

(e) Going back to (b), consider now the case of *low*-energy scattering from an extremely heavy "muon": $|\mathbf{p}_e|/c \ll m_e \ll m_\gamma \ll m_\mu$. Find the differential cross section in the *lab* frame (muon at rest), assuming the muon does not recoil appreciably. Compare the Rutherford formula (Example 7.7), and calculate the total cross section.

7.44. **(a)** Find the amplitude, $\mathcal{M}$, for pair annihilation, $e^+ + e^- \rightarrow \gamma + \gamma$), in this theory.

(b) Determine $\langle|\mathcal{M}|^2\rangle$, assuming the energy is high enough that we can ignore both the electron and the "photon" mass ($m_e,\ m_\gamma \rightarrow 0$).

(c) Evaluate your result, in (b), in the *CM* system. Express your answer in terms of the incident electron energy, E, and the scattering angle, θ.

(d) Find the differential cross section for pair annihilation, in the *CM* system, still assuming $m_e = m_\gamma = 0$. Is the total cross section finite?

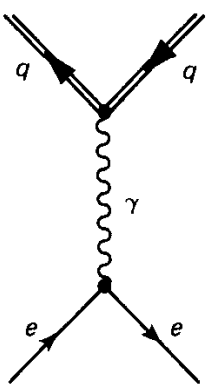
Chapter 8

Electrodynamics of Quarks and Hadrons

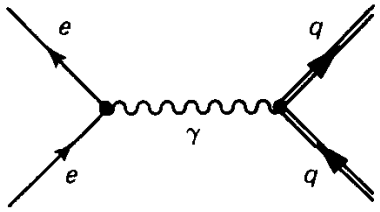
Because the electromagnetic interactions of electrons are well understood, they serve as useful probes of the structure of mesons and baryons. In this chapter we investigate the two most important examples: production of hadrons in e^+e^- scattering, and "deep inelastic scattering" of electrons and protons. Both were crucial in establishing the color/flavor model of quarks. This material will not be used in subsequent chapters, but the first two sections, at least, should be studied with care.

8.1 ELECTRON-QUARK INTERACTIONS

Everything I said in Chapter 7 about the electrodynamics of electrons and muons applies just as well to quarks (using, of course, the appropriate charge: $\frac{2}{3}e$ or $-\frac{1}{3}e$). However, the experimental situation is complicated by the fact that the quarks themselves never see the light of day, and we are obliged to *infer* from the observed behavior of mesons and baryons what their constituents are up to. In this chapter we shall consider two important examples: the production of hadrons in electron-positron scattering, for which the underlying electromagnetic process is



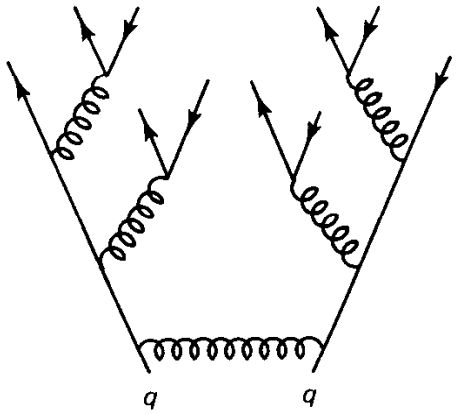
and high-energy electron-proton scattering ("deep inelastic scattering"), for which the basic diagram is



In the former case the quark-antiquark pair subsequently "hadronizes," producing the mesons and baryons we actually observe; in the latter case the quark is accompanied by "spectator" quarks to make up the proton. However, *that* part of the problem is *chromo*dynamics, not *electro*dynamics; for the moment, we are concerned only with the *electrodynamic* part of the process.

8.2 HADRON PRODUCTION IN e^+e^- SCATTERING

Electrons and positrons do not participate in the strong interactions; at present energies, the only way an e^+e^- collision can produce strongly interacting particles is through a virtual photon: $e^+ + e^- \rightarrow \gamma \rightarrow q + \bar{q} \rightarrow$ hadrons. (By 1987 the Linear Collider at SLAC should be operating at energies high enough to produce virtual Z^0's copiously; the dominant mechanism will then be the *weak* interaction: $e^+ + e^- \rightarrow Z^0 \rightarrow q + \bar{q} \rightarrow$ hadrons.) For a brief moment the quarks fly apart as free particles, but when they reach a separation distance of around 10^{-15} m (the diameter of a hadron), their (strong) interaction is so great that new quark-antiquark pairs are produced—this time mainly from gluons:



These quarks and antiquarks, literally *dozens* of them, in a typical modern experiment, join together in myriad combinations to make the mesons and baryons that are actually recorded at the detector. In all the debris there is one unmistakable footprint left behind by the original quark-antiquark pair: the hadrons emerge in two back-to-back "jets," one along the direction of the primordial quark, the other marking the direction of the antiquark (Fig. 8.1). [Occasionally one sees a three-jet event (Fig. 8.2), indicating that a gluon carrying a substantial fraction of the total energy was emitted in conjuntion with the original $q\bar{q}$ production:

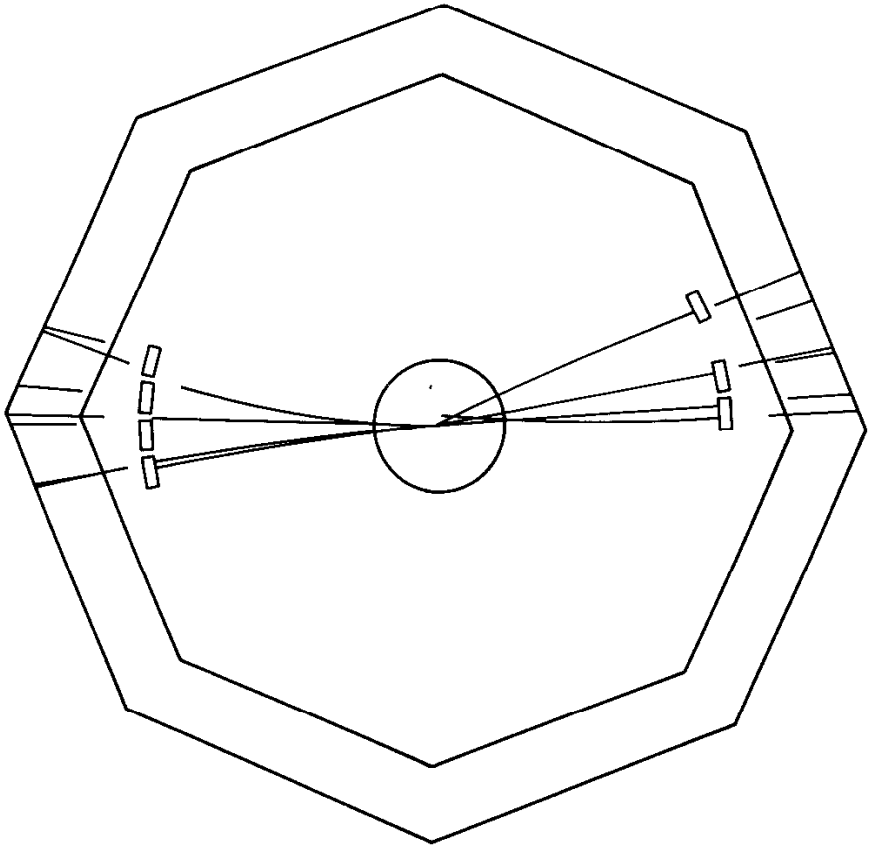


Figure 8.1 A typical two-jet event. (Courtesy J. Dorfan, SLAC.)

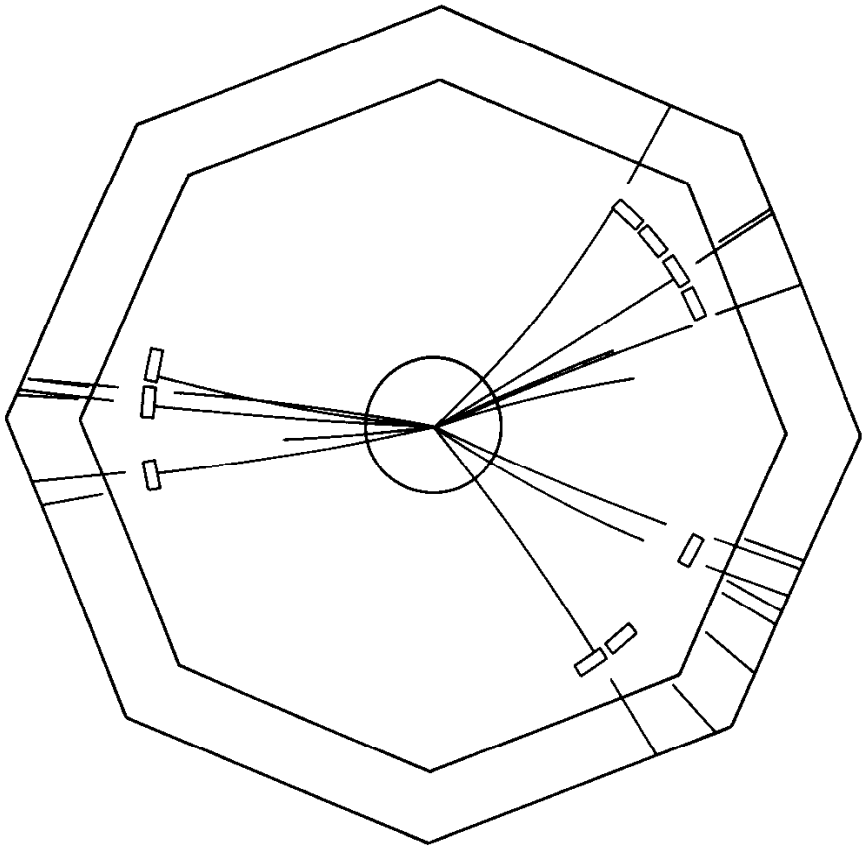
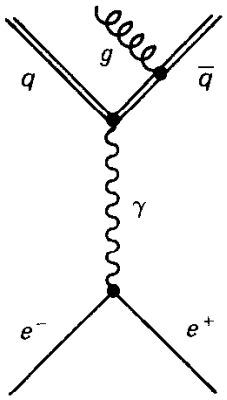
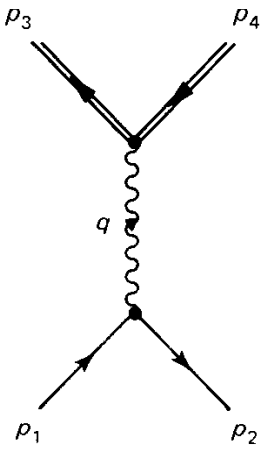


Figure 8.2 A three-jet event. (Courtesy J. Dorfan, SLAC.)



Indeed, the observation of three-jet events is generally regarded as our most direct evidence for the existence of gluons.]

Now, the *first* stage in this process ($e^+ + e^- \to \gamma \to q + \bar{q}$) is ordinary QED; the calculation is exactly the same as for $e^+ + e^- \to \gamma \to \mu^+ + \mu^-$:



The amplitude is

$$\mathcal{M} = \frac{Qg_e^2}{(p_1 + p_2)^2}[\bar{v}(p_2)\gamma^\mu u(p_1)][\bar{u}(p_3)\gamma_\mu v(p_4)] \tag{8.1}$$

where Q is the quark charge, in units of e ($\frac{2}{3}$, for u, c, and t; $-\frac{1}{3}$ for d, s, and b). Exploiting Casimir's trick, we obtain

$$\langle|\mathcal{M}|^2\rangle = \frac{1}{4}\left[\frac{Qg_e^2}{(p_1 + p_2)^2}\right]^2 Tr[\gamma^\mu(\not{p}_1 + mc)\gamma^\nu(\not{p}_2 - mc)] \times Tr[\gamma^\mu(\not{p}_4 - Mc)\gamma^\nu(\not{p}_3 + Mc)] \tag{8.2}$$

where m is the mass of the electron and M is that of the quark. (See Problem 8.1.) Invoking the trace theorems of Chapter 7, we can reduce this to

$$\langle|\mathcal{M}|^2\rangle = 8\left[\frac{Qg_e^2}{(p_1 + p_2)^2}\right]^2 [(p_1 \cdot p_3)(p_2 \cdot p_4) + (p_1 \cdot p_4)(p_2 \cdot p_3) + (mc)^2(p_3 \cdot p_4) + (Mc)^2(p_1 \cdot p_2) + 2(mc)^2(Mc)^2] \tag{8.3}$$

Or, in terms of the incident (CM) electron energy E and the angle θ between the incoming electron and the outgoing quark:

$$\langle|\mathcal{M}|^2\rangle = Q^2 g_e^4 \left\{1 + \left(\frac{mc^2}{E}\right)^2 + \left(\frac{Mc^2}{E}\right)^2 + \left[1 - \left(\frac{mc^2}{E}\right)^2\right]\left[1 - \left(\frac{Mc^2}{E}\right)^2\right]\cos^2\theta\right\} \tag{8.4}$$

The differential scattering cross section is given by equation (6.42); integrating over θ and ϕ, we obtain the total cross section (Problem 8.2):

$$\sigma = \frac{\pi Q^2}{3}\left(\frac{\hbar c\alpha}{E}\right)^2 \sqrt{\frac{1 - (Mc^2/E)^2}{1 - (mc^2/E)^2}}\left[1 + \frac{1}{2}\left(\frac{Mc^2}{E}\right)^2\right]\left[1 + \frac{1}{2}\left(\frac{mc^2}{E}\right)^2\right] \tag{8.5}$$

Notice the threshold at $E = Mc^2$; for energies less than this the square root is imaginary, reflecting the fact that the process is kinematically forbidden when there is not enough energy to create the $q\bar{q}$ pair. If we are substantially above threshold ($E > Mc^2 \gg mc^2$), equation (8.5) simplifies considerably:*

$$\sigma = \frac{\pi}{3}\left(\frac{\hbar Qc\alpha}{E}\right)^2 \tag{8.6}$$

As we crank up the beam energy, we encounter a succession of such thresholds—first the muon and the light quarks, later (at about 1500 MeV) the charm quark, the tau (at 1784 MeV), the bottom quark (4700 MeV) and eventually (one hopes) the top quark. There is a beautiful way to display this structure: suppose we examine the *ratio* of the rate of hadron production to that for muon pairs:

$$R \equiv \frac{\sigma(e^+e^- \to \text{hadrons})}{\sigma(e^+e^- \to \mu^+\mu^-)} \tag{8.7}$$

Since the numerator includes all the quark-antiquark events,† equation (8.6) gives

$$R(E) = 3\sum Q_i^2 \tag{8.8}$$

in which the sum is over all quark flavors with thresholds below E. Notice the 3 in front—it records the fact that there are three *colors* for each flavor. We anticipate a "staircase" graph for $R(E)$, then, ascending one step at each new

* This approximation is actually better than it looks, because of a lucky algebraic cancellation: expanding the radical, $\sqrt{1 - (Mc^2/E)^2}\,[1 + \frac{1}{2}(Mc^2/E)^2] = 1 - \frac{3}{8}(Mc^2/E)^4 \ldots$, so the error is of order $(Mc^2/E)^4$, not $(Mc^2/E)^2$. As for the *electron* mass terms, these are smaller to begin with, though there is a second-order correction; however, these terms cancel exactly in the calculation of R [eq. (8.7)].

† The τ lepton decays predominantly into hadrons, and this should add something (less than 1) to R, above 1784 MeV. This presumably explains why the experimental numbers are somewhat higher than the "$u + d + s + c$" line in Figure 8.3. The fact that the (unanticipated) τ threshold comes so soon after the c led to some confusion at first, and the theory seemed to be in trouble, but the discovery of the τ, and the flatness of the graph above the Υ restored confidence in this simple picture.

quark threshold, with the height of the rise determined by the quark's charge. At low energy where only the *u, d,* and *s* quarks contribute, we expect

$$R = 3[(\tfrac{2}{3})^2 + (-\tfrac{1}{3})^2 + (-\tfrac{1}{3})^2] = 2 \tag{8.9}$$

Between the *c* threshold and the *b* threshold we should have

$$R = 2 + 3(\tfrac{2}{3})^2 = \tfrac{10}{3} \tag{8.10}$$

and above the *b*

$$R = \tfrac{10}{3} + 3(-\tfrac{1}{3})^2 = \tfrac{11}{3} \tag{8.11}$$

The experimental results are shown in Figure 8.3. (The top quark should produce a jump to $R = 5$; evidently its mass exceeds 18 GeV/c^2, since no such rise is apparent in these data.) The agreement between theory and experiment is pretty good, especially at high energy. But you may well ask why it is not *perfect.* Apart from the approximation in going from equation (8.5) to equation (8.6) (which artificially sharpens the corners at each threshold), and the neglect of the tau (which leads to an underestimate in the region above 1784 GeV), we have made a fundamental oversimplification in assuming that we could treat the process as a sequence of two independent operations: $e^+e^- \rightarrow q\bar{q}$ (QED) followed by $q\bar{q} \rightarrow$ hadrons (QCD). In point of fact, the quarks produced in the first step are *not* free particles, obeying the Dirac equation; rather, they are *virtual* particles, on their way to a second interaction. This is particularly critical when the energy is right for formation of a bound state ($\phi = s\bar{s}$, $\psi = c\bar{c}$, $\Upsilon = b\bar{b}$); in the vicinity of such a "resonance" the interaction of the two quarks can scarcely be ignored. Hence the sharp spikes in the graph, which typically occur just below each threshold. In the broad plateau regions, however, the flatness of the graph is an indication that our naïve picture is not too bad. Like the jet structure, this is a manifestation of asymptotic freedom, and not surprisingly it works better at high energies.

But, really, all this is quibbling anyway, for the importance of Figure 8.3 lies not in what the small discrepancies whisper, but in what the overall agreement shouts: the factor of 3 in equation (8.8) clearly belongs there. Without it the theory would be wildly off (look at the dashed line in Figure 8.3)—and not just as isolated resonances, but across-the-board. That 3, remember, counts the number of colors. Here, then, is compelling *experimental* evidence for the color hypothesis—a hypothesis that was introduced originally for esoteric theoretical reasons but is now, of course, an indispensable ingredient in the whole story of strong interactions.

8.3 ELASTIC ELECTRON-PROTON SCATTERING

We now turn to electron-proton scattering, our best probe of the internal structure of the proton. If the proton were a simple point charge, obeying the Dirac equation, we could just copy our analysis of electron-muon scattering, with M now the mass of the proton. The lowest-order Feynman diagram would be

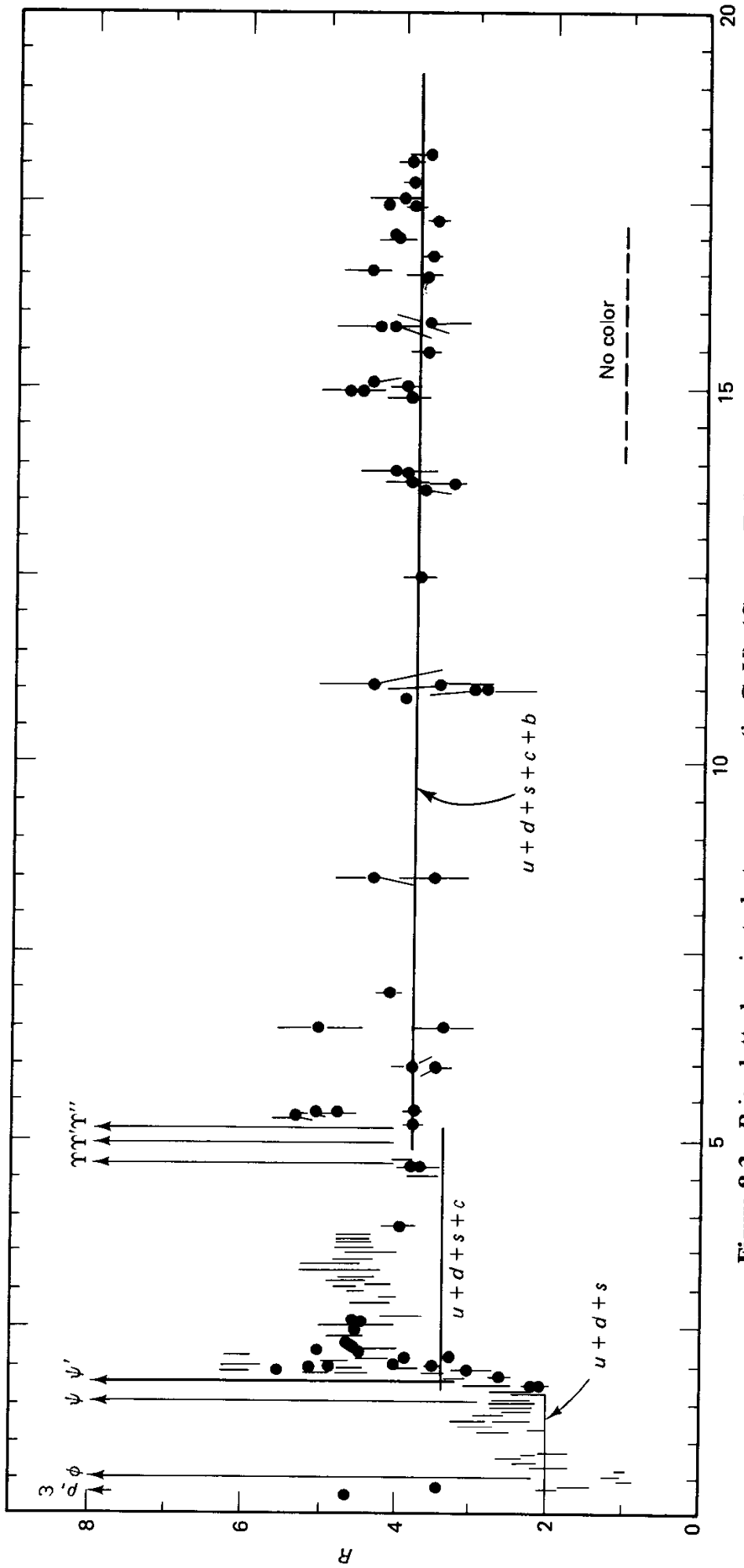
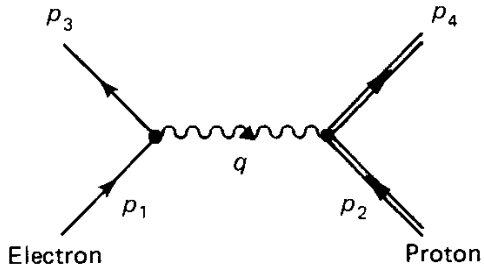


Figure 8.3 R is plotted against electron energy (in GeV). (*Source:* F. Halzen and A. D. Martin, *Quarks and Leptons* (New York: Wiley, copyright © 1984, p. 229. Reprinted by permission of John Wiley & Sons, Inc.)



and the (spin-averaged) amplitude would be [eq. (7.124)]

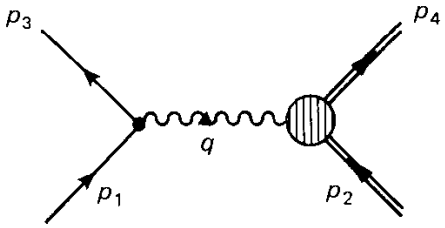
$$\langle|\mathcal{M}|^2\rangle = \frac{g_e^4}{q^4} L^{\mu\nu}_{\text{electron}} L_{\mu\nu\ \text{proton}} \tag{8.12}$$

where $q = p_1 - p_3$ and [eq. (7.125)]

$$L^{\mu\nu}_{\text{electron}} = 2\{p_1^\mu p_3^\nu + p_1^\nu p_3^\mu + g^{\mu\nu}[(mc)^2 - (p_1 \cdot p_3)]\} \tag{8.13}$$

(and a similar expression for $L^{\mu\nu}_{\text{proton}}$, only with $m \to M$ and $1, 3 \to 2, 4$). We used these results in Example 7.7 to derive the Mott and Rutherford scattering formulas.

But the proton is *not* a simple point charge, and so, long before the advent of the quark model, a more flexible formalism was introduced for describing electron-proton scattering. We might represent the process, in lowest-order QED, by a diagram like this:



where the blob on the right serves to remind us that we don't really know how the (virtual) photon interacts with the proton. (We *do* assume, however, that the scattering is *elastic:* $e + p \to e + p$; we shall consider *inelastic* processes, $e + p \to e + X$, in the next section.) Now, the essential point is that the *electron* vertex and the *photon propagator* are *unchanged,* and therefore, since $\langle|\mathcal{M}|^2\rangle$ neatly factors [eq. (8.12)],

$$\langle|\mathcal{M}|^2\rangle = \frac{g_e^4}{q^4} L^{\mu\nu}_{\text{electron}} K_{\mu\nu\ \text{proton}} \tag{8.14}$$

where $K_{\mu\nu}$ is an unknown quantity describing the photon-proton vertex.

Well . . . not *completely* unknown, for this much we *can* say: it is certainly a second-rank tensor, and the only variables that it can possibly depend on are p_2, p_4, and q. Since $q = p_4 - p_2$, these three are not independent, and we are free to use any two of them; the customary choice is q and p_2 (I'll drop the subscript from here on: $p \equiv p_2$ is the initial proton momentum). Now, there aren't many tensors that can be *constructed* out of just two four-vectors; the most general possible form is

$$K^{\mu\nu}_{\text{proton}} = -K_1 g^{\mu\nu} + \frac{K_2}{(Mc)^2} p^\mu p^\nu + \frac{K_4}{(Mc)^2} q^\mu q^\nu + \frac{K_5}{(Mc)^2}(p^\mu q^\nu + p^\nu q^\mu) \tag{8.15}$$

where the K_i are (unknown) functions of the only scalar variable in the problem: q^2. [Notice that $p^2 = (Mc)^2$ is a constant, and $q \cdot p = -q^2/2$]. The factors $(Mc)^{-2}$ have been pulled out, in defining K_2, K_4, and K_5, just so all the K's will have the same dimensions.* In principle, we could add an antisymmetric combination $(p^\mu q^\nu - p^\nu q^\mu)$, but since $L^{\mu\nu}$ is symmetric (8.13), such a term would contribute nothing to $\langle|\mathcal{M}|^2\rangle$. Now, these four functions are not independent; it can be shown (Problem 8.4) that

$$q_\mu K^{\mu\nu} = 1 \tag{8.16}$$

from which it follows (Problem 8.5) that

$$K_4 = \frac{(Mc)^2}{q^2} K_1 + \frac{1}{4} K_2 \quad \text{and} \quad K_5 = \frac{1}{2} K_2 \tag{8.17}$$

Thus $K^{\mu\nu}$ can be expressed in terms of just *two* (unknown) functions, $K_1(q^2)$ and $K_2(q^2)$:

$$K^{\mu\nu}_{\text{proton}} = K_1\left(-g^{\mu\nu} + \frac{q^\mu q^\nu}{q^2}\right) + \frac{K_2}{(Mc)^2}\left(p^\mu + \frac{1}{2} q^\mu\right)\left(p^\nu + \frac{1}{2} q^\nu\right) \tag{8.18}$$

A fundamental problem for any theory of proton structure is to determine these two functions. They are easy enough to measure experimentally, for they are directly related to the electron-proton elastic scattering cross section. According to equations (8.13) and (8.18), (Problem 8.7)

$$\langle|\mathcal{M}|^2\rangle = \left(\frac{2g_e^2}{q^2}\right)^2 \left\{K_1[(p_1 \cdot p_3) - 2(mc)^2] + K_2\left[\frac{(p_1 \cdot p)(p_3 \cdot p)}{(Mc)^2} + \frac{q^2}{4}\right]\right\} \tag{8.19}$$

We shall work in the laboratory frame, with the target proton at rest, $p = (Mc, 0, 0, 0)$. An electron with incident energy E scatters at an angle θ, emerging with energy E'. Let us assume it's a moderately energetic collision $(E, E') \gg mc^2$, so that we can safely ignore the mass of the electron (set $m = 0$);† then $p_1 = \frac{E}{c}(1, \hat{\mathbf{p}}_i)$ and $p_3 = \frac{E'}{c}(1, \hat{\mathbf{p}}_f)$, with $\hat{\mathbf{p}}_i \cdot \hat{\mathbf{p}}_f = \cos\theta$, and we find (Problem 8.8)

$$\langle|\mathcal{M}|^2\rangle = \frac{g_e^4 c^2}{4EE' \sin^4(\theta/2)}\left(2K_1 \sin^2\frac{\theta}{2} + K_2 \cos^2\frac{\theta}{2}\right) \tag{8.20}$$

The outgoing electron energy, E', is not an independent variable; it is kinematically determined by E and θ (Problem 8.9):

$$E' = \frac{E}{1 + (2E/Mc^2)\sin^2(\theta/2)} \tag{8.21}$$

* The subscript 3 is traditionally reserved for a term that enters in the corresponding analysis of *neutrino*-proton scattering, but does not occur here.

† The Mott formula (7.127) neglects proton structure and proton recoil; it applies to the regime $E \ll Mc^2$, but it does not assume $E \gg mc^2$. We now work in the regime $E \gg mc^2$, but do *not* ignore proton structure and recoil (i.e., we do *not* assume $E \ll Mc^2$). In the intermediate range, $mc^2 \ll E \ll Mc^2$, the two results agree. (See Problem 8.10.)

For a massless incident particle we have (Problem 6.10)

$$\frac{d\sigma}{d\Omega} = \left(\frac{\hbar E'}{8\pi McE}\right)^2 \langle |\mathcal{M}|^2 \rangle \tag{8.22}$$

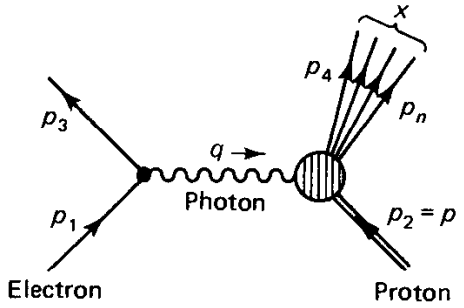
and so, for elastic electron-proton scattering

$$\frac{d\sigma}{d\Omega} = \left(\frac{\alpha\hbar}{4ME \sin^2(\theta/2)}\right)^2 \frac{E'}{E} [2K_1 \sin^2(\theta/2) + K_2 \cos^2(\theta/2)] \tag{8.23}$$

where E' is given by equation (8.21). This is known as the *Rosenbluth formula;* it was first derived in 1950.[1] By counting the number of electrons scattered in a given direction, for a range of incident energies, we can determine the "form factors" $K_1(q^2)$ and $K_2(q^2)$, and compare the results with theoretical predictions[2] (see Fig. 8.4).

8.4 INELASTIC ELECTRON-PROTON SCATTERING

At modest energies, electron-proton scattering is necessarily elastic ($e + p \rightarrow e + p$); the proton recoils, but it's still just a proton. But if the incident electron carries enough energy, all sorts of other stuff may come out—pions, kaons, deltas, you name it. We describe such an inelastic process ($e + p \rightarrow e + X$) by a diagram of the form



Again, the blob veils our ignorance about the photon-proton vertex; the extra outgoing lines represent the hadronic "shrapnel" from the exploded proton. As before, the *electron* vertex is unaffected by all the excitement at the proton end, and so the (spin-averaged) amplitude, for a given final state X, takes the form

$$\langle |\mathcal{M}|^2 \rangle = \frac{g_e^4}{q^4} L^{\mu\nu}_{\text{electron}} K_{\mu\nu}(X) \tag{8.24}$$

where $K_{\mu\nu}$ is some (unknown) quantity describing the subprocess $\gamma + p \rightarrow X$; it depends on $q \equiv (p_1 - p_3)$, $p \equiv p_2$, and the various outgoing momenta p_4, p_5, . . . , p_n. The scattering cross section is determined by the "Golden Rule" [eq. (6.34)]:

$$d\sigma = \frac{\hbar^2 \langle |\mathcal{M}|^2 \rangle}{4\sqrt{(p_1 \cdot p_2)^2 - (m_1 m_2 c^2)^2}} \left(\frac{c\, d^3\mathbf{p}_3}{(2\pi)^3 2E_3}\right)\left(\frac{c\, d^3\mathbf{p}_4}{(2\pi)^3 2E_4}\right) \cdots \left(\frac{c\, d^3\mathbf{p}_n}{(2\pi)^3 2E_n}\right)$$
$$\times (2\pi)^4\, \delta^4(p_1 + p_2 - p_3 - p_4 - \cdots - p_n) \tag{8.25}$$

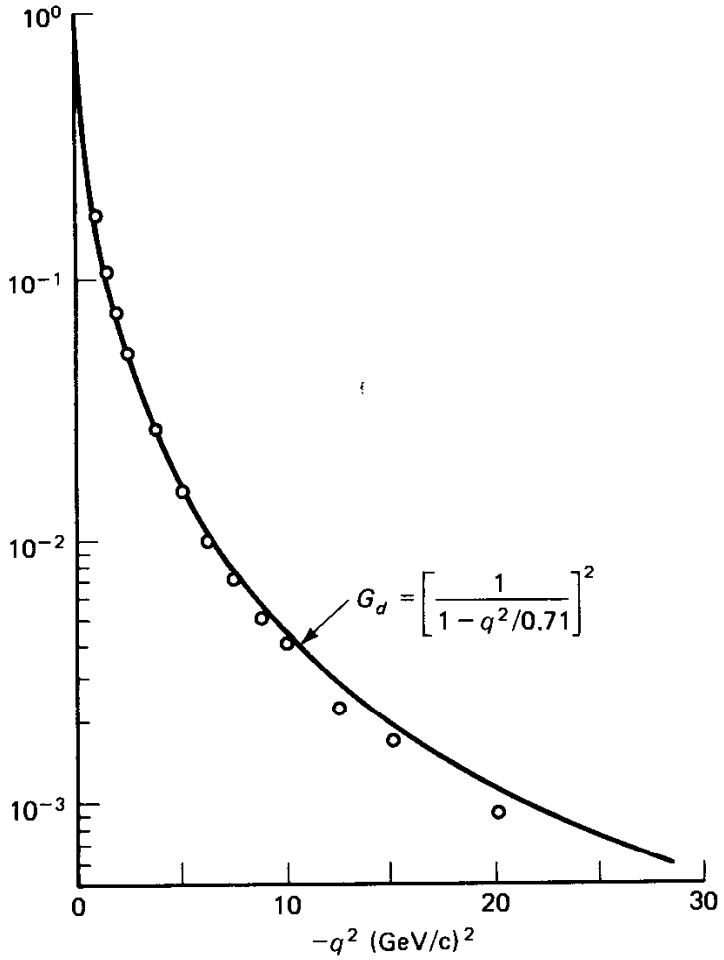


Figure 8.4 Proton elastic form factors. Apart from an overall constant $(1 + K)$, the electric and magnetic form factors G_E and G_M (ref. 2) are practically identical, and [at least, up to about 10 $(\text{GeV}/c)^2$] are well fit by the phenomenological "dipole" function G_d (solid line). Circles are experimental values of $G_M/(1 + K)(\approx G_E)$. [*Source:* H. Frauenfelder and E. M. Henley, *Subatomic Physics* (Englewood Cliffs, NJ: Prentice-Hall, copyright © 1974), p. 127. Based on data of P. N. Kirk et al., *Phys. Rev.* **D8,** 63 (1973).]

However, in a typical experiment only the momentum of the scattered electron ($\mathbf{p}_3$) is recorded. What we measure is the so-called "inclusive" cross section, in which *all* accessible final states X, and all possible outgoing momenta, are included. To obtain the inclusive cross section from equation (8.25), we sum over X and integrate over $\mathbf{p}_4, \mathbf{p}_5, \ldots, \mathbf{p}_n$:

$$d\sigma = \frac{\hbar^2 g_e^4 L^{\mu\nu}}{4q^4\sqrt{(p_1 \cdot p_2)^2 - (m_1 m_2 c^2)^2}} \left(\frac{c\, d^3\mathbf{p}_3}{(2\pi)^3 2E_3}\right) 4\pi M W_{\mu\nu} \tag{8.26}$$

where

$$W_{\mu\nu} \equiv \frac{1}{4\pi M} \sum_X \int \cdots \int K_{\mu\nu}(X) \left(\frac{c\, d^3\mathbf{p}_4}{(2\pi)^3 2E_4}\right) \cdots \left(\frac{c\, d^3\mathbf{p}_n}{(2\pi)^3 2E_n}\right)$$

$$\times (2\pi)^4\, \delta^4(q + p - p_4 - \cdots - p_n) \tag{8.27}$$

For a massless electron of energy E striking a stationary proton of mass M, the square root is just ME. Meanwhile, $d^3\mathbf{p}_3 = |\mathbf{p}_3|^2\, d|\mathbf{p}_3| d\Omega$, and (again setting $m = 0$) $|\mathbf{p}_3| = E'/c$, where $E'(=E_3)$ is the outgoing electron energy. Thus

$$\frac{d\sigma}{dE'\,d\Omega} = \left(\frac{\alpha\hbar}{cq^2}\right)^2 \frac{E'}{E} L^{\mu\nu} W_{\mu\nu} \tag{8.28}$$

Notice that E' is *not* kinematically determined by E and θ, in the inelastic case, for the outgoing hadrons can soak up a range of energies. [More precisely, the total hadronic momentum, $p_{\text{tot}} = p_4 + p_5 + \cdots + p_n$, is no longer constrained by the condition $p_{\text{tot}}^2 = M^2c^2$, as it would be for a single proton. Therefore $p_3 = p_1 + p_2 - p_{\text{tot}}$ is also unconstrained, and equation (8.21) no longer follows.] Accordingly, what we are interested in is the differential cross section for scattering *in a particular energy range dE'*, and this is what equation (8.28) gives us.

From here on the story is familiar: $W_{\mu\nu}$ is a second-rank tensor, which can only depend on q and p, all other momenta having been integrated out. As before, it can be written in the form

$$W^{\mu\nu} = -W_1 g^{\mu\nu} + \frac{W_2}{(Mc)^2} p^\mu p^\nu + \frac{W_4}{(Mc)^2} q^\mu q^\nu + \frac{W_5}{(Mc)^2}(p^\mu q^\nu + p^\nu q^\mu) \tag{8.29}$$

However, this time the W_i are functions of *two* independent scalars, q^2 and $q \cdot p$ (because p_{tot}^2 is no longer constrained, $q \cdot p$ is no longer related to q^2). Once again, we have (Problem 8.4)

$$q_\mu W^{\mu\nu} = 0 \tag{8.30}$$

from which it follows that (Problem 8.11)

$$W_4 = \frac{(Mc)^2}{q^2} W_1 + \left(\frac{q \cdot p}{q^2}\right)^2 W_2 \quad \text{and} \quad W_5 = -\frac{(q \cdot p)}{q^2} W_2 \tag{8.31}$$

and therefore $W^{\mu\nu}$ can be expressed in terms of just *two* "structure functions":

$$W^{\mu\nu} = W_1\left(-g^{\mu\nu} + \frac{q^\mu q^\nu}{q^2}\right) + \frac{W_2}{(Mc)^2}\left[p^\mu - \left(\frac{q \cdot p}{q^2}\right)q^\mu\right]\left[p^\nu - \left(\frac{q \cdot p}{q^2}\right)q^\nu\right] \tag{8.32}$$

Putting equations (8.13) and (8.32) into equation (8.28) (and again setting $m = 0$) we conclude

$$\frac{d\sigma}{dE'\,d\Omega} = \left(\frac{\alpha\hbar}{2E \sin^2(\theta/2)}\right)^2 [2W_1 \sin^2(\theta/2) + W_2 \cos^2(\theta/2)] \tag{8.33}$$

Equation (8.33) is the fundamental result for inclusive inelastic electron-proton scattering; it is the analog to the Rosenbluth formula (8.23)

$$\frac{d\sigma}{d\Omega} = \left(\frac{\alpha\hbar}{4ME \sin^2(\theta/2)}\right)^2 \frac{E'}{E} [2K_1 \sin^2(\theta/2) + K_2 \cos^2(\theta/2)] \tag{8.23}$$

which describes *elastic* scattering. Remember that the structure functions (W_1 and W_2) depend on *two* independent variables, for a given incident energy (E). The experimentalist would use E' and θ, whereas the theorist would generally prefer the Lorentz-invariant quantities q^2 and $q \cdot p$—or better (for reasons you'll see in the next section)—q^2 and

$$x \equiv -\frac{q^2}{2q \cdot p} \tag{8.34}$$

By contrast, the *elastic* form factors (K_1 and K_2) depend on only *one* variable (θ, for the experimentalist, q^2 for the theorist)—in this case E' is determined by equation (8.21) and x is fixed ($x = 1$). In a formal sense, elastic scattering is a special case of inelastic scattering, in which we impose an extra constraint ($p_{\text{tot}}^2 = M^2c^2$) on the outgoing hadron momenta. It should be possible, therefore, to obtain the Rosenbluth formula from equation (8.33), by appropriate choice of the W's. You can check for yourself that

$$W_{1,2}(q^2, x) = -\frac{K_{1,2}(q^2)}{2Mq^2}\,\delta(x - 1) \tag{8.35}$$

does the job (see Problem 8.12).

I should point out that there is precious little *physics* in all of this; what we have done is to *set the agenda* for a theory of proton structure. A successful theory must enable us to *calculate* the structure functions and form factors, which at this stage are completely arbitrary. The most naïve model treats the proton as a simple point charge; in this case (Problem 8.6)

$$K_1 = -q^2 \quad \text{and} \quad K_2 = (2Mc)^2 \tag{8.36}$$

It's not a bad model at low energies, where only elastic scattering occurs, and the electron never gets close enough to "see" the structure of the proton. But it is grossly inadequate at high energies (see Fig. 8.5). In the next two sections we shall see what the quark model has to say about the high-energy régime.

8.5 THE PARTON MODEL AND BJORKEN SCALING

In the late sixties, Bjorken predicted that at very high energy the dependence of the inelastic structure functions on q^2 fades away, and they become functions of x alone. More precisely, he suggested that[3]

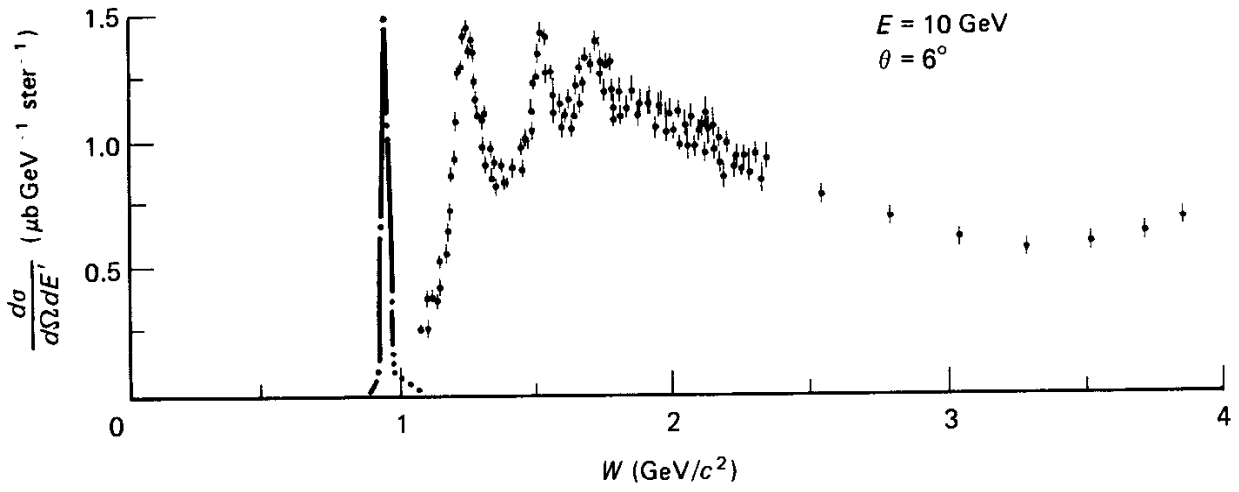


Figure 8.5 Cross section for inclusive inelastic electron-proton scattering, as a function of "missing mass" ($W = \sqrt{p_{\text{tot}}^2}/c$). The elastic peak at $W = M$ has been reduced by factor 8.5, to fit it on the graph. (*Source:* J. I. Friedman and H. W. Kendall; reproduced, with permission, from the Annual Review of Nuclear and Particle Science, Volume 22, © 1972 by Annual Reviews Inc.; page 210.)

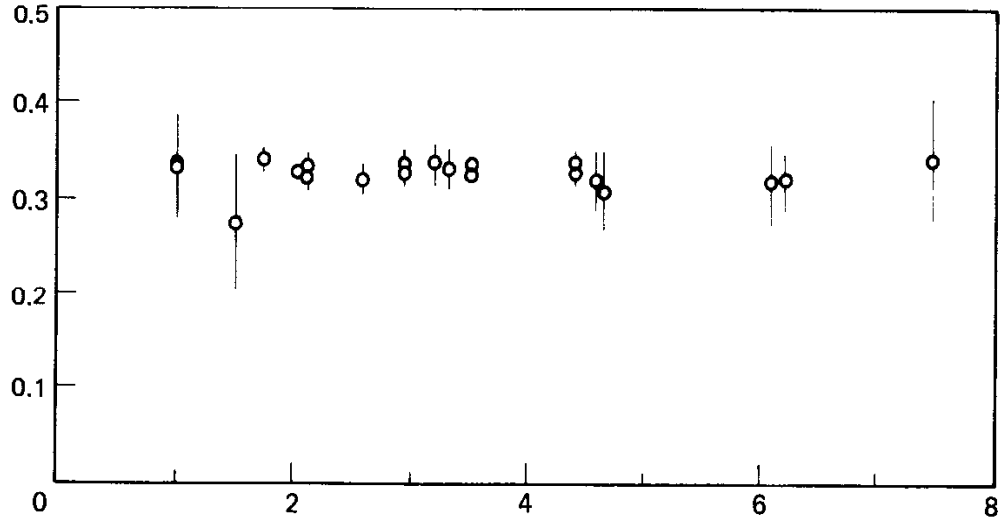


Figure 8.6 Scaling behavior of the structure function W_2 in deep inelastic scattering. Here the quantity $-(q^2/2Mc^2x)W_2(q^2, x)$ is plotted against $-q^2$ (in GeV/c)^2, for $x = 0.25$. (*Source:* J. I. Friedman and H. W. Kendall; reproduced, with permission, from the Annual Review of Nuclear and Particle Science, Volume 22, © 1972 by Annual Reviews Inc.: page 227.)

$$MW_1(q^2, x) \to F_1(x) \tag{8.37}$$

$$\frac{-q^2}{2Mc^2x} W_2(q^2, x) \to F_2(x) \tag{8.38}$$

in the so-called "deep inelastic scattering" régime, where $-q^2[=(4EE'/c^2) \times \sin^2(\theta/2)]$ and $q \cdot p[=M(E - E')]$ are both large, but their *ratio* $(2x = -q^2/q \cdot p)$ is *not.** This behavior is known as "scaling," and it was dramatically confirmed by experiments at SLAC in the early seventies (Fig. 8.6). As we shall see in a moment, scaling is a consequence of the fact that the proton is made of pointlike constituents ("partons"—hideous term—they were called in those days, when it was unfashionable to take quarks and gluons too seriously). In 1969, Callan and Gross[4] suggested that Bjorken's scaling functions are related:

$$2xF_1(x) = F_2(x) \tag{8.39}$$

This, too, has been confirmed experimentally (Fig. 8.7). The Callan–Gross relation reflects the fact that the charged constituents of the proton carry spin $\frac{1}{2}$ (for spin 0 one predicts $2xF_1/F_2 = 0$, instead of 1, and this is clearly inconsistent with the data). The experimental verification of Bjorken scaling and the Callan–Gross relation in deep inelastic scattering provided the first compelling evidence for the existence of quarks.

There are several ways to derive equations (8.37), (8.38), and (8.39), but the crucial point is that *at high energies the virtual photon interacts with a single essentially free quark.*† We can treat this scattering, therefore, using the old

* Experimentally, Bjorken scaling sets in for $-q^2 \gtrsim 1$ (GeV/c)^2 and $(q \cdot p) \gtrsim 3.5$ (GeV/c)^2. Notice that x is restricted to the range $0 \leq x \leq 1$ (see Problem 8.13).

† At low energies (long wavelengths) the photon "sees" the whole proton as a simple point charge—that gives Mott scattering. At high energies (short wavelengths) the photon "sees" a single quark—this gives Bjorken scaling and the Callan–Gross relation. At intermediate energies the photon "sees" the proton in all its complex structure—this will clearly be the hardest case to analyse.

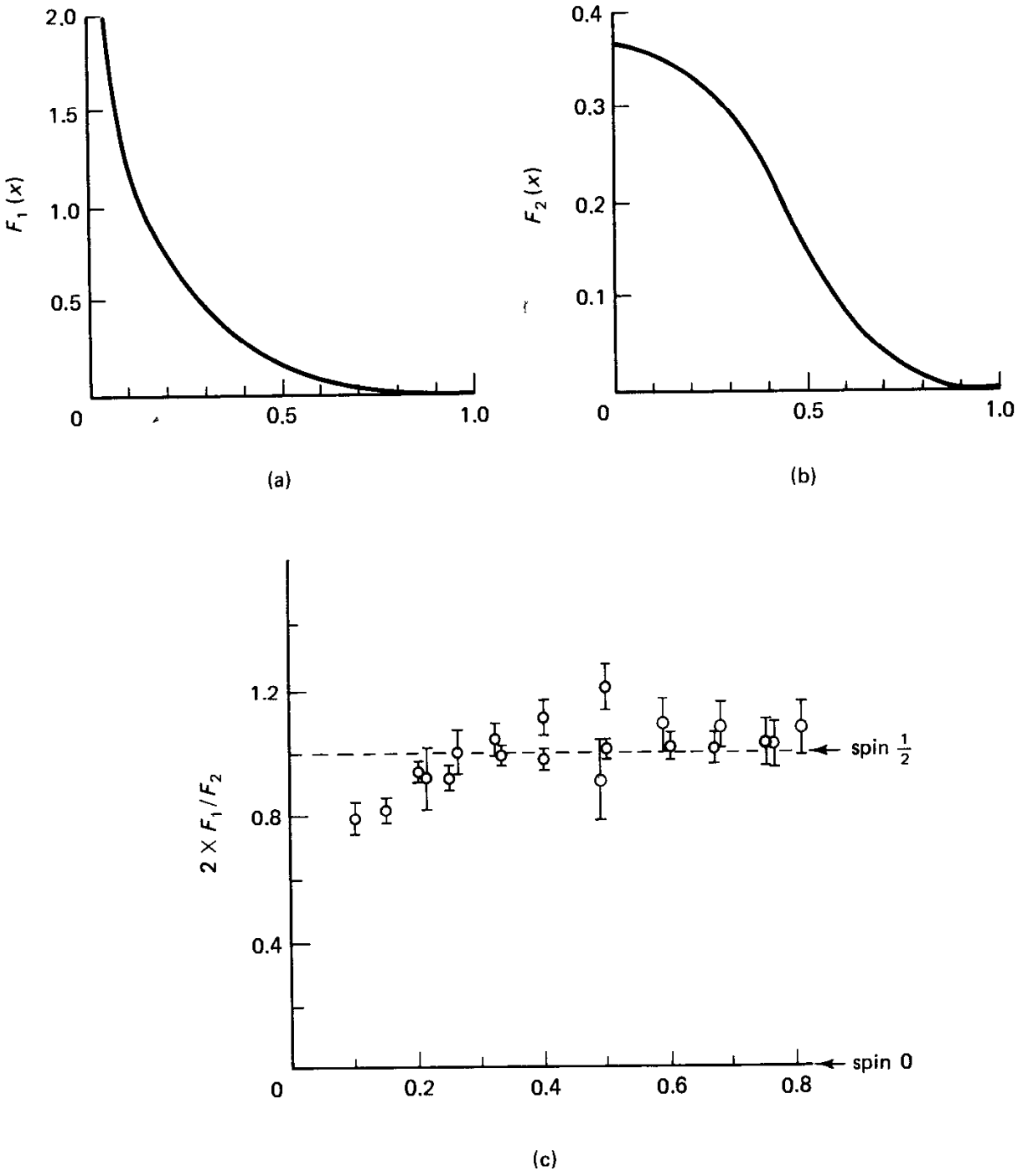


Figure 8.7 Scaling functions and the Callan–Gross relation. In (a) and (b) I plot the experimental measurements of $F_1(x)$ and $F_2(x)$. In (c) the ratio $2xF_1/F_2$ is plotted against x, as a test of the Callan–Gross relation, which evidently holds well for $x \gtrsim 0.2$. [Data from A. Bodek et al., *Phys. Rev.* **D20**, 1471 (1979).]

electron-muon results. According to equations (8.35) and (8.36), the structure functions for scattering off a quark of flavor i are

$$W_1^i = \frac{Q_i^2}{2m_i}\,\delta(x_i - 1), \qquad W_2^i = -\frac{2m_i c^2 Q_i^2}{q^2}\,\delta(x_i - 1) \tag{8.40}$$

Here m_i is the mass of the quark and

$$x_i = -\frac{q^2}{2q \cdot p_i} \tag{8.41}$$

where p_i is the quark's momentum. As before, Q_i is the quark charge, in units of e ($\frac{2}{3}$ for the u, $-\frac{1}{3}$ for d and s); we include a factor of Q_i^2 in W^i so that we don't have to touch the rest of the cross-section formula (8.33). The trouble is, although the *proton* momentum is p, we do not know the momenta of the individual quarks. Let us suppose that z_i is the *fraction* of the total momentum carried by quark i, so that

$$p_i = z_i p \tag{8.42}$$

[This may sound reasonable, but it is actually a pretty slippery proposition, for it assumes that each component p_i^μ gets the *same* fraction of p^μ. There is no room here for the quarks to move around within the proton; if the proton is at rest, so too are all the quarks. In particular, it follows (since $p_i^2 = m_i^2 c^2$ and $p^2 = M^2 c^2$) that

$$m_i = z_i M \tag{8.43}$$

which implies that if z_i is variable, so too is the quark mass!*] Equation (8.42) implies that

$$x_i = \frac{x}{z_i} \tag{8.44}$$

and hence

$$W_1^i = \frac{Q_i^2}{2M}\,\delta(x - z_i), \qquad W_2^i = -\frac{2x^2 Mc^2}{q^2}\,Q_i^2\,\delta(x - z_i) \tag{8.45}$$

Finally, let $f_i(z_i)$ be the *probability* that the ith quark carries momentum fraction z_i. Integrating over z_i, and summing over all the quarks in the proton, we conclude that

$$W_1 = \sum_i \int_0^1 \frac{Q_i^2}{2M}\,\delta(x - z_i) f_i(z_i)\,dz_i = \frac{1}{2M}\sum_i Q_i^2 f_i(x) \tag{8.46}$$

$$W_2 = \sum_i \int_0^1 \left(-\frac{2x^2 Mc^2}{q^2}\right) Q_i^2\,\delta(x - z_i) f_i(z_i)\,dz_i = -\frac{2Mc^2}{q^2}\,x^2 \sum_i Q_i^2 f_i(x) \tag{8.47}$$

Thus

$$MW_1 = \frac{1}{2}\sum_i Q_i^2 f_i(x) \equiv F_1(x) \tag{8.48}$$

$$-\frac{q^2}{2Mc^2 x}\,W_2 = x\sum_i Q_i^2 f_i(x) \equiv F_2(x) \tag{8.49}$$

confirming the Bjorken scaling law. Comparing the two expressions, we obtain

$$F_2(x) = 2xF_1(x) \tag{8.50}$$

which is the Callan–Gross relation.

The importance of Bjorken scaling and the Callan–Gross formula lies in the fact that they provide for a clear experimental test of the quark-parton model,

* More rigorous derivations of the scaling equations avoid this problem by working in the "infinite momentum frame," in which the proton energy is so much greater than its mass that the proton and the quarks can be treated as massless particles, and (8.43) is trivially satisfied.

a test that was passed with flying colors in SLAC's deep inelastic scattering experiments. Protons *really do* have charged constituents, and those constituents *really do* behave as pointlike Dirac particles, carrying spin $\frac{1}{2}$. One modest dividend is a radical simplification in the analysis of high-energy electron-proton scattering, for in place of two unknown functions of two variables [$W_1(q^2, x)$ and $W_2(q^2, x)$] we now have only *one* unknown function of *one* variable to contend with [$F_1(x)$]. If we put equations (8.37), (8.38), and (8.39) back into (8.33), we find

$$\frac{d\sigma}{dE'\,d\Omega} = \frac{F_1(x)}{2M}\left(\frac{\alpha\hbar}{E\sin(\theta/2)}\right)^2\left[1 + \frac{2EE'}{(E-E')^2}\cos^2\frac{\theta}{2}\right] \tag{8.51}$$

Moreover, equation (8.48) shows that to finish the job we need to know the probability functions $f_i(x)$, for then

$$F_1(x) = \tfrac{1}{2}\sum_i Q_i^2 f_i(x) \tag{8.52}$$

We consider this problem in the following section.

8.6 QUARK DISTRIBUTION FUNCTIONS

If we take equation (8.43) at face value, the momentum fraction carried by the ith quark is proportional to its mass, and the probability density f_i must therefore be a delta function:*

$$f_i(z_i) = \delta\left(\frac{m_i}{M} - z_i\right) \tag{8.53}$$

If, moreover, the proton consists simply of two up quarks and a down quark, then equation (8.48) says

$$F_1(x) = \frac{1}{2}\left\{\left(\frac{2}{3}\right)^2\delta\left(\frac{m_u}{M} - x\right) + \left(\frac{2}{3}\right)^2\delta\left(\frac{m_u}{M} - x\right) + \left(\frac{-1}{3}\right)^2\delta\left(\frac{m_d}{M} - x\right)\right\} \tag{8.54}$$

and if $m_u = m_d$, we obtain the simple result†

$$F_1(x) = \frac{1}{2}\,\delta\left(\frac{m_u}{M} - x\right), \qquad F_2(x) = x\,\delta\left(\frac{m_u}{M} - x\right) \tag{8.55}$$

In this model the cross section reduces to the electron-muon form, with the quark mass in place of the muon mass. The proton as such has disappeared from the problem; we simply have elastic scattering of electrons from free quarks.

What's *wrong* with this naïve picture (which, as you can see from Figure 8.7, is totally incompatible with the experimental data)? Basically, we have taken too literally the idea that the quarks inside a proton are *free*. It is true that they

* I assume that $\int_0^1 f_i(x)dx = 1$, since this is the total probability of finding quark i with *some* fraction of the proton momentum.

† Consistency requires that $m_u = \frac{1}{3}M$, for if these are the only constituents of the proton, we must have $\Sigma_i \int_0^1 x f_i(x)dx = 1$. However, this naive picture ignores the contribution of gluons and "sea" quarks, as we shall see momentarily.

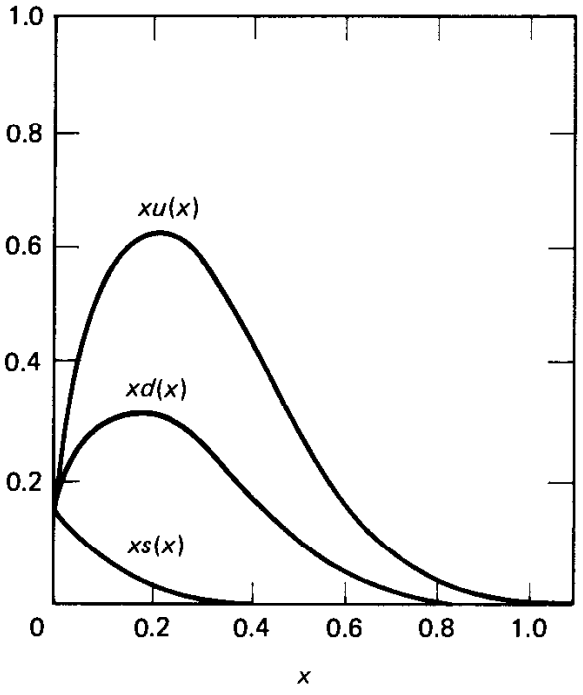


Figure 8.8 Quark structure functions. [*Source:* F. Halzen and A. D. Martin, *Quarks and Leptons* (New York: Wiley, copyright © 1984), p. 203. Reprinted by permission of John Wiley & Sons, Inc.]

behave as free particles in their interaction with the virtual photon, but on a longer time scale they are obviously *not* free; after all, they are bound together by the confining force of QCD. In particular, the "mass" of a quark within a hadron is not a very well-defined notion (as we have discovered on several previous occasions—see Chap. 4, Sect. 4.5 and Chap. 5, Sects. 5.8 and 5.10). Since they are continually interacting with one another, the quarks are really *virtual* particles, and do not lie on their mass shells. In this sense it is appropriate after all to regard m_i as a continuous variable, in equation (8.43), and the delta function in (8.53) is smeared out accordingly. Exactly what *shape* it assumes is a problem for quantum *chromodynamics,* which we are not presently in a position to address.

Moreover, it is an oversimplification to say that the proton consists only of quarks. Let's say that $u(x)$ is the probability (density) that momentum fraction x is carried by a u quark, and $d(x)$ is the corresponding probability for a d quark,* so that

$$F_2(x) = x\{(\tfrac{2}{3})^2 u(x) + (\tfrac{1}{3})^2\, d(x)\} \tag{8.56}$$

[In the naïve model $u(x) = 2\delta(m_u/M - x)$ and $d(x) = \delta(m_d/M - x)$.] One is tempted to guess that $u(x) = 2d(x)$, but this assumption is not sustained by the data (see Fig. 8.8), at least, not near $x = 0$ and $x = 1$. (A quark carrying 90% of the proton's momentum is almost certain to be a u, whereas at the 10% level it is only slightly more likely to be a u.) Nevertheless, the *average* momentum carried by up quarks ($\int_0^1 pxu(x)dx$) is surely twice the *average* carried by the down quark, since there are twice as many *of* them, and they all weigh about the same:

$$\int_0^1 xu(x)dx = 2\int_0^1 x\, d(x)dx \tag{8.57}$$

* More precisely, if you had a large sample of protons, $u(x)dx$ is the average number of up quarks (per proton) carrying a momentum fraction between x and $(x + dx)$.

[And this is confirmed by data on electron-*neutron* scattering (see Problem 8.15).] Combining equations (8.56) and (8.57), we find

$$\int_0^1 x\, d(x)dx = \int_0^1 F_2(x)dx \tag{8.58}$$

If you measure the area under the experimental curve (Fig. 8.7b), you'll find that this integral is about 0.18, and hence

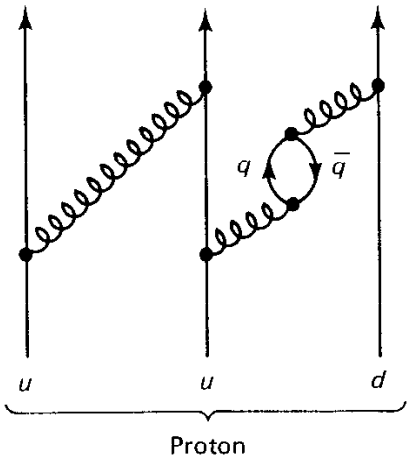
$$\int_0^1 x\, d(x)dx = 0.18, \qquad \int_0^1 xu(x)dx = 0.36 \tag{8.59}$$

These numbers have a remarkable implication: the average *total* momentum carried by the quarks is

$$\int_0^1 xpu(x)dx + \int_0^1 xpd(x)dx = p(0.18 + 0.36) = 0.54p \tag{8.60}$$

On the average, then, only 54% of the proton's momentum is accounted for by the quarks. Who's got the rest of it? *Answer:* the *gluons.* They're uncharged, so they don't contribute to electron-proton scattering, but they *do* carry a share of the proton's momentum. The structure functions allow us to determine the momentum carried by *charged* partons; whatever is left over must be ascribed to *uncharged* partons. Thus, in an indirect way, the deep inelastic scattering experiments provide substantial evidence for the existence of gluons, as well as quarks.

Finally, even the *quark* content of the proton is more complicated than I have suggested. For the gluons can produce quark-antiquark pairs:



At any given moment the proton might actually contain an extra $u\bar{u}$ pair, or $d\bar{d}$, or $s\bar{s}$, or even several such pairs. (In principle, it could even have a heavy quark pair—$c\bar{c}$, $b\bar{b}$, $t\bar{t}$—but this is far less likely, because of the large mass term in the denominator of the quark propagator.) We call the "original" quarks (u, u, d, for the proton) "valence" quarks, and the "extra" ones "sea" quarks. It is possible for the virtual photon, in electron-proton scattering, to couple to

one of the sea quarks, so they should really be included in our formulas for F_1 and F_2:

$$F_1(x) = \tfrac{1}{2}\{(\tfrac{2}{3})^2[u(x) + \bar{u}(x)] + (\tfrac{1}{3})^2[d(x) + \bar{d}(x) + s(x) + \bar{s}(x)]\} \quad (8.61)$$

Not much progress here: we seem to have traded *one* unknown function for *six* unknown functions! Fortunately, the same "quark structure functions"* occur (with different coefficients) in electron-*neutron* scattering, and in *neutrino*-proton scattering, so we have a certain amount of independent experimental information about them. In addition, they are constrained by several "sum rules" (see Problem 8.16). Because the sea quarks are all produced by the same mechanism, and carry roughly the same mass, it is reasonable to assume that

$$\bar{u}(x) \cong \bar{d}(x) \cong \bar{s}(x) \cong s(x) \quad (8.62)$$

while the up and down quark distributions can be separated into a valence part and a sea part [the latter presumably again equal to $s(x)$]

$$u(x) = u_v(x) + s(x), \qquad d(x) = d_v(x) + s(x) \quad (8.63)$$

This reduces the problem down to *three* unknown functions:

$$F_1(x) = \tfrac{1}{18}\{4u_v(x) + d_v(x) + 12s(x)\} \quad (8.64)$$

The shape of these functions, as inferred from experiments, is indicated in Figure 8.8. Notice that the sea contributes only at relatively low x (that is, sea quarks typically carry only a small fraction of the proton's momentum—that's why I could safely ignore them in calculating the gluon contribution). Now, the number of valence u quarks is certainly 2, and for d quarks, 1, so

$$\int_0^1 u_v(x)dx = 2, \qquad \int_0^1 d_v(x)dx = 1 \quad (8.65)$$

It follows that

$$\int_0^1 s(x)dx = \tfrac{3}{4}\left(2\int_0^1 F_1(x)dx - 1\right) \quad (8.66)$$

and therefore, in principle, we can determine the average number of sea quarks of each species, by measuring the area under the $F_1(x)$ graph (Fig. 8.7a). Unfortunately, the area is extremely sensitive to the behavior of the function as $x \to 0$, and depending on how you extrapolate the experimental curve, it is possible to produce any number between 0.5 and infinity. (As a matter of fact, some theories predict that the proton contains enormous numbers of very low-energy sea quarks; hence, I suppose, the name "sea").

* In this business *everything* is called a "structure function": W_1, W_2, F_1, and F_2 are "proton structure functions"; the f_i's and $u(x)$, $\bar{u}(x)$, $d(x)$, $\bar{d}(x)$, $s(x)$, $\bar{s}(x)$, $u_v(x)$, and $d_v(x)$ are "quark structure functions." I prefer the term "distribution functions" for the latter.

REFERENCES AND NOTES

1. M. N. Rosenbluth, Phys. Rev. **79,** 615 (1950).
2. See, for example, H. Frauenfelder and E. M. Henley, *Subatomic Physics,* (Englewood Cliffs, N.J.: Prentice-Hall, 1974), Chapter 6. I should warn you that this subject is a notational nightmare in the literature. There are only two variables in the problem—the incident electron energy (E) and the scattering angle (θ)—but it is common to encounter a more or less random mixture of E, E', θ, q^2, $Q^2 \equiv -q^2$, $\tau \equiv -q^2/4M^2c^2$, $\nu \equiv p\cdot q/Mc$, $\omega \equiv -2p\cdot q/q^2$, $W \equiv \sqrt{(q_i + p)^2}$, $x \equiv -q^2/2p\cdot q$, and $y \equiv p\cdot q/p\cdot p_1$. Moreover, although there are only two independent form factors involved, there are many different ways to express them. Some authors favor F_1 and F_2, with $K_1 = -q^2(F_1 + KF_2)^2$ and $K_2 = (2Mc)^2F_1^2 - K^2q^2F_2^2$ ($K = 1.7928$ is the "anomalous" contribution to the proton's magnetic moment); others prefer $G_E = F_1 - K\tau F_2$, $G_M = F_1 + KF_2$. (The latter are related to the Fourier transforms of the charge and magnetic moment distributions, respectively; see F. Halzen and A. D. Martin, *Quarks and Leptons* (New York: Wiley, 1984), Sect. 8.2.) Anyone can play this game; K_1, and K_2 are my own contributions.
3. J. D. Bjorken, *Phys. Rev.* **179,** 1547 (1969); *Phys. Rev.* **163,** 1767 (1967). (Bjorken's functions F_1 and F_2 have no relation to the elastic form factors F_1 and F_2 of ref. 2. Sorry about that.)
4. C. G. Callan and D. Gross, *Phys. Rev. Lett.* **22,** 156 (1969).

PROBLEMS

8.1. **(a)** Derive equation (8.1), from the Feynman rules for QED.
(b) Obtain equation (8.2) from equation (8.1).
(c) Derive equation (8.3) from equation (8.2).
(d) Derive equation (8.4) from equation (8.3).

8.2. Derive equation (8.5), starting with equation (8.4).

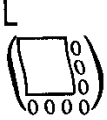
8.3. Why don't we use $\sigma(e^+e^- \to e^+e^-)$ in the denominator, to define R [eq. (8.7)]? [*Answer:* For one thing, we would have to include the crossed diagram



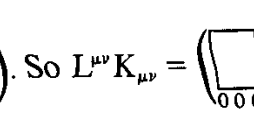
and the kinematic factors would no longer cancel.]

8.4. Prove equation (8.16). [*Hint:* First show that $q_\mu L^{\mu\nu} = 0$. Then argue that we *may as well* take $K^{\mu\nu}$ such that $q_\mu K^{\mu\nu} = 0$, in the sense that any term in $K^{\mu\nu}$ that does *not* obey $q_\mu K^{\mu\nu} = 0$ will contribute nothing to $L^{\mu\nu}K_{\mu\nu}$.] *Comment:* Equation (8.16) actually follows more simply and generally from charge conservation at the proton vertex, but I have not developed the formalism here to make this argument (see Halzen and Martin, ref. 2, Sects. 8.2 and 8.3).

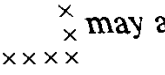
[One way to proceed is as follows. Take $q^\mu = (0, 0, 0, q)$; then $q_\mu L^{\mu\nu} = 0 \Rightarrow L^{\mu\nu} =$



. So $L^{\mu\nu}K_{\mu\nu} =$



, and



may as well be zero.]

8.5. Prove equation (8.17), from equation (8.16). [*Hint:* First contract with q_ν, then with p_ν.]

8.6. Find K_1 and K_2 for a "Dirac" proton. [*Answer:* compare (8.13) and (8.18): $K_1 = -q^2$, $K_2 = (2Mc)^2$.]

8.7. Derive equation (8.19).

8.8. Derive equation (8.20).

8.9. Derive equation (8.21).

8.10. Check that the Rosenbluth formula (8.23) agrees with the Mott formula (7.128) in the intermediate-energy régime ($mc^2 \ll E \ll Mc^2$). Use the expressions for K_1 and K_2 appropriate to a "Dirac" proton (Problem 8.6).

8.11. Derive equation (8.31). [Note the reduction to equation (8.17) in the elastic case, where $p \cdot q = -q^2/2$.)

8.12. Put equation (8.35) into equation (8.33), and carry out the E' integration (holding θ constant) to recover equation (8.23).

8.13. Show that x [eq. (8.34)] is restricted to the range $0 \leq x \leq 1$. What sort of collision would have $x = 1$? What sort of collision has $x = 0$? [*Answers:* Elastic; $E \gg E'$.]

8.14. Suppose the proton were a "Dirac" charge; would the structure functions scale? If so, what are $F_1(x)$ and $F_2(x)$ in this case? Is the Callan–Gross relation satisfied? [*Answer:* $F_1(x) = \frac{1}{2}\delta(x-1)$; $F_2(x) = \delta(x-1)$; yes.]

8.15. Electron-neutron scattering experiments are harder to do than electron-proton experiments, because you cannot make a target of free neutrons. Nevertheless, the essential data can be inferred from electron-*deuteron* scattering, and it is found that

$$\int_0^1 F_2^{\text{neutron}}\, dx = 0.12$$

Use this, together with the proton result

$$\int_0^1 F_2^{\text{proton}}\, dx = 0.18$$

to confirm equation (8.57). [*Hint:* How do you suppose $u^n(x)$ and $d^n(x)$ are related to the corresponding functions for the proton?]

8.16. From the known flavor content of the proton, find the value of $\int_0^1 [u(x) - \bar{u}(x)]dx$. State corresponding "sum rules" for d and s.

8.17 Are the data in Figures 8.7b and 8.8 compatible with equation (8.56)?

Chapter 9

Quantum Chromodynamics

In this chapter I develop the Feynman rules for quantum chromodynamics, the theory of strong interactions. Some suggestive results in perturbative QCD are derived (in particular, the effective interquark potential in various meson and baryon configurations), and the essential ideas underlying asymptotic freedom are sketched. This material relies heavily on Chapter 7, and also on the last three sections of Chapter 5. It is not necessary as background for Chapters 10 and 11.

9.1 FEYNMAN RULES FOR CHROMODYNAMICS

In the last two chapters we have seen how quantum *electro*dynamics (QED) describes the interactions of *charged* particles; in the present chapter we look at how quantum *chromo*dynamics (QCD) describes the interactions of *colored* particles. Electromagnetic interactions are mediated by *photons,* chromodynamic interactions by *gluons.* The strength of the electromagnetic force is set by the coupling constant

$$g_e = \sqrt{4\pi\alpha} \tag{9.1}$$

In appropriate units g_e is the fundamental charge (the charge of the positron). The strength of the chromodynamic force is set by the "strong" coupling constant

$$g_s = \sqrt{4\pi\alpha_s} \tag{9.2}$$

which may be thought of as the fundamental unit of color. Quarks come in three colors,* "red" (r), "blue" (b), and "green" (g). Thus the specification of a quark

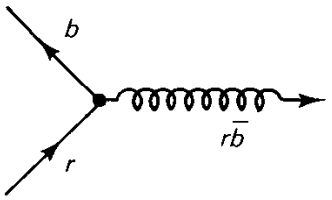
* Quarks also come in different *flavors,* of course, but this is irrelevant in QCD, except insofar as the different quark flavors carry different *masses.* Just as QED only looks at the *charge* of a particle, QCD cares only about its *color.*

state in QCD requires not only the Dirac spinor $u^{(s)}(p)$, giving its momentum and spin, but also a three-element column vector c, giving its color:

$$c = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} \text{ for red,} \qquad \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} \text{ for blue,} \qquad \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} \text{ for green} \tag{9.3}$$

(I'll label the elements of c by a Roman subscript near the middle of the alphabet—c_i, for example—so that $i, j, k, \ldots$ run from 1 to 3 over quark colors.)*

Typically, the quark color changes at a quark-gluon vertex, and the difference is carried off by the gluon. For example:



(In this diagram a red quark turned into a blue quark, emitting a red-antiblue gluon.) Each gluon carries one unit of color and one of anticolor. It would appear, then, that there should be nine species of gluons—$r\bar{r}$, $r\bar{b}$, $r\bar{g}$, $b\bar{r}$, $b\bar{b}$, $b\bar{g}$, $g\bar{r}$, $g\bar{b}$, $g\bar{g}$. Such a nine-gluon theory is perfectly possible in *principle*, but it would describe a world very different from our own. In terms of color $SU(3)$ symmetry (on which, as we shall see, QCD is based), these nine states constitute a "color octet":

$$\left\{ \begin{array}{ll} |1\rangle = (r\bar{b} + b\bar{r})/\sqrt{2} & |5\rangle = -i(r\bar{g} - g\bar{r})/\sqrt{2} \\ |2\rangle = -i(r\bar{b} - b\bar{r})/\sqrt{2} & |6\rangle = (b\bar{g} + g\bar{b})/\sqrt{2} \\ |3\rangle = (r\bar{r} - b\bar{b})/\sqrt{2} & |7\rangle = -i(b\bar{g} - g\bar{b})/\sqrt{2} \\ |4\rangle = (r\bar{g} + g\bar{r})/\sqrt{2} & |8\rangle = (r\bar{r} + b\bar{b} - 2g\bar{g})/\sqrt{6} \end{array} \right\} \tag{9.4}$$

and a "color singlet":

$$|9\rangle = (r\bar{r} + b\bar{b} + g\bar{g})/\sqrt{3} \tag{9.5}$$

(See Section 5.8; there we were concerned with *flavor*, not color, but the *mathematics* is identical—just let $u, d, s \rightarrow r, b, g$. We're not concerned with isotopic spin, here, and I have used different linear combinations of states within the octet. This simplifies the notation later on.) If the singlet gluon existed, it would be as common and conspicuous as the photon.† Confinement requires that all naturally occurring particles be color singlets, and this "explains" why the octet gluons never appear as free particles.‡ But $|9\rangle$ *is* a color singlet, and if it exists

* I should perhaps warn you that most books do not specify quark color states explicitly; they are "implied," or "understood to be contained in $u(p)$." I think it is wiser at this stage to write them out explicitly, even at the cost of some extra notational complexity.

† Maybe the "ninth gluon" *is* the photon! That would make for a beautiful unification of the strong and electromagnetic interactions. Of course, the coupling strength isn't quite right, but that's a problem with all unification schemes, and could presumably be handled. There's a much more serious difficulty with this idea, which I'll let you figure out (see Problem 9.1).

‡ Notice the distinction between "colorless" and "color singlet." Gluons $|3\rangle$ and $|8\rangle$ are colorless. in the sense that the net amount of each color is zero, but they are not color singlets. The situation

as a mediator it should also occur as a free particle. Moreover, it could be exchanged between two color singlets (a proton and a neutron, say), giving rise to a long-range force with strong coupling,* whereas in fact we know that the strong force is of very *short* range. In our world, then, there are evidently only eight kinds of gluons.†

Like the photon, gluons are massless particles of spin 1; they are represented by a polarization vector, ϵ^μ, which is orthogonal to the gluon momentum, p:

$$\epsilon^\mu p_\mu = 0 \qquad \text{(Lorentz condition)} \tag{9.6}$$

As before, we adopt the Coulomb gauge:‡

$$\epsilon^0 = 0, \qquad \text{so that } \boldsymbol{\epsilon} \cdot \mathbf{p} = 0 \tag{9.7}$$

This spoils manifest Lorentz covariance, but it cannot be helped (see Sect. 7.4). To describe the color state of the gluon, we need in addition an eight-element column vector, a:

$$a = \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix} \text{for } |1\rangle, \qquad \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 1 \\ 0 \end{pmatrix} \text{for } |7\rangle, \qquad \text{and so on} \tag{9.8}$$

[Elements of a will be labeled by a Greek superscript near the front of the alphabet (a^α); $\alpha, \beta, \gamma, \ldots$ run from 1 to 8 over gluon color states.] Because the gluons themselves carry color (in contrast to the photon, which is electrically neutral), they couple directly to one another. In fact, there is a three-gluon vertex and a four-gluon vertex:

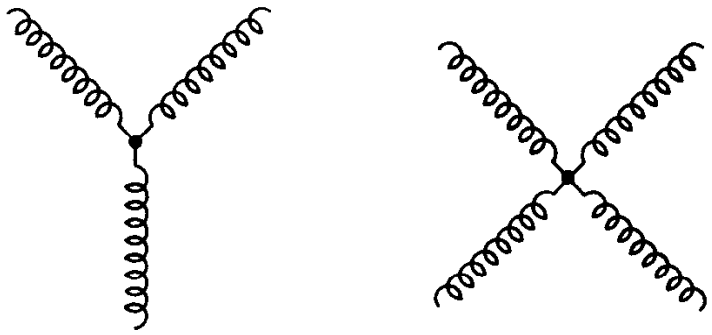
---

is analogous to spin: We can have a state with $S_z = 0$, but this does not prove it has spin 0 (although spin 0 certainly implies $S_z = 0$, and by the same token a color singlet is necessarily colorless). Many authors use the word "colorless" to *mean* "color singlet," but this can lead to misunderstanding. (I was sloppy myself, back in Chapters 1 and 2, because at that stage it was not possible to explain the idea of a color singlet.) You might prefer the word "color-invariant" (instead of "color singlet"), or even "color scalar"; the essential point is that such a state is unaffected by the transformations of color $SU(3)$ (see Problem 9.2).

* Because gluons are massless, they mediate a force of infinite range (the same as electrodynamics). In this sense the force between two quarks is actually *long* range. However, confinement, and the absence of a singlet gluon, conceals this from us. A singlet state (such as the proton) can only emit and absorb a singlet (such as the pion), so individual gluons cannot be exchanged between a proton and a neutron. That's why the force we *observe* is of *short* range. If the singlet gluon existed, it *could* be exchanged between singlets, and the strong force would have a component of infinite range.

† In group theoretical terms, the issue here is whether the symmetry of QCD is $U(3)$ (which would require all nine gluons) or $SU(3)$ (which calls for only eight). The experimental situation resolves the question decisively in favor of the latter.

‡ There is a subtle problem here, because gauge transformations in chromodynamics are more complicated than equation (7.81), and in fact the Coulomb gauge *cannot* be consistently imposed. However, the *correction* to equation (7.81) contains a factor of g_s, and hence, in the Feynman calculus, the "error" introduced by using the Coulomb gauge can be compensated for by appropriate modification of the rules for computing higher-order (loop) diagrams.



Before I can state the Feynman rules for QCD, I need to introduce two items of notation. First, the Gell-Mann "λ-matrices," which are to $SU(3)$ what the Pauli spin matrices are to $SU(2)$:

$$\lambda^1 = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \quad \lambda^2 = \begin{pmatrix} 0 & -i & 0 \\ i & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \quad \lambda^3 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

$$\lambda^4 = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix} \quad \lambda^5 = \begin{pmatrix} 0 & 0 & -i \\ 0 & 0 & 0 \\ i & 0 & 0 \end{pmatrix} \quad \lambda^6 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix} \tag{9.9}$$

$$\lambda^7 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & -i \\ 0 & i & 0 \end{pmatrix} \quad \lambda^8 = \frac{1}{\sqrt{3}} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -2 \end{pmatrix}$$

Second, the *commutators* of the λ matrices define the "structure constants" ($f^{\alpha\beta\gamma}$) of the group $SU(3)$:

$$[\lambda^\alpha, \lambda^\beta] = 2if^{\alpha\beta\gamma}\lambda^\gamma \tag{9.10}$$

(summation over γ—from 1 to 8—implied by the repeated index). The structure constants are completely antisymmetric, $f^{\beta\alpha\gamma} = f^{\alpha\gamma\beta} = -f^{\alpha\beta\gamma}$. You can work them out for yourself (Problem 9.5). Since each index runs from 1 to 8, there are $8 \times 8 \times 8 = 512$ structure constants in all, but most of them are zero, and the rest can be obtained by antisymmetry from the following set:

$$\begin{aligned} &f^{123} = 1, \quad f^{147} = f^{246} = f^{257} = f^{345} = f^{516} = f^{637} = \tfrac{1}{2}, \\ &f^{458} = f^{678} = \sqrt{3}/2 \end{aligned} \tag{9.11}$$

I can now state the Feynman rules for evaluating tree-level diagrams* in QCD:

1. *External Lines.* For an external quark with momentum p, spin s, and color c:

$$\text{Quark} \begin{cases} \text{incoming} \left(\ \nearrow\ \right): u^{(s)}(p)c \\ \text{outgoing} \left(\ \nearrow\ \right): \bar{u}^{(s)}(p)c^\dagger \end{cases} \tag{9.12}$$

(note that $c^\dagger = \tilde{c}^*$ will be a *row* matrix). For an external *anti*quark:

* Loop diagrams in QCD require special rules, including the introduction of so-called "Faddeev–Popov ghosts." These are deep waters, into which we shall not venture here.[1]

$$\text{Antiquark}\begin{cases}\text{incoming } (\ \swarrow\ ): \bar{v}^{(s)}(p)c^{\dagger} \\ \text{outgoing } (\ \nearrow\ ): v^{(s)}(p)c\end{cases} \tag{9.13}$$

where c represents the color of the corresponding *quark*. For an external gluon of momentum p, polarization ϵ, and color a, include a factor

$$\text{Gluon}\begin{cases}\text{incoming } (\ \overset{\alpha,\mu}{\sim}\ ): & \epsilon_{\mu}(p)a^{\alpha} \\ \text{outgoing } (\ \overset{\alpha,\mu}{\sim}\ ): & \epsilon_{\mu}^{*}(p)a^{\alpha *}\end{cases} \tag{9.14}$$

To avoid confusion it is helpful to indicate on the diagram the indices (space–time and color) you are using for each gluon.

2. *Propagators.* Each internal line contributes a factor

$$\text{Quark-antiquark } (\ \underset{q}{\longrightarrow}\ ): \frac{i(\not{q} + mc)}{q^2 - m^2c^2} \tag{9.15}$$

$$\text{Gluon } (\ \overset{\alpha,\mu}{\bullet}\!\!\sim\!\!\underset{q}{\rightarrow}\!\!\sim\!\!\overset{\beta,\nu}{\bullet}\ ): \frac{-ig_{\mu\nu}\,\delta^{\alpha\beta}}{q^2} \tag{9.16}$$

3. *Vertices.* Each vertex introduces a factor

$$\text{Quark-gluon } (\ \underset{\alpha,\mu}{\sim}\!\!\rightarrow\ ): \frac{-ig_s}{2}\lambda^{\alpha}\gamma^{\mu} \tag{9.17}$$

Three gluon (β,ν; k_2; γ,λ; k_3; k_1; α,μ):

$$-g_s f^{\alpha\beta\gamma}[g_{\mu\nu}(k_1 - k_2)_{\lambda} + g_{\nu\lambda}(k_2 - k_3)_{\mu} + g_{\lambda\mu}(k_3 - k_1)_{\nu}] \tag{9.18}$$

Here the gluon momenta (k_1, k_2, k_3) are assumed to point *into* the vertex; if any point *outward* in your diagram, change their signs.

Four gluon (β,ν; γ,λ; α,μ; δ,ρ):

$$-ig_s^2[f^{\alpha\beta\eta}f^{\gamma\delta\eta}(g_{\mu\lambda}g_{\nu\rho} - g_{\mu\rho}g_{\nu\lambda}) + f^{\alpha\delta\eta}f^{\beta\gamma\eta}(g_{\mu\nu}g_{\lambda\rho} - g_{\mu\lambda}g_{\nu\rho}) + f^{\alpha\gamma\eta}f^{\delta\beta\eta}(g_{\mu\rho}g_{\nu\lambda} - g_{\mu\nu}g_{\lambda\rho})] \quad (9.19)$$

(summation over η implied.

Everything else is the same as for QED: Impose conservation of energy and momentum at each vertex to determine the internal four momenta; follow each fermion line "backward" along the arrow, erase the overall delta function, and set the result equal to $-i\mathcal{M}$. In the next two sections I'll work out some examples. to show you how it goes.

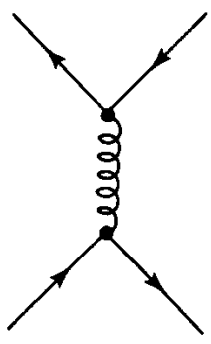
9.2 THE QUARK-QUARK INTERACTION

In this section we consider the interaction between two quarks (also a quark and an antiquark) in lowest-order QCD. Of course, we cannot observe quark-quark scattering directly in the laboratory (although hadron-hadron scattering is an indirect manifestation), so we won't be looking for cross sections here. Instead, we concentrate on the effective *potentials* between quarks—the QCD analog of the Coulomb potential in electrodynamics. We used such potentials, with a promise to derive them later, back in Chapter 5, in the analysis of quarkonium. Bear in mind that this is a perturbation theory calculation, valid only insofar as the coupling α_s is small. We cannot hope to get the confining term in the potential by this route—we are implicitly relying on asymptotic freedom, and all we're going to find is the short-range behavior. Nevertheless, we will obtain a very suggestive result: *Quarks attract one another most strongly when they are in the color singlet configuration* (indeed, in other arrangements they generally repel). At very short range, then, the color singlet is the "maximally attractive channel"—an indication that binding is more likely, at least, for singlet states.*

Case 1: Quark and Antiquark Consider first the interaction of a quark and an antiquark, in QCD. We shall assume that they have different flavors, so the only diagram (in lowest order) is the one in Figure 9.1,† representing, for instance, $u + \bar{d} \to u + \bar{d}$. The amplitude is given by

* This is a very pleasing conclusion, but it does not prove that binding *must* occur in the color singlet, or that it *cannot* occur in other configurations. For this we would have to know the *long*-range behavior of the potential, about which, at present, we can only speculate.

† In *principle*, for the *same* flavor (e.g., $u + u \to u + u$) we should include a second diagram:



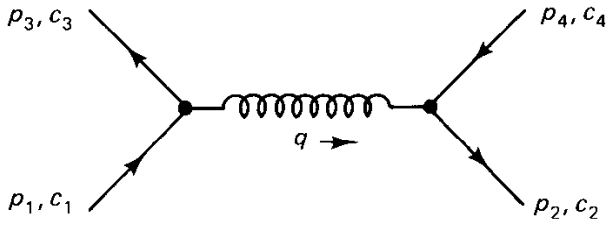


Figure 9.1 The quark-antiquark interaction.

$$-i\mathcal{M} = [\bar{u}(3)c_3^\dagger]\left[-i\frac{g_s}{2}\lambda^\alpha\gamma^\mu\right][u(1)c_1]\left[\frac{-ig_{\mu\nu}\,\delta^{\alpha\beta}}{q^2}\right]$$
$$\times\,[\bar{v}(2)c_2^\dagger]\left[-i\frac{g_s}{2}\lambda^\beta\gamma^\nu\right][v(4)c_4] \tag{9.20}$$

Thus
$$\mathcal{M} = \frac{-g_s^2}{4}\frac{1}{q^2}\,[\bar{u}(3)\gamma^\mu u(1)][\bar{v}(2)\gamma_\mu v(4)](c_3^\dagger\lambda^\alpha c_1)(c_2^\dagger\lambda^\alpha c_4) \tag{9.21}$$

(summation over α implied). This is exactly what we had for electron-positron scattering (7.106), except that g_e is replaced by g_s (of course), and we have in addition the "color factor"

$$f = \tfrac{1}{4}(c_3^\dagger\lambda^\alpha c_1)(c_2^\dagger\lambda^\alpha c_4) \tag{9.22}$$

The *potential* describing the $q\bar{q}$ interaction is, therefore, the same as that acting in electrodynamics between two opposite charges (to wit: the Coulomb potential), only with α replaced by $f\alpha_s$:

$$V_{q\bar{q}}(r) = -f\frac{(\alpha_s\hbar c)}{r} \tag{9.23}$$

Now, the color factor itself depends on the color state of the interacting quarks. From a quark and an antiquark we can make a color singlet (9.5) and a color octet (9.4) (all members of which yield the same f). I'll calculate the octet color factor first, because it's a little easier.[2]

EXAMPLE 9.1 Color Factor for Octet Configuration

A typical octet state (9.4) is $r\bar{b}$ (any of the others would do just as well; see Problem 9.6). Here the incoming quark is red, and the incoming antiquark is antiblue. Because color is conserved, the outgoing quark must also be red and the antiquark antiblue. Thus

$$c_1 = c_3 = \begin{pmatrix}1\\0\\0\end{pmatrix}, \qquad c_2 = c_4 = \begin{pmatrix}0\\1\\0\end{pmatrix}$$

and hence
$$f = \tfrac{1}{4}\left[(1\,0\,0)\lambda^\alpha\begin{pmatrix}1\\0\\0\end{pmatrix}\right]\left[(0\,1\,0)\lambda^\alpha\begin{pmatrix}0\\1\\0\end{pmatrix}\right] = \tfrac{1}{4}\lambda_{11}^\alpha\lambda_{22}^\alpha$$

However, in the nonrelativistic limit of interest here this second diagram does not contribute anyway (see footnote, page 234), so in *practice* what we're doing applies just as well *whatever* the quark flavors. (See also Problem 9.7.)

A glance at the λ matrices reveals that the only ones with entries in the 11 and 22 positions are λ^3 and λ^8. So

$$f = \tfrac{1}{4}(\lambda^3_{11}\lambda^3_{22} + \lambda^8_{11}\lambda^8_{22}) = \tfrac{1}{4}[(1)(-1) + (1/\sqrt{3})(1/\sqrt{3})] = -\tfrac{1}{6} \quad (9.24)$$

EXAMPLE 9.2 Color Factor for Singlet Configuration

The color singlet state is (9.5)

$$(1/\sqrt{3})(r\bar{r} + b\bar{b} + g\bar{g})$$

If the incoming quarks are in the singlet state (as they would be for a meson, say) the color factor is a sum of three terms:

$$f = \frac{1}{4}\cdot\frac{1}{\sqrt{3}}\left\{\left[c_3^\dagger\lambda^\alpha\begin{pmatrix}1\\0\\0\end{pmatrix}\right][(1\,0\,0)\lambda^\alpha c_4] + \left[c_3^\dagger\lambda^\alpha\begin{pmatrix}0\\1\\0\end{pmatrix}\right][(0\,1\,0)\lambda^\alpha c_4] + \left[c_3^\dagger\lambda^\alpha\begin{pmatrix}0\\0\\1\end{pmatrix}\right][(001)\lambda^\alpha c_4]\right\}$$

The outgoing quarks are necessarily also in the singlet state, and we get *nine* terms in all, which can be written compactly as follows:

$$f = \frac{1}{4}\cdot\frac{1}{\sqrt{3}}\cdot\frac{1}{\sqrt{3}}(\lambda^\alpha_{ij}\lambda^\alpha_{ji}) = \frac{1}{12}Tr(\lambda^\alpha\lambda^\alpha) \quad (9.25)$$

(summation over i and j, from 1 to 3, implied in the second expression). Now

$$Tr(\lambda^\alpha\lambda^\beta) = 2\delta^{\alpha\beta} \quad (9.26)$$

(Problem 9.3), so, with the summation over α,

$$Tr(\lambda^\alpha\lambda^\alpha) = 16 \quad (9.27)$$

Evidently, then, for the color singlet

$$f = \tfrac{4}{3} \quad (9.28)$$

Putting equations (9.24) and (9.28) into equation (9.23), we conclude that the quark-antiquark potentials are

$$V_{q\bar{q}}(r) = -\frac{4}{3}\frac{(\alpha_s\hbar c)}{r} \qquad \text{(color singlet)} \quad (9.29)$$

$$V_{q\bar{q}}(r) = \frac{1}{6}\frac{(\alpha_s\hbar c)}{r} \qquad \text{(color octet)} \quad (9.30)$$

From the signs we see that the force is *attractive* in the color singlet but *repulsive* for the octet. This helps to explain why quark-antiquark binding (to form mesons) occurs in the singlet configuration but not in the octet (which would have produced colored mesons).

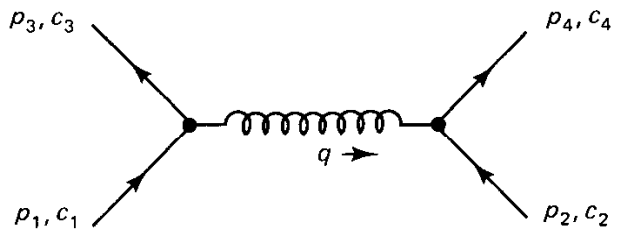


Figure 9.2 The quark-quark interaction.

Case 2: Quark and Quark We turn now to the interaction of two quarks. Again, we shall assume that they have different flavors, so the only diagram (in lowest order) is the one indicated in Figure 9.2,* representing, say, $u + d \rightarrow u + d$. The amplitude is

$$\mathcal{M} = \frac{-g_s^2}{4}\frac{1}{q^2}[\bar{u}(3)\gamma^\mu u(1)][\bar{u}(4)\gamma_\mu u(2)](c_3^\dagger \lambda^\alpha c_1)(c_4^\dagger \lambda^\alpha c_2) \tag{9.31}$$

This is the same as for electron-muon scattering (7.104), except that g_e is replaced by g_s, and there is a color factor

$$f = \tfrac{1}{4}(c_3^\dagger \lambda^\alpha c_1)(c_4^\dagger \lambda^\alpha c_2) \tag{9.32}$$

The potential, therefore, takes the same form as that for like charges in electrodynamics:

$$V_{qq}(r) = f\frac{(\alpha_s \hbar c)}{r} \tag{9.33}$$

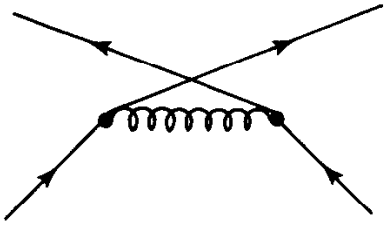
Again, the color factor depends on the configuration of the quarks. From two quarks, however, you can't make a singlet and an octet (as for $q\bar{q}$)—rather, we obtain a triplet (the antisymmetric combinations):

$$\left\{\begin{array}{l} (rb - br)/\sqrt{2} \\ bg - gb)/\sqrt{2} \\ (gr - rg)/\sqrt{2} \end{array}\right\} \qquad \text{(triplet)} \tag{9.34}$$

and a sextet (the symmetric combinations):†

$$\left\{\begin{array}{l} rr, \quad bb, \quad gg, \\ (rb + br)/\sqrt{2}, \quad (bg + gb)/\sqrt{2}, \quad (gr + rg)/\sqrt{2} \end{array}\right\} \qquad \text{(sextet)} \tag{9.35}$$

* For identical quarks there is also the "crossed" diagram:



However, inclusion of this diagram, together with the statistical factor S in the cross-section formula, leads to the same nonrelativistic limit (see footnote page 234), so in fact our potentials are correct even for same-flavor quarks.

† In group theoretical language, $3 \otimes \bar{3} = 1 \oplus 8$, but $3 \otimes 3 = \bar{3} \oplus 6$.

EXAMPLE 9.3 Color Factor for Sextet Configuration

A typical sextet state is rr (use any of the others if you prefer—you'll get the same result for f). In this case

$$c_1 = c_2 = c_3 = c_4 = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}$$

$$\text{and hence} \quad f = \frac{1}{4}\left[(1\,0\,0)\lambda^\alpha \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}\right]\left[(1\,0\,0)\lambda^\alpha \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}\right] = \frac{1}{4}(\lambda^\alpha_{11}\lambda^\alpha_{11})$$

$$= \frac{1}{4}[\lambda^3_{11}\lambda^3_{11} + \lambda^8_{11}\lambda^8_{11}] = \frac{1}{4}\left[(1)(1) + \left(\frac{1}{\sqrt{3}}\right)\left(\frac{1}{\sqrt{3}}\right)\right]$$

$$= \frac{1}{3} \tag{9.36}$$

EXAMPLE 9.4 Color Factor for Triplet Configuration

A typical triplet state is $(rb - br)/\sqrt{2}$, so

$$f = \frac{1}{4}\cdot\frac{1}{\sqrt{2}}\cdot\frac{1}{\sqrt{2}}\left\{ \left[(1\,0\,0)\lambda^\alpha \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}\right] \left[(0\,1\,0)\lambda^\alpha \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}\right] \right.$$
$$- \left[(0\,1\,0)\lambda^\alpha \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}\right] \left[(1\,0\,0)\lambda^\alpha \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}\right]$$
$$- \left[(1\,0\,0)\lambda^\alpha \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}\right] \left[(0\,1\,0)\lambda^\alpha \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}\right]$$
$$\left. - \left[(0\,1\,0)\lambda^\alpha \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}\right] \left[(1\,0\,0)\lambda^\alpha \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}\right] \right\}$$

$$= \tfrac{1}{8}\{\lambda^\alpha_{11}\lambda^\alpha_{22} - \lambda^\alpha_{21}\lambda^\alpha_{12} - \lambda^\alpha_{12}\lambda^\alpha_{21} + \lambda^\alpha_{22}\lambda^\alpha_{11}\}$$
$$= \tfrac{1}{4}(\lambda^\alpha_{11}\lambda^\alpha_{22} - \lambda^\alpha_{12}\lambda^\alpha_{21})$$
$$= \tfrac{1}{4}(\lambda^3_{11}\lambda^3_{22} + \lambda^8_{11}\lambda^8_{22} - \lambda^1_{12}\lambda^1_{21} - \lambda^2_{12}\lambda^2_{21})$$
$$= \tfrac{1}{4}(-1 + \tfrac{1}{3} - 1 - 1) = -\tfrac{2}{3} \tag{9.37}$$

Putting equations (9.36) and (9.37) into equation (9.33), we conclude that the quark-quark potentials are

$$V_{qq}(r) = -\frac{2}{3}\frac{(\alpha_s \hbar c)}{r} \qquad \text{(color triplet)} \tag{9.38}$$

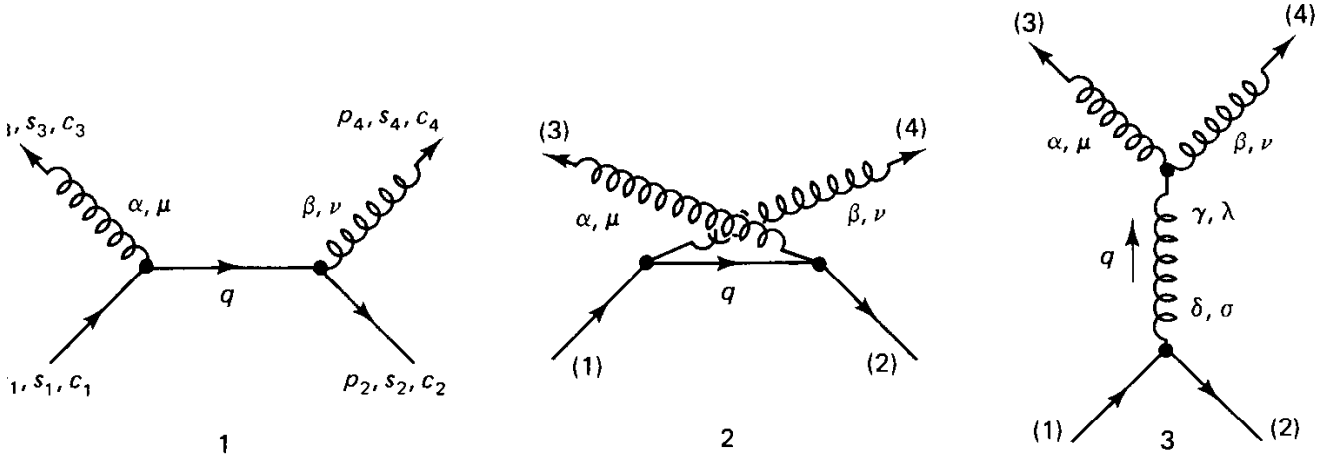
$$V_{qq}(r) = \frac{1}{3}\frac{(\alpha_s \hbar c)}{r} \qquad \text{(color sextet)} \tag{9.39}$$

In particular, the signs indicate that the force is *attractive* for the triplet and *repulsive* for the sextet. Of course, that's not too helpful as it stands, because

neither combination occurs in nature.* However, it does have interesting implications for the binding of *three* quarks. This time we can make a singlet (completely antisymmetric), a decuplet (completely symmetric), and two octets (of mixed symmetry), as we found in Section 5.9 of Chapter 5.† Since the singlet is completely antisymmetric, every *pair* of quarks is in the (antisymmetric) triplet state—the *attractive* channel. In the decuplet, every pair is in the (symmetric) sextet state—they *repel.* As for the two octets, some pairs are triplet and some are sextet; we expect some attraction, then, and some repulsion. *Only in the singlet configuration,* though, *do we get complete mutual attraction* of the three quarks. Again, this is a comforting result: as in the case of mesons, the potential is most favorable for binding when the quarks are in the color singlet configuration.

9.3 PAIR ANNIHILATION IN QCD

In this section we consider the process quark plus antiquark → two gluons—the QCD analog of pair annihilation. The calculation is quite similar to Example 7.8; however, in QCD there are *three* contributing diagrams, in lowest order:



The amplitude for diagram 1 is given by

$$-i\mathcal{M}_1 = \bar{v}(2)c_2^\dagger\left[-i\frac{g_s}{2}\lambda^\beta\gamma^\nu\right][\epsilon_4^*{}_\nu a_4^{\beta *}]\left[\frac{i(\not{q}+mc)}{q^2-m^2c^2}\right]$$
$$\times\left[-i\frac{g_s}{2}\lambda^\alpha\gamma^\mu\right][\epsilon_3^*{}_\mu a_3^{\alpha *}]u(1)c_1 \qquad (9.40)$$

(To simplify the already overburdened notation I'll leave the * off the gluon polarization vectors and color states until the end.) Here $q = p_1 - p_3$, so

$$q^2 - m^2c^2 = p_1^2 - 2p_1\cdot p_3 + p_3^2 - m^2c^2 = -2p_1\cdot p_3 \qquad (9.41)$$

* If you don't heed the warning in footnote (*) on p. 284, you may be alarmed to find that two quarks in the triplet state attract one another. There is some comfort in the observation that the singlet $q\bar{q}$ coupling is *twice as strong;* but still, if this were the *whole* story we might very well expect triplet qq binding to occur, leading to free "diquark" states. There has, in fact, been some speculation about the possible existence of diquarks within nuclei.[3]

† In Chapter 5 we were dealing with *flavor,* not *color,* but the *mathematics* is the same. Group theoretically, $3 \otimes 3 \otimes 3 = 1 \oplus 8 \oplus 8 \oplus 10$.

and hence
$$\mathcal{M}_1 = \frac{-g_s^2}{8}\frac{1}{p_1\cdot p_3}\bar{v}(2)[\not\epsilon_4(\not p_1 - \not p_3 + mc)\not\epsilon_3]u(1) \times a_3^\alpha a_4^\beta(c_2^\dagger\lambda^\beta\lambda^\alpha c_1) \tag{9.42}$$

Similarly, for diagram 2

$$\mathcal{M}_2 = \frac{-g_s^2}{8}\frac{1}{p_1\cdot p_4}\bar{v}(2)[\epsilon_3(\not p_1 - \not p_4 + mc)\epsilon_4]u(1)\, a_3^\alpha a_4^\beta(c_2^\dagger\lambda^\alpha\lambda^\beta c_1) \tag{9.43}$$

Notice that the λ's appear this time in the opposite order. Finally, for diagram 3:

$$-i\mathcal{M}_3 = \bar{v}(2)c_2^\dagger\left[-i\frac{g_s}{2}\lambda^\delta\gamma_\sigma\right]u(1)c_1\left[-i\frac{g^{\sigma\lambda}\,\delta^{\delta\gamma}}{q^2}\right]\cdot\{-g_s f^{\alpha\beta\gamma}[g_{\mu\nu}(-p_3 + p_4)_\lambda + g_{\nu\lambda}(-p_4 - q)_\mu + g_{\lambda\mu}(q + p_3)_\nu]\}[\epsilon_3^\mu a_3^\alpha][\epsilon_4^\nu a_4^\beta] \tag{9.44}$$

In this case $q = p_3 + p_4$, so $q^2 = 2p_3\cdot p_4$; simplifying (and using $\epsilon_3\cdot p_3 = \epsilon_4\cdot p_4 = 0$), we find (Problem 9.10):

$$\mathcal{M}_3 = i\frac{g_s^2}{4}\frac{1}{p_3\cdot p_4}\bar{v}(2)[(\epsilon_3\cdot\epsilon_4)(\not p_4 - \not p_3) + 2(p_3\cdot\epsilon_4)\epsilon_3 - 2(p_4\cdot\epsilon_3)\epsilon_4]u(1) \times f^{\alpha\beta\gamma}a_3^\alpha a_4^\beta(c_2^\dagger\lambda^\gamma c_1) \tag{9.45}$$

So far, this is all completely general (and rather messy). To make things more manageable, let's assume (as we did in our study of e^+e^- annihilation) that the initial particles are at rest:

$$p_1 = p_2 = (mc, \mathbf{0}), \qquad p_3 = (mc, \mathbf{p}), \qquad p_4 = (mc, -\mathbf{p}) \tag{9.46}$$

Then
$$p_1\cdot p_3 = p_1\cdot p_4 = (mc)^2 \quad \text{and} \quad p_3\cdot p_4 = 2(mc)^2 \tag{9.47}$$

Meanwhile, in the Coulomb gauge, equation (9.7)

$$p_3\cdot\epsilon_4 = -\mathbf{p}\cdot\epsilon_4 = -p_4\cdot\epsilon_4 = 0 \tag{9.48}$$

(likewise $p_4\cdot\epsilon_3 = 0$), so two terms in $\mathcal{M}_3$ drop out. Using equations (7.137) and (7.138) to simplify $\mathcal{M}_1$ and $\mathcal{M}_2$, we find that the total amplitude ($\mathcal{M} = \mathcal{M}_1 + \mathcal{M}_2 + \mathcal{M}_3$) can be written

$$\mathcal{M} = -\frac{g_s^2}{8(mc)^2}a_3^\alpha a_4^\beta\,\bar{v}(2)c_2^\dagger[\not\epsilon_3\not\epsilon_4\not p_4\lambda^\alpha\lambda^\beta + \not\epsilon_4\not\epsilon_3\not p_3\lambda^\beta\lambda^\alpha - i(\epsilon_3\cdot\epsilon_4)(\not p_4 - \not p_3)f^{\alpha\beta\gamma}\lambda^\gamma]c_1u(1) \tag{9.49}$$

We may as well orient our coordinates so that the z axis lies along $\mathbf{p}$; then

$$\not p_3 = mc(\gamma^0 - \gamma^3), \qquad \not p_4 = mc(\gamma^0 + \gamma^3), \qquad \not p_4 - \not p_3 = 2mc\gamma^3 \tag{9.50}$$

From equations (7.142) and (7.143) we have

$$\not\epsilon_3\not\epsilon_4 = -(\epsilon_3\cdot\epsilon_4) - i(\epsilon_3\times\epsilon_4)\cdot\boldsymbol{\Sigma}, \qquad \not\epsilon_4\not\epsilon_3 = -(\epsilon_3\cdot\epsilon_4) + i(\epsilon_3\times\epsilon_4)\cdot\boldsymbol{\Sigma} \tag{9.51}$$

Putting this into equation (9.49), and exploiting the commutation relation (9.10) for the λ's, we obtain

$$\mathcal{M} = \frac{g_s^2}{8mc} a_3^\alpha a_4^\beta \bar{v}(2)c_2^\dagger((\epsilon_3 \cdot \epsilon_4)\{\lambda^\alpha, \lambda^\beta\}\gamma^0 + i(\epsilon_3 \times \epsilon_4)\cdot \boldsymbol{\Sigma}([\lambda^\alpha, \lambda^\beta]\gamma^0 + \{\lambda^\alpha, \lambda^\beta\}\gamma^3))c_1 u(1) \tag{9.52}$$

where curly brackets denote the anticommutator: $\{A, B\} \equiv AB + BA$. (You might compare this result with the corresponding expression in QED (7.146), to which it reduces if you set all the λ's equal to 1, drop the color states a and c, and let $g_s/2 \to g_e$.)

Suppose now we put the quarks into a *spin-0* (singlet) state (the triplet state cannot go to two gluons anyway; it needs at least *three*):

$$\mathcal{M} = (\mathcal{M}_{\uparrow\downarrow} - \mathcal{M}_{\downarrow\uparrow})/\sqrt{2} \tag{9.53}$$

For $\mathcal{M}_{\uparrow\downarrow}$ we have [see eqs. (7.150) and (7.151)]

$$\bar{v}(2)\gamma^0 u(1) = \bar{v}(2)\boldsymbol{\Sigma}\gamma^0 u(1) = 0, \qquad \bar{v}(2)\boldsymbol{\Sigma}\gamma^3 u(1) = -2mc\hat{z} \tag{9.54}$$

As before, $\mathcal{M}_{\downarrow\uparrow} = -\mathcal{M}_{\uparrow\downarrow}$, and we are left with*

$$\mathcal{M} = -i\sqrt{2}\,\frac{g_s^2}{4}\,(\epsilon_3 \times \epsilon_4)_z\, a_3^\alpha a_4^\beta (c_2^\dagger\{\lambda^\alpha, \lambda^\beta\}c_1) \qquad \text{(spin singlet)} \tag{9.55}$$

Once again, we have obtained a result that is identical to the one in QED (eq. 7.155), except that $g_e \to g_s$, and there is a color factor

$$f = \tfrac{1}{8} a_3^\alpha a_4^\beta (c_2^\dagger\{\lambda^\alpha, \lambda^\beta\}c_1) \tag{9.56}$$

In particular, if the quarks occupy the color singlet state, $(1/\sqrt{3})(r\bar{r} + b\bar{b} + g\bar{g})$, then

$$f = \frac{1}{8} a_3^\alpha a_4^\beta \frac{1}{\sqrt{3}} \left\{ (1\,0\,0)\{\lambda^\alpha, \lambda^\beta\}\begin{pmatrix}1\\0\\0\end{pmatrix} + (0\,1\,0)\{\lambda^\alpha, \lambda^\beta\}\begin{pmatrix}0\\1\\0\end{pmatrix} + (0\,0\,1)\{\lambda^\alpha, \lambda^\beta\}\begin{pmatrix}0\\0\\1\end{pmatrix} \right\}$$

$$= \frac{1}{8\sqrt{3}} a_3^\alpha a_4^\beta Tr\{\lambda^\alpha, \lambda^\beta\} \tag{9.57}$$

But

$$Tr\{\lambda^\alpha, \lambda^\beta\} = 2Tr(\lambda^\alpha\lambda^\beta) = 4\delta^{\alpha\beta} \tag{9.58}$$

(Problem 9.3), so

$$f = \frac{1}{2\sqrt{3}} a_3^\alpha a_4^\alpha \qquad \text{(color singlet)} \tag{9.59}$$

Now, the singlet state for two gluons (see Problem 9.12) is

$$|\text{singlet}\rangle = \frac{1}{\sqrt{8}} \sum_{n=1}^{8} |n\rangle_1 |n\rangle_2 \tag{9.60}$$

* At this stage all terms in $\epsilon_3 \cdot \epsilon_4$ drop out. The fact that $\mathcal{M}_3$ is proportional to $\epsilon_3 \cdot \epsilon_4$ [eq. (9.49)] means that the diagram containing a three-gluon vertex makes no contribution, when the quarks are at rest in the spin singlet configuration. Most books simply ignore it from the start, but in principle it should be included (see also Problem 9.11).

Evidently $$a_3^\alpha a_4^\alpha = \frac{1}{\sqrt{8}}(8) = 2\sqrt{2} \tag{9.61}$$

and hence $$f = \sqrt{2/3} \tag{9.62}$$

Conclusion: for $q + \bar{q} \to g + g$ in the spin singlet, color singlet configuration, with the quarks at rest, the amplitude is

$$\mathcal{M} = -4\sqrt{2/3}g_s^2 \tag{9.63}$$

[see eq. (7.160)], and the cross section is

$$\sigma = \frac{2}{3}\frac{4\pi}{cv}\left(\frac{\hbar\alpha_s}{m}\right)^2 \tag{9.64}$$

[see eq. (7.165)]. Just as the cross section for $e^+ + e^- \to \gamma + \gamma$ indicated, the positronium decay rate

$$\Gamma = \sigma v|\psi(0)|^2 \tag{9.65}$$

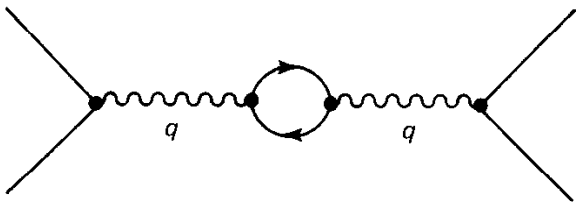
[eq. (7.168)], so we can now give a formula for the decay of a spin-0 quarkonium state (such as η_c—note that ψ and Υ themselves carry spin 1, and go to *three* gluons):

$$\Gamma(\eta_c \to 2g) = \frac{8\pi}{3c}\left(\frac{\hbar\alpha_s}{m}\right)^2|\psi(0)|^2 \tag{9.66}$$

As it stands, this is not terribly useful, since we don't know $\psi(0)$. However, the electromagnetic decay $\eta_c \to 2\gamma$ involves the same factor, and we *can* derive a clean expression for the branching ratio (see Problem 9.13).

9.4 ASYMPTOTIC FREEDOM

In the last section of Chapter 7 we found that the loop diagram

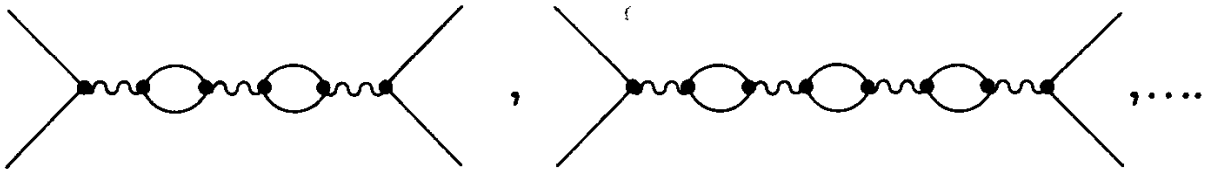


in quantum electrodynamics makes the effective charge of the electron a function of the momentum transfer q:*

$$\alpha(|q^2|) = \alpha(0)\left\{1 + \frac{\alpha(0)}{3\pi}\ln(|q^2|/(mc)^2)\right\} \qquad (|q^2| = -q^2 \gg (mc)^2) \tag{9.67}$$

* It *also* introduces a *divergent* term, which we soak up in the "renormalized" charge [eq. (7.185)]. But that's an entirely different problem, one which (however troublesome you may find it) has no observable consequences, and once the appropriate words have been said is of no further importance. The perfectly *finite* dependence of α on q^2 is the *significant* matter, for it carries direct and measurable implications.

The coupling strength *increases* as the charges get closer together (larger $|q^2|$), a fact that we interpret physically as a consequence of "vacuum polarization": the vacuum functions as a kind of dielectric medium, partially screening the charge. The closer we approach, the less complete is the screening, and the greater is the effective charge. Of course, equation (9.67) is valid only to order $\alpha(0)^2$. There are higher-order corrections, of which the dominant ones come from *chains* of bubbles:

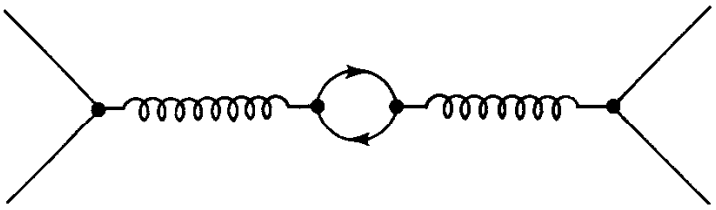


As it happens, these can be summed explicitly, and the result is*

$$\alpha(|q^2|) = \frac{\alpha(0)}{1 - (\alpha(0)/3\pi)\ln(|q^2|/(mc)^2)} \qquad (|q^2| \gg (mc)^2) \qquad (9.68)$$

Ostensibly, the coupling *blows up* at $\ln(|q^2|/(mc)^2) = 3\pi/\alpha(0)$. However, this is not to be taken too seriously, since it occurs at an energy of about 10^{280} MeV, which (to put it mildly) is not an accessible region (see Problem 9.15).

Much the same thing happens in QCD: quark-antiquark bubbles

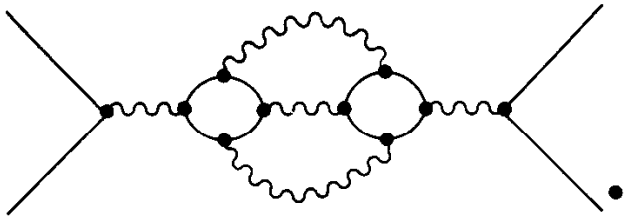


lead to a screening of the quark color which (modulo appropriate color factors) is the same as equation (9.67). However, there is a new twist to the story, for in QCD we also have virtual *gluon* bubbles

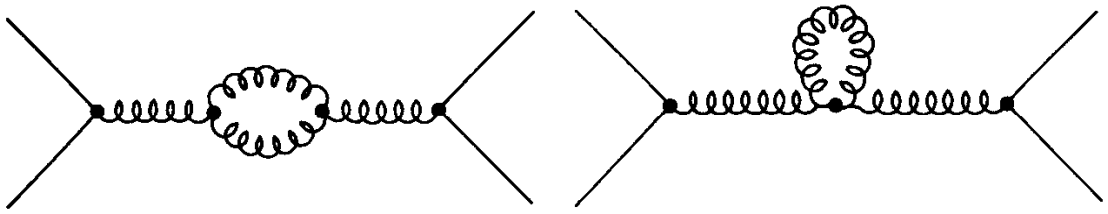
* This is perhaps not so surprising. What we have, in effect, is the geometric series

$$1 + x + x^2 + x^3 + \cdots = \frac{1}{1 - x}$$

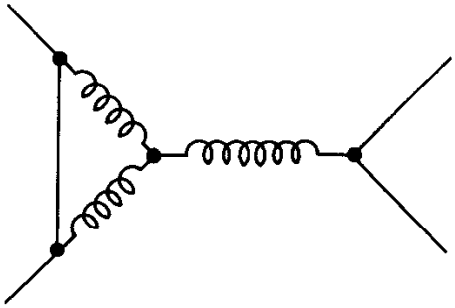
where x is for one bubble, x^2 is for two, and so on. Although equation (9.68) is correct to all orders in $\alpha(0)$, it is not *exact*, since we are ignoring diagrams such as



These can be shown to make a much smaller contribution in the limit $|q^2| \gg (mc)^2$. Equation (9.68) is called the "leading log" approximation.



as well as diagrams of the form



It turns out[4] that the gluon contribution works in the other direction, producing "antiscreening," or "camouflage." I do not know of a persuasive qualitative explanation of this effect[5]—suffice it to say that the formula for the running coupling constant in QCD [analogous to eq. (9.68)] is

$$\alpha_s(|q^2|) = \frac{\alpha_s(\mu^2)}{1 + (\alpha_s(\mu^2)/12\pi)(11n - 2f)\ln(|q^2|/\mu^2)} \qquad (|q^2| \gg \mu^2) \qquad (9.69)$$

where n is the number of colors (3, in the Standard Model), and f is the number of flavors (6, in the Standard Model). In any theory for which $11n > 2f$, antiscreening will dominate, and the coupling constant will *decrease* with increasing $|q^2|$; at short distances the "strong" force becomes relatively weak. This, of course, is the basis of *asymptotic freedom,* on which so much of what we can say quantitatively about the hadrons is predicated. Asymptotic freedom is what allows us to treat partons as essentially free particles, leading to Bjorken scaling; it is what licenses the use of the Feynman calculus in QCD to calculate interquark potentials; it is a basic ingredient in the theory of quarkonium; and it is presumably responsible for the OZI rule. Chromodynamics would have gone out of business if it had not been for the timely discovery of asymptotic freedom.[6]

You may have noticed the appearance of a new parameter, μ, in equation (9.69). In electrodynamics it is natural to define "the charge" of a particle as the long-range (fully screened) value—that's what Coulomb and Millikan measured, and it's what an engineer or a chemist or even an atomic physicist (unless he's measuring the Lamb shift) is concerned with. Thus $\alpha(0)$ is the "good old" fine structure constant, $\frac{1}{137}$, and it is the sensible parameter in terms of which to do perturbation expansions. But we don't *have* to do it this way; we *could* work from any other value of q^2 [provided only that we stay well below the singularity in (9.68), where $\alpha(|q^2|)$ runs larger than 1, and perturbation theory breaks down]. In QCD, however, we *cannot* work from $q^2 = 0$, because that's where α_s is *large.* We must use as a reference some place where α_s is small enough to justify a perturbation expansion. That's why equation (9.69) is expressed in terms of $\alpha_s(\mu^2)$, instead of $\alpha_s(0)$. Provided that it's large enough so that $\alpha_s(\mu^2) < 1$, it

doesn't matter *what* value of μ we use (see Problem 9.16). Indeed, if we introduce a new variable Λ, defined by

$$\ln \Lambda^2 = \ln \mu^2 - 12\pi/[(11n - 2f)\alpha_s(\mu^2)] \tag{9.70}$$

the running coupling constant can be expressed in terms of a *single* parameter:

$$\alpha_s(|q^2|) = \frac{12\pi}{(11n - 2f)\ln(|q^2|/\Lambda^2)} \qquad (|q^2| \gg \Lambda^2) \tag{9.71}$$

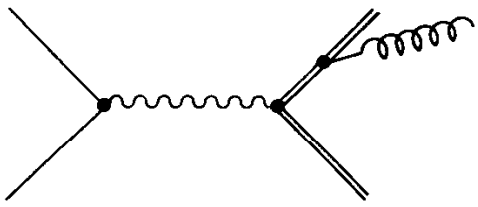
(see Problem 9.17). This compact result tells us explicitly the value of the strong coupling at any $|q^2|$, in terms of the constant Λ. Unfortunately, it is hard to determine Λ precisely from experimental data, but Λc appears to lie somewhere in the range

$$100 \text{ MeV} < \Lambda c < 500 \text{ MeV}. \tag{9.72}$$

Notice that whereas the QED coupling varies only minutely over the accessible energy range (Problem 9.15), variation in the QCD coupling is substantial (Problem 9.18).

9.5 APPLICATIONS OF QCD

It must be admitted that the number of things one can actually *calculate* in QCD is, at this stage, embarrassingly meager. I have not, of course, shown you *everything* that has been done: violations of scaling, due to gluon emission processes such as



have been analyzed in detail, and the agreement with experiment is impressive; QCD corrections to the R formula [eq. (8.8)] can be computed; you can work out the distribution of jets in e^+e^- and e^-p scattering; and the "Drell–Yan" process, $p + p \to \mu^+ + \mu^- + x$, can be studied for scaling violations.[7] All this is *perturbative*, based on asymptotic freedom, and limited to the short-distance region. At the other extreme, much work has been devoted to an understanding of confinement in QCD, but this, involving as it must the *long*-range behavior of the interquark force, *cannot* be done perturbatively. The most promising technique is "lattice gauge theory," in which the space–time continuum is replaced by a finite lattice of discrete points, and the equations of QCD are solved numerically. One hopes to achieve realistic results in the limit as the lattice spacing shrinks to zero.[8] The trouble is that *any* theory, even QED, exhibits confinement on a finite lattice. The delicate question is whether this behavior persists in the continuum limit. In QED we find a kind of "phase transition" at which the system flips over to a nonconfining mode; computer studies indicate

that *no* such phase transition occurs in QCD, and so the theory remains confining. This is tantalizingly close to a confirmation of confinement in QCD, although a rigorous proof would, of course, be preferable.

But even confinement, important though it is, has more to do with the internal consistency of the model than with experimental data. If QCD is the correct theory of strong interactions, where is its solution to the *classic* problems in hadron physics? Why don't we now calculate the neutron-proton mass difference, or the force between two protons, or the cross section for pion-nucleon scattering, or the binding energy of the deuteron? The trouble is that all these simple-sounding questions involve complicated many-body problems. I suppose that in the course of time we will find ways of handling such matters using QCD, just as physical chemists have learned to apply quantum mechanics to large molecules. But for the moment we must make do with more modest achievements.[9]

REFERENCES AND NOTES

1. The interested reader should consult the classic treatise by E. S. Abers and B. W. Lee, *Phys. Rep.* **9C,** (1973), 1.
2. People in the know seem to be able to calculate color factors on their fingers; everyone has his or her own tricks. See D. H. Perkins, *Introduction to High-Energy Physics,* 2d Ed. (Reading, MA: Addison-Wesley, 1982), App. G; F. Halzen and A. D. Martin, *Quarks and Leptons,* (New York: Wiley, 1984), Sect. 2.15 and p. 211; C. Quigg, *Gauge Theories of the Strong, Weak, and Electromagnetic Interactions* (Reading, MA: Benjamin/Cummings, 1983), pp. 198–199. I prefer to do it by the more pedestrian method shown here, which is close in spirit to the approach of G. L. Kane in *Color Symmetry and Quark Confinement,* Proceedings of the 12th Rencontre de Mariond, Vol. III, J. Tran Thanh Van, ed. (1977), p. 9.
3. See F. Close, "Demon Nuclei," in *Nature* **296,** 305 (1982).
4. C. Quigg, ref. 2, Sect. 8.3.
5. See, however, C. Quigg, ref. 2, p. 223 and *Sci. Am.* (April 1985).
6. H. D. Politzer, *Phys. Rev. Lett.* **30,** 1346 (1973); *Phys. Rep.* **14C,** 130 (1974); D. J. Gross and F. Wilczek, *Phys. Rev. Lett.* **30,** 1343 (1973).
7. Halzen and Martin, ref. 2, Chaps. 10 and 11; Quigg, ref. 2, Sects. 8.4 and 8.5.
8. C. Rebbi, *Sci. Am.* (February 1983).
9. A summary of the experimental evidence in support of QCD is given by P. Söding and G. Wolf, *Ann. Rev. Nucl. Part. Sci.* **31,** 231 (1981).

PROBLEMS

9.1. Why can't the "ninth gluon" be the photon? [*Answer:* The gluon would couple to all baryons with the same strength, not (as the photon does) in proportion to their charge. Since the mass and baryon numbers are approximately proportional in bulk matter, such a force would, in fact, look very much like an extra contribution to gravity. There was a flurry of interest in this possibility in early 1986. (E. Fischbach et al., *Phys. Rev. Lett.* **56,** 3 (1986). See, however, the comments in *Phys. Rev. Lett.* **56,** 2423 (1986).]

9.2. Color $SU(3)$ transformations relabel "red," "blue," and "green" according to the transformation rule

$$c \longrightarrow c' = Uc$$

where U is any unitary ($UU^\dagger = 1$) 3×3 matrix of determinant 1, and c is a three-element column vector. For example

$$U = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix}$$

would take $r \longrightarrow b$, $b \longrightarrow g$, $g \longrightarrow r$. The ninth gluon ($|9\rangle$) is obviously invariant under U, but the octet gluons are *not*. Show that $|7\rangle$ and $|8\rangle$ go into linear combinations of one another:

$$|7'\rangle = \alpha|7\rangle + \beta|8\rangle, \qquad |8'\rangle = \gamma|7\rangle + \delta|8\rangle$$

Find the numbers α, β, γ, and δ.

9.3. Show that

$$Tr(\lambda^\alpha \lambda^\beta) = 2\delta^{\alpha\beta}$$

(Notice that all the λ matrices are traceless.)

9.4. What are the structure constants for $SU(2)$? That is, what are the numbers f^{ijk} in

$$[\sigma^i, \sigma^j] = 2if^{ijk}\sigma^k$$

9.5. **(a)** Given that $f^{\alpha\beta\gamma}$ is completely antisymmetric (so that $f^{112} = 0$ automatically, and having calculated f^{123}, we don't need to bother with f^{213}, f^{231}, etc.) how many *distinct* nontrivial structure constants remain?

$$\left[\textit{Answer: } \frac{8 \cdot 7 \cdot 6}{3 \cdot 2 \cdot 1} = 56\right]$$

(Of these, it turns out that only nine are nonzero—those listed in equation (9.11)—and among these there are only three different *numbers*.)

(b) Work out $[\lambda^1, \lambda^2]$, and confirm that $f^{12\gamma} = 0$ for all γ except 3, while $f^{123} = 1$.

(c) Similarly, compute $[\lambda^1, \lambda^3]$ and $[\lambda^4, \lambda^5]$, and determine the resulting structure constants.

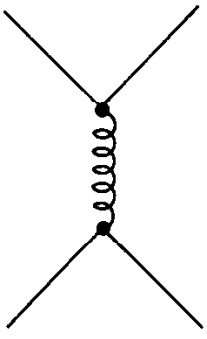
9.6. Calculate the octet $q\bar{q}$ color factor using the state

(a) $b\bar{g}$

(b) $(r\bar{r} - b\bar{b})\sqrt{2}$

(c) $(r\bar{r} + b\bar{b} - 2g\bar{g})\sqrt{6}$

9.7. Find the amplitude $\mathcal{M}$ for the diagram



What is the color factor [analogous to eq. (9.22)] in this case? Evaluate f in the color singlet configuration. Can you explain this answer? [*Answer:* it's zero; a singlet cannot couple to an octet (gluon).]

9.8. Calculate the sextet qq color factor using the state $(rb + br)\sqrt{2}$.

9.9. Color factors always involve expressions of the form $\lambda_{ij}^{\alpha}\lambda_{kl}^{\alpha}$ (summed over α). There is a simple formula for this quantity, which shortens the arithmetic:

$$\lambda_{ij}^{\alpha}\lambda_{kl}^{\alpha} = 2\delta_{il}\,\delta_{jk} - \tfrac{2}{3}\delta_{ij}\,\delta_{kl}$$

[See Kane, ref. 2]. Check this theorem for

(a) $i = j = k = l = 1$ [see eq. (9.36)]
(b) $i = j = 1, \quad k = l = 2$ [see eq. (9.24)]
(c) $i = l = 1, \quad j = k = 2$ [see eq. (9.37)]

and

(d) Use it to confirm equation (9.27).

9.10. Derive equation (9.45), starting from equation (9.44).

9.11. There is a simple test for the gauge invariance of an amplitude ($\mathcal{M}$) in QCD (or QED): Replace any gluon (or photon) polarization vector by its momentum ($\epsilon_3 \to p_3$, say), and you must get zero (see Problem 7.21). Show using this criterion that $\mathcal{M} = \mathcal{M}_1 + \mathcal{M}_2 + \mathcal{M}_3$ is gauge-invariant, but $\mathcal{M}_1 + \mathcal{M}_2$ alone is *not*. [Thus the three-gluon vertex is essential in QCD to preserve gauge invariance. Notice, by contrast, that $\mathcal{M}_1 + \mathcal{M}_2$ alone is gauge-invariant in QED (Example 7.8). The fact that λ matrices do not commute makes the difference.]

9.12. Construct the color singlet combination of two gluons (9.60). [One method is as follows:

Let
$$c = \begin{pmatrix} r \\ b \\ g \end{pmatrix}$$

Under $SU(3)$, $c \to c' = Uc$, where U is a unitary ($UU^\dagger = 1$) matrix of determinant 1. Similarly, let $d^\dagger = (\bar{r}, \bar{b}, \bar{g})$, transforming by the rule $d^\dagger \to d'^\dagger = d^\dagger U^\dagger$. Form the matrix

$$M \equiv cd^\dagger = \begin{pmatrix} r\bar{r} & r\bar{b} & r\bar{g} \\ b\bar{r} & b\bar{b} & b\bar{g} \\ g\bar{r} & g\bar{b} & g\bar{g} \end{pmatrix}$$

Note that $M' = c'd'^\dagger = UMU^\dagger$.
Remove the trace:

$$N \equiv M - \tfrac{1}{3}[Tr(M)], \text{ so that } Tr(N) = 0$$

[Note that $Tr(M') = Tr(M) = (r\bar{r} + b\bar{b} + g\bar{g})$, so this combination is $SU(3)$-invariant; it is the *singlet* combination in $3 \otimes \bar{3} = 1 \oplus 8$, N is the octet.]

$$N' = M' - \tfrac{1}{3}(Tr(M')] = UMU^\dagger - \tfrac{1}{3}[Tr(M)]UU^\dagger = UNU^\dagger$$

The question is how to put together two 8's to make a 1; that is, how to make something bilinear in N_1 and N_2 which is invariant under U. The solution is

$$s \equiv Tr(N_1 N_2)$$

For

$$s' = Tr(N_1'N_2') = Tr(UN_1U^\dagger UN_2U^\dagger) = Tr(U^\dagger UN_1N_2) = Tr(N_1N_2) = s$$

It remains to figure out what s is in terms of the elements of M_1 and M_2:

$$\begin{aligned} Tr(N_1N_2) &= Tr\{(M_1 - \tfrac{1}{3}[Tr(M_1)](M_2 - \tfrac{1}{3}[Tr(M_2)]\} \\ &= Tr(M_1M_2) - \tfrac{1}{3}[Tr(M_1)][Tr(M_2)] \\ &= \tfrac{2}{3}[(r\bar{r})_1(r\bar{r})_2 + (b\bar{b})_1(b\bar{b})_2 + (g\bar{g})_1(g\bar{g})_2] \\ &\quad - \tfrac{1}{3}[(r\bar{r})_1(b\bar{b})_2 + (r\bar{r})_1(g\bar{g})_2 + (b\bar{b})_1(r\bar{r})_2 + (b\bar{b})_1(g\bar{g})_2 \\ &\quad + (g\bar{g})_1(r\bar{r})_2 + (g\bar{g})_1(b\bar{b})_2] + [(r\bar{b})_1(b\bar{r})_2 + (r\bar{g})_1(g\bar{r})_2 \\ &\quad + (b\bar{r})_1(r\bar{b})_2 + (b\bar{g})_1(g\bar{b})_2 + (g\bar{r})_1(r\bar{g})_2 + (g\bar{b})_1(b\bar{g})_2] \\ &= |1\rangle_1|1\rangle_2 + |2\rangle_1|2\rangle_2 + |3\rangle_1|3\rangle_2 + |4\rangle_1|4\rangle_2 \\ &\quad + |5\rangle_1|5\rangle_2 + |6\rangle_1|6\rangle_2 + |7\rangle_1|7\rangle_2 + |8\rangle_1|8\rangle_2 = \sum_{n=1}^{8} |n\rangle_1|n\rangle_2 \end{aligned}$$

This—the invariant product of two octets—is the $SU(3)$ analog to the dot product of two 3-vectors in $SU(2)$.]

9.13. Determine the branching ratio $\Gamma(\eta_c \to 2g)/\Gamma(\eta_c \to 2\gamma)$. [*Hint:* Use equation (9.66) for the numerator, and a suitable modification of equations (7.165) and (7.168) for the denominator. There are two modifications: (i) the quark change is Qe, and (ii) there is a color factor of 3, for quarks in the singlet state (9.5). *Answer:* $\frac{2}{8}(\alpha_s/\alpha)^2$.]

9.14. (Gluon-gluon scattering)

(a) Draw the lowest-order diagrams (there are four of them) representing the interaction of two gluons.

(b) Write down the corresponding amplitudes.

(c) Put the incoming gluons into the color singlet state; do the same for the outgoing gluons. Compute the resulting amplitudes.

(d) Go to the *CM* frame, in which each gluon has energy E; express all the kinematic factors in terms of E and the scattering angle θ. Add the amplitudes to get the total, $\mathcal{M}$.

(e) Find the differential scattering cross section.

(f) Determine whether the force is attractive or repulsive (if it is the former, this may be a likely glueball configuration).

9.15. **(a)** Calculate the energy ($\sqrt{|q^2|c^2}$) at which the QED coupling constant (9.68) blows up. (Remember, $\alpha(0) = \frac{1}{137}$, the fine structure constant.)

(b) At what energy do we get a 1% departure from $\alpha(0)$? Is this an accessible energy?

9.16. Prove that the value of μ in equation (9.69) is arbitrary. [That is, suppose physicist A uses the value μ_a, and physicist B uses a different value, μ_b. *Assume* A's version of equation (9.69) is correct, and *prove* that B's is also correct.]

9.17. Derive equation (9.71) from equations (9.69) and (9.70).

9.18. Calculate α_s at 10 and 100 GeV. Assume $\Lambda c = 0.3$ GeV. What if $\Lambda c = 1$ GeV? How about $\Lambda c = 0.1$ GeV?

Chapter 10

Weak Interactions

This chapter surveys the theory of weak interactions. It relies heavily on Chapter 7, but not on Chapters 8 and 9; Section 4.6 of Chapter 4 would be useful background. I begin by stating the Feynman Rules for the coupling of leptons to $W^{\pm}$, and treat three classic problems in some detail: the beta decays of the muon, the neutron, and the charged pion. Next, we consider the coupling of quarks to $W^{\pm}$, which brings in the Cabibbo angle, the GIM mechanism, and the Kobayashi–Maskawa matrix. In Section 10.6 I state the Feynman rules for coupling quarks and leptons to the Z^0, and the final section (probably the most difficult in this book) shows how all electromagnetic and weak vertex factors can be derived, in the Glashow–Weinberg–Salam electroweak theory.

10.1 CHARGED LEPTONIC WEAK INTERACTIONS

The mediators of weak interactions (analogous to photons in QED and gluons in QCD) are the W's (W^+ and W^-) and the Z^0. Unlike the photon and gluons, which are massless, these "intermediate vector bosons" are extremely heavy—by far the heaviest elementary particles yet detected. Experimentally,

$$M_W = 82 \pm 2 \text{ GeV}/c^2, \qquad M_Z = 92 \pm 2 \text{ GeV}/c^2 \tag{10.1}$$

Now, a *massive* particle of spin 1 has *three* allowed polarization states ($m_s = 1$, 0, -1), whereas a free *massless* particle has only *two* (if z is the direction of motion, the "longitudinal" polarization $m_s = 0$ does not occur). Thus for photons and gluons, we imposed both the Lorentz condition

$$\epsilon^{\mu} p_{\mu} = 0 \tag{10.2}$$

(reducing the number of independent components in ϵ^{μ} from 4 to 3) and also

the Coulomb gauge ($\epsilon^0 = 0$, so that $\boldsymbol{\epsilon} \cdot \mathbf{p} = 0$, which reduces it further from 3 to 2). However, for the W's and the Z the Lorentz condition alone exhausts the gauge freedom, and we do *not* invoke the Coulomb gauge. Moreover, the propagator for massive spin-1 particles is no longer simply $-ig_{\mu\nu}/q^2$, but rather,*

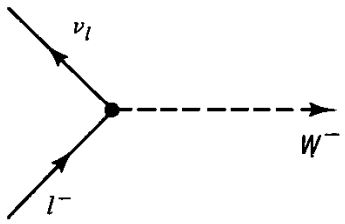
$$\frac{-i(g_{\mu\nu} - q_\mu q_\nu/M^2c^2)}{q^2 - M^2c^2} \qquad \text{(propagator for } W \text{ and } Z\text{)} \tag{10.3}$$

where M is M_W or M_Z, as the case may be. In practice, q^2 is ordinarily so much smaller than $(Mc)^2$ that we may safely use

$$\frac{ig_{\mu\nu}}{(Mc)^2} \qquad \text{(propagator for } q^2 \ll (Mc)^2\text{)} \tag{10.4}$$

However, when a process involves energies that are comparable to Mc^2 we must, of course, revert to the exact expression.

The theory of "charged" weak interactions (mediated by the W's) is simpler than that for "neutral" ones (mediated by the Z), so for the moment I shall concentrate on the former. In this section we consider the coupling of W's to *leptons;* in the next section we'll discuss their coupling to quarks and hadrons. The fundamental leptonic vertex is



Here an electron, muon, or tau is converted into the associated neutrino, with emission of a W^- (or absorption of W^+). The reverse process ($\nu_l \longrightarrow l^- + W^+$) is also possible, of course, as well as the "crossed" reactions involving antileptons. The Feynman rules are the same as for QED (apart from the modifications already mentioned to accommodate the massive mediator), except for the vertex factor

$$\frac{-ig_w}{2\sqrt{2}}\,\gamma^\mu(1 - \gamma^5) \qquad \text{(weak vertex factor)} \tag{10.5}$$

The various 2's are purely conventional, and $g_w = \sqrt{4\pi\alpha_w}$ is the "weak coupling constant" (analogous to g_e in QED and g_s in QCD). The factor $(1 - \gamma^5)$, however, is of profound importance, for γ^μ alone would yield a *vector* coupling (like QED or QCD), whereas $\gamma^\mu\gamma^5$ gives an *axial* vector [see eq. (7.68)]. A theory that adds

* It might bother you that this does not reduce to the photon propagator as $M \longrightarrow 0$. For particles of spin 1 (or higher) the massless limit is notoriously treacherous, because in one critical respect it is not a continuous procedure. The number of degrees of freedom (that is, the number of allowed spin orientations) drops abruptly from $2s + 1$ (for $M \neq 0$) to 2 (for $M = 0$). There are ways of formulating the theory that allow a smooth transition to $M = 0$, but only at the cost of introducing spurious nonphysical states.

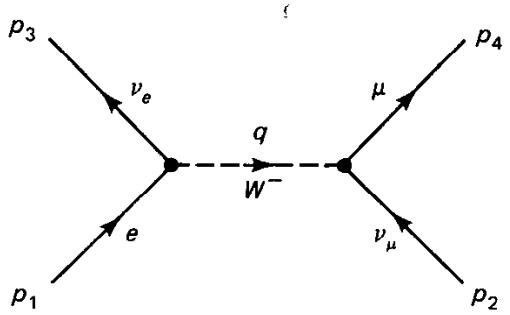
a *vector* to an *axial* vector is bound to violate the conservation of parity, and this is precisely what happens in the weak interactions (Chap. 4, Sect. 4.6).*

EXAMPLE 10.1 Inverse Muon Decay

Consider the process

$$\nu_\mu + e^- \longrightarrow \mu^- + \nu_e$$

represented (in lowest order) by the diagram



Here $q = p_1 - p_3$, and for any experiment likely in the near future $q^2 \ll M_W^2 c^2$, so we can safely use the simplified propagator (10.4), and the amplitude is

$$\mathcal{M} = \frac{g_w^2}{8(M_W c)^2}\,[\bar{u}(3)\gamma^\mu(1-\gamma^5)u(1)][\bar{u}(4)\gamma_\mu(1-\gamma^5)u(2)] \tag{10.6}$$

Applying Casimir's trick (7.123), we find

$$\sum_{\text{spins}} |\mathcal{M}|^2 = \left(\frac{g_w^2}{8(M_W c)^2}\right)^2 Tr[\gamma^\mu(1-\gamma^5)(\not{p}_1 + m_e c)\gamma^\nu(1-\gamma^5)\not{p}_3]$$
$$\times\, Tr[\gamma_\mu(1-\gamma^5)\not{p}_2\gamma_\nu(1-\gamma^5)(\not{p}_4 + m_\mu c)] \tag{10.7}$$

The trace theorems of Chapter 7, Section 7.7, yield

$$8[p_1^\mu p_3^\nu + p_1^\nu p_3^\mu - g^{\mu\nu}(p_1 \cdot p_3) - i\epsilon^{\mu\nu\lambda\sigma}p_{1_\lambda}p_{3_\sigma}] \tag{10.8}$$

for the first trace, and

$$8[p_{2_\mu}p_{4_\nu} + p_{2_\nu}p_{4_\mu} - g_{\mu\nu}(p_2 \cdot p_4) - i\epsilon_{\mu\nu\kappa\tau}p_2^\kappa p_4^\tau] \tag{10.9}$$

for the second. It follows that†

$$\sum_{\text{spins}} |\mathcal{M}|^2 = 4\left(\frac{g_w}{M_W c}\right)^4 (p_1 \cdot p_2)(p_3 \cdot p_4) \tag{10.10}$$

Actually, we want the sum over final spins but the *average* over initial

* In fact, the violation is "maximal," in the sense that the two terms are equally large. When parity violation was first considered, a factor of the form $(1 + \epsilon\gamma^5)$ was used, but experiments soon dictated that $\epsilon = -1$. (See Problem 10.1.) We call it a "*V–A*" ("vector minus axial vector") coupling. Fermi's original theory of beta decay was a pure *vector* theory (like QED), and although others proposed scalar, pseudoscalar, tensor, or pure axial couplings, it was not until 1956 that anyone seriously contemplated *mixing* terms of different parity.

† Note that $\epsilon^{\mu\nu\lambda\sigma}\epsilon_{\mu\nu\kappa\tau} = -2(\delta_\kappa^\lambda\delta_\tau^\sigma - \delta_\tau^\lambda\delta_\kappa^\sigma)$. (See Problem 7.33.)

spins. The electron has *two* spin states, but the neutrino (as we learned in Chapter 4, Section 4.6) has only *one* (if you like, the incident neutrinos are *always* polarized, since they only come "left-handed"). So

$$\langle|\mathcal{M}|^2\rangle = 2\left(\frac{g_w}{M_W c}\right)^4 (p_1 \cdot p_2)(p_3 \cdot p_4) \tag{10.11}$$

If we now go to the *CM* frame, and neglect the mass of the electron

$$\langle|\mathcal{M}|^2\rangle = 8\left(\frac{g_w E}{M_W c^2}\right)^4 \left\{1 - \left(\frac{m_\mu c^2}{2E}\right)^2\right\} \tag{10.12}$$

where E is the incident electron (or neutrino) energy. The differential scattering cross section [eq. (6.42)] is isotropic (all scattering angles equally likely)

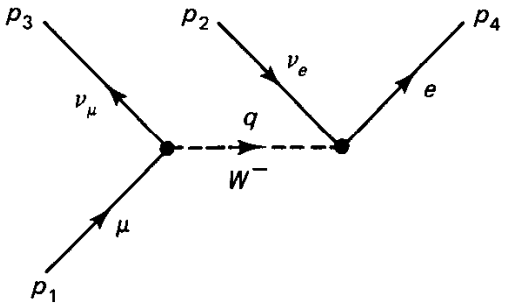
$$\frac{d\sigma}{d\Omega} = \frac{1}{2}\left(\frac{\hbar c g_w^2 E}{4\pi (M_W c^2)^2}\right)^2 \left\{1 - \left(\frac{m_\mu c^2}{2E}\right)^2\right\}^2 \tag{10.13}$$

and the total cross section is

$$\sigma = \frac{1}{8\pi}\left[\left(\frac{g_w}{M_W c^2}\right)^2 \hbar c E\right]^2 \left\{1 - \left(\frac{m_\mu c^2}{2E}\right)^2\right\}^2 \tag{10.14}$$

10.2 DECAY OF THE MUON

Electron-neutrino scattering is not the easiest thing in the world to study experimentally, but the closely related process, muon decay ($\mu \longrightarrow e + \nu_\mu + \bar{\nu}_e$), is the cleanest of all weak interaction phenomena, theoretically and experimentally. The Feynman diagram



leads to the amplitude

$$\mathcal{M} = \frac{g_w^2}{8(M_W c)^2}[\bar{u}(3)\gamma^\mu(1 - \gamma^5)u(1)][\bar{u}(4)\gamma_\mu(1 - \gamma^5)v(2)] \tag{10.15}$$

from which we obtain, as before,

$$\langle|\mathcal{M}|^2\rangle = 2\left(\frac{g_w}{M_W c}\right)^4 (p_1 \cdot p_2)(p_3 \cdot p_4) \tag{10.16}$$

In the muon rest frame, $p_1 = (m_\mu c, \mathbf{o})$, we have

$$p_1 \cdot p_2 = m_\mu E_2 \tag{10.17}$$

and since $p_1 = p_2 + p_3 + p_4$

$$\begin{aligned}(p_3 + p_4)^2 &= p_3^2 + p_4^2 + 2p_3 \cdot p_4 = m_e^2 c^2 + 2p_3 \cdot p_4 \\ &= (p_1 - p_2)^2 = p_1^2 + p_2^2 - 2p_1 \cdot p_2 = m_\mu^2 c^2 - 2p_1 \cdot p_2\end{aligned} \tag{10.18}$$

from which it follows that

$$p_3 \cdot p_4 = \frac{(m_\mu^2 - m_e^2)c^2}{2} - m_\mu E_2 \tag{10.19}$$

The algebra will be simpler later on, at no significant cost in accuracy, if we set $m_e = 0$, so that

$$\langle |\mathcal{M}|^2 \rangle = \left(\frac{g_w}{M_w c}\right)^4 m_\mu^2 E_2 (m_\mu c^2 - 2E_2) \tag{10.20}$$

Now, the decay rate is given by the Golden Rule (6.15):*

$$d\Gamma = \frac{\langle |\mathcal{M}|^2 \rangle}{2\hbar m_\mu} \left(\frac{c\, d^3\mathbf{p}_2}{(2\pi)^3 2E_2}\right)\left(\frac{c\, d^3\mathbf{p}_3}{(2\pi)^3 2E_3}\right)\left(\frac{c\, d^3\mathbf{p}_4}{(2\pi)^3 2E_4}\right)(2\pi)^4\, \delta^4(p_1 - p_2 - p_3 - p_4) \tag{10.21}$$

where $E_2 = |\mathbf{p}_2|c$, $E_3 = |\mathbf{p}_3|c$, and $E_4 = |\mathbf{p}_4|c$. To begin with, we peel apart the delta function:

$$\delta^4(p_1 - p_2 - p_3 - p_4) = \delta\left(m_\mu c - \frac{E_2}{c} - \frac{E_3}{c} - \frac{E_4}{c}\right)\delta^3(\mathbf{p}_2 + \mathbf{p}_3 + \mathbf{p}_4) \tag{10.22}$$

and perform the $\mathbf{p}_3$ integral:

$$d\Gamma = \frac{\langle |\mathcal{M}|^2 \rangle c^3}{16(2\pi)^5 \hbar m_\mu} \frac{(d^3\mathbf{p}_2)(d^3\mathbf{p}_4)}{E_2 E_3 E_4}\, \delta\left(m_\mu c - \frac{E_2}{c} - \frac{E_3}{c} - \frac{E_4}{c}\right) \tag{10.23}$$

where E_3 now stands for $|\mathbf{p}_2 + \mathbf{p}_4|c$. Next we'll do the $\mathbf{p}_2$ integral. Setting the polar axis along $\mathbf{p}_4$ (which is *fixed,* for the purposes of the $\mathbf{p}_2$ integration), we have

$$\begin{aligned}\left(\frac{E_3}{c}\right)^2 &= |\mathbf{p}_2 + \mathbf{p}_4|^2 = \mathbf{p}_2^2 + \mathbf{p}_4^2 + 2\mathbf{p}_2 \cdot \mathbf{p}_4 \\ &= \frac{1}{c^2}(E_2^2 + E_4^2 + 2E_2 E_4 \cos\theta)\end{aligned} \tag{10.24}$$

and

$$d^3\mathbf{p}_2 = \left(\frac{E_2}{c}\right)^2 \frac{dE_2}{c} \sin\theta\, d\theta\, d\phi \tag{10.25}$$

The ϕ integral is trivial ($\int d\phi = 2\pi$); to carry out the θ integration, let

$$x \equiv \frac{1}{c}\sqrt{E_2^2 + E_4^2 + 2E_2 E_4 \cos\theta} = \frac{E_3}{c} \tag{10.26}$$

* Note that this is a *three* body decay, so we have to go all the way back to the Golden Rule.

so that
$$dx = -\frac{E_2 E_4 \sin\theta\, d\theta}{cE_3} \tag{10.27}$$

Then
$$\int_0^\pi \frac{\sin\theta\, d\theta}{E_3}\,\delta\left(m_\mu c - \frac{E_2}{c} - \frac{E_3}{c} - \frac{E_4}{c}\right)$$
$$= \frac{c}{E_2 E_4}\int_{x_-}^{x_+} \delta\left(m_\mu c - x - \frac{E_2}{c} - \frac{E_4}{c}\right)dx$$
$$= \begin{cases} \dfrac{c}{E_2 E_4}, & \text{if } x_- < \left(\, m_\mu c - \dfrac{E_2}{c} - \dfrac{E_4}{c}\,\right) < x_+ \\ 0, & \text{otherwise} \end{cases} \tag{10.28}$$

where
$$x_\pm \equiv \frac{1}{c}\sqrt{E_2^2 + E_4^2 \pm 2E_2E_4} = \frac{1}{c}\,|E_2 \pm E_4| \tag{10.29}$$

The inequality in equation (10.28) can be expressed more neatly:

$$|E_2 - E_4| < (m_\mu c^2 - E_2 - E_4) < E_2 + E_4 \tag{10.30}$$

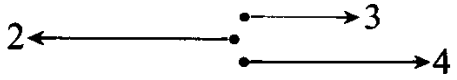
or, adding $(E_2 + E_4)$ and dividing through by 2:

$$\tfrac{1}{2}\{|E_2 - E_4| + E_2 + E_4\} < \tfrac{1}{2}m_\mu c^2 < (E_2 + E_4) \tag{10.31}$$

The term on the left is simply the larger of E_2 and E_4; the other one is necessarily even smaller, so expression (10.31) is equivalent to *three* inequalities:

$$\begin{cases} E_2 < \tfrac{1}{2}m_\mu c^2 \\ E_4 < \tfrac{1}{2}m_\mu c^2 \\ (E_2 + E_4) > \tfrac{1}{2}m_\mu c^2 \end{cases} \tag{10.32}$$

[These constraints make good sense kinematically: Particle 2, for example, gets the maximum possible energy when 3 and 4 emerge diametrically opposite to it:



In this case 2 picks up half the available energy ($\frac{1}{2}m_\mu c^2$), while 3 and 4 share the other half. If there is a nonzero angle between 3 and 4, 2 gets less, and 3 plus 4 get correspondingly more. Thus $\frac{1}{2}m_\mu c^2$ is the maximum energy for any *individual* outgoing particle, and the *minimum* total for any *pair*.]

The inequalities (10.32) specify the limits on the E_2 and E_4 integrals: E_2 runs from $\frac{1}{2}m_\mu c^2 - E_4$ up to $\frac{1}{2}m_\mu c^2$, and E_4 will go from 0 to $\frac{1}{2}m_\mu c^2$. The θ and ϕ integrals leave us with

$$d\Gamma = \frac{\langle |M|^2\rangle c}{(4\pi)^4 \hbar m_\mu}\, dE_2\, \frac{d^3\mathbf{p}_4}{E_4^2} \tag{10.33}$$

Putting in equation (10.20) and carrying out the E_2 integral, we have

$$d\Gamma = \left(\frac{g_w}{4\pi M_W c}\right)^4 \frac{m_\mu c}{\hbar} \frac{d^3\mathbf{p}_4}{E_4^2} \int_{1/2m_\mu c^2 - E_4}^{1/2m_\mu c^2} E_2(m_\mu c^2 - 2E_2)dE_2$$

$$= \left(\frac{g_w}{4\pi M_W c}\right)^4 \frac{m_\mu c}{\hbar} \left(\frac{m_\mu c^2}{2} - \frac{2}{3}E_4\right) d^3\mathbf{p}_4 \tag{10.34}$$

Finally, writing

$$d^3\mathbf{p}_4 = 4\pi\left(\frac{E_4}{c}\right)^2 \frac{dE_4}{c}$$

and dropping the subscript ($E \equiv E_4$ is the electron energy), we obtain

$$\frac{d\Gamma}{dE} = \left(\frac{g_w}{M_W c}\right)^4 \frac{m_\mu^2 E^2}{2\hbar(4\pi)^3}\left(1 - \frac{4E}{3m_\mu c^2}\right) \tag{10.35}$$

This tells us the energy distribution of the electrons emitted in muon decay. It fits the experimental spectrum perfectly (Fig. 10.1). The total decay rate is

$$\Gamma = \left(\frac{g_w}{M_W c}\right)^4 \frac{m_\mu^2}{2\hbar(4\pi)^3} \int_0^{1/2m_\mu c^2} E^2\left(1 - \frac{4E}{3m_\mu c^2}\right) dE$$

$$= \left(\frac{m_\mu g_w}{M_W}\right)^4 \frac{m_\mu c^2}{12\hbar(8\pi)^3} \tag{10.36}$$

and hence the lifetime of the muon is

$$\tau = \frac{1}{\Gamma} = \left(\frac{M_W}{m_\mu g_w}\right)^4 \frac{12\hbar(8\pi)^3}{m_\mu c^2} \tag{10.37}$$

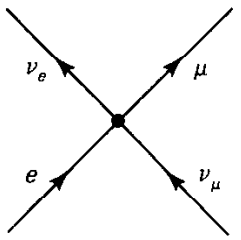
Notice that g_w and M_W do not appear *separately,* either in the muon lifetime formula or in the electron-neutrino scattering cross section; only their *ratio* occurs. It is traditional, in fact, to express weak interaction formulas in terms of the "Fermi coupling constant"

$$G_F \equiv \frac{\sqrt{2}}{8}\left(\frac{g_w}{M_W c^2}\right)^2 (\hbar c)^3 \tag{10.38}$$

Thus the muon lifetime is written

$$\tau = \frac{192\pi^3\hbar^7}{G^2 m_\mu^5 c^4} \tag{10.39}$$

In Fermi's original theory of beta decay (1933) there *was* no W; the interaction was supposed to be a direct four-particle coupling, represented in the Feynman language by a diagram of the form



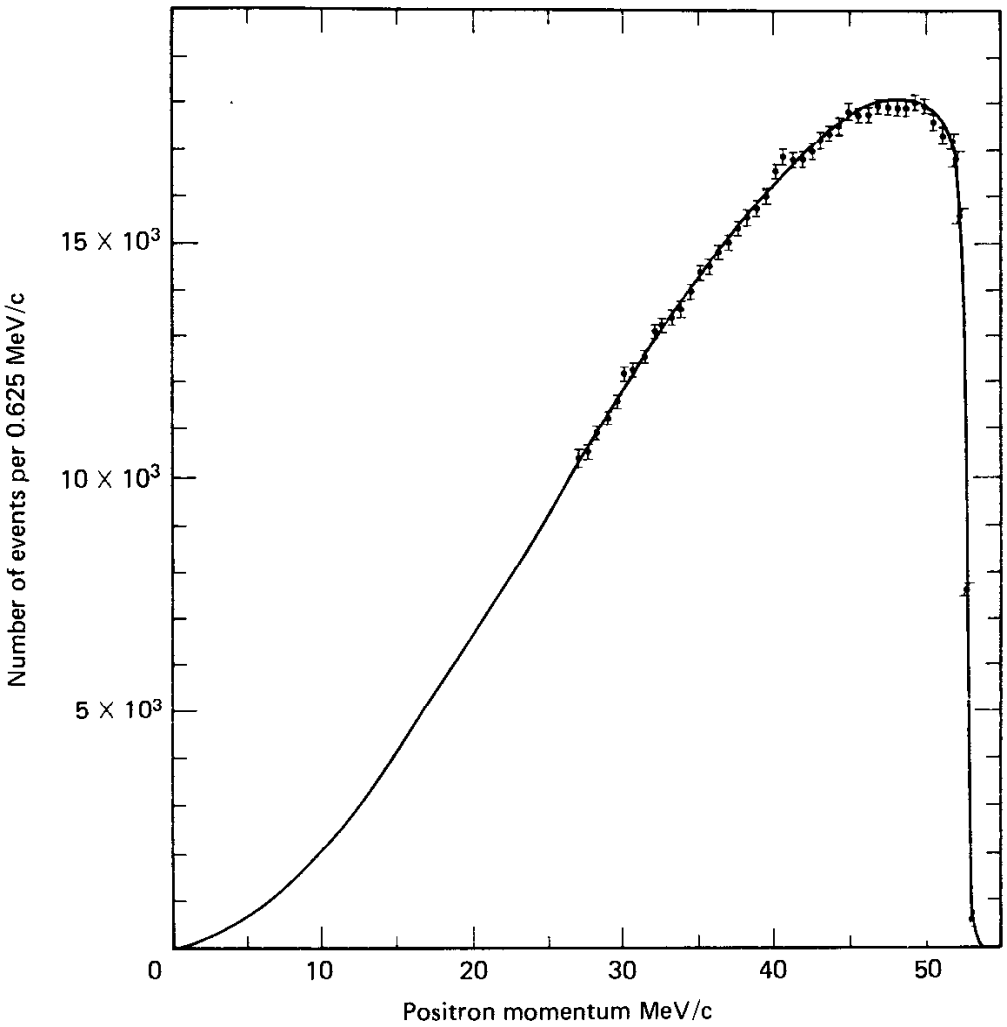
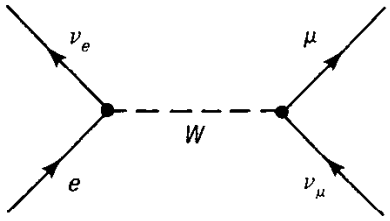


Figure 10.1 Experimental spectrum of positrons in $\mu^+ \rightarrow e^+ + \nu_e + \nu_\mu$. The solid line is the theoretically predicted spectrum based on equation (10.35), corrected for electromagnetic effects. (*Source:* M. Bardon et al., *Phys. Rev. Lett.* **14,** 449 (1965).)

From the modern perspective, Fermi's theory combined the W propagator with the two vertex factors, in the diagram



to make an effective four-particle coupling constant G_F. It worked, but only because the W is so heavy that expression (10.4) is a good approximation to the true propagator (10.3),* and in fact it was recognized already in the fifties that Fermi's theory could not be valid at high energies. The idea of a weak mediator (analogous to the photon) was suggested by O. Klein as far back as 1938.

* Fermi also thought the coupling was pure vector, as I mentioned in the footnote (*) on p. 303. Despite these defects (for which Fermi could scarcely be blamed; after all, he invented the theory at a time when the neutrino was a wild speculation and the Dirac equation itself was quite new) Fermi's theory was astonishingly prescient, and all subsequent developments have been relatively small adjustments to it.

If we put in the observed muon lifetime and mass, we find that

$$G_F/(\hbar c)^3 = \frac{\sqrt{2}}{8}\left(\frac{g_w}{M_W c^2}\right)^2 = 1.166 \times 10^{-5}/\text{GeV}^2 \tag{10.40}$$

The corresponding value of g_w (less accurately known, at present, because of the experimental uncertainty in M_W) is

$$g_w = 0.66 \tag{10.41}$$

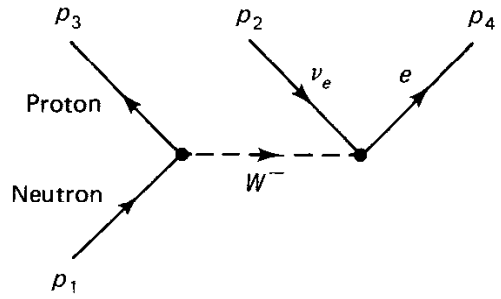
and hence the "weak fine structure constant" is

$$\alpha_w = \frac{g_w^2}{4\pi} = \frac{1}{29} \tag{10.42}$$

This number should come as something of a shock: It is *larger* than the electromagnetic fine structure constant ($\alpha = \frac{1}{137}$), by a factor of nearly five! Weak interactions are feeble *not* because the *intrinsic* coupling is small (it *isn't*), but because the *mediators* are so *massive*—or, more precisely, because we typically work at energies so far below the W mass that the denominator in the propagator $(q^2 - M_W^2 c^2)$ is extremely large. New machines are presently under construction that will run at energies close to $M_W c^2$, and in this regime the "weak" interactions will far surpass the electromagnetic ones in strength.

10.3 DECAY OF THE NEUTRON

The success of the muon decay formula (10.35) encourages us to apply the same methods to the decay of the neutron, $n \to p + e + \bar{\nu}_e$. Of course, the neutron and proton are composite particles, but just as the Mott and Rutherford cross sections (which treat the proton as an elementary "Dirac" particle) give a good account of low-energy electron-proton scattering, so we might hope that the diagram



(the same as for muon decay, only with $n \to p + W^-$ in place of $\mu \to \nu_\mu + W^-$) will afford a reasonable approximation to neutron beta decay. From a calculational point of view the only new feature is that 3 is now a massive particle (a proton, instead of a neutrino). As it happens (Problem 10.4) this does not change the amplitude:

$$\langle|\mathcal{M}|^2\rangle = 2\left(\frac{g_w}{M_W c}\right)^4 (p_1 \cdot p_2)(p_3 \cdot p_4) \tag{10.43}$$

which is the same as (10.16). In the rest frame of the neutron, we find

$$\langle|\mathcal{M}|^2\rangle = \left(\frac{g_w}{M_W c}\right)^4 m_n E_2[(m_n^2 - m_p^2 - m_e^2)c^2 - 2m_n E_2] \tag{10.44}$$

In this case the electron rest energy is a substantial fraction of the total energy released, $(m_n - m_p - m_e)c^2$, so we *cannot* afford to ignore the electron mass.

The decay rate calculation proceeds as before:

$$d\Gamma = \frac{\langle|\mathcal{M}|^2\rangle}{2\hbar m_n}\left(\frac{c\,d^3\mathbf{p}_2}{(2\pi)^3 2E_2}\right)\left(\frac{c\,d^3\mathbf{p}_3}{(2\pi)^3 2E_3}\right)\left(\frac{c\,d^3\mathbf{p}_4}{(2\pi)^3 2E_4}\right)(2\pi)^4\,\delta^4(p_1 - p_2 - p_3 - p_4) \tag{10.45}$$

where $E_2 = c|\mathbf{p}_2|, \quad E_3 = c\sqrt{\mathbf{p}_3^2 + m_p^2 c^2}, \quad E_4 = c\sqrt{\mathbf{p}_4^2 + m_e^2 c^2}$ (10.46)

The $\mathbf{p}_3$ integral yields

$$d\Gamma = \frac{2c^3\langle|\mathcal{M}|^2\rangle}{(4\pi)^5\hbar m_n}\frac{d^3\mathbf{p}_2\,d^3\mathbf{p}_4}{E_2 E_3 E_4}\,\delta\left(m_n c - \frac{E_2}{c} - \frac{E_3}{c} - \frac{E_4}{c}\right) \tag{10.47}$$

which is the same as equation (10.23), except that this time

$$E_3 = c\sqrt{(\mathbf{p}_2 + \mathbf{p}_4)^2 + m_p^2 c^2} \tag{10.48}$$

To carry out the $\mathbf{p}_2$ integral, we set

$$d^3\mathbf{p}_2 = |\mathbf{p}_2|^2\,d|\mathbf{p}_2|\,\sin\theta\,d\theta\,d\phi = \frac{1}{c^3}E_2^2\,dE_2\,\sin\theta\,d\theta\,d\phi \tag{10.49}$$

and orient the coordinates so that the z axis lies along $\mathbf{p}_4$ (which is fixed, for purposes of the $\mathbf{p}_2$ integral); then

$$E_3 = c\sqrt{|\mathbf{p}_2|^2 + |\mathbf{p}_4|^2 + 2|\mathbf{p}_2||\mathbf{p}_4|\cos\theta + m_p^2 c^2} \equiv cx \tag{10.50}$$

and

$$\frac{E_2\sin\theta\,d\theta}{E_3} = -\frac{dx}{|\mathbf{p}_4|} \tag{10.51}$$

The ϕ and θ (or rather, x) integrals yield

$$d\Gamma = \frac{\langle|\mathcal{M}|^2\rangle}{(4\pi)^4\hbar m_n}\frac{dE_2\,d^3\mathbf{p}_4}{E_4|\mathbf{p}_4|}\,I \tag{10.52}$$

where $I \equiv \int_{x_-}^{x_+}\delta\left(m_n c - x - \frac{E_2}{c} - \frac{E_4}{c}\right)dx$

$$= \begin{cases} 1, & \text{if } x_- < \left(m_n c - \frac{E_2}{c} - \frac{E_4}{c}\right) < x_+ \\ 0, & \text{otherwise} \end{cases} \tag{10.53}$$

and the limits [from (10.50)] are

$$x_\pm = \sqrt{(|\mathbf{p}_2| \pm |\mathbf{p}_4|)^2 + m_p^2 c^2} \tag{10.54}$$

As before, equation (10.53) defines the range of the E_2 integral; I'll let you work out the algebra (Problem 10.5):

$$E_{\pm} = \frac{\frac{1}{2}(m_n^2 - m_p^2 + m_e^2)c^2 - m_n E_4}{m_n - E_4/c^2 \mp |\mathbf{p}_4|/c} \tag{10.55}$$

The E_2 integral is thus

$$\int_{E_-}^{E_+} E_2[(m_n^2 - m_p^2 - m_e^2)c^2 - 2m_n E_2]dE_2 \equiv J(E_4) \tag{10.56}$$

and since

$$d^3\mathbf{p}_4 = 4\pi|\mathbf{p}_4|^2\, d|\mathbf{p}_4| = \frac{4\pi}{c^2}\,|\mathbf{p}_4|E_4\, dE_4 \tag{10.57}$$

we conclude that

$$\frac{d\Gamma}{dE} = \frac{1}{\hbar c^2(4\pi)^3}\left(\frac{g_w}{M_W c}\right)^4 J(E) \tag{10.58}$$

(since there is no further occasion for ambiguity, I'm eliminating the subscript on E_4; from now on E is the electron energy).

Equation (10.58) is *exact* (use it, if you like, to rederive equation (10.35), setting $m_n \to m_\mu$ and m_p, $m_e \to 0$), but $J(E)$ is a rather cumbersome function. From the definition (10.56):

$$J(E) = \frac{1}{2}(m_n^2 - m_p^2 - m_e^2)c^2(E_+^2 - E_-^2) - \frac{2m_n}{3}(E_+^3 - E_-^3) \tag{10.59}$$

where $E_\pm$ are given by equation (10.55). It pays to approximate, at this stage, recognizing that there are four small numbers here:

$$\epsilon \equiv \frac{m_n - m_p}{m_n} = 0.0014, \qquad \delta \equiv \frac{m_e}{m_n} = 0.0005,$$
$$\eta \equiv \frac{E_4}{m_n c^2} \quad (\delta < \eta < \epsilon), \qquad \phi \equiv \frac{|\mathbf{p}_4|}{m_n c} \quad (0 < \phi < \eta) \tag{10.60}$$

(The last of these is not independent, of course: $\phi^2 = \eta^2 - \delta^2$.) Expanding to lowest order (Problem 10.5), we obtain

$$J \cong 4m_n^4 c^6 \eta\phi(\epsilon - \eta)^2 = \frac{4}{c^2} E\sqrt{E^2 - m_e^2 c^4}[(m_n - m_p)c^2 - E]^2 \tag{10.61}$$

So the distribution of electron energies is given by

$$\frac{d\Gamma}{dE} = \frac{1}{\pi^3\hbar}\left(\frac{g_w}{2M_W c^2}\right)^4 E\sqrt{E^2 - m_e^2 c^4}[(m_n - m_p)c^2 - E]^2 \tag{10.62}$$

The experimental results are shown in Figure 10.2. The electron energies range from $m_e c^2$ up to about $(m_n - m_p)c^2$ (Problem 10.6). Integrating over E, we get the total decay rate (Problem 10.7):

$$\Gamma = \frac{1}{4\pi^3\hbar}\left(\frac{g_w}{2M_W c^2}\right)^4 (m_e c^2)^5$$
$$\times\left[\frac{1}{15}(2a^4 - 9a^2 - 8)\sqrt{a^2 - 1} + a\ln(a + \sqrt{a^2 - 1})\right] \tag{10.63}$$

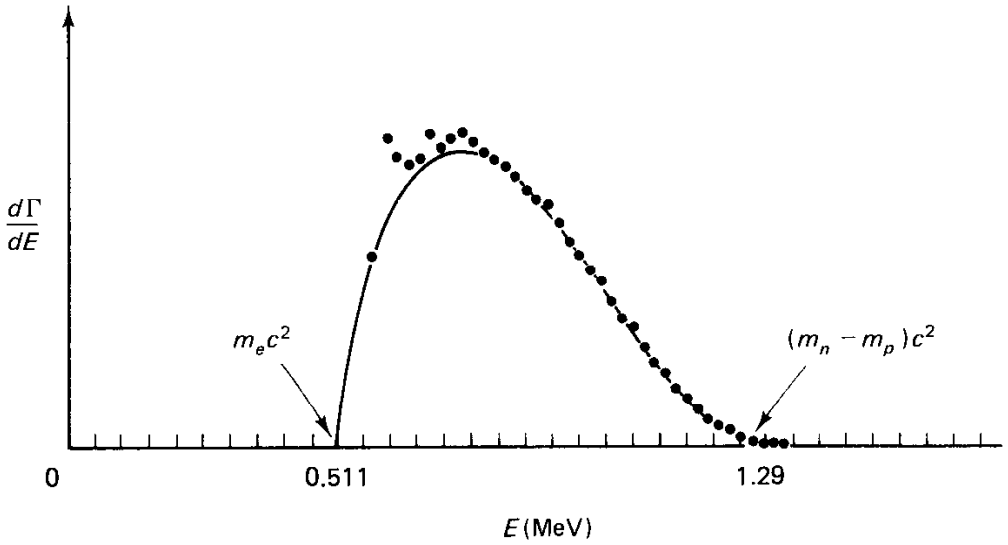


Figure 10.2 Electron energy distribution from neutron beta decay. (Solid line is the theoretical curve; dots are experimental data.) [*Source:* C. J. Christensen et al., *Phys. Rev.* **D5**, 1628 (1972), Figure 4.]

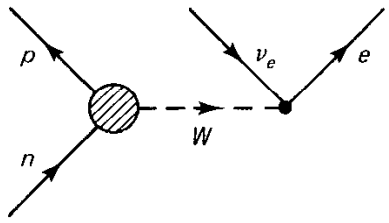
where
$$a \equiv \frac{m_n - m_p}{m_e} \tag{10.64}$$

Putting in the numbers, we find (Problem 10.8)

$$\tau = \frac{1}{\Gamma} = 1316 \text{ sec} \tag{10.65}$$

This is in the right ball park, as they say: The experimental neutron lifetime* is 898 ± 16 sec, and given that weak decays range from 15 min down to 10^{-13} sec, perhaps we should be pleased to get the right order of magnitude. But why isn't the agreement *perfect?*

The main problem is that we have treated the proton and neutron as though they were simple point particles, interacting with the W in exactly the same way as leptons do. To be honest about it, we should go back to the beginning, admit that we do not really know how the W couples to composite structures, draw in a blob on the Feynman diagram (to symbolize our ignorance)



and express the amplitude in terms of various unknown "form factors," whose

* The number is taken from the Particle Data Booklet. Free neutrons are hard to work with, and the "official" neutron lifetime has changed substantially over time, dropping by more than 10% in the last 15 years. Note also that nuclear physicists tend to list the *half-life* ($t_{1/2} = \tau \ln 2$), and beta-decay specialists often quote the "comparative half-life"—the so-called "*ft*" value—which has certain kinematic and Coulombic factors removed. (For the neutron the correction factor f is about 1.7.) All this is just to warn you that the numbers quoted in the literature for the neutron "lifetime" are scattered all over the map, and it pays to read the fine print and check the date.

structure is limited only by Lorentz covariance—just as we did in Chapter 8 for the proton-*photon* vertex. Only when a mature QCD can provide us with the detailed structure of the nucleons will we be in a position to perfect the neutron lifetime calculation.

And yet, the Mott formula works well for low-energy electron-proton scattering: Why does essentially the same procedure give us the right answer in electrodynamics, but not in the weak interactions? In both cases the wavelength of the "probe" (γ or W, as the case may be) is much larger than the diameter of the "target" (p or n) (see Problem 10.9); the nucleon's internal structure is not "resolved," and it behaves as a point particle. The *crucial* question, though, is: *What is the net coupling strength of this object?* Of course, the net *charge* of the proton is simply e (the same as for the μ^+, say.) It doesn't *matter* what complicated processes are going on inside—valence quarks emitting virtual gluons, gluons producing quark-antiquark pairs, these "sea" quarks recombining, and so on—because *all this frenzied activity conserves charge.* From the perspective of a long wavelength photon it just looks like a point, and the net charge of the composite nucleon is just the sum of the charges of the valence quarks. But there is no a priori reason to suppose that the same applies to the *weak* coupling; when a gluon splits into a quark-antiquark pair, the net contribution of this pair to the weak coupling may *not* be zero—who knows? To account for this, we make the following replacement in the $n \to p + W$ vertex factor:

$$(1 - \gamma^5) \to (c_V - c_A\gamma^5) \tag{10.66}$$

where c_V is the correction to the vector "weak charge," and c_A is the correction to the axial vector "weak charge."* Now, the same basic process, $n \to p + e + \bar{\nu}_e$, occurs not only for the free neutron, but also within radioactive nuclei, so we have in principle many independent opportunities to measure c_V and c_A.† The experimental results are as follows:

$$c_V = 1.000 \pm 0.003, \qquad c_A = 1.26 \pm 0.02 \tag{10.67}$$

Surprisingly, the *vector* weak charge is *not* modified by the strong interactions within the nucleon. Presumably, like electric charge, it is "protected" by a conservation law; we call this the "Conserved Vector Current" (CVC) hypothesis. Even the axial term is not altered *much;* evidently, it is "almost" conserved. We call this the "Partially Conserved Axial Current" (PCAC) hypothesis.

The effect of the substitution (10.66) on the neutron lifetime is something you can calculate for yourself, if you have the stamina; to good approximation, the decay rate is increased by a factor of

$$\tfrac{1}{4}(c_V^2 + 3c_A^2) = 1.44 \tag{10.68}$$

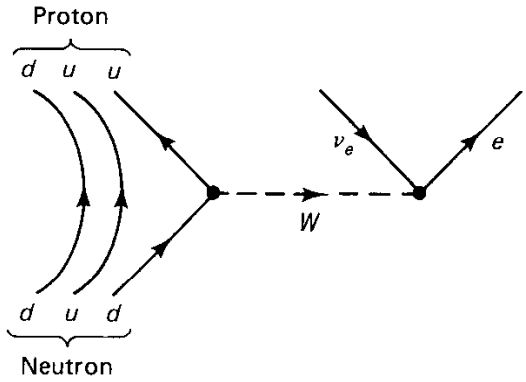
and the lifetime is decreased in the same ratio:

* c_V and c_A are related to the $q^2 = 0$ limits of the corresponding weak form factors.

† A particular favorite is $^{14}O \to {}^{14}N$, which is known (from the observed spin and parity of the initial and final states) to involve only *vector* coupling. It affords a direct and relatively precise measure of c_V.

$$\tau = \frac{1316 \text{ sec}}{1.44} = 914 \text{ sec} \tag{10.69}$$

This is now within the error bars of the experimental value. Unfortunately, the agreement is deceptive, for there is yet *another* correction to be made. The underlying quark process here is $d \rightarrow u + W$ (with two spectators):



and this quark vertex carries a factor of $\cos \theta_C$, where

$$\theta_C = 13.1° \tag{10.70}$$

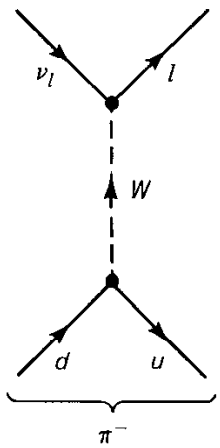
is the Cabibbo angle. I'll have more to say about this in the next section, but the essential point for now is that our theoretical value for the neutron lifetime, corrected for nonconservation of the axial charge and modified by the Cabibbo angle, is

$$\tau = \frac{914 \text{ sec}}{\cos^2 \theta_C} = 963 \text{ sec} \tag{10.71}$$

Two steps forward, one step back!*

10.4 DECAY OF THE PION

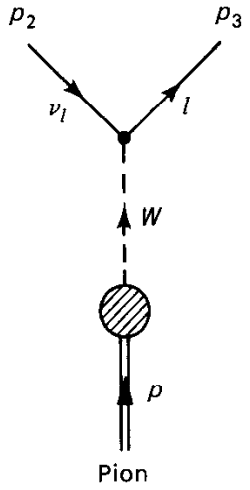
According to the quark model, the decay of a charged pion ($\pi^- \rightarrow l^- + \bar{\nu}_l$, where l is a muon or an electron) is really a *scattering* event in which the incident quarks happen to be bound together:



* This isn't the end of the story; there is a small Coulomb correction, (due to the attraction of the electron and proton in the final state); there is presumably some q^2 dependence in the form factors even near $q^2 = 0$; and there may yet be inaccuracies in the experimental data. In particular,

In this sense it is a weak-interaction analog to positronium decay ($e^+ + e^- \to \gamma + \gamma$) or η_c decay ($c + \bar{c} \to g + g$)—electromagnetic and strong processes, respectively. We *could* analyze it this way, following the methods of Example 7.8 and Chapter 9, Section 9.3 (see Problem 10.10.), but in the end we would be stuck with a factor of $|\psi(0)|^2$, and at this stage we have no idea what the wave function (ψ) of the quarks within a pion looks like. Given that such a calculation will carry this undetermined multiplicative factor *anyway*, it is simpler to proceed as follows.

Redraw the Feynman diagram, with a blob to represent the coupling of π^- to W^-:



We may not know how the W couples to the pion, but we *do* know how it couples to the leptons, so the amplitude must have the general form

$$\mathcal{M} = \frac{g_w^2}{8(M_W c)^2}[\bar{u}(3)\gamma_\mu(1 - \gamma^5)v(2)]F^\mu \tag{10.72}$$

where F^μ is a "form factor" describing the $\pi \to W$ blob. It has to be a four-vector, to contract with the γ_μ in the lepton factor. But the pion has spin zero; the only vector associated with it, out of which we might construct F^μ, is its momentum, p^μ.* (I won't bother with a subscript on the pion's momentum: $p \equiv p_1$.) So F^μ must be some scalar quantity times p^μ:

$$F^\mu = f_\pi p^\mu \tag{10.73}$$

(In principle, f_π is a function of p^2—the only available scalar—but since the pion is on its mass shell, $p^2 = m_\pi^2 c^2$ and hence f_π is simply a fixed number. We call it the "pion decay constant".)†

---

the neutron lifetime is very sensitive to uncertainties in c_A. But we are within 6% of the experimental result, and it is time to move on.

* Notice that we introduce the (weak) *pion* form factor at the level of $\mathcal{M}$, whereas for the (electromagnetic) proton form factors and structure functions we waited until the $\langle|\mathcal{M}|^2\rangle$ stage. The reason is that the proton has a spin, and we would have to include that in the roster of available vectors; it is only after we have *averaged* over the spins that the list reduces to two, and the problem becomes manageable. The pion, however, has no spin, so we can afford to introduce the form factor directly in $\mathcal{M}$, where it is only a *vector* quantity, instead of a tensor.

† The pion decay constant evidently contains the factor of $|\psi(0)|^2$ alluded to earlier; we have simply wrapped our ignorance in a more convenient package. (See Problem 10.10.)

Summing over the outgoing spins, we have

$$\langle|\mathcal{M}|^2\rangle = \left[\frac{f_\pi}{8}\left(\frac{g_w}{M_W c}\right)^2\right]^2 p_\mu p_\nu \, Tr[\gamma^\mu(1-\gamma^5)\not{p}_2\gamma^\nu(1-\gamma^5)(\not{p}_3 + m_l c)]$$

$$= \frac{1}{8}\left[f_\pi\left(\frac{g_w}{M_W c}\right)^2\right]^2 [2(p\cdot p_2)(p\cdot p_3) - p^2(p_2\cdot p_3)] \tag{10.74}$$

[the trace was already calculated in equation (10.9)]. But $p = p_2 + p_3$, so

$$p\cdot p_2 = p_2\cdot p_3, \qquad p\cdot p_3 = m_l^2 c^2 + p_2\cdot p_3 \tag{10.75}$$

and $$p^2 = p_2^2 + p_3^2 + 2p_2\cdot p_3, \quad \text{so that} \quad 2p_2\cdot p_3 = (m_\pi^2 - m_l^2)c^2 \tag{10.76}$$

Thus $$\langle|\mathcal{M}|^2\rangle = \left(\frac{g_w}{2M_W}\right)^4 f_\pi^2 m_l^2 (m_\pi^2 - m_l^2) \tag{10.77}$$

(a *constant*).

The decay rate is given by the standard formula (6.32):

$$\Gamma = \frac{|\mathbf{p}_2|}{8\pi\hbar m_\pi^2 c}\langle|\mathcal{M}|^2\rangle \tag{10.78}$$

and the outgoing momentum is [see eq. (6.31) or Problem 3.16]

$$|\mathbf{p}_2| = \frac{c}{2m_\pi}(m_\pi^2 - m_l^2) \tag{10.79}$$

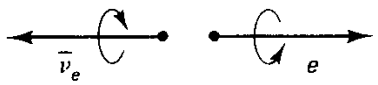
Thus $$\Gamma = \frac{f_\pi^2}{\pi\hbar m_\pi^3}\left(\frac{g_w}{4M_W}\right)^4 m_l^2 (m_\pi^2 - m_l^2)^2 \tag{10.80}$$

Of course, without knowing the decay constant, f_π, we cannot calculate the pion lifetime.* Nevertheless, we *are* able to determine the branching ratio

$$\frac{\Gamma(\pi^- \to e^- + \bar{\nu}_e)}{\Gamma(\pi^- \to \mu^- + \bar{\nu}_\mu)} = \frac{m_e^2(m_\pi^2 - m_e^2)^2}{m_\mu^2(m_\pi^2 - m_\mu^2)^2} = 1.28 \times 10^{-4} \tag{10.81}$$

The experimental number is $1.23 \pm 0.02 \times 10^{-4}$. At first glance this is a very surprising result, for it predicts (correctly) that the pion prefers the *muon* mode, in spite of the fact that the electron is much lighter. Phase space considerations favor decays for which the mass decrease is as large as possible, and unless some conservation law intervenes, we ordinarily find that the lightest final state is the most common one. Pion decay is the notorious exception, and it calls for some special dynamical explanation. A clue is suggested by equation (10.80): Notice that if the electron were *massless*, the $\pi^- \to e^- + \bar{\nu}_e$ mode would be forbidden *completely*. Can we understand this limiting case? Yes. The pion has spin 0, so the electron and the antineutrino must emerge with opposite spins, and hence equal helicities:

* It is a rather striking fact that if you put in $f_\pi = m_\pi c$ (or, better yet, $m_\pi c \cos\theta_C$) you come out very close to the $\pi^- \to \mu^- + \bar{\nu}_\mu$ lifetime, but I know of no persuasive theoretical justification for this ansatz.



The antineutrino is always right-handed, so the electron must be right-handed as well. But if the electron were truly massless, then (like the neutrino) it would only exist as a left-handed particle. More precisely, the $1 - \gamma^5$ in the weak vertex factor would couple only to left-handed electrons, just as it couples only to left-handed neutrinos (see Problem 10.11). That's why if the electron were massless, the decay $\pi^- \to e^- + \bar{\nu}_e$ could not occur at *all*, and why (the physical electron being very *close* to massless) the decay is so heavily suppressed.

10.5 CHARGED WEAK INTERACTIONS OF QUARKS

In the case of leptons, the coupling to $W^\pm$ takes place strictly *within a particular generation:*

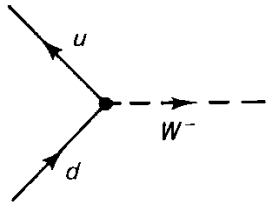
$$\begin{pmatrix} \nu_e \\ e \end{pmatrix}, \quad \begin{pmatrix} \nu_\mu \\ \mu \end{pmatrix}, \quad \begin{pmatrix} \nu_\tau \\ \tau \end{pmatrix} \qquad \text{(lepton generations)}$$

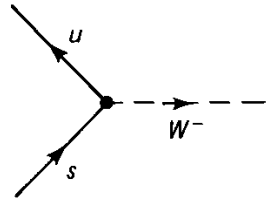
That is, $e^- \to \nu_e + W^-$, $\mu^- \to \nu_\mu + W^-$, $\tau^- \to \nu_\tau + W^-$, but there is no cross-generational coupling, of the form $e^- \to \nu_\mu + W^-$, for example. This observation, in fact, is enshrined in the laws of conservation of electron number, muon number, and tau number. The coupling of W to *quarks* is not quite so simple, for although the generation structure is similar

$$\begin{pmatrix} u \\ d \end{pmatrix}, \quad \begin{pmatrix} c \\ s \end{pmatrix}, \quad \begin{pmatrix} t \\ b \end{pmatrix} \qquad \text{(quark generations)}$$

the weak interactions do not strictly respect their separate identities. There are, to be sure, interactions of the form $d \to u + W^-$ (the process that underlies neutron decay, $n \to p + e + \bar{\nu}_e$), but there exist as well cross-generational couplings, such as $s \to u + W^-$ (seen, for example, in the decay $\Lambda \to p + e + \bar{\nu}_e$). Indeed, if this were *not* the case, we would have three absolute "flavor-conservation" laws: conservation of "upness-plus-downness," "charm-plus-strangeness," and "truth-plus-beauty"—analogous to the three lepton number conservation laws. As a result, the lightest strange particle (K^-) would be absolutely stable, and so would the B meson (the lightest beautiful particle); our world would be a quite different place.

In 1963 (when u, d, and s were the only quarks known) Cabibbo[1] suggested that the $d \to u + W^-$ vertex carries a factor of $\cos\theta_C$, whereas $s \to u + W^-$ carries a factor of $\sin\theta_C$; apart from that they are identical to the leptonic couplings [eq. (10.5)]:





$$\frac{-ig_w}{2\sqrt{2}}\gamma^\mu(1-\gamma^5)\cos\theta_C \qquad \frac{-ig_w}{2\sqrt{2}}\gamma^\mu(1-\gamma^5)\sin\theta_C \qquad (10.82)$$

The strangeness-changing process ($s \to u + W^-$) is conspicuously *weaker* than the strangeness-conserving one ($d \to u + W^-$), so evidently the "Cabibbo angle" θ_C is rather small. Experimentally

$$\theta_C = 13.1° \tag{10.83}$$

The weak interactions *almost* respect quark generations . . . but not *quite.*

EXAMPLE 10.2 Leptonic Decays

Consider the decay $K^- \to l^- + \bar{\nu}_l$, where l is an electron or a muon. This is the analog to π^- decay (Sect. 10.4), but now the quark vertex is $s + \bar{u} \to W^-$, instead of $d + \bar{u} \to W^-$. From equation (10.80) we have

$$\Gamma = \frac{f_K^2}{\pi \hbar m_K^3}\left(\frac{g_w}{4M_W}\right)^4 m_l^2(m_K^2 - m_l^2)^2$$

The coupling strength is presumably the same as before, except that where f_π contained a factor of $\cos\theta_C$, f_K carries a factor of $\sin\theta_C$. Accordingly,

$$\frac{\Gamma(K^- \to l^- + \bar{\nu}_l)}{\Gamma(\pi^- \to l^- + \bar{\nu}_l)} = \tan^2\theta_C\left(\frac{m_\pi}{m_K}\right)^3\left(\frac{m_K^2 - m_l^2}{m_\pi^2 - m_l^2}\right)^2 \tag{10.84}$$

Putting in the appropriate numbers, we get 0.96 for the muon mode ($l = \mu$) and 0.19 for the electron mode ($l = e$). [The experimental ratios are 1.34 and 0.26, respectively, corresponding to a Cabibbo angle of 15.4°. These decays are pure axial-vector, and as we discovered earlier—see eq. (10.67)—perfect agreement is not to be expected.]

For obvious reasons, processes of the kind considered in Example 10.2 are called *leptonic* decays. There also exist *semileptonic* decays, such as $\pi^- \to \pi^0 + e^- + \bar{\nu}_e$, $\bar{K}^0 \to \pi^+ + \mu^- + \bar{\nu}_\mu$ (Fig. 10.3a), or for that matter the beta decay of the neutron: $n \to p^+ + e^- + \bar{\nu}_e$. Finally, there are *nonleptonic* weak interactions, such as $K^- \to \pi^0 + \pi^-$ or $\Lambda \to p^+ + \pi^-$ (Fig. 10.3b). Generally speaking, the latter are the hardest to analyze, because there is strong interaction contamination at both ends of the W line. We shall not consider nonleptonic weak processes in this book.[2]

EXAMPLE 10.3 Semileptonic Decays

In the case of neutron decay ($n \to p + e + \bar{\nu}_e$) the basic quark process is $d \to u + W^-$ (with two spectators). However, there are *two* d quarks in the neutron, and *either one* could couple to the W; the net amplitude for the process is the *sum.* The simplest way to keep track of the numbers is to use the quark wave functions of Chapter 5, Section 5.9; the flavor states ψ_{12}, for instance, give $n = (ud - du)d/\sqrt{2}$, from which (with $d \to u$) we get $[(uu - uu)d + (ud - du)u]/\sqrt{2} = (ud - du)u/\sqrt{2} = p$. The overall coefficient is then simply $\cos\theta_C$ (as I claimed at the end of Section 10.3).

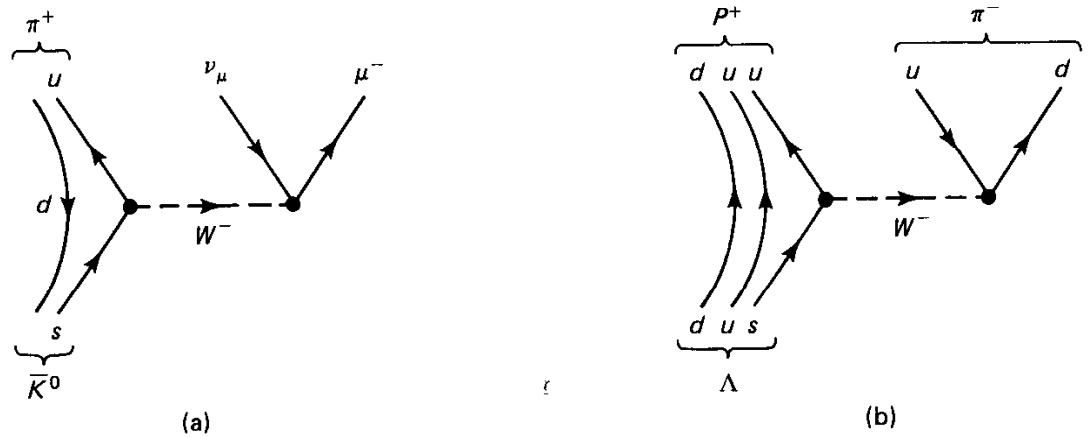


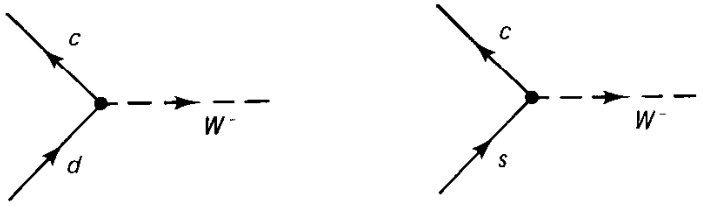
Figure 10.3 (a) A typical semileptonic decay ($\bar{K}^0 \to \pi^+ + \mu^- + \bar{\nu}_u$). (b) A typical nonleptonic weak decay ($\Lambda \to p^+ + \pi^-$).

By contrast, in the decay $\Sigma^0 \to \Sigma^+ + e + \bar{\nu}_e$, the quark process is still $d \to u$, but here $\Sigma^0 = [(us - su)d + (ds - sd)u]/2 \to [(us - su)u + (us - su)u]/2 = (us - su)u = \sqrt{2}\Sigma^+$, and hence the amplitude carries a factor of $\sqrt{2}\cos\theta_C$.* The decay rate is given by equation (10.63), which reduces (in the case $a \gg 1$) to the form

$$\Gamma = \frac{1}{30\pi^3\hbar}\left(\frac{g_w}{2M_Wc^2}\right)^4(\Delta mc^2)^5X^2$$

where Δm is the baryon mass decrease and X is the Cabibbo factor ($\cos\theta_C$, for neutron decay; $\sqrt{2}\cos\theta_C$, for $\Sigma^0 \to \Sigma^+ + e + \bar{\nu}_e$; etc.). I'll let you work out the numbers for yourself (Problem 10.13).†

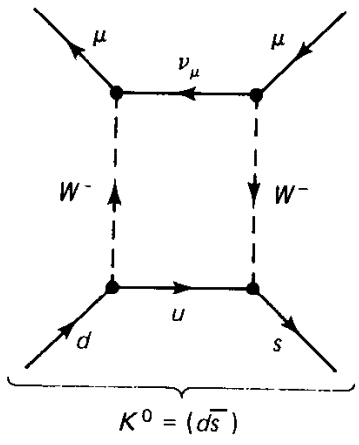
Cabibbo's theory was very successful in correlating dozens of decay rates, but there remained a disturbing problem: this picture allowed the K^0 to decay into a $\mu^+\mu^-$ pair (see Fig. 10.4). The amplitude should be proportional to $\sin\theta_C\cos\theta_C$, but the calculated rate is far greater than the experimental limits allow. A solution to this paradox was proposed in 1970 by Glashow, Iliopoulos, and Maiani (GIM).[3] They introduced a fourth quark (c)—remember, this was four years before the "November Revolution" produced the first direct experimental evidence for charm—whose couplings to s and d carry factors of $\cos\theta_C$ and $-\sin\theta_C$, respectively:



$$\frac{-ig_w}{2\sqrt{2}}\gamma^\mu(1-\gamma^5)(-\sin\theta_C) \qquad \frac{-ig_w}{2\sqrt{2}}\gamma^\mu(1-\gamma^5)\cos\theta_C \tag{10.85}$$

* Actually, there is a technical difference here: The active quark is bound to the spectator in a *spin singlet* state. Fortunately, this does not affect the lifetime.

† This procedure includes only the valence quarks, and hence is insensitive to the nonconservation of the axial coupling. As we found in equation (10.68), PCAC can lead to a correction of nearly 50%, so one does not expect fine precision in the lifetimes. Cabibbo's theory included a way of calculating the axial couplings, but I shall not go into that here.



Fignre 10.4 The decay $K^0 \to \mu^+ + \mu^-$.

In the "GIM mechanism," the diagram in Fig. 10.4 is canceled by the corresponding diagram with c in place of u (Fig. 10.5), for this time the amplitude is proportional to $-\sin\theta_C \cos\theta_C$.*

The Cabibbo–GIM scheme invites a simple interpretation: Instead of the physical quarks d and s, the "correct" states to use in the weak interactions are d' and s', given by

$$d' = d\cos\theta_C + s\sin\theta_C, \qquad s' = -d\sin\theta_C + s\cos\theta_C \tag{10.86}$$

or, in matrix form

$$\begin{pmatrix} d' \\ s' \end{pmatrix} = \begin{pmatrix} \cos\theta_C & \sin\theta_C \\ -\sin\theta_C & \cos\theta_C \end{pmatrix} \begin{pmatrix} d \\ s \end{pmatrix} \tag{10.87}$$

The W's couple to the "Cabibbo-rotated" states

$$\begin{pmatrix} u \\ d' \end{pmatrix}, \qquad \begin{pmatrix} c \\ s' \end{pmatrix}$$

in exactly the same way that they couple to lepton pairs, $\begin{pmatrix} \nu_e \\ e \end{pmatrix}$ and $\begin{pmatrix} \nu_\mu \\ \mu \end{pmatrix}$; their couplings to the *physical* particles (states of specific flavor) are then given by

$$\begin{pmatrix} u \\ d' \end{pmatrix} = \begin{pmatrix} u \\ d\cos\theta_C + s\sin\theta_C \end{pmatrix}, \qquad \begin{pmatrix} c \\ s' \end{pmatrix} = \begin{pmatrix} c \\ -d\sin\theta_C + s\cos\theta_C \end{pmatrix}$$

That is, $d \to u + W^-$ carries a factor $\cos\theta_c$, $s \to u + W^-$ a factor $\sin\theta_c$, and so on.†

* The cancellation is not *perfect,* because the mass of the c is not the same as the mass of the u. However, these virtual particles are so far off the mass shell that both propagators are essentially just iq/q^2. (In calculating $\mathcal{M}$ we shall be integrating over the one remaining internal momentum which is not fixed by the conservation laws. This is essentially the momentum "circulating around the loop." Because of the two W propagators, the main contribution will come in the region of the W mass, which is so much greater than the c or u mass that the latter can, to good approximation, be neglected. Actually, the decay *does* occur, it's just extremely slow, and if you *include* the effects of u/c mass difference, the calculation is consistent with the observed rate.)

† It is purely conventional that we "rotate" d and s, rather than u and c; we could accomplish the same purpose by introducing $u' = u\cos\theta_C - c\sin\theta_C$ and $c' = u\sin\theta_C + c\cos\theta_C$. Incidentally,

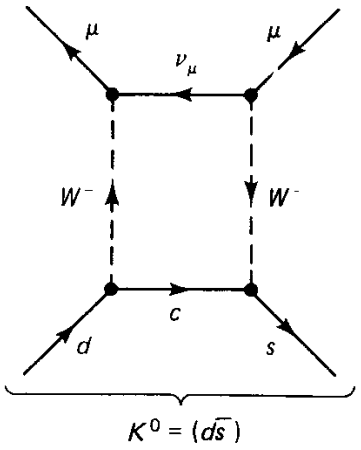


Figure 10.5 The GIM mechanism. [This diagram cancels (10.4). Note the virtual c quark replacing the u.]

At the time, the GIM mechanism seemed a little extravagant—introducing a new quark just to fix a rather esoteric technical defect in a largely untested theory. But the skeptics were silenced by the discovery of the $\psi(c\bar{c})$ in 1974. Meanwhile, Kobayashi and Maskawa[4] had generalized the Cabibbo–GIM scheme to handle *three* generations of quarks.* The "weak interaction generations,"

$$\begin{pmatrix} u \\ d' \end{pmatrix}, \quad \begin{pmatrix} c \\ s' \end{pmatrix}, \quad \begin{pmatrix} t \\ b' \end{pmatrix} \qquad \text{(weak interaction quark generations)} \qquad (10.88)$$

are related to the physical quark states by the *Kobayashi–Maskawa matrix:*

$$\begin{pmatrix} d' \\ s' \\ b' \end{pmatrix} = \begin{pmatrix} U_{ud} & U_{us} & U_{ub} \\ U_{cd} & U_{cs} & U_{cb} \\ U_{td} & U_{ts} & U_{tb} \end{pmatrix} \begin{pmatrix} d \\ s \\ b \end{pmatrix} \qquad (10.89)$$

where U_{ud}, for example, specifies the coupling of u to $d(d \to u + W^-)$. There are nine entries in the KM matrix, but they are not all independent (see Problem 10.14). U can be reduced to a kind of "canonical form," in which there remain just three "generalized Cabibbo angles," $(\theta_1, \theta_2, \theta_3)$ and one phase factor (δ):[5]

$$U = \begin{pmatrix} c_1 & s_1c_3 & s_1s_3 \\ -s_1c_2 & c_1c_2c_3 - s_2s_3\,e^{i\delta} & c_1c_2s_3 + s_2c_3\,e^{i\delta} \\ -s_1s_2 & c_1s_2c_3 + c_2s_3\,e^{i\delta} & c_1s_2s_3 - c_2c_3\,e^{i\delta} \end{pmatrix} \qquad (10.90)$$

Here c_i stands for $\cos\theta_i$, and s_i for $\sin\theta_i$. If $\theta_2 = \theta_3 = 0$, the third generation does not mix with the other two, and we recover the original Cabibbo–GIM picture, with $\theta_1 = \theta_C$. However, there is compelling evidence (namely, the ob-

---

you might ask why a similar rotation does not occur in the *lepton* sector. The answer is, it already *has*—or rather, we'd never notice it if it *did*. The point is that all neutrinos are massless, and any linear combination of them is *still* massless. So there is no "tag" to identify the "unrotated" states, and what we call ν_e, for example, *is* "the neutrino paired with the e in weak interactions," just as d' is "the quark paired with u in weak interactions."

* It is interesting to note that Kobayashi and Maskawa proposed a third quark generation before the *second* was complete, and long before there was any experimental evidence for a third generation of leptons *or* quarks. They were motivated by a desire to explain *CP* violation within the Cabibbo–GIM scheme. It turned out that for this purpose they needed a complex number in the "rotation" matrix (10.87), but such a term could always be eliminated, by suitable redefinition of the quark phases, unless they went to a 3×3 matrix, and hence to three generations. It remains to be seen whether the δ term in the KM matrix is the actual source of *CP* violation in nature, but the precocious prediction of three generations has, of course, been richly confirmed.

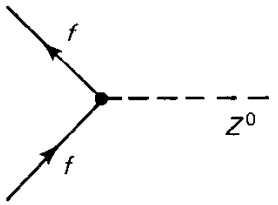
served decay of the $B^-(b\bar{u})$ meson) for *some* third-generation mixing, although it must be fairly small in order to account for the success of the original Cabibbo–GIM scheme. The Standard Model offers no insight into the KM matrix (indeed, this is one of the most conspicuous weaknesses of the Standard Model); for the moment, we simply take the values of the matrix elements from experiment. There is a small industry devoted to the accurate measurement of these parameters. So far, only their *magnitudes* are known with any precision:[6]

$$|U_{ij}| = \begin{pmatrix} 0.9705 \text{ to } 0.9770 & 0.21 \text{ to } 0.24 & 0. \text{ to } 0.014 \\ 0.21 \text{ to } 0.24 & 0.971 \text{ to } 0.973 & 0.036 \text{ to } 0.070 \\ 0. \text{ to } 0.024 & 0.036 \text{ to } 0.069 & 0.997 \text{ to } 0.999 \end{pmatrix} \tag{10.91}$$

The third generation mixing (the off-diagonal elements in the third row and column) turns out to be very small indeed, as we learned from the surprisingly long lifetime of the B meson (10^{-12} sec).

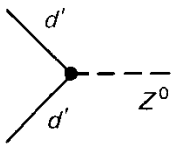
10.6 NEUTRAL WEAK INTERACTIONS

In 1958, Bludman[7] suggested that there might exist *neutral* weak interactions, mediated by an uncharged partner of the W's—the Z^0:

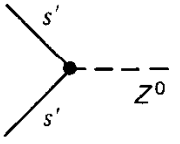


Here f stands for any lepton or any quark. Notice, however, that the *same* Fermion comes out as went in (just as in QED and QCD). We do not allow couplings of the form $\mu^- \to e^- + Z^0$, for example (this would violate conservation of muon and electron number), nor of the form $s \to d + Z^0$ (such a strangeness-changing neutral process would lead to $K^0 \to \mu^+ + \mu^-$, which, as I have already remarked, is strongly suppressed).* In 1961, Glashow[8] published the first paper on unification of weak and electromagnetic interactions; his theory required the existence

* In the case of *neutral* processes, it doesn't matter whether you use the physical states (d, s, b) or the "Cabibbo-rotated" states (d', s', b'). Schematically, the argument runs as follows:



gives $\mathcal{M} \sim \bar{d}'d' = \bar{d}d \cos^2\theta_C + \bar{s}s \sin^2\theta_C + (\bar{d}s + \bar{s}d) \sin\theta_C \cos\theta_C$



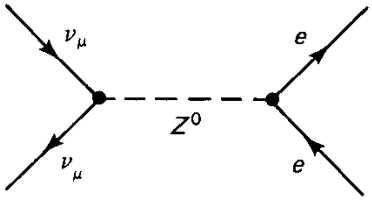
gives $\mathcal{M} \sim \bar{s}'s' = \bar{d}d \sin^2\theta_C + \bar{s}s \cos^2\theta_C - (\bar{d}s + \bar{s}d) \sin\theta_C \cos\theta_C$

So the *sum* of the two is $\mathcal{M}_{\text{tot}} \sim \bar{d}'d' + \bar{s}'s' = \bar{d}d + \bar{s}s$. Thus the *net* amplitude, once both diagrams are combined, is the same whichever states we use. (The same argument generalizes to three generations, as long as the KM matrix is unitary.)

of neutral weak processes, and specified their structure (see Sect. 10.7). In 1967, Weinberg and Salam[9] formulated Glashow's model as a "spontaneously broken gauge theory," and in 1971, 't Hooft[10] demonstrated that the Glashow–Weinberg–Salam scheme is renormalizable. Thus there were increasingly persuasive theoretical reasons for thinking that neutral weak interactions occur in nature, but for a long time there were no *experimental* data to support this hope. Finally, in 1973,[11] a bubble chamber photograph at CERN (Fig. 10.6) revealed the first convincing evidence for the reaction

$$\bar{\nu}_\mu + e \longrightarrow \bar{\nu}_\mu + e$$

suggesting mediation by the Z^0:



The same series of experiments also witnessed the corresponding neutrino-*quark* process, in the form of inclusive neutrino-nucleon scattering:

$$\bar{\nu}_\mu + N \longrightarrow \bar{\nu}_\mu + X$$
$$\nu_\mu + N \longrightarrow \nu_\mu + X$$

Their cross sections were about a third as large as those of the related *charged* events ($\bar{\nu}_\mu + N \longrightarrow \mu^+ + X$ and $\nu_\mu + N \longrightarrow \mu^- + X$), indicating that this was indeed a new kind of weak interaction, and not simply a higher-order process

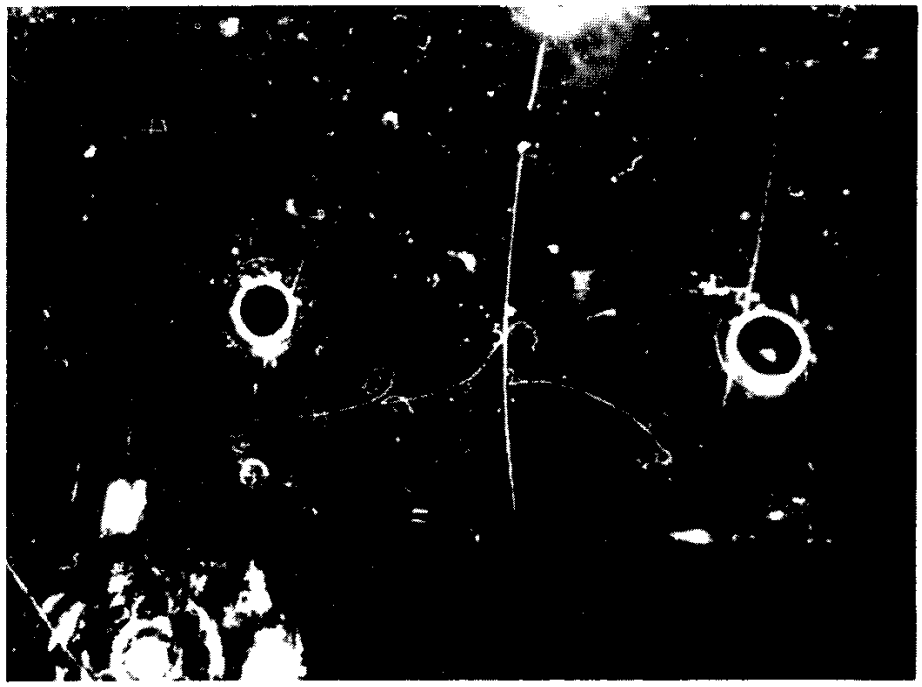
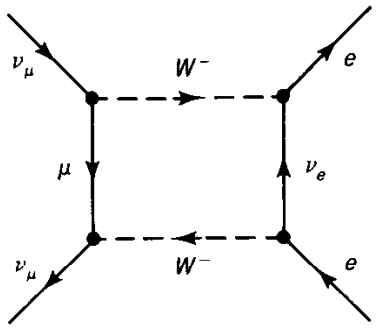


Figure 10.6 The first picture of a neutral weak process ($\bar{\nu}_\mu + e^- \longrightarrow \bar{\nu}_\mu + e^-$). The neutrino enters from the left (leaving no track), and strikes an electron, which moves off horizontally to the right, emitting two photons (which show up in the picture only when they subsequently produce electron-positron pairs) as it slows down and spirals inward in the superimposed magnetic field. (Photo courtesy CERN.)



(which would yield a far smaller cross section). The CERN results came as welcome encouragement to electroweak theorists, who had been out on a limb now for several years. Meanwhile a series of deep inelastic neutrino scattering experiments was conducted (also at CERN) that confirmed not only the basic structure of charged and neutral weak interactions, but also the quark-parton model itself—which carries over directly from *electron*-nucleon scattering, (an electromagnetic process) to *neutrino*-nucleon scattering (a weak process).[12]

As we have seen, the coupling of quarks and leptons to $W^\pm$ is a universal "V–A" form; the vertex factor is always

$$\frac{-ig_w}{2\sqrt{2}}\gamma^\mu(1-\gamma^5) \qquad (W^\pm \text{ vertex factor}) \tag{10.92}$$

(It is true that the axial coupling to *composite* structures, such as the proton, is modified, but that is a result of strong interaction contamination—the underlying quark process is pure V–A). The coupling of the Z^0 is not so simple:

$$\frac{-ig_z}{2}\gamma^\mu(c_V^f - c_A^f\gamma^5) \qquad (Z^0 \text{ vertex factor}) \tag{10.93}$$

where g_z is the neutral coupling constant, and the coefficients c_V^f and c_A^f depend on the particular quark or lepton (f) involved. In the GWS model, *all* these numbers are determined by a single fundamental parameter θ_w, called the "weak mixing angle" (or "Weinberg angle"). See Table 10.1. Moreover, the weak coupling constants are related to the basic unit of electric charge:

$$g_w = \frac{g_e}{\sin\theta_w}, \qquad g_z = \frac{g_e}{\sin\theta_w\cos\theta_w} \tag{10.94}$$

where g_e is the electromagnetic coupling constant (in appropriate units, the charge

TABLE 10.1 NEUTRAL VECTOR AND AXIAL VECTOR COUPLINGS IN THE GWS MODEL

f	c_V	c_A
ν_e, ν_μ, ν_τ	$\frac{1}{2}$	$\frac{1}{2}$
e^-, μ^-, τ^-	$-\frac{1}{2} + 2\sin\theta_w$	$-\frac{1}{2}$
u, c, t	$\frac{1}{2} - \frac{4}{3}\sin^2\theta_w$	$\frac{1}{2}$
d, s, b	$-\frac{1}{2} + \frac{2}{3}\sin^2\theta_w$	$-\frac{1}{2}$

of the positron). You'll see where these predictions come from in the following section. The Standard Model provides no way to calculate θ_w itself; like the KM matrix, its value is taken from experiment:

$$\theta_w = 28.7° \qquad (\sin^2 \theta_w = 0.23) \tag{10.95}$$

Derivation of θ_w stands as a major challenge for any theory going beyond the Standard Model. Meanwhile, the Z^0 propagator is [eq. (10.3)]

$$\frac{-i(g_{\mu\nu} - q_\mu q_\nu / M_Z^2 c^2)}{q^2 - M_Z^2 c^2} \tag{10.96}$$

In the typical case $q^2 \ll M_Z^2 c^2$, it reduces to

$$\frac{i g_{\mu\nu}}{(M_Z c)^2} \tag{10.97}$$

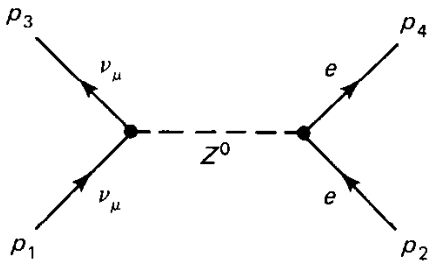
Finally, the $W^\pm$ and Z^0 masses are related by

$$M_W = M_Z \cos \theta_w \tag{10.98}$$

Equations (10.93)–(10.98) are the basic predictions of the GWS model. Given the weak mixing angle, we can now *calculate* the W and Z masses (see Problem 10.17). Their discovery by Rubbia at CERN in 1983, at $M_W = 82$ GeV/c^2 and $M_Z = 92$ GeV/c^2 (as predicted) was persuasive evidence for the GWS model.[13]

EXAMPLE 10.4 Elastic Neutrino-Electron Scattering

In Example 10.1 we calculated the cross section for the W-mediated process $\nu_\mu + e \rightarrow \nu_e + \mu$. We now consider the related Z^0-mediated reaction $\nu_\mu + e \rightarrow \nu_\mu + e$:



The amplitude is

$$\mathcal{M} = \frac{g_z^2}{8(M_Z c)^2} [\bar{u}(3)\gamma^\mu(1 - \gamma^5)u(1)][\bar{u}(4)\gamma_\mu(c_V - c_A\gamma^5)u(2)] \tag{10.99}$$

and hence

$$\begin{aligned}\langle|\mathcal{M}|^2\rangle &= 2\left(\frac{g_z}{4M_Z c}\right)^4 Tr\{\gamma^\mu(1 - \gamma^5)\not{p}_1\gamma^\nu(1 - \gamma^5)\not{p}_3\} \\ &\quad \times Tr\{\gamma_\mu(c_V - c_A\gamma^5)(\not{p}_2 + mc)\gamma_\nu(c_V - c_A\gamma^5)(\not{p}_4 + mc)\} \\ &= \frac{1}{2}\left(\frac{g_z}{M_Z c}\right)^4 \{(c_V + c_A)^2(p_1 \cdot p_2)(p_3 \cdot p_4) \\ &\quad + (c_V - c_A)^2(p_1 \cdot p_4)(p_2 \cdot p_3) - (mc)^2(c_V^2 - c_A^2)(p_1 \cdot p_3)\}\end{aligned} \tag{10.100}$$

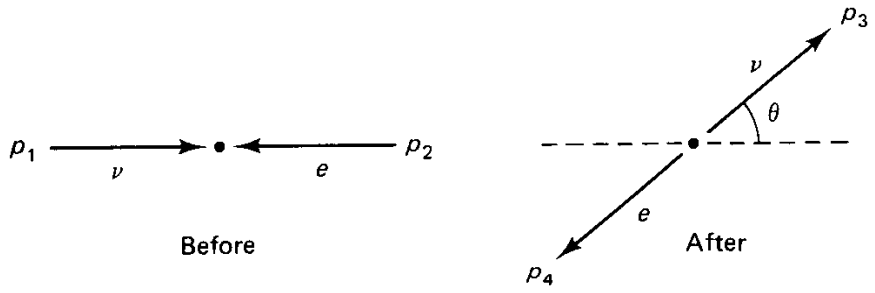


Figure 10.7 Elastic neutrino-electron scattering in the *CM*.

where m is the mass of the electron, and c_V and c_A are the neutral weak couplings for the electron. If we now go to the CM frame, and ignore the electron mass (i.e., set $m \to 0$), we find

$$\langle |\mathcal{M}|^2 \rangle = 2\left(\frac{g_z E}{M_Z c^2}\right)^4 \left[(c_V + c_A)^2 + (c_V - c_A)^2 \cos^4 \frac{\theta}{2}\right] \tag{10.101}$$

where E is the electron (or neutrino) energy, and θ is the scattering angle (Fig. 10.7). The differential scattering cross section [eq. (6.42)] is

$$\frac{d\sigma}{d\Omega} = 2\left(\frac{\hbar c}{\pi}\right)^2 \left(\frac{g_z}{4M_Z c^2}\right)^4 E^2 \left[(c_V + c_A)^2 + (c_V - c_A)^2 \cos^4 \frac{\theta}{2}\right] \tag{10.102}$$

and the total cross section (integrating over all angles) is

$$\sigma = \frac{2}{3\pi} (\hbar c)^2 \left(\frac{g_z}{2M_Z c^2}\right)^4 E^2 (c_V^2 + c_A^2 + c_V c_A) \tag{10.103}$$

Putting in the GWS values for c_V and c_A (from Table 10.1), and comparing the result of Example 10.1 [eq. (10.14)], we find that for energies substantially above the muon mass

$$\frac{\sigma(\nu_\mu + e^- \to \nu_\mu + e^-)}{\sigma(\nu_\mu + e \to \nu_e + \mu^-)} = \frac{1}{4} - \sin^2 \theta_w + \frac{4}{3} \sin^4 \theta_w = 0.09 \tag{10.104}$$

The current experimental value[14] is 0.08, which, given the 10% uncertainties in the measurements, as well as the (somewhat smaller) uncertainty in θ_w, is excellent agreement.

You might well ask why it took so long for neutral weak interactions to be detected in the laboratory; after all, 15 years separate Bludman's original speculations from the definitive experiments at CERN. The reason is that *most* neutral processes are "masked" by competing electromagnetic ones. For example, $e^+ + e^- \to \mu^+ + \mu^-$ can occur *either* by a virtual Z^0 *or* by a virtual γ (Fig. 10.8); at low energies the photon mechanism overwhelmingly dominates.* That's why

* It is interesting to note, however, that there is a weak contamination in *every* electromagnetic process, since the Z^0 couples to everything the γ does (and then some). For example, the Coulomb potential binding the electrons to the nucleus in an atom is slightly modified by Z^0 exchange, and this is observable in atomic spectra. Similarly, there is a weak contribution to electron-proton scattering. Although these effects are minute, they carry a tell-tale signature: parity violation. (See ref. 15.)

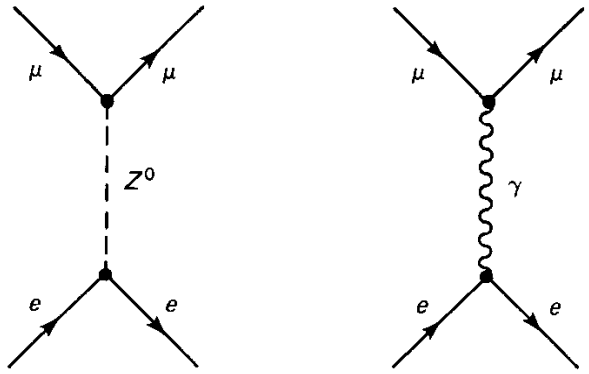


Figure 10.8 Weak and electromagnetic contributions to $e^+ + e^- \to \mu^+ + \mu^-$.

neutrino scattering was originally used to confirm the existence of neutral weak interactions; neutrinos have no electromagnetic coupling, so the weak effects are not obscured. But neutrino experiments are notoriously difficult; hence the long delay. An alternative would be to work at extremely high energies—specifically, in the neighborhood of the Z^0 mass, where the denominator of the Z^0 propagator is small, and the "weak" interaction is correspondingly large. In the early days it was hard to estimate θ_w, and hence the Z^0 mass was quite uncertain. But by the late seventies a variety of experimental data pointed to $\theta_w \approx 29°$, and hence to $M_Z = 90\ \text{GeV}/c^2$ (see Problem 10.17). This prediction was stunningly confirmed in 1983,[13] and inspired a major effort to build electron-positron colliders that would operate at the Z^0 peak. [Two such facilities, the Stanford Linear Collider (SLC) at SLAC, and LEP at CERN, are presently under construction, and scheduled to begin operation in 1987 and 1989, respectively.]

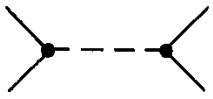
EXAMPLE 10.5 Electron-Positron Scattering Near the Z^0 Pole

Consider the process $e^+ + e^- \to f + \bar{f}$ (Fig. 10.9), where f is any quark or lepton.* This time we shall *not* use the approximate form of the Z^0 propagator [eq. (10.97)], for we are interested precisely in the regime $q^2 \approx (M_Z c)^2$. The amplitude is

$$\mathcal{M} = -\frac{g_z^2}{4(q^2 - (M_Z c)^2)}[\bar{u}(4)\gamma^\mu(c_V^f - c_A^f\gamma^5)v(3)] \times \left(g_{\mu\nu} - \frac{q_\mu q_\nu}{(M_Z c)^2}\right)[\bar{v}(2)\gamma^\nu(c_V^e - c_A^e\gamma^5)u(1)] \qquad (10.105)$$

where $q = p_1 + p_2 = p_3 + p_4$. Since we are working in the vicinity of 90 GeV, we can afford to ignore the lepton and quark masses.† In this case the second term in the propagator contributes nothing, for q_μ contracts with γ^μ to give

* *Not* an electron, however, for then we would have to include the diagram



† I assume $m_f \ll M_Z$, which is safe unless perhaps f is a top quark. All we know at present is that $m_t > 23$ GeV.

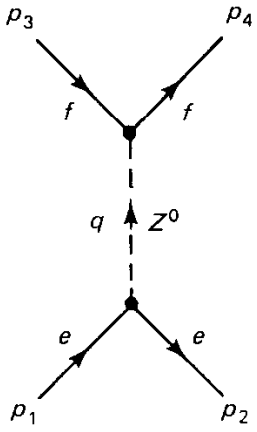


Figure 10.9 Electron-positron scattering near the Z^0 pole.

$$\bar{u}(4)\not{q}(c_V - c_A\gamma^5)v(3)$$

but $\not{q} = \not{p}_3 + \not{p}_4$, and $\bar{u}(4)\not{p}_4 = 0$ (the Dirac equation (7.94) for a massless particle), and

$$\not{p}_3(c_V - c_A\gamma^5)v(3) = (c_V + c_A\gamma^5)\not{p}_3 v(3) = 0$$

for the same reason. Thus

$$\mathcal{M} = -\frac{g_z^2}{4[q^2 - (M_Z c)^2]}[\bar{u}(4)\gamma^\mu(c_V^f - c_A^f\gamma^5)v(3)] \times [\bar{v}(2)\gamma_\mu(c_V^e - c_A^e\gamma^5)u(1)] \tag{10.106}$$

and it follows that

$$\langle|\mathcal{M}|^2\rangle = \left[\frac{g_z^2}{8(q^2 - (M_Z c)^2)}\right]^2 Tr\{\gamma^\mu(c_V^f - c_A^f\gamma^5)\not{p}_3\gamma^\nu(c_V^f - c_A^f\gamma^5)\not{p}_4\} \times Tr\{\gamma_\mu(c_V^e - c_A^e\gamma^5)\not{p}_1\gamma_\nu(c_V^e - c_A^e\gamma^5)\not{p}_2\} \tag{10.107}$$

Now, the first trace is

$$(c_V^2 + c_A^2)[p_3^\mu p_4^\nu + p_3^\nu p_4^\mu - g^{\mu\nu}(p_3 \cdot p_4)] - 2ic_V c_A \epsilon^{\mu\nu\lambda\sigma} p_{3\lambda} p_{4\sigma} \tag{10.108}$$

and there is the corresponding expression for the second trace, so we obtain

$$\langle|\mathcal{M}|^2\rangle = \frac{1}{2}\left[\frac{g_z^2}{q^2 - (M_Z c)^2}\right]^2 \{[(c_V^f)^2 + (c_A^f)^2][(c_V^e)^2 + (c_A^e)^2] \times [(p_1 \cdot p_3)(p_2 \cdot p_4) + (p_1 \cdot p_4)(p_2 \cdot p_3)] + 4c_V^f c_A^f c_V^e c_A^e[(p_1 \cdot p_3)(p_2 \cdot p_4) - (p_1 \cdot p_4)(p_2 \cdot p_3)]\} \tag{10.109}$$

In the CM frame this reduces to

$$\langle|\mathcal{M}|^2\rangle = \left[\frac{g_z^2 E^2}{(2E)^2 - (M_Z c^2)^2}\right]^2 \{[(c_V^f)^2 + (c_A^f)^2][(c_V^e)^2 + (c_A^e)^2](1 + \cos^2\theta) - 8c_V^f c_A^f c_V^e c_A^e \cos\theta\} \tag{10.110}$$

where E is the energy of each particle and θ is the scattering angle. The differential scattering cross section (6.42) is therefore

$$\frac{d\sigma}{d\Omega} = \left(\frac{\hbar c g_z^2 E}{16\pi[(2E)^2 - (M_Z c^2)^2]}\right)^2 \{[(c_V^f)^2 + (c_A^f)^2][(c_V^e)^2 + (c_A^e)^2](1 + \cos^2\theta) - 8c_V^f c_A^f c_V^e c_A^e \cos\theta\} \tag{10.111}$$

and the total cross section is

$$\sigma = \frac{1}{3\pi}\left(\frac{\hbar c g_z^2 E}{4[(2E)^2 - (M_Z c^2)^2]}\right)^2 [(c_V^f)^2 + (c_A^f)^2][(c_V^e)^2 + (c_A^e)^2] \tag{10.112}$$

As it stands, σ blows up at the Z^0 pole—that is, when the total energy $(2E)$ hits the value $M_Z c^2$ (just right to put the Z^0 on its mass shell). The problem is that we have treated the Z^0 as a *stable* particle, which it is not. It has a finite lifetime τ_Z (not yet accurately measured), which has the effect of "smearing out" its mass. We can account for this by modifying the propagator[16]

$$\frac{1}{q^2 - (M_Z c)^2} \to \frac{1}{q^2 - (M_Z c)^2 + i\hbar M_Z \Gamma_Z} \tag{10.113}$$

where Γ_Z is the decay rate ($\Gamma_Z = 1/\tau_Z$). With this adjustment, the cross section becomes

$$\sigma = \frac{(\hbar c g_z^2 E)^2}{48\pi} \frac{[(c_V^f)^2 + (c_A^f)^2][(c_V^e)^2 + (c_A^e)^2]}{[(2E)^2 - (M_Z c^2)^2]^2 + (\hbar M_Z c^2 \Gamma_Z)^2} \tag{10.114}$$

Because $\hbar\Gamma_Z \ll M_Z c^2$, the correction for finite Z^0 lifetime is negligible except in the immediate vicinity of the Z^0 pole, where it has the effect of softening the infinite spike.

In Chapter 8 we calculated the cross section for the same process when mediated by a *photon* [eq. (8.6)]:

$$\sigma = \frac{(\hbar c g_e^2)^2}{48\pi} \frac{(Q^f)^2}{E^2} \tag{10.115}$$

(where Q^f is the charge of f, in units of e). Thus the ratio of weak to electromagnetic rates in (for example) muon production, is

$$\frac{\sigma(e^+e^- \to Z^0 \to \mu^+\mu^-)}{\sigma(e^+e^- \to \gamma \to \mu^+\mu^-)} = \left\{\frac{(\frac{1}{2} - 2\sin^2\theta_w + 4\sin^4\theta_w)^2}{(\sin\theta_w \cos\theta_w)^4}\right\} \times \frac{E^4}{[(2E)^2 - (M_Z c^2)^2]^2 + (\hbar\Gamma_Z M_Z c^2)^2} \tag{10.116}$$

The factor in curly brackets is approximately 2, if we use the current value of the weak mixing angle (10.95). Substantially *below* the Z^0 pole ($2E \ll M_Z c^2$), then,

$$\frac{\sigma_Z}{\sigma_\gamma} \cong 2\left(\frac{E}{M_Z c^2}\right)^4 \tag{10.117}$$

and the electromagnetic route dominates (at $2E = \frac{1}{2}M_Z c^2$, for instance, the weak contribution is less than 1%). But right *on* the Z^0 pole ($2E = M_Z c^2$)

$$\frac{\sigma_Z}{\sigma_\gamma} \cong \frac{1}{8}\left(\frac{M_Z c^2}{\hbar\Gamma_Z}\right) \tag{10.118}$$

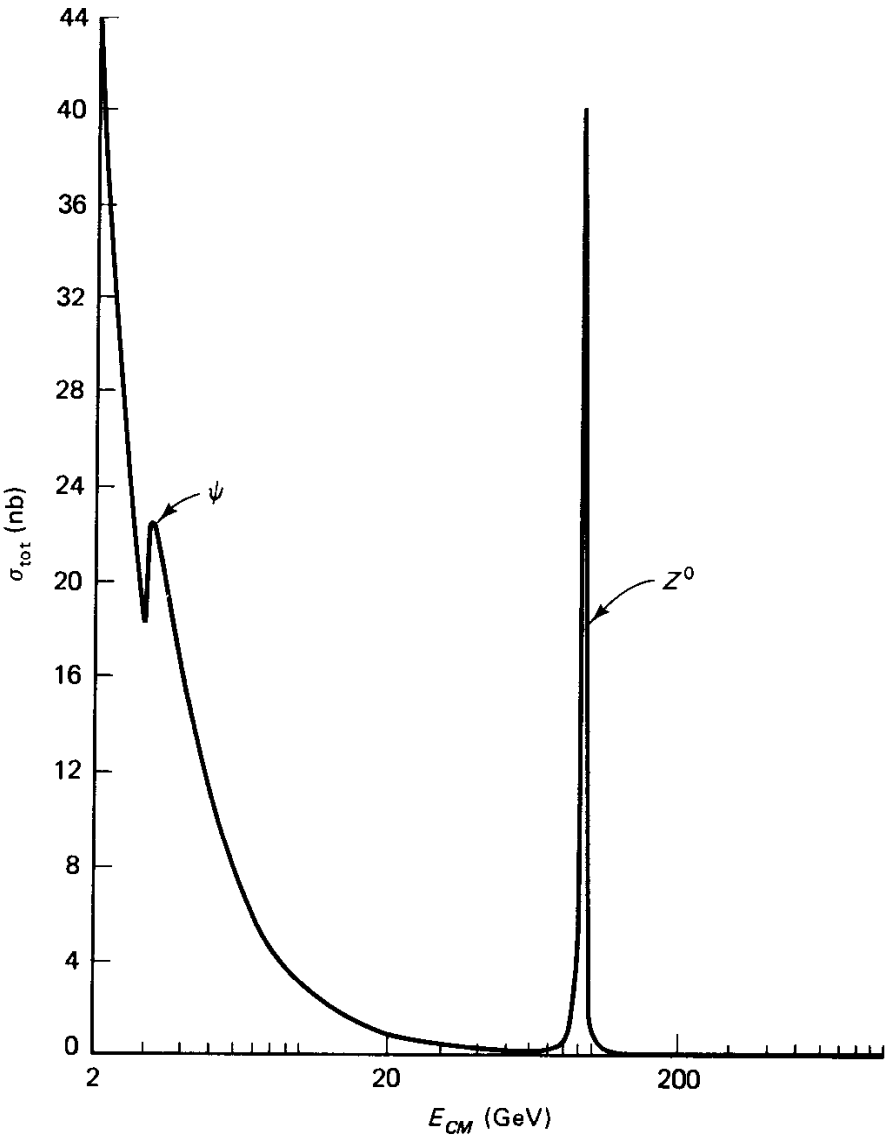


Figure 10.10 Electron-positron scattering in the neighborhood of the Z^0 pole.

The lifetime of the Z^0 is easily estimated (Problem 10.20); using $\hbar\Gamma_Z = 2.5$ GeV, we obtain $\sigma_Z/\sigma_\gamma \cong 200$. At the Z^0 pole, therefore, the *weak* mechanism is favored, by a factor of well over 100, and possibly by as much as 1000.* (See Fig. 10.10.)

10.7 ELECTROWEAK UNIFICATION

10.7.1 Chiral Fermion States

All the cards are now on the table;† it remains only to explain where the GWS parameters in Table 10.1 and equations (10.94) and (10.98) *come* from. Glashow's original aim was to *unify* the weak and electromagnetic interactions, to combine

* Equally interesting is the electromagnetic-weak "interference" that occurs when the two amplitudes are combined: $|\mathcal{M}_\gamma + \mathcal{M}_Z|^2 = |\mathcal{M}_\gamma|^2 + |\mathcal{M}_Z|^2 + 2\,\mathrm{Re}(\mathcal{M}_\gamma \mathcal{M}_Z)$. We have calculated $|\mathcal{M}_Z|^2$ and (in Chap. 8) $|\mathcal{M}_\gamma|^2$, but the cross term provides a sensitive test of the GWS theory, even at energies substantially below the Z^0 pole. (See Halzen and Martin, ref. 12, Sect. 13.6, and ref. 15.) Indeed, it was the success of the electroweak interference experiments in 1978 that convinced most theorists that the GWS model is correct. For a contemporary account, see *Physics Today*, September 1978, p. 17.

† I have not discussed the couplings of W's and Z^0's to one another (or of W's to the photon). The rules are similar to those for gluon-gluon coupling in QCD, and are listed in Appendix D.

them into a single theoretical system, in which they would appear not as unrelated phenomena, but rather as different manifestations of one fundamental "electroweak" interaction. This was a bold proposition, in 1961.[17] In the first place, there was the enormous disparity in strength between weak and electromagnetic forces. However, as Glashow and others recognized, this could be accounted for if the weak interactions were mediated by extremely massive particles. Of course, this immediately begs the second question: If it's really all one basic interaction, how come the electromagnetic mediator (γ) is massless, when the weak mediators ($W^{\pm}$ and Z^0) are so heavy? Glashow had no particularly good answer ("It is a stumbling block we must overlook," he said coyly). The solution was provided by Weinberg and Salam, in 1967 (see refs. 8 and 9) in the form of the "Higgs mechanism" (Chapter 11). Finally, there is a structural difference between the electromagnetic and weak vertex factors, which at first glance would seem to preclude any possibility of unification: The former are purely vectorial (γ^μ), whereas the latter contain vector and axial vector parts. In particular, the $W^{\pm}$ coupling is "maximally" mixed V–A in character ($\gamma^\mu(1 - \gamma^5)$).

This last difficulty is overcome by the ingenious device of absorbing the matrix $(1 - \gamma^5)$ into the particle spinor itself. Specifically, we define

$$u_L(p) \equiv \frac{(1 - \gamma^5)}{2} u(p) \tag{10.119}$$

The subscript (L) stands for "left-handed," and is supposed to make you think "helicity -1." However, this is somewhat misleading, since u_L is not, in general, a helicity eigenstate. In fact, as you can easily show (see Problem 10.23)

$$\gamma^5 u(p) = \begin{pmatrix} \dfrac{c(\mathbf{p}\cdot\boldsymbol{\sigma})}{E + mc^2} & 0 \\ 0 & \dfrac{c(\mathbf{p}\cdot\boldsymbol{\sigma})}{E - mc^2} \end{pmatrix} u(p) \tag{10.120}$$

If the particle in question is massless, then $E = |\mathbf{p}|c$, and

$$\gamma^5 u(p) = (\hat{p}\cdot\boldsymbol{\Sigma})u(p) \tag{10.121}$$

where

$$\boldsymbol{\Sigma} = \begin{pmatrix} \boldsymbol{\sigma} & 0 \\ 0 & \boldsymbol{\sigma} \end{pmatrix} \tag{10.122}$$

as before. Recall [eq. (7.48)] that $\hbar/2\boldsymbol{\Sigma}$ is the spin matrix for a Dirac particle, and hence $(\hat{p}\cdot\boldsymbol{\Sigma})$ is the helicity, with eigenvalues ± 1. Accordingly

$$\tfrac{1}{2}(1 - \gamma^5)u(p) = \begin{cases} 0, & \text{if } u(p) \text{ carries helicity } +1 \\ u(p), & \text{if } u(p) \text{ carries helicity } -1 \end{cases} \quad (\text{for } m = 0 \text{ only}) \tag{10.123}$$

More generally, $\frac{1}{2}(1 - \gamma^5)$ functions as a "projection operator," picking out the helicity -1 component of $u(p)$. On the other hand, *if the particle is not massless,* it is only in the ultrarelativistic regime ($E \gg mc^2$) that equation (10.121) holds (approximately), and hence only in this limit that u_L (as defined by equation (10.119) carries helicity -1. Nevertheless, everybody *calls* u_L a "left-handed"

TABLE 10.2 CHIRAL SPINORS

Particles	Antiparticles
$u_L = \frac{1}{2}(1 - \gamma^5)u$	$v_L = \frac{1}{2}(1 + \gamma^5)v$
$u_R = \frac{1}{2}(1 + \gamma^5)u$	$v_R = \frac{1}{2}(1 - \gamma^5)v$
$\bar{u}_L = \bar{u}\frac{1}{2}(1 + \gamma^5)$	$\bar{v}_L = \bar{v}\frac{1}{2}(1 - \gamma^5)$
$\bar{u}_R = \bar{u}\frac{1}{2}(1 - \gamma^5)$	$\bar{v}_R = \bar{v}\frac{1}{2}(1 + \gamma^5)$

R and L correspond to helicity $+1$ and -1 if $m = 0$, and *approximately* so if $E \gg mc^2$.

state, and I shall stick to the customary language.* Meanwhile, for an *anti*particle equation (10.121) reads

$$\gamma^5 v(p) = -(\hat{p} \cdot \boldsymbol{\Sigma})v(p) \qquad (\text{if } m = 0) \tag{10.124}$$

(see Problem 10.23), and for this reason we define

$$v_L(p) \equiv \frac{(1 + \gamma^5)}{2} v(p) \tag{10.125}$$

The corresponding "right-handed" spinors are

$$u_R(p) \equiv \frac{(1 + \gamma^5)}{2} u(p), \qquad v_R(p) \equiv \frac{(1 - \gamma^5)}{2} v(p) \tag{10.126}$$

As for the adjoint spinors, we have

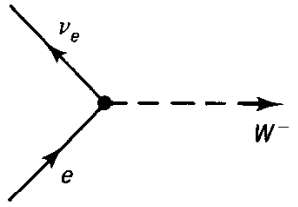
$$\bar{u}_L = u_L^\dagger \gamma^0 = u^\dagger \frac{(1 - \gamma^5)}{2} \gamma^0 = u^\dagger \gamma^0 \frac{(1 + \gamma^5)}{2} = \bar{u} \frac{(1 + \gamma^5)}{2} \tag{10.127}$$

[Recall that γ^5 is Hermitian ($\gamma^{5\dagger} = \gamma^5$), and it anticommutes with γ^μ ($\gamma^\mu \gamma^5 = -\gamma^5 \gamma^\mu$).] Similarly

$$\bar{v}_L = \bar{v} \frac{(1 - \gamma^5)}{2}, \qquad \bar{u}_R = \bar{u} \frac{(1 - \gamma^5)}{2}, \qquad \bar{v}_R = \bar{v} \frac{(1 + \gamma^5)}{2} \tag{10.128}$$

We call these various spinors (summarized in Table 10.2) "chiral" fermion states (from the Greek word for "hand"—same root as "chiropractor").

I emphasize that all this is only notation and terminology; it is *useful* because it allows us to recast the weak and electromagnetic interactions in a form that facilitates their unification. Consider, to begin with, the coupling of an electron and a neutrino to the W^- (as it occurs, say, in inverse beta decay, Example 10.1):



* Please understand that equation (10.119) is a *definition* of u_L—nobody's arguing about that. I'm only worrying about the potentially misleading *name*. "Left-handed" does *not* mean "helicity -1," except in cases where the particle's mass is negligible.

The contribution to $\mathcal{M}$ from this vertex is given by

$$j_\mu^- = \bar{\nu}\gamma_\mu\left(\frac{1-\gamma^5}{2}\right)e \tag{10.129}$$

(here e and ν stand for the particle spinors; for a while we need to keep careful track of the different particle species, and u_e, u_{ν_e}, etc. just gets too cumbersome). This quantity is called the (negatively charged) weak "current"; as we shall see, it plays a role somewhat analogous to the electric current in QED. Now

$$\left(\frac{1-\gamma^5}{2}\right)^2 = \frac{1}{4}[1 - 2\gamma^5 + (\gamma^5)^2] = \left(\frac{1-\gamma^5}{2}\right) \tag{10.130}$$

and

$$\gamma_\mu\left(\frac{1-\gamma^5}{2}\right) = \left(\frac{1+\gamma^5}{2}\right)\gamma_\mu \tag{10.131}$$

so

$$\gamma_\mu\left(\frac{1-\gamma^5}{2}\right) = \left(\frac{1+\gamma^5}{2}\right)\gamma_\mu\left(\frac{1-\gamma^5}{2}\right) \tag{10.132}$$

This may not look like much of an improvement, but it enables us to rewrite equation (10.129) more neatly, in terms of the chiral spinors:

$$j_\mu^- = \bar{\nu}_L\gamma_\mu e_L \tag{10.133}$$

The weak vertex factor is now *purely vectorial*—but it couples only *left-handed* electrons to *left-handed* neutrinos. In the latter sense it is still structurally different from the fundamental vertex in QED; however, we can play a similar game there, too. Notice that

$$u = \left(\frac{1-\gamma^5}{2}\right)u + \left(\frac{1+\gamma^5}{2}\right)u = u_L + u_R \tag{10.134}$$

(similarly $\bar{u} = \bar{u}_L + \bar{u}_R$), so the electromagnetic "current" can itself be written in terms of chiral spinors:

$$j_\mu^{em} = -\bar{e}\gamma_\mu e = -(\bar{e}_L + \bar{e}_R)\gamma_\mu(e_L + e_R) = -\bar{e}_L\gamma_\mu e_L - \bar{e}_R\gamma_\mu e_R \tag{10.135}$$

(For future purposes it is best to build in a factor of -1, to account for the negative charge of the electron). Observe that the "cross terms" vanish:

$$\bar{e}_L\gamma_\mu e_R = \bar{e}\left(\frac{1+\gamma^5}{2}\right)\gamma_\mu\left(\frac{1+\gamma^5}{2}\right)e = \bar{e}\gamma_\mu\left(\frac{1-\gamma^5}{2}\right)\left(\frac{1+\gamma^5}{2}\right)e \tag{10.136}$$

but

$$(1-\gamma^5)(1+\gamma^5) = 1 - (\gamma^5)^2 = 0 \tag{10.137}$$

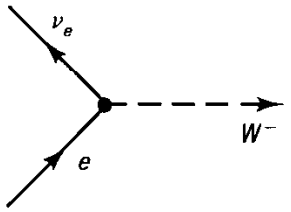
Equations (10.133) and (10.135) are beginning to look like the stuff of which one might build a unified theory. It is true that the weak current only couples left-handed states, whereas the electromagnetic current couples both types, but apart from that they are strikingly similar. So attractive is this formulation that physicists have come to regard left- and right-handed fermions almost as different

particles.* In this view, the factor $(1 - \gamma^5)/2$ in the charged weak coupling characterizes the participating particles, rather than the interaction itself; the latter is vectorial in *all* cases—strong, electromagnetic, and weak alike.

10.7.2 Weak Isospin and Hypercharge

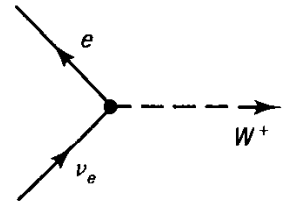
In addition to the negatively charged weak current

$$j_\mu^- = \bar{\nu}_L \gamma_\mu e_L$$



describing the process $e^- \to \nu_e + W^-$, there is also, of course, a positively charged current

$$j_\mu^+ = \bar{e}_L \gamma_\mu \nu_L$$



representing the process $\nu_e \to e^- + W^+$. We can express them both in a more compact notation by introducing the left-handed doublet

$$\chi_L = \begin{pmatrix} \nu_e \\ e \end{pmatrix}_L \tag{10.138}$$

and the 2×2 matrices

$$\tau^+ \equiv \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}, \qquad \tau^- \equiv \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix} \tag{10.139}$$

so that

$$j_\mu^\pm = \bar{\chi}_L \gamma_\mu \tau^\pm \chi_L \tag{10.140}$$

The matrices $\tau^\pm$ are linear combinations of the first two Pauli spin matrices [eq. (4.26)]:

$$\tau^\pm = \tfrac{1}{2}(\tau^1 \pm i\tau^2) \tag{10.141}$$

* There is a danger in carrying this *too* far. You may find yourself wondering, for example, whether the left-handed electron necessarily has the same *mass* as the right-handed electron; or, noting that no vector interaction can couple a left-handed particle to a right-handed one [see eqs. (10.136) and (10.137)], you may ask how the two "worlds" communicate at all. Both questions are based on a misunderstanding of u_L and u_R. The problem is that, useful as it is in describing particle *interactions,* handedness is not *conserved* in the propagation of a free particle (unless its mass is zero). (Formally, γ^5 does not commute with the free particle Hamiltonian.) In fact, u_L and u_R do not satisfy the Dirac equation (see Problem 10.24). A particle that starts out left-handed will soon evolve a right-handed component. (By contrast, *helicity is* conserved in free-particle propagation.) Only for *massless* fermions can left- and right-handed species be considered distinct particles in the full sense of the word; and, of course, left- and right-handed neutrinos *are* distinct: as far as we know right-handed neutrinos do not exist at all.

(I use the letter τ here, instead of σ, to avoid possible confusion with ordinary spin.) This is all very reminiscent of *isospin;* in Chapter 4, Section 4.5, we put the proton and neutron into a doublet similar to (10.138). Indeed, we could contemplate a full "weak isospin" symmetry, if only there were a *third* weak current, corresponding to $\frac{1}{2}\tau^3 = \frac{1}{2}\begin{pmatrix}1 & 0\\ 0 & -1\end{pmatrix}$:

$$j_\mu^3 = \bar{\chi}_L \gamma_\mu \tfrac{1}{2}\tau^3 \chi_L = \tfrac{1}{2}\bar{\nu}_L \gamma_\mu \nu_L - \tfrac{1}{2}\bar{e}_L \gamma_\mu e_L \tag{10.142}$$

"Perfect!" (I hear you exclaim.) "There's the *neutral* weak current!" Not so fast. *This* current only couples *left*-handed particles; in the older language it is pure *V–A,* whereas the neutral weak interaction involves right-handed components as well. But hang on—we're *almost* there.

Building on the parallel with isospin, we are led to consider a weak analog of *hypercharge* (Y),* which is related to electric charge (Q, in units of e) and the third component of isospin (I^3), by the Gell-Mann–Nishijima formula [eq. (4.37)]:

$$Q = I^3 + \tfrac{1}{2}Y \tag{10.143}$$

We introduce, then, the "weak hypercharge" current

$$j_\mu^Y = 2j_\mu^{em} - 2j_\mu^3 = -2\bar{e}_R \gamma_\mu e_R - \bar{e}_L \gamma_\mu e_L - \bar{\nu}_L \gamma_\mu \nu_L \tag{10.144}$$

This is an *invariant* construct, as far as weak isospin is concerned, for the latter does not touch *right*-handed components at all, and the combination

$$\bar{e}_L \gamma_\mu e_L + \bar{\nu}_L \gamma_\mu \nu_L = \bar{\chi}_L \gamma_\mu \chi_L$$

is itself invariant.† The underlying symmetry group is called $SU(2)_L \otimes U(1)$; $SU(2)_L$ refers to the weak isospin (with a subscript to remind us that it involves left-handed states only), and $U(1)$ refers to weak hypercharge (involving both chiralities).

I have developed all this in terms of the electron and its neutrino, but it is a trivial matter to extend it to the other leptons and quarks. From the left-handed doublets (Cabibbo-rotated, in the case of the quarks)

$$\chi_L \to \begin{pmatrix}\nu_e\\ e\end{pmatrix}_L, \quad \begin{pmatrix}\nu_\mu\\ \mu\end{pmatrix}_L, \quad \begin{pmatrix}\nu_\tau\\ \tau\end{pmatrix}_L, \quad \begin{pmatrix}u\\ d'\end{pmatrix}_L, \quad \begin{pmatrix}c\\ s'\end{pmatrix}_L, \quad \begin{pmatrix}t\\ b'\end{pmatrix}_L \tag{10.145}$$

we construct three weak isospin currents

$$\mathbf{j}_\mu = \tfrac{1}{2}\bar{\chi}_L \gamma_\mu \boldsymbol{\tau} \chi_L \tag{10.146}$$

and a weak hypercharge current

$$j_\mu^Y = 2j_\mu^{em} - 2j_\mu^3 \tag{10.147}$$

* You have probably forgotten this word, but *hypercharge* is essentially the same as strangeness, only shifted, in the case of baryons, so that the center row of Eightfold Way diagrams will always carry $Y = 0$. Specifically, $Y = S + A$, where A is the baryon number.

† If you care to think of it this way, what we have done is to combine two weak-isospin doublets to make an isotriplet, $\bar{\nu}_L e_L$, $(\bar{\nu}_L \nu_L - \bar{e}_L e_L)$, $\bar{e}_L \nu_L$ [analogous to (5.89)], and an isosinglet $(\bar{\nu}_L \nu_L + e_L e_L)$ [analogous to (5.90)]. The first three go to make the weak isospin currents $j^\pm$ and j^3; the last, together with a right-handed piece, makes the weak hypercharge current, j^Y.

where j_μ^{em} is the electric current:

$$j_\mu^{em} = \sum_{i=1}^{2} Q_i(\bar{u}_{i_L}\gamma_\mu u_{i_L} + \bar{u}_{i_R}\gamma_\mu u_{i_R}) \tag{10.148}$$

(summed over the particles in the doublet, with Q_i the electric charge).*

10.7.3 Electro-Weak Mixing

Now, the GWS model asserts that the three weak isospin currents couple, with strength g_w, to a weak isotriplet of intermediate vector bosons, **W**, whereas the weak hypercharge current couples with strength $g'/2$ to an isosinglet intermediate vector boson, B:

$$-i\left[g_w \mathbf{j}_\mu \cdot \mathbf{W}^\mu + \frac{g'}{2} j_\mu^Y B^\mu\right] \tag{10.149}$$

Within this fundamental structure is contained all of electrodynamics and all of the weak interactions. The arrow denotes a three-vector in weak isospin space; the dot product can be written out explicitly:

$$\mathbf{j}_\mu \cdot \mathbf{W}^\mu = j_\mu^1 W^{\mu 1} + j_\mu^2 W^{\mu 2} + j_\mu^3 W^{\mu 3} \tag{10.150}$$

or, in terms of the charged currents, $j_\mu^\pm = j_\mu^1 \pm ij_\mu^2$:

$$\mathbf{j}_\mu \cdot \mathbf{W}^\mu = (1/\sqrt{2})j_\mu^+ W^{\mu +} + (1/\sqrt{2})j_\mu^- W^{\mu -} + j_\mu^3 W^{\mu 3} \tag{10.151}$$

where

$$W_\mu^\pm \equiv (1/\sqrt{2})(W_\mu^1 \mp iW_\mu^2) \tag{10.152}$$

are the wave functions representing the $W^\pm$ particles.

The couplings to $W^\pm$ can now be read off, from the coefficients of $W_\mu^\pm$ in expression (10.149). For example, in the process $e^- \to \nu_e + W^-$ we have $j_\mu^- = \bar{\nu}_L\gamma_\mu e_L = \bar{\nu}\gamma_\mu[(1 - \gamma^5)/2]e$ [see eq. (10.129)], giving a term

$$-ig_w(1/\sqrt{2})j_\mu^- W^{\mu -} = -\frac{ig_w}{2\sqrt{2}}[\bar{\nu}\gamma_\mu(1 - \gamma^5)e]W^{\mu -} \tag{10.153}$$

The vertex factor is

$$\frac{-ig_w}{2\sqrt{2}}\gamma_\mu(1 - \gamma^5) \tag{10.154}$$

which is exactly what we started with [eq. (10.5)].

* You might ask what the *difference* is between *weak* isospin (and hypercharge) and their ordinary ("strong") counterparts. The question is particularly pertinent when you come to the light quarks: The weak isospin doublet is $\begin{pmatrix} u \\ d' \end{pmatrix}_L$, whereas the strong isospin doublet is $\begin{pmatrix} u \\ d \end{pmatrix}$. Pretty similar . . . is there anything to this? Nope. After all, (i) weak isospin applies to leptons as well as quarks (and to all three quark generations); (ii) weak isospin involves only the *left-handed* chiralities, (all right-handed states are *singlets*—i.e., invariant—as far as weak isospin is concerned); (iii) weak isodoublets are Cabibbo-rotated. To put it plainly, strong isospin and weak isospin have nothing to do with one another, save for a common mathematical structure (which, for that matter, they share with many other systems, such as ordinary spin $\frac{1}{2}$) and the (perhaps unfortunate) similarity in their names.

But the underlying $SU(2)_L \otimes U(1)$ symmetry is "broken," in GWS theory: The two *neutral* states, W^3 and B, "mix," producing one massless linear combination (the photon), and an orthogonal massive combination (the Z^0):

$$A_\mu = B_\mu \cos\theta_w + W^3_\mu \sin\theta_w$$
$$Z_\mu = -B_\mu \sin\theta_w + W^3_\mu \cos\theta_w \tag{10.155}$$

(You see now why θ_w is called the "weak mixing angle.") In terms of the *physical* states (A^μ and Z^μ), then, the neutral portion of the electro-weak interaction (10.149) reads:

$$-i\left[g_w j^3_\mu W^{\mu 3} + \frac{g'}{2} j^Y_\mu B^\mu\right] = -i\left\{\left[g_w \sin\theta_w j^3_\mu + \frac{g'}{2}\cos\theta_w j^Y_\mu\right]A^\mu + \left[g_w \cos\theta_w j^3_\mu - \frac{g'}{2}\sin\theta_w j^Y_\mu\right]Z^\mu\right\} \tag{10.156}$$

Of course, we *know* the electromagnetic coupling; in the present language it is

$$-ig_e j^{em}_\mu A^\mu \tag{10.157}$$

Meanwhile, from equation (10.147), $j^{em}_\mu = j^3_\mu + \frac{1}{2}j^Y_\mu$. Evidently consistency of the unified electro-weak theory with ordinary QED requires

$$g_w \sin\theta_w = g' \cos\theta_w = g_e \tag{10.158}$$

The weak and electromagnetic coupling constants are *not independent.*

There remains the weak coupling to the Z^0. Using equations (10.147), (10.156), and (10.158), we obtain

$$-ig_z(j^3_\mu - \sin^2\theta_w j^{em}_\mu)Z^\mu \tag{10.159}$$

where

$$g_z = \frac{g_e}{\sin\theta_w \cos\theta_w} \tag{10.160}$$

From expression (10.159) we can pick out the neutral weak couplings. For example, the process $\nu_e \longrightarrow \nu_e + Z^0$ comes exclusively from the j^3_μ term; referring back to equation (10.142), we have

$$-i\frac{g_z}{2}(\bar{\nu}_L\gamma_\mu\nu_L)Z^\mu = -\frac{ig_z}{2}\left[\bar{\nu}\gamma_\mu\left(\frac{1-\gamma^5}{2}\right)\nu\right]Z^\mu$$

and hence the vector and axial vector couplings [eq. (10.93)] are $c^\nu_V = c^\nu_A = \frac{1}{2}$. I'll leave it for you to work out the other entries in Table 10.1* (Problem 10.26).

Finally, there is the obvious question: *why* is the underlying $SU(2)_2 \otimes U(1)$ symmetry of the electroweak interactions "broken"—why do the B and W^3 states "mix"? [eq. (10.155)]—to form the Z^0 and the photon? If weak and electromagnetic interactions are, deep down, both manifestations of a single electroweak force, how come the weak mediators ($W^\pm$ and Z^0) are so heavy, while the electromagnetic mediator (γ) is massless? We shall address these matters in the next chapter.

* Since the weak mixing angle is undetermined, in the GWS model, there remain in effect *two* independent coupling constants (g_e and g_w, say, or g_e and g_z); in this sense it is not a completely *unified* theory, but rather an *integrated* theory of weak and electromagnetic interactions.

REFERENCES AND NOTES

1. N. Cabibbo, *Phys. Rev. Lett.* **10,** 531 (1963).
2. For more detailed calculations in weak interaction theory consult the classic treatise by R. E. Marshak, Riazuddin, and C. P. Ryan: *Theory of Weak Interactions in Particle Physics* (New York: Wiley, 1969), or the briefer account by E. D. Commins: *Weak Interactions* (New York: McGraw-Hill, 1973). For an up-to-date review of weak interactions in the quark model, see J. F. Donoghue, E. Golowich, and B. Holstein, *Phys. Rep.* **131,** 319 (1986) or E. D. Commins and P. H. Bucksbaum: *Weak Interactions of Leptons and Quarks* (Cambridge: Cambridge Univ. Press, 1983).
3. S. L. Glashow, J. Iliopoulos, and L. Maiani, *Phys. Rev.* **D2,** 1585 (1970). This and the other fundamental papers in modern weak interaction theory are reprinted in *Gauge Theory of Weak and Electromagnetic Interactions,* C. H. Lai, ed. (Singapore: World Scientific, 1981).
4. M. Kobayashi and K. Maskawa, *Prog. Theor. Phys.* **49,** 652 (1973).
5. Different authors use different conventions in writing the KM matrix; I follow the notation of L.-L. Chau, *Phys. Rep.* **95,** 2 (1983).
6. F. J. Gilman, *Rev. Mod. Phys.* **56,** S297 (1984). For a useful discussion of the KM matrix, see T.-P. Cheng and L.-F. Li, *Gauge Theory of Elementary Particle Physics* (New York: Oxford, 1984), Sect. 12.2.
7. S. A. Bludman, *Nuovo Cimento* **9,** 443 (1958). In a sense the process shown in Figure (10.4) is already a neutral weak interaction, since d and s have the same charge. However, I shall reserve the word "neutral" for weak interactions mediated by the Z^0.
8. S. L. Glashow, *Nucl. Phys.* **22,** 579 (1961). Reprinted in Lai, ref. 3.
9. S. Weinberg, *Phys. Rev. Lett.* **19,** 1264 (1967); A. Salam, in *Elementary Particle Theory,* N. Svartholm, ed. (Stockholm: Almquist and Wiksell, 1968). Reprinted in Lai, ref. 3.
10. G. 't Hooft, *Nucl. Phys.* **B33,** 173 (1971) and **B35,** 167 (1971). Reprinted in Lai, ref. 3.
11. F. J. Hasert et al., *Phys. Lett.* **46B,** 138 (1973); *Nucl. Phys.* **B73,** 1 (1974).
12. See, for example, D. H. Perkins, *Introduction to High-Energy Physics,* 2d Ed. (Reading, MA: Addison-Wesley, 1982), Sect. 8.8.2; I. J. R. Aitchison and A. J. G. Hey, *Gauge Theories in Particle Physics* (Bristol: Adam Hilger, 1982) Sect. 6.2; F. Halzen and A. D. Martin, *Quarks and Leptons* (New York: Wiley, 1984), Sects. 12.7 and 12.10; F. E. Close, *An Introduction to Quarks and Partons* (London: Academic, 1979), Sect. 11.3.
13. G. Arnison, et al., *Phys. Lett.* **122B,** 103 (1983) and **126B,** 398 (1983). For a review of these discoveries, see E. Radermacher, *Prog. Part. Nucl. Phys.* **14,** 231 (1985).
14. Data on $\nu_\mu + e^- \rightarrow \nu_e + \mu^-$ is from F. Bergsma et al., *Phys. Lett.* **122B,** 465 (1983); data on $\nu_\mu + e^- \rightarrow \nu_\mu + e^-$ is from R. H. Heisterberg et al., *Phys. Rev. Lett.* **44,** 635 (1980) (earlier data by P. Alibran et al., *Phys. Lett.* **74B,** 422 (1978), which was inconsistent with the GWS model, turned out to be wrong, although it caused some consternation at the time; it was corrected by N. Armenise et al., *Phys. Lett.* **86B,** 225 (1979).
15. E. N. Fortson and L. Wilets, *Advances in Atomic and Molecular Physics,* **16,** 319 (1980); C. Y. Prescott et al., *Phys. Lett.* **77B,** 347 (1978) and **84B,** 524 (1979); S. L. Wu, *Phys. Rep.* **107,** 229 (1984).
16. See, for example, H. Frauenfelder and E. M. Henley, *Subatomic Physics,* (Englewood Cliffs, NJ: Prentice-Hall, 1974), Sect. 5.7.

17. Even bolder in 1957, when Schwinger laid the essential groundwork for such a theory: J. Schwinger, *Ann. Phys.* (NY) **2**, 407 (1957). Reprinted in Lai, Ref. 3.

PROBLEMS

10.1. **(a)** Calculate $\langle|\mathcal{M}|^2\rangle$ for $\nu_\mu + e^- \rightarrow \mu^- + \nu_e$ using the more general coupling $\gamma^\mu(1 + \epsilon\gamma^5)$. Check that your answer reduces to equation (10.11) in the case $\epsilon = -1$.

$$\left[\textit{Answer: } \sum_{\text{spins}} |\mathcal{M}|^2 = \frac{1}{2}\left(\frac{g_w}{M_W c}\right)^4 [(1-\epsilon^2)^2(p_1\cdot p_4)(p_2\cdot p_3) + (1+6\epsilon^2+\epsilon^4)(p_1\cdot p_2)(p_3\cdot p_4)]\right]$$

(b) Let $m_e = m_\mu = 0$, and calculate the *CM* differential scattering cross section. Also, find the *total* cross section.

(c) If you had accurate experimental data on this reaction, how could you determine ϵ?

10.2. Calculate the lifetime of the τ lepton. Compare the experimental result. (Assume that the muon mass can be neglected, in comparison with m_τ. Do the experimental data support this approximation?)

10.3. Suppose the weak interaction were *pure vector* (as Fermi supposed). Would you still get the graph shown in Figure 10.1?

10.4. Using the coupling $\gamma^\mu(1 + \epsilon\gamma^5)$ for $n \rightarrow p + W$, but $\gamma^\mu(1 - \gamma^5)$ for the leptons, calculate the spin-averaged amplitude for neutron beta decay. Show that your result reduces to equation (10.43) when $\epsilon = -1$.

$$\left[\textit{Answer: } \langle|\mathcal{M}|^2\rangle = \frac{1}{2}\left(\frac{g_w}{M_W c}\right)^4 [(p_1\cdot p_2)(p_3\cdot p_4)(1-\epsilon)^2 + (p_1\cdot p_4)(p_2\cdot p_3)(1+\epsilon)^2 - (1-\epsilon^2)m_p m_n c^2(p_2\cdot p_4)]\right]$$

10.5. **(a)** Derive equation (10.55). (b) Derive equation (10.61).

10.6. In the text I said that electron energies in neutron decay range up to about $(m_n - m_p)c^2$. This is not *exact*, since it ignores the *kinetic* energy of the proton and the neutrino. What kinematic configuration gives the maximum electron energy? Apply conservation of energy and momentum to determine the *exact* maximum electron energy.

$$[\textit{Answer: } (m_n^2 - m_p^2 + m_e^2)c^2/2m_n.]$$

How far off is the approximate answer (give the percent error)?

10.7. **(a)** Integrate equation (10.62) to get equation (10.63).

(b) Approximate as suitable for $m_e \ll \Delta m = (m_n - m_p)$. Note that m_e now drops out.

10.8. Obtain equation (10.65).

10.9. Find the minimum de Broglie wavelength ($\lambda = h/p$) of the W in neutron decay, and compare it with the diameter of the neutron ($\sim 10^{-13}$ cm). [*Answer:* maximum

$|\mathbf{p}| = 1.18$ MeV/c, occurring when p and e emerge back to back, so the minimum $\lambda = 10^{-10}$ cm]

10.10. Analyze π^- decay as a scattering process, using the methods of Example 7.8 and Section 9.3. Calculate the decay rate, and, by comparing your answer with the one in the text, obtain the formula for f_π in terms of $|\psi(0)|^2$. Assume $m_u = m_d = m$.

$$\left[\textit{Answer: } f_\pi^2 = \frac{4\hbar^3 m}{m_\pi^2 c}\cos^2\theta_C|\psi(0)|^2\right]$$

10.11. Show that if $mc^2 \ll E$

$$\gamma^5 u \cong \begin{pmatrix} \sigma\cdot\hat{p} & 0 \\ 0 & \sigma\cdot\hat{p} \end{pmatrix} u$$

where u is a particle spinor satisfying the Dirac equation:

$$u = \begin{pmatrix} u_A \\ \dfrac{c(\mathbf{p}\cdot\sigma)}{E + mc^2}\,u_A \end{pmatrix}$$

with $E > 0$ [eq. (7.36)]. Show therefore that the *projection matrix*

$$P_\pm \equiv \tfrac{1}{2}(1 \pm \gamma^5)$$

picks out the helicity ± 1 component of u:

$$\boldsymbol{\Sigma}\cdot\hat{p}(P_\pm u) = \pm(P_\pm u)$$

10.12. Calculate the ratio of the decay rates $K^- \to e^- + \bar{\nu}_e$ and $K^- \to \mu^- + \bar{\nu}_\mu$. The observed K^- lifetime is 1.2×10^{-8} sec, and 64% of all K^- particles decay by the $\mu^- + \bar{\nu}_\mu$ route. Estimate the kaon decay constant f_K.

10.13. Calculate decay rates for the following processes: (a) $\Sigma^0 \to \Sigma^+ + e + \bar{\nu}_e$, (b) $\Sigma^- \to \Lambda + e + \bar{\nu}_e$, (c) $\Xi^- \to \Xi^0 + e + \bar{\nu}_e$, (d) $\Lambda \to p + e + \bar{\nu}_e$, (e) $\Sigma^- \to n + e + \bar{\nu}_e$; (f) $\Xi^0 \to \Sigma^+ + e + \bar{\nu}_e$. Assume the coupling is always $\gamma^\mu(1 - \gamma^5)$—that is, ignore the strong interaction corrections to the axial coupling—but do not forget the Cabibbo factor. Compare the experimental data, where available.

10.14. **(a)** Show that as long as the KM matrix is *unitary* ($U^{-1} = U^\dagger$), the GIM mechanism for eliminating $K^0 \to \mu^+\mu^-$ works for three (or any number of) generations.

(b) How many independent real parameters are there in the general 3×3 unitary matrix? How about $n \times n$?

We are free to change the *phase* of each quark wave function (normalization of u really only determines $|N|^2$; see Problem 7.3), so $2n$ of these parameters are arbitrary—or rather, $(2n - 1)$, since changing the phase of *all* quark wave functions by the same amount has no effect on U. *Question:* Can we thus reduce the KM matrix to a *real* matrix (if it is real and unitary, then it is *orthogonal:* $U^{-1} = \tilde{U}$).

(c) How many independent real parameters are there in the general 3×3 (real) orthogonal matrix? How about $n \times n$?

(d) So, what is the answer? *Can* you reduce the KM matrix to real form? How about for only two generations ($n = 2$)?

10.15. Show that the KM matrix (10.90) is unitary for any (real) numbers θ_1, θ_2, θ_3, and δ.

10.16. Suppose you started with a T^+ meson ($t\bar{d}$). Given equation (10.91), what is the most likely sequence of decays? [*Answer:* Leaving out pions or leptons, we expect $T^+ \to B^0 \to D^+ \to \bar{K}^0 \to \pi^+$.]

10.17. Using the value of the Fermi constant G_F [eq. (10.40)] and of θ_w [eq. (10.95)], "predict" the mass of the $W^\pm$ and the Z^0, in GWS theory. Compare the experimental values.

10.18. In Example 10.4 I used *muon* neutrinos, rather than *electron* neutrinos. As a matter of fact, ν_μ and $\bar{\nu}_\mu$ beams are easier to produce then ν_e and $\bar{\nu}_e$, but there is also a *theoretical* reason why $\nu_\mu + e^- \to \nu_\mu + e^-$ is simpler than $\nu_e + e^- \to \nu_e + e^-$ or $\bar{\nu}_e + e^- \to \bar{\nu}_e + e^-$. Explain.

10.19. **(a)** Calculate the differential and total cross section for $\bar{\nu}_\mu + e^- \to \bar{\nu}_\mu + e^-$ in the GWS model.
[*Answer:* Same as equation (10.103), only with the sign of $c_A c_V$ reversed; see Halzen and Martin, ref. 12, eq. 13.49.]

(b) Find the ratio $\sigma(\bar{\nu}_\mu + e^- \to \bar{\nu}_\mu + e^-)/\sigma(\nu_\mu + e^- \to \nu_\mu + e^-)$. Assume the energy is high enough that you can set $m_e = 0$.

10.20. **(a)** Calculate the decay rate for $Z^0 \to f + \bar{f}$, where f is any quark or any lepton. Assume f is so light (compared to the Z) that its mass can be neglected.

$$\left[\textit{Answer: } \Gamma(Z^0 \to f + \bar{f}) = \frac{g_z^2 M_Z c^2}{48\pi\hbar}\left(|c_V^f|^2 + |c_A^f|^2\right).\right]$$

(b) Assuming these are the dominant decay modes, find the branching ratio for each species of quark and lepton (remember that the quarks come in three colors). Assume that $2m_t < M_Z$, and that the approximation in (a) is valid even for t.
[*Answer:* 3% each for e, μ, τ; 6% each for ν_e, ν_μ, ν_τ; 10% each for u, c, t; 14% each for d, s, b.]

(c) Calculate the lifetime of the Z^0. How would it change if there exists a fourth generation? (Notice that an accurate measurement of the Z^0 lifetime will tell us how many quarks and leptons there can be with masses less than 45 GeV/c^2.)

10.21. Estimate R (the total ratio of quark pair production to muon pair production in e^+e^- scattering), when the process is mediated by Z^0. For the sake of argument assume the top quark is light enough so that equation (10.112) can be used. Don't forget color.

10.22. Graph the ratio, equation (10.116) as a function of total energy ($2E$), using 2 for the expression in brackets, $M_Z c^2 = 90$ GeV, and $\hbar\Gamma_Z = 2.5$ GeV.

10.23. Derive equation (10.120), using equation (7.36). Also derive equation (10.124).

10.24. (a) If $u(p)$ satisfies the Dirac equation (7.34), show that u_L and u_R (Table 10.2) do *not* (unless $m = 0$).

(b) Find the eigenvalues and eigenspinors of the matrices $P_\pm \equiv \frac{1}{2}(1 \pm \gamma^5)$.

(c) Can there exist spinors that are simultaneously eigenstates of P_+ (say) *and* of the Dirac operator ($\not{p} - mc$)?
[*Answer:* No; these operators do not commute.]

10.25. Work out the weak isospin currents $j_\mu^\pm$ and j_μ^3 for the light quark doublet u and d'. Also, construct the electromagnetic current (j_μ^{em}) and the weak hypercharge current (j_μ^Y). (Leave your answers in terms of d'.)

10.26. From expression (10.159), determine the vector and axial vector couplings in Table 10.1.

Chapter 11

Gauge Theories

This chapter introduces the "gauge theories" that are now believed to underlie all elementary particle interactions. I begin with the Lagrangian formulation of classical mechanics, and proceed to Lagrangian field theory, the principle of local gauge invariance, the notion of spontaneous symmetry breaking, and the Higgs mechanism (which accounts for the mass of the W's and the Z). This material is quite abstract (in contrast to previous chapters); it concerns the fundamental quantum field theories from which the Feynman rules derive. It will not help you to calculate any cross sections or lifetimes. On the other hand, the ideas discussed here constitute the foundation on which virtually all modern theories are predicated. To understand this chapter it will help to have studied some Lagrangian mechanics, but more essential is the relativistic notation in Chapter 3, the taste of group theory in Chapter 4, the Feynman calculus from Chapter 6, and the Dirac equation from Chapter 7.

11.1 LAGRANGIAN FORMULATION OF CLASSICAL PARTICLE MECHANICS

According to Newton's second law of motion, a particle of mass m, subjected to a force $\mathbf{F}$, undergoes an acceleration $\mathbf{a}$ given by

$$\mathbf{F} = m\mathbf{a} \tag{11.1}$$

If the force is *conservative*, it can be expressed as the gradient of a scalar potential energy function U:

$$\mathbf{F} = -\nabla U \tag{11.2}$$

and Newton's law reads

$$m\frac{d\mathbf{v}}{dt} = -\nabla U \tag{11.3}$$

where $\mathbf{v}$ is the velocity.[1]

An alternative formulation of classical mechanics begins with the "Lagrangian"

$$L = T - U \tag{11.4}$$

where T is the kinetic energy of the particle:

$$T = \tfrac{1}{2}m\mathbf{v}^2 \tag{11.5}$$

The Lagrangian is a function of the coordinates q_i (say, $q_1 = x$, $q_2 = y$, $q_3 = z$) and their time derivatives $\dot{q}_i(\dot{q}_1 = v_x, \dot{q}_2 = v_y, \dot{q}_3 = v_z)$. In the Lagrangian formulation the fundamental law of motion is the Euler–Lagrange equation:[2]

$$\frac{d}{dt}\left(\frac{\partial L}{\partial \dot{q}_i}\right) = \frac{\partial L}{\partial q_i} \qquad (i = 1, 2, 3) \tag{11.6}$$

Thus in Cartesian coordinates we have

$$\frac{\partial L}{\partial \dot{q}_1} = \frac{\partial T}{\partial v_x} = mv_x \tag{11.7}$$

$$\frac{\partial L}{\partial q_1} = -\frac{\partial U}{\partial x} \tag{11.8}$$

and the Euler–Lagrange equation (for $i = 1$) reproduces the x component of Newton's law, in the form of equation (11.3). The Lagrangian formulation is thus equivalent to Newton's (at least, for conservative systems), but it has certain theoretical advantages, as we shall see in the following sections. (See also Problem 11.1.)

11.2 LAGRANGIANS IN RELATIVISTIC FIELD THEORY

A *particle,* by its nature, is a *localized* entity; in classical particle mechanics we are typically interested in calculating its *position* as a function of time: $x(t)$, $y(t)$, $z(t)$. A *field,* on the other hand, occupies some *region* of space; in field theory our concern is to calculate one or more functions of *position and time:* $\phi_i(x, y, z, t)$. The field variables ϕ_i might be, for example, the temperature at each point in a room, or the electric potential V, or the three components of the magnetic field $\mathbf{B}$. In particle mechanics we introduced a Lagrangian L that was a function of the coordinates q_i and their time derivatives, $\dot{q}_i$; in field theory we start with a Lagrangian (technically, a Lagrangian *density*) $\mathcal{L}$, which is a function of the fields ϕ_i and their x, y, z and t derivatives:

$$\partial_\mu \phi_i \equiv \frac{\partial \phi_i}{\partial x^\mu} \tag{11.9}$$

In the former case, the left side of the Euler–Lagrange equation (11.6) involves only *time* derivatives; a *relativistic* theory must treat space and time coordinates on an equal footing, and the Euler–Lagrange equations generalize as you might expect:

$$\partial_\mu\left(\frac{\partial \mathcal{L}}{\partial(\partial_\mu \phi_i)}\right) = \frac{\partial \mathcal{L}}{\partial \phi_i} \qquad (i = 1, 2, 3, \ldots) \tag{11.10}$$

EXAMPLE 11.1 The Klein–Gordon Lagrangian for a Scalar (Spin-0) Field

Suppose we have a single, scalar field variable ϕ, and the Lagrangian is

$$\mathcal{L} = \frac{1}{2}(\partial_\mu \phi)(\partial^\mu \phi) - \frac{1}{2}\left(\frac{mc}{\hbar}\right)^2 \phi^2 \tag{11.11}$$

In this case

$$\frac{\partial \mathcal{L}}{\partial(\partial_\mu \phi)} = \partial^\mu \phi \tag{11.12}$$

(If this confuses you, write out the Lagrangian "longhand":

$$\mathcal{L} = \frac{1}{2}[\partial_0\phi\ \partial_0\phi - \partial_1\phi\ \partial_1\phi - \partial_2\phi\ \partial_2\phi - \partial_3\phi\ \partial_3\phi] - \frac{1}{2}\left(\frac{mc}{\hbar}\right)^2 \phi^2$$

In this form it is clear that

$$\frac{\partial}{\partial(\partial_0\phi)} = \partial_0\phi = \partial^0\phi, \qquad \frac{\partial}{\partial(\partial_1\phi)} = -\partial_1\phi = \partial^1\phi,$$

and so on.) Meanwhile

$$\frac{\partial \mathcal{L}}{\partial \phi} = -\left(\frac{mc}{\hbar}\right)^2 \phi$$

and hence the Euler–Lagrange formula requires

$$\partial_\mu \partial^\mu \phi + \left(\frac{mc}{\hbar}\right)^2 \phi = 0 \tag{11.13}$$

which is the Klein–Gordon equation [eq (7.9)], describing (in quantum field theory) a particle of spin 0 and mass m.

EXAMPLE 11.2 The Dirac Lagrangian for a Spinor (Spin-$\frac{1}{2}$) Field

Consider now a spinor field ψ, and the Lagrangian

$$\mathcal{L} = i(\hbar c)\bar{\psi}\gamma^\mu\, \partial_\mu \psi - (mc^2)\bar{\psi}\psi \tag{11.14}$$

We treat ψ and the adjoint spinor $\bar{\psi}$ as independent field variables.* Applying the Euler–Lagrange equation to $\bar{\psi}$, we find

* Since ψ is a complex spinor, there are actually *eight* independent fields here (i runs from 1 to 8): the real and imaginary parts of each of the four components of ψ. But in applying the Euler–Lagrange equations any linear combinations of these eight will do just as well, and we choose to use the four components of ψ plus the four components of $\bar{\psi}$.

$$\frac{\partial \mathcal{L}}{\partial(\partial_\mu \bar{\psi})} = 0, \qquad \frac{\partial \mathcal{L}}{\partial \bar{\psi}} = i\hbar c\gamma^\mu \partial_\mu \psi - mc^2\psi$$

so that
$$i\gamma^\mu \partial_\mu \psi - \left(\frac{mc}{\hbar}\right)\psi = 0 \tag{11.15}$$

This is the Dirac equation [eq. (7.20)], describing (in quantum field theory) a particle of spin $\frac{1}{2}$ and mass m. Meanwhile, if we apply the Euler–Lagrange equation to ψ, we obtain

$$\frac{\partial \mathcal{L}}{\partial(\partial_\mu \psi)} = i\hbar c\bar{\psi}\gamma^\mu, \qquad \frac{\partial \mathcal{L}}{\partial \psi} = -mc^2\bar{\psi}$$

and hence
$$i\,\partial_\mu \bar{\psi}\gamma^\mu + \left(\frac{mc}{\hbar}\right)\bar{\psi} = 0$$

which is the adjoint of the Dirac equation (see Problem 7.13).

EXAMPLE 11.3 The Proca Lagrangian for a Vector (Spin-1) Field

Finally, suppose we take a vector field, A^μ, with the Lagrangian

$$\mathcal{L} = \frac{-1}{16\pi}(\partial^\mu A^\nu - \partial^\nu A^\mu)(\partial_\mu A_\nu - \partial_\nu A_\mu) + \frac{1}{8\pi}\left(\frac{mc}{\hbar}\right)^2 A^\nu A_\nu \tag{11.16}$$

Here
$$\frac{\partial \mathcal{L}}{\partial(\partial_\mu A_\nu)} = \frac{-1}{4\pi}(\partial^\mu A^\nu - \partial^\nu A^\mu) \tag{11.17}$$

(see Problem 11.2), and

$$\frac{\partial \mathcal{L}}{\partial A_\nu} = \frac{1}{4\pi}\left(\frac{mc}{\hbar}\right)^2 A^\nu \tag{11.18}$$

so the Euler–Lagrange equation yields

$$\partial_\mu(\partial^\mu A^\nu - \partial^\nu A^\mu) + \left(\frac{mc}{\hbar}\right)^2 A^\nu = 0 \tag{11.19}$$

This is called the Proca equation; it describes a particle of spin 1 and mass m. Incidentally, since the combination $(\partial^\mu A^\nu - \partial^\nu A^\mu)$ occurs repeatedly in this theory, it is useful to introduce the shorthand

$$F^{\mu\nu} \equiv \partial^\mu A^\nu - \partial^\nu A^\mu \tag{11.20}$$

Then the Lagrangian reads

$$\mathcal{L} = -\frac{1}{16\pi}F^{\mu\nu}F_{\mu\nu} + \frac{1}{8\pi}\left(\frac{mc}{\hbar}\right)^2 A^\nu A_\nu \tag{11.21}$$

and the field equation becomes

$$\partial_\mu F^{\mu\nu} + \left(\frac{mc}{\hbar}\right)^2 A^\nu = 0 \tag{11.22}$$

If the notation is beginning to remind you of electrodynamics, it's no accident, for the electromagnetic field *is* precisely a massless vector field; if you set $m = 0$ in equation (11.22) you're left with Maxwell's equations for empty space.

The Lagrangians in these examples came out of thin air (or rather, they were concocted in such a way as to reproduce the desired field equations). In classical particle mechanics L is *derived* ($L = T - U$), but in relativistic field theory $\mathcal{L}$ is usually taken as *axiomatic*—we have to start *somewhere.* The Lagrangian for a particular system is by no means unique; you can always multiply $\mathcal{L}$ by a constant, or add a divergence, ($\partial_\mu M^\mu$, where M^μ is any function of ϕ_i and $\partial_\mu \phi_i$)—such terms cancel out when you apply the Euler–Lagrange equations, so they do not affect the field equations. In this sense the factors of $\frac{1}{2}$ in the Klein–Gordon Lagrangian, for example, are purely conventional.* Apart from that, however, what we have here are *the* Lagrangians for spin 0, spin $\frac{1}{2}$, and spin 1. So far, however, we are talking only of *free* fields, with no sources or interactions.

EXAMPLE 11.4 The Maxwell Lagrangian for a Massless Vector Field with Source J^μ

Suppose

$$\mathcal{L} = \frac{-1}{16\pi} F^{\mu\nu} F_{\mu\nu} - \frac{1}{c} J^\mu A_\mu \tag{11.23}$$

where $F^{\mu\nu}$ (again) stands for $\partial^\mu A^\nu - \partial^\nu A^\mu$, and J^μ is some specified function. The Euler–Lagrange equations yield

$$\partial_\mu F^{\mu\nu} = \frac{4\pi}{c} J^\nu \tag{11.24}$$

which (as we found in Chapter 7, Section 7.4) is the tensor form of Maxwell's equations, describing the electromagnetic fields produced by a current J^μ. Incidentally, it follows from equation (11.24) that

$$\partial_\nu J^\nu = 0 \tag{11.25}$$

That is, the internal consistency of the Maxwell Lagrangian (11.23) requires that the current satisfy the continuity equation (7.74); you can't just put in *any* old function for J^μ—it's got to respect conservation of charge.

* The Lagrangian (L) carries units of energy [eq. (11.4)], and the Lagrangian *density* ($\mathcal{L}$) has the units of energy *per unit volume.* The *fields* carry dimensions as follows:

ϕ (scalar field): $\sqrt{ML}/T$
ψ (spinor field): $L^{-3/2}$
A^μ (vector field): $\sqrt{ML}/T$

These are chosen so that ψ will go over to the Schrodinger wave function (in the nonrelativistic limit) and A^μ to the Maxwell vector potential (in the nonquantum limit). By the way, in Heaviside–Lorentz units the Proca and Maxwell Lagrangians would be multiplied by 4π.

11.3 LOCAL GAUGE INVARIANCE

Notice that the Dirac Lagrangian

$$\mathcal{L} = i\hbar c\bar{\psi}\gamma^\mu \partial_\mu\psi - mc^2\bar{\psi}\psi \tag{11.14}$$

is invariant under the transformation

$$\psi \rightarrow e^{i\theta}\psi \qquad \text{(global gauge transformation)} \tag{11.26}$$

(where θ is any real number), for then $\bar{\psi} \rightarrow e^{-i\theta}\bar{\psi}$, and in the combination $\bar{\psi}\psi$ the exponential factors cancel out. For historical reasons, we call (11.26) a (global) *gauge transformation* ("phase" transformation would be a more sensible term). But what if the phase factor is different at different space-time points; that is, *what if θ is a function of x^μ*:

$$\psi \rightarrow e^{i\theta(x)}\psi \qquad \text{(local gauge transformation)} \tag{11.27}$$

Is the Lagrangian invariant under such a *"local"* gauge transformation? The answer is *no,* for now we pick up an extra term from the derivative of θ:

$$\partial_\mu(e^{i\theta}\psi) = i(\partial_\mu\theta)e^{i\theta}\psi + e^{i\theta}\partial_\mu\psi \tag{11.28}$$

so that

$$\mathcal{L} \rightarrow \mathcal{L} - \hbar c(\partial_\mu\theta)\bar{\psi}\gamma^\mu\psi \tag{11.29}$$

Actually, for what follows it is convenient to pull a factor of $-(q/\hbar c)$ out of θ, letting

$$\lambda(x) \equiv -\frac{\hbar c}{q}\,\theta(x) \tag{11.30}$$

where q is the charge of the particle involved. In terms of λ, then,

$$\mathcal{L} \rightarrow \mathcal{L} + (q\bar{\psi}\gamma^\mu\psi)\partial_\mu\lambda \tag{11.31}$$

under the local gauge transformation

$$\psi \rightarrow e^{-iq\lambda(x)/\hbar c}\psi \tag{11.32}$$

So far, there is nothing particularly new or deep in all this. The crucial point comes when we *demand* that the *complete* Lagrangian be invariant under local gauge transformations.* Since the *free* Dirac Lagrangian (11.14) is *not* locally gauge invariant, we are obliged to *add* something, in order to soak up the extra term in equation (11.31). Suppose

$$\mathcal{L} = [i\hbar c\bar{\psi}\gamma^\mu \partial_\mu\psi - mc^2\bar{\psi}\psi] - (q\bar{\psi}\gamma^\mu\psi)A_\mu \tag{11.33}$$

where A_μ is some *new* field (called a "gauge" field) which transforms under local gauge transformations according to the rule

* I know of no compelling physical argument for why a global invariance *should* necessarily hold locally. If you believe that gauge transformations are in some sense "fundamental," then I suppose one should be able to carry them out independently at spacelike-separated points (which are, after all, out of communication with one another). But I think this begs the question. Better, for the moment at least, to take the requirement of local gauge invariance as a new principle of physics in its own right.

$$A_\mu \rightarrow A_\mu + \partial_\mu \lambda \tag{11.34}$$

This "new, improved" Lagrangian *is* now invariant under local gauge transformations; the price we had to pay was the introduction of a new vector field that couples to ψ through the last term in equation (11.33) (see Problem 11.6). But equation (11.33) isn't the whole story; the *full* Lagrangian must include a "free" term for the gauge field. Since it is a vector field, we look to the Proca Lagrangian

$$\mathcal{L} = \frac{-1}{16\pi} F^{\mu\nu} F_{\mu\nu} + \frac{1}{8\pi} \left(\frac{m_A c}{\hbar} \right)^2 A^\nu A_\nu \tag{11.21}$$

But there is a problem here, for whereas $F^{\mu\nu} \equiv \partial^\mu A^\nu - \partial^\nu A^\mu$ is invariant under (11.34), as you should check for yourself, $A^\nu A_\nu$ is *not. Evidently, the gauge field must be massless* ($m_A = 0$), otherwise local gauge invariance will be lost.

Conclusion: if we start with the Dirac Lagrangian, and impose local gauge invariance, we are forced to introduce a massless vector field (A^μ), and the complete Lagrangian becomes

$$\mathcal{L} = [i\hbar c \bar{\psi} \gamma^\mu \partial_\mu \psi - mc^2 \bar{\psi} \psi] + \left[\frac{-1}{16\pi} F^{\mu\nu} F_{\mu\nu} \right] - [(q \bar{\psi} \gamma^\mu \psi) A_\mu] \tag{11.35}$$

As you will have guessed, A^μ is precisely the electromagnetic potential; the gauge transformation rule for A^μ (11.34) is just what we found back in Chapter 7 [eq. (7.81)], and the last two terms in equation (11.35) reproduce the Maxwell Lagrangian (11.23), with the current density

$$J^\mu = cq(\bar{\psi} \gamma^\mu \psi) \tag{11.36}$$

Thus the requirement of local gauge invariance, applied to the free Dirac Lagrangian, generates all of electrodynamics, and specifies the current produced by Dirac particles.

In case the procedure for invoking local gauge invariance seems mysterious, let's review it, and see what was actually involved. The difference between global and local gauge transformations arises when we calculate *derivatives* of the fields [eq. (11.28)]:

$$\partial_\mu \psi \rightarrow e^{-iq\lambda/\hbar c} \left[\partial_\mu - i \frac{q}{\hbar c} (\partial_\mu \lambda) \right] \psi \tag{11.37}$$

Instead of a simple phase factor, we pick up an extra term involving $\partial_\mu \lambda$. *If in the original (free) Lagrangian we replace every derivative* (∂_μ) *by the so-called "covariant derivative"*

$$\mathcal{D}_\mu \equiv \partial_\mu + i \frac{q}{\hbar c} A_\mu \tag{11.38}$$

the transformation of A_μ [eq. (11.34)] will cancel the offending term in equation (11.37)

$$\mathcal{D}_\mu \psi \rightarrow e^{-iq\lambda/\hbar c} \mathcal{D}_\mu \psi \tag{11.39}$$

and the invariance of $\mathcal{L}$ is restored. The substitution of $\mathcal{D}_\mu$ for ∂_μ, then, is a simple device for converting a *globally* invariant Lagrangian into a *locally* invariant one; we call this the *"minimal coupling rule"* [it's what I used, in fact, to generate the extra term in eq. (11.33)].* But the covariant derivative introduces a new vector field (A_μ), which requires its own *free* Lagrangian; if the latter is not to spoil local gauge invariance, we must take the gauge fields to be massless. This leads to the final expression (11.35), which people in the know would immediately recognize as the Lagrangian for quantum electrodynamics—Dirac fields (electrons and positrons) interacting wth Maxwell fields (photons).

The idea of local gauge invariance goes back to the work of Hermann Weyl in 1919.[3] However, its power and generality were not fully appreciated until the early seventies. Our starting point—the global phase transformation (11.26)—may be thought of as multiplication of ψ by a unitary 1×1 matrix:

$$\psi \rightarrow U\psi, \qquad \text{where } U^\dagger U = 1 \tag{11.40}$$

(Here $U = e^{i\theta}$). The group of *all* such matrices is $U(1)$ (see Table 4.2), and hence the symmetry involved is called "$U(1)$ gauge invariance." This terminology is extravagant for the case at hand (a 1×1 matrix is a *number,* so why not leave it at that?), but in 1954 Yang and Mills[4] applied the same strategy (insisting that a *global* invariance hold locally) to the group $SU(2)$, and later on the idea was extended to color $SU(3)$, producing chromodynamics. In the Standard Model *all* of the fundamental interactions are generated in this way.

11.4 YANG–MILLS THEORY

Suppose now that we have *two* spin-$\frac{1}{2}$ fields, ψ_1 and ψ_2. The Lagrangian, in the absence of any interactions, is

$$\mathcal{L} = [i\hbar c\bar{\psi}_1\gamma^\mu \partial_\mu\psi_1 - m_1c^2\bar{\psi}_1\psi_1] + [i\hbar c\bar{\psi}_2\gamma^\mu\partial_\mu\psi_2 - m_2c^2\bar{\psi}_2\psi_2] \tag{11.41}$$

It's just the *sum* of the two Dirac Lagrangians. (Apply the Euler–Lagrange equations to this $\mathcal{L}$, and you'll find that ψ_1 and ψ_2 both obey the Dirac equation, with the appropriate mass.) But we can write equation (11.41) more compactly by combining ψ_1 and ψ_2 into a two-component column vector:

$$\psi \equiv \begin{pmatrix} \psi_1 \\ \psi_2 \end{pmatrix} \tag{11.42}$$

(Of course, ψ_1 and ψ_2 are themselves four-component Dirac spinors, and you might prefer a double-index notation: $\psi_{\alpha,i}$, where $\alpha = 1, 2$ identifies the *particle*

* The minimal coupling rule is much older than the principle of local gauge invariance. In terms of momentum [$p_\mu \leftrightarrow i\hbar\partial_\mu$, see eq. (7.5)] it reads $p_\mu \rightarrow p_\mu - i(q/c)A_\mu$, and is a well-known trick in classical electrodynamics for obtaining the equation of motion for a charged particle in the presence of electrodynamic fields. See J. D. Jackson, *Classical Electrodynamics,* 2d Ed. (New York: Wiley, 1975), eq. (12.29). It amounts, in this sense, to a sophisticated formulation of the Lorentz force law. In modern particle theory we prefer to regard local gauge invariance as fundamental, and minimal coupling as a device for achieving it.

and $i = 1, 2, 3, 4$ labels the spinor component. However, in the present context we are *only* concerned with the particle index, although the Dirac matrices, of course, act on the spinor indices.) The adjoint spinor is

$$\bar{\psi} = (\bar{\psi}_1 \quad \bar{\psi}_2) \tag{11.43}$$

and the Lagrangian becomes

$$\mathcal{L} = i\hbar c\bar{\psi}\gamma^\mu \partial_\mu \psi - c^2\bar{\psi}M\psi \tag{11.44}$$

where

$$M = \begin{pmatrix} m_1 & 0 \\ 0 & m_2 \end{pmatrix} \tag{11.45}$$

is the "mass matrix." In particular, if the two masses happen to be *equal* equation (11.44) reduces to

$$\mathcal{L} = i\hbar c\bar{\psi}\gamma^\mu \partial_\mu \psi - mc^2\bar{\psi}\psi \tag{11.46}$$

This *looks* just like the one-particle Dirac Lagrangian. However, ψ is now a two-element column vector, and $\mathcal{L}$ admits a more general global invariance than before:

$$\psi \longrightarrow U\psi \tag{11.47}$$

where U is any 2×2 unitary matrix

$$U^\dagger U = 1 \tag{11.48}$$

For under the transformation (11.47),

$$\bar{\psi} \longrightarrow \bar{\psi}U^\dagger \tag{11.49}$$

and hence the combination $\bar{\psi}\psi$ is invariant. Now, just as any complex number of modulus 1 can be written in the form $e^{i\theta}$, with real θ, so any unitary matrix can be written in the form[5]

$$U = e^{iH} \tag{11.50}$$

where H is *Hermitian* ($H^\dagger = H$).* Moreover, the most general Hermitian 2×2 matrix can be expressed in terms of four real numbers, a_1, a_2, a_3, and θ (Problem 11.10):

$$H = \theta 1 + \boldsymbol{\tau} \cdot \mathbf{a} \tag{11.51}$$

where 1 is the 2×2 unit matrix, τ_1, τ_2, τ_3 are the Pauli matrices (4.26), and the dot product is a convenient shorthand for $\tau_1 a_1 + \tau_2 a_2 + \tau_3 a_3$. Thus any unitary 2×2 matrix can be expressed as a product:

$$U = e^{i\theta}e^{i\boldsymbol{\tau}\cdot\mathbf{a}} \tag{11.52}$$

* In matrix theory the natural generalization of complex conjugation ($*$) is Hermitian conjugation ($\dagger$)—*transpose* conjugation. Of course, there's no distinction in the case of 1×1 matrices (complex numbers), but for higher dimensions it is the Hermitian conjugate that shares the most useful properties of ordinary complex conjugation. In this sense the closest analog to a *real* number ($a = a^*$) is a *Hermitian* matrix ($A = A^\dagger$), and the analog to a number of modulus 1 ($a^*a = 1$) is a unitary matrix ($A^\dagger A = 1$).

We have already explored the implications of *phase* transformations ($e^{i\theta}$); in this section we shall concentrate on transformations of the form

$$\psi \rightarrow e^{i\boldsymbol{\tau}\cdot\mathbf{a}}\psi \qquad \text{[global } SU(2) \text{ transformation]} \tag{11.53}$$

The matrix $e^{i\boldsymbol{\tau}\cdot\mathbf{a}}$ has determinant 1 [see Problem 4.22], and therefore belongs to the group $SU(2)$. Generalizing the terminology of Section 11.3, we say that Lagrangian (11.46) is invariant under global $SU(2)$ gauge transformations.* What Yang and Mills did was to promote this *global* invariance to the status of a *local* invariance.

The inspiration and the strategy were similar to Weyl's, but the implementation is more subtle, in fact, it's quite remarkable that it works at all. The first step is to let the parameters $(\mathbf{a})$ be functions of x^μ [as before, equation (11.30), I'll let $\boldsymbol{\lambda}(x) \equiv -(\hbar c/q)\mathbf{a}(x)$, where q is a coupling constant analogous to electric charge]:

$$\psi \rightarrow S\psi, \qquad \text{where } S \equiv e^{-iq\boldsymbol{\tau}\cdot\boldsymbol{\lambda}(x)/\hbar c} \qquad \text{[local } SU(2) \text{ transformation]} \tag{11.54}$$

As it stands, $\mathcal{L}$ is *not* invariant under such a transformation, for the derivative picks up an extra term:

$$\partial_\mu\psi \rightarrow S\,\partial_\mu\psi + (\partial_\mu S)\psi \tag{11.55}$$

The remedy, again, is to replace the derivative in $\mathcal{L}$ by a "covariant derivative," modeled on equation (11.38), but taking into account the structure of equation (11.55):

$$\mathcal{D}_\mu \equiv \partial_\mu + i\frac{q}{\hbar c}\boldsymbol{\tau}\cdot\mathbf{A}_\mu \tag{11.56}$$

and assign to the gauge fields $\mathbf{A}_\mu$ (it takes *three* of them this time) a transformation rule such that

$$\mathcal{D}_\mu\psi \rightarrow S(\mathcal{D}_\mu\psi) \tag{11.57}$$

For then the Lagrangian (11.46) will clearly be invariant.

It is not a trivial matter to deduce the transformation rule for $\mathbf{A}_\mu$ from (11.57).[6] I'll leave it for you to show (Problem 11.11) that $\mathbf{A}_\mu \rightarrow \mathbf{A}'_\mu$, where $\mathbf{A}'_\mu$ is given by

$$\boldsymbol{\tau}\cdot\mathbf{A}'_\mu = S(\boldsymbol{\tau}\cdot\mathbf{A}_\mu)S^{-1} + i\left(\frac{\hbar c}{q}\right)(\partial_\mu S)S^{-1} \tag{11.58}$$

This much is relatively straightforward. But S and S^{-1} in the first term cannot be brought together, because they do not commute with $\boldsymbol{\tau}\cdot\mathbf{A}_\mu$. Nor is the gradient of S simply $-i(q\boldsymbol{\tau}\cdot\partial_\mu\boldsymbol{\lambda}/\hbar c)S$, because S does not commute with $\boldsymbol{\tau}\cdot\partial_\mu\boldsymbol{\lambda}$. You can work out the exact result (using Problems 4.20 and 4.21), if you have the energy,

* It is *also* invariant under the *larger* group $U(2)$. But (11.52) shows that any element of $U(2)$ can be expressed as an element of $SU(2)$ times an appropriate phase factor (in group-theoretical language, $U(2) = U(1) \otimes SU(2)$), and since we have already studied $U(1)$ invariance, the only thing *new* here is the $SU(2)$ symmetry.

but the answer is not particularly illuminating. For our purposes it will suffice to know the *approximate* transformation rule, in the limiting case of very small λ, for which we may expand S and keep only the first-order terms:

$$S \cong 1 - \frac{iq}{\hbar c}\boldsymbol{\tau}\cdot\boldsymbol{\lambda}, \qquad S^{-1} \cong 1 + \frac{iq}{\hbar c}\boldsymbol{\tau}\cdot\boldsymbol{\lambda}, \qquad \partial_\mu S \cong -\frac{iq}{\hbar c}\boldsymbol{\tau}\cdot(\partial_\mu\boldsymbol{\lambda}) \tag{11.59}$$

In this approximation equation (11.58) yields

$$\boldsymbol{\tau}\cdot\mathbf{A}'_\mu \cong \boldsymbol{\tau}\cdot\mathbf{A}_\mu + \frac{iq}{\hbar c}[\boldsymbol{\tau}\cdot\mathbf{A}_\mu, \boldsymbol{\tau}\cdot\boldsymbol{\lambda}] + \boldsymbol{\tau}\cdot\partial_\mu\boldsymbol{\lambda} \tag{11.60}$$

and hence (using Problem 4.20, to evaluate the commutator)

$$\mathbf{A}'_\mu \cong \mathbf{A}_\mu + \partial_\mu\boldsymbol{\lambda} + \frac{2q}{\hbar c}(\boldsymbol{\lambda}\times\mathbf{A}_\mu) \tag{11.61}$$

The resulting Lagrangian

$$\mathcal{L} = i\hbar c\bar{\psi}\gamma^\mu\mathcal{D}_\mu\psi - mc^2\bar{\psi}\psi = [i\hbar c\bar{\psi}\gamma^\mu\partial_\mu\psi - mc^2\bar{\psi}\psi] - (q\bar{\psi}\gamma^\mu\boldsymbol{\tau}\psi)\cdot\mathbf{A}_\mu \tag{11.62}$$

is invariant under local gauge transformations (11.54) and (11.58), but we have been obliged to introduce three new vector fields $\mathbf{A}^\mu = (A_1^\mu, A_2^\mu, A_3^\mu)$, and they will require their own *free* Lagrangian:

$$\mathcal{L}_A = -\frac{1}{16\pi}F_1^{\mu\nu}F_{\mu\nu 1} - \frac{1}{16\pi}F_2^{\mu\nu}F_{\mu\nu 2} - \frac{1}{16\pi}F_3^{\mu\nu}F_{\mu\nu 3} = -\frac{1}{16\pi}\mathbf{F}^{\mu\nu}\cdot\mathbf{F}_{\mu\nu} \tag{11.63}$$

(Again, the three-vector notation pertains to the *particle* indices.) The Proca mass term

$$\frac{1}{8\pi}\left(\frac{m_A c}{\hbar}\right)^2\mathbf{A}^\nu\cdot\mathbf{A}_\nu \tag{11.64}$$

is excluded by local gauge invariance; as before, the gauge fields must be massless. But this time the old association $F^{\mu\nu} = \partial^\mu A^\nu - \partial^\nu A^\mu$ must itself be modified, for with this definition the gauge field Lagrangian (11.63) is not invariant *either* (see Problem 11.12). Rather, we take*

$$\mathbf{F}^{\mu\nu} \equiv \partial^\mu\mathbf{A}^\nu - \partial^\nu\mathbf{A}^\mu - \frac{2q}{\hbar c}(\mathbf{A}^\mu\times\mathbf{A}^\nu) \tag{11.65}$$

Under infinitesimal local gauge transformations (11.61)

$$\mathbf{F}^{\mu\nu} \to \mathbf{F}^{\mu\nu} + \frac{2q}{\hbar c}(\boldsymbol{\lambda}\times\mathbf{F}^{\mu\nu}) \tag{11.66}$$

(Problem 11.13), and hence $\mathcal{L}_A$ is invariant. (See Problem 11.14 for a proof that the invariance extends to *finite* gauge transformations.)

* Definition (11.65) is not as arbitrary as it may seem; the point is that with *three* vector fields there is a second antisymmetric tensor form available, $(\mathbf{A}^\mu\times\mathbf{A}^\nu)$, and the coefficient, $-2q/\hbar c$, is chosen precisely to *make* $\mathcal{L}_A$ invariant. Notice that when the coupling constant q goes to zero we are left with the free Dirac Lagrangian for each spinor field and the free (massless) Proca Lagrangian for each of the three gauge fields.

Conclusion: The complete Yang–Mills Lagrangian is

$$\mathcal{L} = [i\hbar c\bar{\psi}\gamma^\mu \partial_\mu \psi - mc^2\bar{\psi}\psi] - \frac{1}{16\pi}\mathbf{F}^{\mu\nu}\cdot\mathbf{F}_{\mu\nu} - (q\bar{\psi}\gamma^\mu\tau\psi)\cdot\mathbf{A}_\mu \qquad (11.67)$$

with $\mathbf{F}^{\mu\nu}$ defined by equation (11.65); it is invariant under local $SU(2)$ gauge transformations, (11.54) and (11.58), and describes two equal-mass Dirac fields in interaction with three massless vector gauge fields. It all results from insisting that the *global* $SU(2)$ invariance of the original free Lagrangian (11.46) shall hold *locally*. Borrowing the language of electrodynamics, we say that the Dirac fields generate three *currents*

$$\mathbf{J}^\mu \equiv cq(\bar{\psi}\gamma^\mu\tau\psi) \qquad (11.68)$$

which act as *sources* for the gauge fields; the Lagrangian for the gauge fields alone

$$\mathcal{L} = -\frac{1}{16\pi}\mathbf{F}^{\mu\nu}\cdot\mathbf{F}_{\mu\nu} - \frac{1}{c}\mathbf{J}^\mu\cdot\mathbf{A}_\mu \qquad (11.69)$$

is reminiscent of the Maxwell Lagrangian (11.23), and gives rise to a rich and interesting *classical* field theory.[7] (See Problem 11.15.)

Although Yang–Mills theory is inspired by the same *idea* as Weyl's (namely: a global invariance should hold locally), the implementation was more subtle at two points: (1) the local transformation rule for gauge fields, and (2) the expression for $F^{\mu\nu}$ in terms of A^μ. Both complications derive from the fact that the symmetry group in question is non-Abelian (2×2 matrices do not commute, whereas 1×1 matrices—obviously—do). To emphasize the distinction, we refer to the Weyl case as an *Abelian* gauge theory, and Yang–Mills as a *non-Abelian* gauge theory. In contemporary elementary particle physics many symmetry groups have been explored; we shall encounter a few in the remaining sections of this book. However, the *hard* work is over: Extending non-Abelian gauge theory to higher symmetry groups is a straightforward procedure, once the Yang–Mills model is on the table.

Curiously, though, Yang–Mills theory in its original form turned out to be of little use. After all, it starts from the premise that there exist two elementary spin-$\frac{1}{2}$ particles of equal mass, and as far as we know there are no such pairs in nature. Yang and Mills themselves had the nucleon system (proton and neutron) in mind, and thought of their model as a way of implementing Heisenberg's isospin invariance in the strong interactions. The small mass difference between proton and neutron, 1.29 MeV/c^2, would be attributed to electromagnetic symmetry-breaking. For the theory to succeed there had to exist a massless isotriplet of vector (spin-1) particles. The only candidates in sight are the ρ mesons; but they are hardly *massless* ($M_\rho = 770$ MeV/c^2), and this is not a minor discrepancy that can be plausibly blamed on electromagnetic contamination. A number of attempts were made to doctor up Yang–Mills theory to accommodate massive gauge bosons, but by the time they finally bore fruit (through the Higgs mechanism) it was pretty clear that p, n, and ρ are composite particles anyway, and that isospin is just one component of a larger flavor symmetry that is too

drastically broken to play any fundamental role in the strong interactions. When non-Abelian gauge theory finally came into its own, it was in the context of *color* ($SU(3)$) symmetry in the strong interactions and *weak* isospin-hypercharge ($SU(2) \otimes U(1)$) symmetry in the weak interactions. Meanwhile, for more than a decade after 1954 the Yang–Mills model languished—a lovely idea that nature had evidently chosen not to exploit.

11.5 CHROMODYNAMICS

According to the colored quark model, each flavor of quark comes in three colors—red, blue, and green. Although the various *flavors* carry different masses (Table 4.4), the three *colors* of a given flavor are all supposed to weigh the same. Thus the free Lagrangian for a particular flavor reads

$$\mathcal{L} = [i\hbar c\bar{\psi}_r\gamma^\mu \partial_\mu\psi_r - mc^2\bar{\psi}_r\psi_r] + [i\hbar c\bar{\psi}_b\gamma^\mu \partial_\mu\psi_b - mc^2\bar{\psi}_b\psi_b] + [i\hbar c\bar{\psi}_g\gamma^\mu \partial_\mu\psi_g - mc^2\bar{\psi}_g\psi_g] \tag{11.70}$$

As before, we can simplify the notation by introducing

$$\psi \equiv \begin{pmatrix} \psi_r \\ \psi_b \\ \psi_g \end{pmatrix}, \qquad \bar{\psi} = (\bar{\psi}_r\, \bar{\psi}_b\, \bar{\psi}_g) \tag{11.71}$$

so that

$$\mathcal{L} = i\hbar c\bar{\psi}\gamma^\mu \partial_\mu\psi - mc^2\bar{\psi}\psi \tag{11.72}$$

This looks just like the original Dirac Lagrangian, only ψ now stands for a three-component column vector (each element of which is itself a four-component Dirac spinor). Just as the one-particle Dirac Lagrangian (11.14) has (global) $U(1)$ phase invariance, and the (equal mass) two-particle Lagrangian (11.41) admits $U(2)$ invariance, so this (equal mass) three-particle Lagrangian exhibits $U(3)$ symmetry. That is to say, it is invariant under transformations of the form

$$\psi \to U\psi \qquad (\bar{\psi} \to \bar{\psi}U^\dagger) \tag{11.73}$$

where U is any unitary 3×3 matrix:

$$U^\dagger U = 1 \tag{11.74}$$

But remember [eq. (11.50)], any unitary matrix can be written as an exponentiated Hermitian matrix:

$$U = e^{iH}, \qquad \text{with } H^\dagger = H \tag{11.75}$$

Moreover, any 3×3 Hermitian matrix can be expressed in terms of nine real numbers, $a_1, a_2, \ldots, a_8$, and θ (Problem 11.16):

$$H = \theta 1 + \boldsymbol{\lambda}\cdot\mathbf{a} \tag{11.76}$$

where 1 is the 3×3 unit matrix, $\lambda_1, \lambda_2, \ldots, \lambda_8$ are the Gell-Mann matrices [eq. (9.9)], and the dot product now denotes a sum from 1 to 8:

$$\boldsymbol{\lambda}\cdot\mathbf{a} \equiv \lambda_1 a_1 + \lambda_2 a_2 + \cdots + \lambda_8 a_8 \tag{11.77}$$

Thus $$U = e^{i\theta}e^{i\boldsymbol{\lambda}\cdot\mathbf{a}} \tag{11.78}$$

We have already explored *phase* transformations ($e^{i\theta}$); what is *new* is the second term. The matrix $e^{i\boldsymbol{\lambda}\cdot\mathbf{a}}$ has determinant 1 (see Problem 11.17); it belongs to the group $SU(3)$.* So what we are interested in is the invariance of the Lagrangian (11.72) under $SU(3)$ gauge transformations, a global symmetry that we now propose to make local.

That is: we modify $\mathcal{L}$ in such a way as to render it invariant under *local* $SU(3)$ gauge transformations:

$$\psi \rightarrow S\psi, \qquad \text{where } S \equiv e^{-iq\boldsymbol{\lambda}\cdot\boldsymbol{\phi}(x)/\hbar c} \tag{11.79}$$

(again, I let $\boldsymbol{\phi} \equiv -(\hbar c/q)\mathbf{a}$, with the coupling constant q playing a role analogous to electric charge in QED). As always, the trick is to replace the ordinary derivative, ∂_μ, by the "covariant derivative" $\mathcal{D}_\mu$:

$$\mathcal{D}_\mu \equiv \partial_\mu + i\frac{q}{\hbar c}\boldsymbol{\lambda}\cdot\mathbf{A}_\mu \tag{11.80}$$

and assign to the gauge fields $\mathbf{A}_\mu$ (there are *eight* of them, notice) a transformation rule such that

$$\mathcal{D}_\mu\psi \rightarrow S(\mathcal{D}_\mu\psi) \tag{11.81}$$

Again [see eq. (11.58)], this entails

$$\boldsymbol{\lambda}\cdot\mathbf{A}'_\mu = S(\boldsymbol{\lambda}\cdot\mathbf{A}_\mu)S^{-1} + i\left(\frac{\hbar c}{q}\right)(\partial_\mu S)S^{-1} \tag{11.82}$$

which, for the infinitesimal case, yields a formula identical to expression (11.61)

$$\mathbf{A}'_\mu \cong \mathbf{A}_\mu + \partial_\mu\boldsymbol{\phi} + \frac{2q}{\hbar c}(\boldsymbol{\phi}\times\mathbf{A}_\mu) \tag{11.83}$$

However, this time the cross-product notation is shorthand for

$$(\mathbf{B}\times\mathbf{C})_i = \sum_{j,k=1}^{8} f_{ijk}B_jC_k \tag{.84}$$

where f_{ijk} are the structure constants of $SU(3)$ [eq. (9.10)], analogous to ϵ_{ijk} for $SU(2)$. (See Problem 11.18.)

The modified Lagrangian

$$\mathcal{L} = i\hbar c\bar{\psi}\gamma^\mu\mathcal{D}_\mu\psi - mc^2\bar{\psi}\psi = [i\hbar c\bar{\psi}\gamma^\mu\partial_\mu\psi - mc^2\bar{\psi}\psi] - (q\bar{\psi}\gamma^\mu\boldsymbol{\lambda}\psi)\cdot\mathbf{A}_\mu \tag{11.85}$$

is invariant under local $SU(3)$ gauge transformations [(11.79) and (11.82)] but as usual the cost is the introduction of gauge fields $\mathbf{A}^\mu$ (eight of them, this time). In particle language, these correspond to the eight gluons, just as the $U(1)$ gauge field in Weyl's theory represents the photon.† To finish the job we must adjoin the free gluon Lagrangian

* In the language of group theory, we have shown that $U(3) = U(1) \otimes SU(3)$.

† Remember that a "ninth gluon," coupling universally to all quarks, is apparently excluded by experiment (see Problem 9.1).

$$\mathcal{L}_{\text{gluons}} = -\frac{1}{16\pi}\mathbf{F}^{\mu\nu}\cdot\mathbf{F}_{\mu\nu} \tag{11.86}$$

where, as in the Yang–Mills case

$$\mathbf{F}^{\mu\nu} \equiv \partial^{\mu}\mathbf{A}^{\nu} - \partial^{\nu}\mathbf{A}^{\mu} - \frac{2q}{\hbar c}(\mathbf{A}^{\mu}\times\mathbf{A}^{\nu}) \tag{11.87}$$

[with the $SU(3)$ "cross-product" defined by equation (11.84)].

Conclusion: The complete Lagrangian for chromodynamics is

$$\mathcal{L} = [i\hbar c\bar{\psi}\gamma^{\mu}\partial_{\mu}\psi - mc^2\bar{\psi}\psi] - \frac{1}{16\pi}\mathbf{F}^{\mu\nu}\cdot\mathbf{F}_{\mu\nu} - (q\bar{\psi}\gamma^{\mu}\lambda\psi)\cdot\mathbf{A}_{\mu} \tag{11.88}$$

Of course, we need six replicas of equation (11.88), each with the appropriate mass, to handle the six quark flavors. $\mathcal{L}$ is invariant under local $SU(3)$ gauge transformations, and describes three equal mass Dirac fields (the three colors of a given quark flavor) in interaction with eight massless vector fields (the gluons). It derives from the requirement that the *global* $SU(3)$ symmetry of the original Lagrangian (11.70) should hold *locally*. The Dirac fields constitute eight color *currents*

$$\mathbf{J}^{\mu} \equiv cq(\bar{\psi}\gamma^{\mu}\lambda\psi) \tag{11.89}$$

which act as *sources* for the color fields ($\mathbf{A}_{\mu}$), in the same way that *electric* currents act as sources for the *electromagnetic* field. The theory described here is very close in structure to that of Yang and Mills. In this case, however, we believe it to be the correct description of a phenomenon realized in nature: the strong interaction.

11.6 FEYNMAN RULES

Up to this point the Lagrangians we have considered might just as well describe *classical* fields as *quantum* ones; indeed, the Maxwell Lagrangian will be found in any textbook on classical electrodynamics. The passage from a classical field theory to the corresponding quantum field theory does not involve modification of the Lagrangian or the field equations, but rather a *reinterpretation* of the field variables; the fields are "quantized," and *particles* emerge as quanta of the associated fields. Thus the photon is the quantum of the electrodynamic field, A^{μ}; leptons and quarks are quanta of Dirac fields; gluons are quanta of the eight $SU(3)$ gauge fields; and $W^{\pm}$ and Z^0 are quanta of the appropriate Proca fields. The quantization procedure itself is recondite, and this is not the place to go into it;[8] for our purposes the essential point is that *each Lagrangian determines a particular set of Feynman rules.* What we need, then, is a prescription for working out the Feynman rules dictated by a given Lagrangian.

To begin with, notice that $\mathcal{L}$ consists of two kinds of terms: the *free* Lagrangian for each participating field, plus various *interaction* terms ($\mathcal{L}_{\text{int}}$). The former—Klein–Gordon, for spin 0; Dirac, for spin $\frac{1}{2}$; Proca, for spin 1; or something more exotic, for a theory with higher spin—determines the *propagator;*

the latter—obtained by invoking local gauge invariance, or by some other means—determine the *vertex factors:*

$$\text{Free Lagrangian} \Rightarrow \text{propagator}$$
$$\text{Interaction terms} \Rightarrow \text{vertex factors}$$

Let us consider the propagators first.

Application of the Euler–Lagrange equation to the free Lagrangian yields the free field equations:

$$\left[\partial^\mu \partial_\mu + \left(\frac{mc}{\hbar}\right)^2\right]\phi = 0 \qquad \text{(Klein–Gordon, for spin 0)} \qquad (11.13)$$

$$\left[i\gamma^\mu \partial_\mu - \left(\frac{mc}{\hbar}\right)\right]\psi = 0 \qquad \text{(Dirac, for spin } \tfrac{1}{2}\text{)} \qquad (11.15)$$

$$\left[\partial_\mu(\partial^\mu A^\nu - \partial^\nu A^\mu) + \left(\frac{mc}{\hbar}\right)^2 A^\nu\right] = 0 \qquad \text{(Proca, for spin 1)} \qquad (11.22)$$

The corresponding "momentum-space" equations are obtained by the standard prescription [eq. (7.5)] $p_\mu \leftrightarrow i\hbar\,\partial_\mu$:

$$[p^2 - (mc)^2]\phi = 0 \qquad (11.90)$$

$$[\not{p} - (mc)]\psi = 0 \qquad (11.91)$$

$$[(-p^2 + (mc)^2)g_{\mu\nu} + p_\mu p_\nu]A^\nu = 0 \qquad (11.92)$$

The propagator is simply (i times) the *inverse* of the factor in square brackets:

$$\text{Spin-0 propagator: } \frac{i}{p^2 - (mc)^2} \qquad (11.93)$$

$$\text{Spin-}\tfrac{1}{2}\text{ propagator: } \frac{i}{\not{p} - mc} = i\frac{(\not{p} + mc)}{p^2 - (mc)^2} \qquad (11.94)$$

$$\text{Spin-1 propagator: } \frac{-i}{p^2 - (mc)^2}\left[g_{\mu\nu} - \frac{p_\mu p_\nu}{(mc)^2}\right] \qquad (11.95)$$

Note that in the second case this factor is a 4×4 *matrix,* and we want the *matrix* inverse; in the third case the factor is a second-rank tensor ($T_{\mu\nu}$), and we want the *tensor* inverse $(T^{-1})_{\mu\nu}$, such that $T_{\mu\lambda}(T^{-1})^{\lambda\nu} = \delta_\mu^\nu$. (See Problem 11.19.) These are precisely the propagators we used in Chapters 6, 7, and 10.* Since we obviously cannot set $m \to 0$ in the Proca propagator (11.95), we must go back to the free field equation (11.22) to work out the photon propagator:

$$\partial_\mu(\partial^\mu A^\nu - \partial^\nu A^\mu) = 0 \qquad \text{(Maxwell, for massless spin 1)} \qquad (11.96)$$

* Actually, this procedure only determines the propagator up to a multiplicative constant, since the field equations, (11.90), (11.91), and (11.92), can always be multiplied by such a factor. In the "canonical" form of these equations the coefficient of mc or $(mc)^2$ is taken to be ± 1, with the sign matching that of the mass term in $\mathcal{L}$. Other conventions lead to a slightly different set of Feynman rules, but do not, of course, change the calculated reaction amplitudes.

As I have remarked before, this equation does not uniquely determine A^μ; if we impose the Lorentz condition

$$\partial_\mu A^\mu = 0 \tag{7.82}$$

then (11.96) reduces to

$$\partial^2 A^\nu = 0 \tag{11.97}$$

which, in momentum space, can be written

$$(-p^2 g_{\mu\nu})\tilde{A}^\nu = 0 \tag{11.98}$$

So the photon propagator is

$$\text{Massless spin-1 propagator: } -i\frac{g_{\mu\nu}}{p^2} \tag{11.99}$$

To get the vertex factors, first write down $i\mathcal{L}_{\text{int}}$ in momentum space ($i\hbar\,\partial_\mu \to p_\mu$), and examine the *fields* involved; these determine the qualitative structure of the interaction. For example, in the case of the QED Lagrangian (11.35)

$$i\mathcal{L}_{\text{int}} = -i(q\bar{\psi}\gamma^\mu\psi)A_\mu \tag{11.100}$$

there are *three* fields involved ($\bar{\psi}$, ψ, and A_μ), and this defines a vertex in which three lines are joined—an incoming fermion, an outgoing fermion, and a photon. To obtain the vertex factor itself, *simply rub out the field variables:*

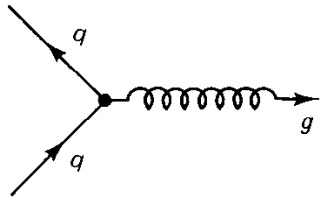
$$-i\sqrt{\frac{4\pi}{\hbar c}}\,q\gamma^\mu = ig_e\gamma^\mu$$

(QED vertex factor for negatively charged particle) (11.101)

(In the case of the photon, what we actually rub out is $\sqrt{\hbar c/4\pi}A^\mu$; the extra factor is due to our use of cgs units which are, for this purpose, a little cumbersome.) The same goes for chromodynamics (11.88): The quark-gluon coupling

$$\mathcal{L}_{\text{int}} = -(q\bar{\psi}\gamma^\mu\boldsymbol{\lambda}\psi)\cdot\mathbf{A}_\mu \tag{11.102}$$

yields a vertex of the form



with the vertex factor

$$-i\frac{g_s}{2}\gamma^\mu\boldsymbol{\lambda} \tag{11.103}$$

(The strong coupling constant is traditionally defined with a factor of 2: $g_s \equiv 2\sqrt{4\pi/\hbar c}q$, where q is the "strong charge" appearing in the Lagrangian). However, there are also direct gluon-gluon couplings, coming from the $\mathbf{F}^{\mu\nu}\cdot\mathbf{F}_{\mu\nu}$ term in $\mathcal{L}$,

since $\mathbf{F}^{\mu\nu}$ contains not only the "free" part, $\partial^\mu \mathbf{A}^\nu - \partial^\nu \mathbf{A}^\mu$, but also an interaction term $-2q/\hbar c(\mathbf{A}^\mu \times \mathbf{A}^\nu)$ [eq. (11.87)]. Squaring it out, we find:

$$\mathcal{L}_{\text{int}} = \left(\frac{q}{8\pi\hbar c}\right)[(\partial^\mu \mathbf{A}^\nu - \partial^\nu \mathbf{A}^\mu)\cdot(\mathbf{A}_\mu \times \mathbf{A}_\nu) + (\mathbf{A}^\mu \times \mathbf{A}^\nu)\cdot(\partial_\mu \mathbf{A}_\nu - \partial_\nu \mathbf{A}_\mu)]$$
$$- \frac{q^2}{4\pi(\hbar c)^2}(\mathbf{A}^\mu \times \mathbf{A}^\nu)\cdot(\mathbf{A}_\mu \times \mathbf{A}_\nu) \tag{11.104}$$

The first term carries three factors of $\mathbf{A}^\mu$, and leads to the three-gluon vertex (9.18); the second term carries *four* factors of $\mathbf{A}^\mu$, and gives the four-gluon vertex (9.19). (For some practice in extracting Feynman rules from Lagrangians, see Problems 11.20 and 11.21.)

11.7 THE MASS TERM

The principle of local gauge invariance works beautifully for the strong and electromagnetic interactions. In the first place, it gives us a *machine* for determining the couplings (in the "old days" the construction of $\mathcal{L}_{\text{int}}$ was a purely ad hoc guess). Moreover, as 't Hooft and others proved in the early seventies,[9] gauge theories are automatically renormalizable. But the application to *weak* interactions was stymied by the fact that gauge fields have to be massless. Remember, the mass term in the Proca Lagrangian is not locally gauge invariant, and whereas the photon and the gluons *are* massless, the W's and the Z^0 certainly are *not*. So the question arises, Can we doctor up gauge theory in such a way as to accommodate massive gauge fields? The answer is *yes*, but the procedure—exploiting spontaneous symmetry-breaking and the Higgs mechanism—is diabolically subtle, and it pays to begin by thinking very carefully about how one identifies the mass term in a Lagrangian.

Suppose, for instance, you were given the following Lagrangian for a scalar field ϕ:

$$\mathcal{L} = \tfrac{1}{2}(\partial_\mu \phi)(\partial^\mu \phi) + e^{-(\alpha\phi)^2} \tag{11.105}$$

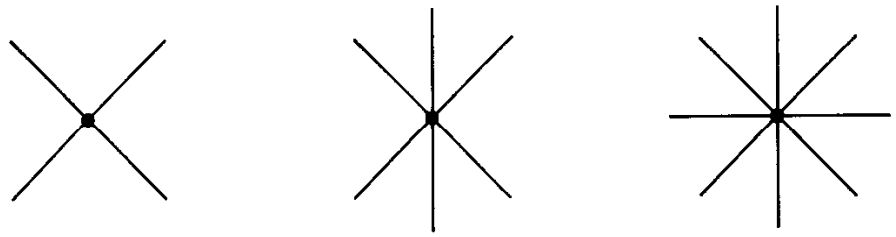
where α is some (real) constant. Where is the mass term here? At first glance there's no sign of one, and you might conclude that this is a *massless* field. But that is incorrect, for if you expand the exponential, $\mathcal{L}$ takes the form

$$\mathcal{L} = \tfrac{1}{2}(\partial_\mu \phi)(\partial^\mu \phi) + 1 - \alpha^2\phi^2 + \tfrac{1}{2}\alpha^4\phi^4 - \tfrac{1}{6}\alpha^6\phi^6 + \cdots \tag{11.106}$$

The 1 is irrelevant (a constant term in $\mathcal{L}$ has no affect on the field equations), but the second term looks just like the mass term in the Klein–Gordon Lagrangian (11.11), with $\alpha^2 = \frac{1}{2}(mc/\hbar)^2$. Evidently this Lagrangian describes a particle of mass

$$m = \sqrt{2}\alpha\hbar/c \tag{11.107}$$

The higher-order terms represent couplings, of the form



and so on. This is not supposed to be a realistic theory, of course—I offer it only as an example of how the mass term in a Lagrangian may be "disguised." To identify it, we expand $\mathcal{L}$ in powers of ϕ and pick out the term proportional to ϕ^2 (in general, it's the term of *second order* in the fields—ϕ, ψ, A^μ, or whatever).

But there is a deeper subtlety lurking here, which I illustrate with the following Lagrangian:

$$\mathcal{L} = \tfrac{1}{2}(\partial_\mu\phi)(\partial^\mu\phi) + \tfrac{1}{2}\mu^2\phi^2 - \tfrac{1}{4}\lambda^2\phi^4 \tag{11.108}$$

Here μ and λ are (real) constants. The second term looks like a mass, and the third like an interaction. But wait! The *sign* is wrong [compare equation (11.11)]—if that's a *mass* term, then m is *imaginary,* which is nonsense. How, then, *should* we interpret this Lagrangian? To answer this question, we must understand that the Feynman calculus is really a perturbation procedure, in which we start from the ground state (the "vacuum"), and treat the fields as fluctuations about that state. For the Lagrangians we have considered so far, the ground state—the field configuration of minimum energy—has always been the trivial one: $\phi = 0$. But for the Lagrangian (11.108), $\phi = 0$ is *not* the ground state. To determine the *true* ground state, we write $\mathcal{L}$ as a "kinetic" term ($\frac{1}{2}\,\partial_\mu\phi\,\partial^\mu\phi$) minus a "potential" term [inspired by the classical Lagrangian (11.4)]:

$$\mathcal{L} = \mathcal{T} - \mathcal{U} \tag{11.109}$$

and look for the minimum of $\mathcal{U}$. In the present case

$$\mathcal{U}(\phi) = -\tfrac{1}{2}\mu^2\phi^2 + \tfrac{1}{4}\lambda^2\phi^4 \tag{11.110}$$

and the minimum occurs at

$$\phi = \pm\mu/\lambda \tag{11.111}$$

(see Fig. 11.1). For this Lagrangian, the Feynman calculus must be formulated in terms of deviations from one or the other of these ground states. This suggests that we introduce a *new* field variable, η, defined by

$$\eta \equiv \phi \pm \frac{\mu}{\lambda} \tag{11.112}$$

In terms of η, the Lagrangian reads

$$\mathcal{L} = \tfrac{1}{2}(\partial_\mu\eta)(\partial^\mu\eta) - \mu^2\eta^2 \pm \mu\lambda\eta^3 - \tfrac{1}{4}\lambda^2\eta^4 + \tfrac{1}{4}(\mu^2/\lambda)^2 \tag{11.113}$$

The second quantity is now a mass term with the *correct* sign, and we discover [comparing eq. (11.11)] that the mass of the particle is

$$m = \sqrt{2}\mu\hbar/c \tag{11.114}$$

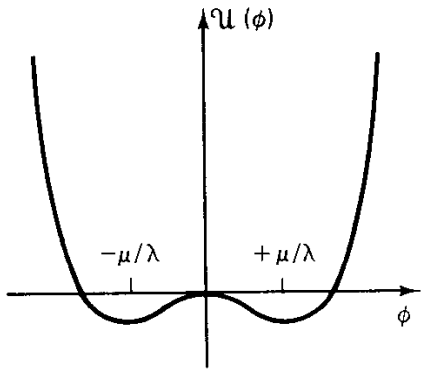
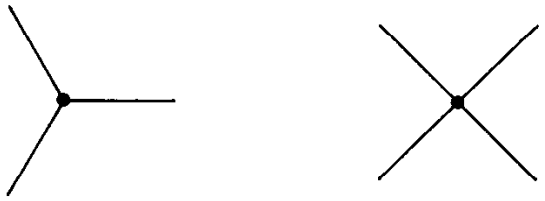


Figure 11.1 Graph of $\mathcal{U}(\phi)$, equation (11.110).

Meanwhile, the third and fourth terms represent couplings of the form



(the last term is a constant, signifying nothing).

I emphasize that Lagrangians (11.108) and (11.113) represent exactly the same physical system; all we have done is to change the *notation* (11.112). But the first version is not suited to the Feynman calculus (technically, a perturbation series in ϕ would not converge, because it is an expansion about an unstable point); only in the second formulation can we read off the mass and the vertex factors.

Conclusion: To identify the mass term in a Lagrangian, we first locate the ground state [the field configuration for which $\mathcal{U}(\phi)$ is a minimum] and reexpress $\mathcal{L}$ as a function of the deviation, η, from this minimum. Expanding in powers of η, we obtain the mass from the coefficient of the η^2 term.

11.8 SPONTANEOUS SYMMETRY-BREAKING

The example we have just considered illustrates another phenomenon of importance: *spontaneous symmetry-breaking.* The original Lagrangian (11.108) is *even* in ϕ: It is invariant as $\phi \rightarrow -\phi$. But the reformulated Lagrangian (11.113) is *not* even in η; the symmetry has been "broken." How did this happen? It happened because the "vacuum" (whichever of the two ground states we care to work with) does not share the symmetry of the Lagrangian. (The collection of *all* ground states, of course, *does,* but to set up the Feynman formalism we are obliged to work with one or the other of them, and that spoils the symmetry.) We call this "spontaneous" symmetry-breaking because no *external* agency is responsible (as, for example, gravity breaks the three-dimensional symmetry in this room, making "up" and "down" quite different from "left" and "right"). To put it the other way around, the true symmetry of the system is "hidden" by the arbitrary selection of a particular (asymmetrical) ground state. There are examples of spontaneous symmetry-breaking in many branches of physics. Take,

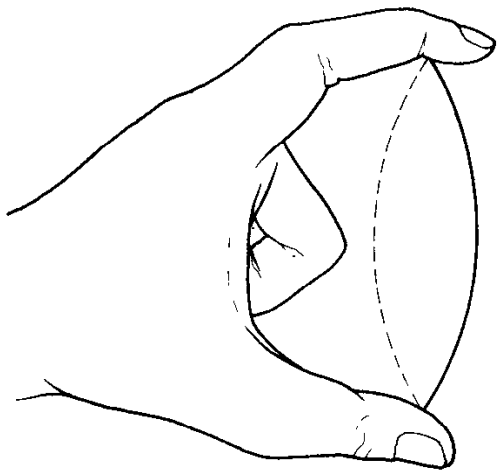


Figure 11.2 Spontaneous symmetry-breaking in a plastic strip.

for instance, a thin plastic strip (say, a short ruler): If you squeeze the ends together, it will snap into a curved configuration, but it can just as well buckle to the left as to the right—both are ground states for the system, and either one breaks the left–right symmetry (see Fig. 11.2).

But the spontaneously broken symmetry we have just considered was a *discrete* symmetry, with just *two* ground states. More interesting things happen when we consider *continuous* symmetries. (Replace the plastic *strip* in Figure 11.2 with a plastic *rod*—say, a knitting needle. Then it can buckle in *any* direction, not just left or right.*) It is easy to construct a Lagrangian with spontaneously broken continuous symmetry. For example

$$\mathcal{L} = \tfrac{1}{2}(\partial_\mu\phi_1)(\partial^\mu\phi_1) + \tfrac{1}{2}(\partial_\mu\phi_2)(\partial^\mu\phi_2) + \tfrac{1}{2}\mu^2(\phi_1^2 + \phi_2^2) - \tfrac{1}{4}\lambda^2(\phi_1^2 + \phi_2^2)^2 \qquad (11.115)$$

This is identical to equation (11.108), except that now there are *two* fields, ϕ_1 and ϕ_2, and because $\mathcal{L}$ involves only the sum of the squares, it is invariant under *rotations* in ϕ_1, ϕ_2 space.†

This time the "potential energy" function is

$$\mathcal{U} = -\tfrac{1}{2}\mu^2(\phi_1^2 + \phi_2^2) + \tfrac{1}{4}\lambda^2(\phi_1^2 + \phi_2^2)^2 \qquad (11.116)$$

and the minima lie on a circle of radius μ/λ:

$$\phi_{1\min}^2 + \phi_{2\min}^2 = \mu^2/\lambda^2 \qquad (11.117)$$

(see Fig. 11.3). To apply the Feynman calculus, we have to expand about a *particular* ground state ("the vacuum")—we may as well pick

$$\phi_{1\min} = \mu/\lambda; \qquad \phi_{2\min} = 0 \qquad (11.118)$$

As before, we introduce new fields, η and ξ, which are the fluctuations about this vacuum state:

$$\eta \equiv \phi_1 - \mu/\lambda; \qquad \xi \equiv \phi_2 \qquad (11.119)$$

* A more sophisticated example is the ferromagnet: In the ground state all the electron spins are aligned, but the *direction* of alignment is an accident of history. The *theory* is symmetrical, but a given piece of iron has to pick a particular orientation, and that breaks the symmetry.

† Group theoretically, it is invariant under $SO(2)$: $\phi_1 \to \phi_1 \cos\theta + \phi_2 \sin\theta$; $\phi_2 \to -\phi_1 \sin\theta + \phi_2 \cos\theta$, for any "rotation angle" θ. (See Problem 4.6.)

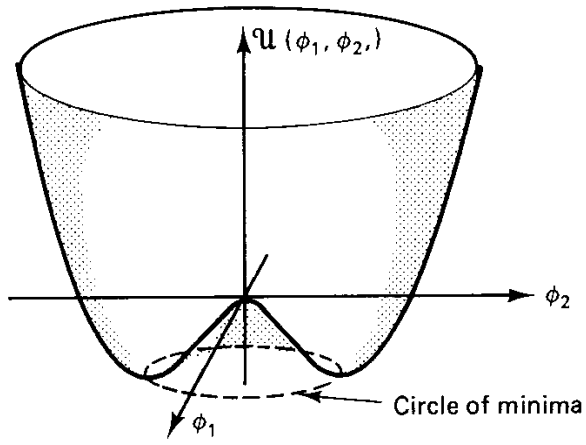


Figure 11.3 The potential function (11.116).

Rewriting the Lagrangian in terms of these new field variables, we find (Problem 11.22):

$$\mathcal{L} = \left[\frac{1}{2}(\partial_\mu\eta)(\partial^\mu\eta) - \mu^2\eta^2\right] + \left[\frac{1}{2}(\partial_\mu\xi)(\partial^\mu\xi)\right] + \left[\mu\lambda(\eta^3 + \eta\xi^2) - \frac{\lambda^2}{4}(\eta^4 + \xi^4 + 2\eta^2\xi^2)\right] + \mu^4/(4\lambda^2) \qquad (11.120)$$

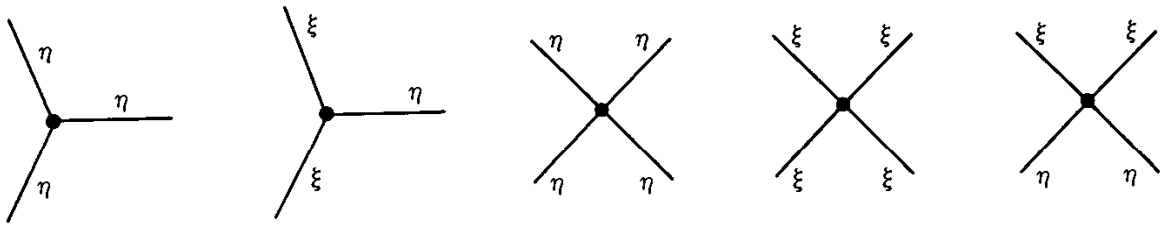
The first term is a free Klein–Gordon Lagrangian (11.11) for the field η, which evidently carries a mass

$$m_\eta = \sqrt{2}\mu\hbar/c \qquad (11.121)$$

[the same as before, see eq. (11.114)]; the second term is a free Lagrangian for the field ξ, which is *massless:*

$$m_\xi = 0 \qquad (11.122)$$

and the third term defines five couplings:



(the final constant, of course, is irrelevant). In this form the Lagrangian doesn't look symmetrical at all; the symmetry of (11.115) has been broken (or rather, "hidden") by the selection of a particular vacuum state.

The important thing to notice here is that one of the fields (ξ) is *automatically massless.* This is no accident. It can be shown (Goldstone's theorem[10]) that spontaneous breaking of a *continuous* global symmetry is *always* accompanied by the appearance of one or more massless scalar (spin-0) particles (we call them "Goldstone bosons").* Well, this is a disaster; we were hoping to use

* Intuitively, this is related to the fact that there is no resistance to excitations in the ξ direction. Flick the bent knitting needle and it will spin freely about the axis, whereas *radial* excitations encounter a restoring force, and the system oscillates.

the mechanism of spontaneous symmetry-breaking to account for the mass of the weak interaction gauge fields, but now we find that this introduces a massless scalar boson, and there is no such thing on the roster of known elementary particles.* But hold on, for there is one final incredible twist in the story. It comes when we apply the idea of spontaneous symmetry-breaking to the case of *local* gauge invariance.

11.9 THE HIGGS MECHANISM

The Lagrangian we studied in Section 11.8 can be written more neatly if we combine the two real fields, ϕ_1 and ϕ_2, into a single complex field:

$$\phi \equiv \phi_1 + i\phi_2 \tag{11.123}$$

so that

$$\phi^*\phi = \phi_1^2 + \phi_2^2 \tag{11.124}$$

In this notation (and it *is* nothing but notation) the Lagrangian (11.115) reads

$$\mathcal{L} = \tfrac{1}{2}(\partial_\mu\phi)^*(\partial^\mu\phi) + \tfrac{1}{2}\mu^2(\phi^*\phi) - \tfrac{1}{4}\lambda^2(\phi^*\phi)^2 \tag{11.125}$$

and the rotational ($SO(2)$) symmetry that was spontaneously broken becomes invariance under ($U(1)$) phase transformations:

$$\phi \longrightarrow e^{i\theta}\phi \tag{11.126}$$

This is precisely the kind of symmetry we considered back in Section 11.3, except that now we are working with scalar fields instead of with spinors. We can make the system invariant under *local* gauge transformations

$$\phi \longrightarrow e^{i\theta(x)}\phi \tag{11.127}$$

by the usual device of introducing a massless gauge field A^μ, and replacing the derivatives in equation (11.125) with *covariant* derivatives (11.38):

$$\mathcal{D}_\mu = \partial_\mu + i\frac{q}{\hbar c}A_\mu \tag{11.128}$$

Thus

$$\mathcal{L} = \frac{1}{2}\left[\left(\partial_\mu - \frac{iq}{\hbar c}A_\mu\right)\phi^*\right]\left[\left(\partial^\mu + \frac{iq}{\hbar c}A^\mu\right)\phi\right] + \frac{1}{2}\mu^2(\phi^*\phi) - \frac{1}{4}\lambda^2(\phi^*\phi)^2 - \frac{1}{16\pi}F^{\mu\nu}F_{\mu\nu} \tag{11.129}$$

Now we simply retrace our steps in Section 11.8, applying them to the *locally* invariant Lagrangian (11.129). Defining the new fields

$$\eta \equiv \phi_1 - \mu/\lambda, \qquad \xi \equiv \phi_2 \tag{11.130}$$

* It is hard to imagine that such a particle could have escaped detection. With *heavy* particles this is always a possibility—maybe you just didn't have enough energy to *produce* it—but a *massless* particle would surely have shown up somewhere, if only in the form of "missing" energy and momentum.

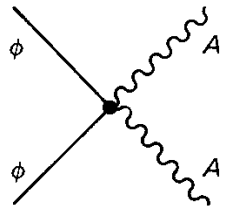
[compare eq. (11.119)], the Lagrangian becomes (see Problem 11.25):

$$\mathcal{L} = \left[\frac{1}{2}(\partial_\mu\eta)(\partial^\mu\eta) - \mu^2\eta^2\right] + \left[\frac{1}{2}(\partial_\mu\xi)(\partial^\mu\xi)\right]$$
$$+ \left[-\frac{1}{16\pi}F^{\mu\nu}F_{\mu\nu} + \frac{1}{2}\left(\frac{q}{\hbar c}\frac{\mu}{\lambda}\right)^2 A_\mu A^\mu\right] - 2i\left(\frac{\mu}{\lambda}\frac{q}{\hbar c}\right)(\partial_\mu\xi)A^\mu$$
$$+ \left\{\frac{q}{\hbar c}[\eta(\partial_\mu\xi) - \xi(\partial_\mu\eta)]A^\mu + \frac{\mu}{\lambda}\left(\frac{q}{\hbar c}\right)^2\eta(A_\mu A^\mu) + \frac{1}{2}\left(\frac{q}{\hbar c}\right)^2(\xi^2 + \eta^2)(A_\mu A^\mu)\right.$$
$$\left. - \lambda\mu(\eta^3 + \eta\xi^2) - \frac{1}{4}\lambda^2(\eta^4 + 2\eta^2\xi^2 + \xi^4)\right\} + \left(\frac{\mu^2}{2\lambda}\right)^2 \qquad (11.131)$$

The first line is the same as before, equation (11.120); it describes a scalar particle (η) of mass $\sqrt{2}\mu\hbar/c$ and a massless Goldstone boson (ξ). The second line describes the free gauge field A^μ, but—*mirabile dictu!*—it has acquired a mass:

$$m_A = 2\sqrt{\pi}\left(\frac{q\mu}{\lambda c^2}\right) \qquad (11.132)$$

[compare the Proca Lagrangian (11.21)]. The term in curly brackets specifies various couplings of ξ, η, and A^μ (see Problem 11.26). It is easy to see where the mass of A^μ came from: The original Lagrangian (11.129) contains a term of the form $\phi^*\phi A_\mu A^\mu$, which—absent spontaneous symmetry-breaking—would represent a *coupling:*



But when the ground state moves "off center," and the field μ/λ picks up a constant [(eq. (11.130)], this piece of the Lagrangian takes the form of the Proca mass term.

However, we still have that unwanted Goldstone boson (ξ). Moreover, there is a suspicious-looking quantity in $\mathcal{L}$:

$$-2i\left(\frac{\mu}{\lambda}\frac{q}{\hbar c}\right)(\partial_\mu\xi)A^\mu \qquad (11.133)$$

What are we to make of *this?* If we read it as an *interaction,* it leads to a vertex of the form



in which the ξ turns into an A. Any such term, bilinear in two different fields, indicates that we have incorrectly identified the fundamental particles in the theory (see Problem 11.23). Both difficulties involve the field $\xi = \phi_2$, and both

can be resolved exploiting the local gauge invariance of $\mathcal{L}$ [in the original form (11.129)] to *transform this field away entirely!* Writing equation (11.126) in terms of its real and imaginary parts:

$$\begin{aligned}\phi \longrightarrow \phi' &= (\cos\theta + i\sin\theta)(\phi_1 + i\phi_2) \\ &= (\phi_1\cos\theta - \phi_2\sin\theta) + i(\phi_1\sin\theta + \phi_2\cos\theta)\end{aligned} \tag{11.134}$$

we see that picking

$$\theta = -\tan^{-1}(\phi_2/\phi_1) \tag{11.135}$$

will render ϕ' real, which is to say that $\phi'_2 = 0$. The gauge field A^μ will transform accordingly (11.34), but the Lagrangian will take the same form in terms of the new field variables as it did in terms of the old ones (that's what it *means* to say that $\mathcal{L}$ is invariant). The only difference is that ξ is now zero. *In this particular gauge,* then, the Lagrangian (11.131) reduces to

$$\begin{aligned}\mathcal{L} = &\left[\frac{1}{2}(\partial_\mu\eta)(\partial^\mu\eta) - \mu^2\eta^2\right] + \left[-\frac{1}{16\pi}F^{\mu\nu}F_{\mu\nu} + \frac{1}{2}\left(\frac{q}{\hbar c}\frac{\mu}{\lambda}\right)^2 A_\mu A^\mu\right] \\ &+ \left\{\frac{\mu}{\lambda}\left(\frac{q}{\hbar c}\right)^2\eta(A_\mu A^\mu) + \frac{1}{2}\left(\frac{q}{\hbar c}\right)^2\eta^2(A_\mu A^\mu) - \lambda\mu\eta^3 - \frac{1}{4}\lambda^2\eta^4\right\} \\ &+ \left(\frac{\mu^2}{2\lambda}\right)^2\end{aligned} \tag{11.136}$$

By an astute choice of gauge, we have eliminated the Goldstone boson and the offending term in $\mathcal{L}$; we are left with a single *massive* scalar η (the "Higgs" particle) and a *massive* gauge field A^μ.

Please understand that Lagrangians (11.129) and (11.136) describe *exactly the same physical system;* all we have done is to select a convenient gauge (11.135) and rewrite the fields in terms of fluctuations about a particular ground state (11.130). We have sacrificed the manifest symmetry of (11.129) in favor of a notation that makes the physical content more transparent, and allows us to extract the Feynman rules more directly. There is an illuminating way to think of this: A *massless* vector field carries *two* degrees of freedom (transverse polarizations); when A^μ acquires mass, it picks up a third degree of freedom (longitudinal polarization). Where did this extra degree of freedom come from? Answer: *it came from the Goldstone boson, which meanwhile disappeared from the theory.* The gauge field "ate" the Goldstone boson, thereby acquiring both a mass and a third polarization state.* This is the famous *Higgs mechanism,* built on the union of gauge invariance and spontaneous symmetry-breaking.

According to the Standard Model, the Higgs mechanism is responsible for the masses of the weak interaction gauge bosons ($W^\pm$ and Z^0). The details are still matters of speculation—the Higgs particle has never been seen in the laboratory (presumably it is just too heavy to make with any existing accelerator),†

* We don't *have* to adopt any particular gauge. However, if we do not, the theory will contain a nonphysical "ghost" particle, and it is simplest to eliminate it explicitly from the start.

† Many particle physicists are presently campaigning for the construction of a Superconducting Supercollider (SSC), whose main purpose would be to search for the Higgs particle.

and the Higgs "potential" [$\mathcal{U}(\phi)$] is completely unknown (I used $\mathcal{U} = -\frac{1}{2}\mu^2(\phi^*\phi) + \frac{1}{4}\lambda^2(\phi^*\phi)^2$ just for the sake of argument). There may in fact be many Higgs particles, or it may be a composite structure, but never mind: The important thing is that we have found a way *in principle* of imparting mass to the gauge fields,* and that is our license to believe that all the fundamental interactions—weak as well as strong and electromagnetic—can be described by local gauge theories.[11]

REFERENCES AND NOTES

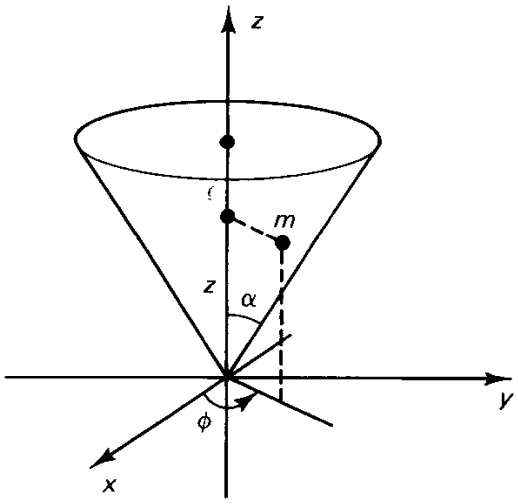
1. See, for example, R. Resnick and D. Halliday, *Physics,* 3d Ed., (New York: Wiley, 1977), Part I, Chap. 8.
2. The Euler–Lagrange equation can be obtained, in turn, from the Principle of Least Action. For further discussion of the Lagrangian method in classical particle mechanics, see J. B. Marion, *Classical Dynamics of Particles and Systems,* 2d Ed. (New York: Academic Press, 1970) or H. Goldstein, *Classical Mechanics,* 2d Ed. (Reading, MA: Addison-Wesley, 1980).
3. H. Weyl, *Ann. Phys.* **59,** 101 (1919). See also K. Moriyasu, *An Elementary Primer for Gauge Theory* (Singapore: World Scientific, 1983).
4. C. N. Yang and R. L. Mills, *Phys. Rev.* **96,** 191 (1954).
5. See, for example, C. Chevalley, *Theory of Lie Groups* (Princeton, NJ: Princeton University Press, 1946).
6. This, and many other aspects of local gauge theories, are treated with extraordinary clarity and beauty in the unpublished research notes "Classical Chromodynamics" and "Bare Bones of the Classical Theory of Gauge Fields," by N. A. Wheeler (Portland, OR: Reed College, 1981).
7. A. Actor, *Rev. Mod. Phys.* **51,** 461 (1979).
8. The details will be found in any treatise on quantum field theory, such as those cited in the Introduction.
9. G. 't Hooft, *Nucl. Phys.* **B33,** 173 (1971), **B35,** 167 (1971). Reprinted in C. H. Lai, ed. (Singapore: World Scientific, 1981).
10. J. Goldstone, *Nuovo Cimento* **19,** 154 (1961); J. Goldstone, A. Salam, and S. Weinberg, *Phys. Rev.* **127,** 965 (1962).
11. For detailed application of the Higgs mechanism in the electroweak theory of Glashow, Weinberg, and Salam, see, for instance, Chapter 15 of F. Halzen and A. D. Martin, *Quarks and Leptons* (New York: Wiley, 1984).

PROBLEMS

11.1. One advantage of the Lagrangian formulation is that it does not commit us to any particular coordinate system—the q's in equation (11.6) could be Cartesian

* In the Standard Model the Higgs particle is *also* responsible for the masses of quarks and leptons; they are initially taken to be massless, but are assumed to have Yukawa couplings (see Problem 11.21) to the Higgs particle. When the latter is "shifted," by spontaneous symmetry-breaking (11.130), the Yukawa coupling splits into two parts, one of which is a true interaction, and the other a *mass* term for the field ψ. This is a nice *idea,* but it does not help us to *calculate* the fermion masses, since the Yukawa coupling constants themselves are unknown. Only when (and *if*) the Higgs particle is actually *found* will it be possible to confirm all this empirically.

coordinates or polar coordinates or any other variables we might use to designate the particle's position. Suppose, for example, we want to analyze the motion of a particle that slides frictionlessly on the inside surface of a cone mounted with its axis pointing upward, as shown.



(a) Express T and U in terms of the variables z and ϕ and the constants α (the opening angle of the cone), m (the mass of the particle), and g (the acceleration of gravity).

(b) Construct the Lagrangian, and apply the Euler–Lagrange equations to obtain differential equations for $z(t)$ and $\phi(t)$.

(c) Show that $L = (m \tan^2 \alpha) z^2 \dot{\phi}$ is a constant of the motion. What *is* this quantity, physically?

(d) Use the result in (c) to eliminate ϕ from the z equation. (You are left with a *second*-order differential equation for $z(t)$; if you want to pursue the problem further, it is easiest to invoke conservation of *energy*, which yields a *first*-order equation for z.)

11.2. Derive equation (11.17).

11.3. Starting with equation (11.19), show that $\partial_\mu A^\mu = 0$, and hence that each component of A^μ satisfies the Klein–Gordon equation: $\Box A^\nu + (mc/\hbar)^2 A^\nu = 0$.

11.4. As it stands, the Dirac Lagrangian (11.14) treats ψ and $\bar{\psi}$ asymmetrically. Some people prefer to deal with them on an equal footing, using the modified Lagrangian

$$\mathcal{L} = \frac{i\hbar c}{2} [\bar{\psi}\gamma^\mu(\partial_\mu \psi) - (\partial_\mu \bar{\psi})\gamma^\mu \psi] - (mc^2)\bar{\psi}\psi$$

Apply the Euler–Lagrange equations to this $\mathcal{L}$, and show that you get the Dirac equation (11.15) and its adjoint.

11.5. The Klein–Gordon Lagrangian for a *complex* field would be

$$\mathcal{L} = \tfrac{1}{2}(\partial_\mu \phi)^*(\partial^\mu \phi) - \tfrac{1}{2}(mc/\hbar)^2 \phi^* \phi$$

Treating ϕ and ϕ^* as independent field variables, deduce the field equations for each, and show that these field equations are consistent (i.e., one is the complex conjugate of the other).

11.6. Apply the Euler–Lagrange equations to (11.33) to find the Dirac equation with electromagnetic coupling.

11.7. Show that the Dirac current (11.36) satisfies the continuity equation (11.25).

11.8. The complex Klein–Gordon Lagrangian (Problem 11.5) is invariant under the global gauge transformation $\phi \rightarrow e^{i\theta}\phi$. Impose local gauge invariance to construct the complete gauge-invariant Lagrangian, and determine the current density J^μ. Check that this current obeys the continuity equation (11.25).

11.9. **(a)** Suppose the field variables (ϕ_i) are subjected to an infinitesimal global transformation $\delta\phi_i$. Show that the Lagrangian $\mathcal{L}(\phi_i, \partial_\mu\phi_i)$ changes by an amount

$$\delta\mathcal{L} = \partial_\mu\left\{\frac{\partial\mathcal{L}}{\partial(\partial_\mu\phi_i)}\,\delta\phi_i\right\}$$

In particular, if the Lagrangian is *invariant* under the transformation in question, then $\delta\mathcal{L} = 0$, and the term in curly brackets constitutes a conserved current (that is, it obeys the continuity equation). This is the essence of Noether's theorem, relating symmetries of the Lagrangian to conservation laws.

(b) Apply Noether's theorem to the Dirac Lagrangian (11.14), to construct the conserved current associated with global phase invariance [equation (11.26)]. Compare the electric current (11.36).

(c) Do the same for the complex Klein–Gordon Lagrangian in Problem 11.8.

11.10. Derive equation (11.51).

11.11. Deduce equation (11.58) from equation (11.57), using (11.54), (11.55), and (11.56).

11.12. Suppose we were to define

$$\mathbf{F}^{\mu\nu} \equiv \partial^\mu\mathbf{A}^\nu - \partial^\nu\mathbf{A}^\mu$$

in Yang–Mills theory.

(a) Find the transformation rule for this $\mathbf{F}^{\mu\nu}$, under infinitesimal gauge transformations (11.61).

(b) Determine the infinitesimal transformation rule for $\mathcal{L}_A$ (11.63), in this case. Is the Lagrangian invariant?

[*Answers:* **(a)** $\mathbf{F}^{\mu\nu} \rightarrow \mathbf{F}^{\mu\nu} + \dfrac{2g}{\hbar c}[\boldsymbol{\lambda} \times \mathbf{F}^{\mu\nu} + \mathbf{A}^\mu \times \partial^\nu\boldsymbol{\lambda} - \mathbf{A}^\nu \times \partial^\mu\boldsymbol{\lambda}$

(b) $\mathbf{F}^{\mu\nu} \cdot \mathbf{F}_{\mu\nu} \rightarrow \mathbf{F}^{\mu\nu} \cdot \mathbf{F}_{\mu\nu} + \dfrac{8g}{\hbar c}(\mathbf{A}_\nu \times \mathbf{F}^{\mu\nu}) \cdot \partial_\mu\boldsymbol{\lambda}$]

11.13. Derive equation (11.66), starting with (11.61) and (11.65).

11.14. Prove that gauge field Lagrangian (11.63) is invariant under finite local gauge transformations, as follows:

(a) Using expressions (11.58) and (11.65), show that

$$\boldsymbol{\tau} \cdot \mathbf{F}^{\mu\nu\prime} = S(\boldsymbol{\tau} \cdot \mathbf{F}^{\mu\nu})S^{-1}$$

[Note that $\partial_\mu(S^{-1}S) = 0 \Rightarrow (\partial_\mu S^{-1})S = -S^{-1}(\partial_\mu S)$.]

(b) Show, therefore, that

$$Tr[(\boldsymbol{\tau} \cdot \mathbf{F}^{\mu\nu})(\boldsymbol{\tau} \cdot \mathbf{F}_{\mu\nu})]$$

is invariant.

(c) Using Problem 4.20(c), show that the trace in (b) is equal to $2\mathbf{F}^{\mu\nu} \cdot \mathbf{F}_{\mu\nu}$.

11.15. Apply the Euler–Lagrange equations to Lagrangian (11.69). Using the standard associations (7.71), (7.72), and (7.79), obtain "Maxwell's equations" for classical Yang–Mills theory. [Note that there are *three* charge densities, three current den-

sities, three scalar potentials, three vector potentials, three "electric" fields, and three "magnetic" fields, in this theory.] (Unlike electrodynamics, your expressions for the divergence and curl of the **E**'s and **B**'s will inevitably involve the *potentials*.)

11.16. Show that any Hermitian 3×3 matrix can be written as a linear combination of the unit matrix and the eight Gell-Mann matrices [eq. (11.76)].

11.17. **(a)** Show that $\det(e^A) = e^{\mathrm{Tr}(A)}$, for any matrix A. [*Hint:* Check it first for a *diagonal* matrix. Then extend the proof to any *diagonalizable* matrix ($S^{-1}AS = D$, where D is diagonal, for some matrix S)—show that $\mathrm{Tr}\,(A) = \mathrm{Tr}\,(D)$ and $S^{-1}e^AS = e^D$, so that $\det(e^A) = \det(e^D)$. Of course, not all matrices *are* diagonalizable; however, every matrix can be brought into *Jordan canonical form* ($S^{-1}AS = J$, where J is diagonal except for some 1's immediately below the main diagonal). Take it from there.]

(b) Show that $e^{i\boldsymbol{\lambda}\cdot\mathbf{a}}$ [in equation (11.78)] has determinant 1.

11.18. Starting with equation (11.81), derive (11.82) and (11.83).

11.19. Confirm that the Proca propagator (11.95) is the inverse of the tensor in equation (11.92), in the sense explained in the text.

11.20. Construct the Lagrangian for *ABC* theory (Chap. 6).

11.21. Give a physical interpretation of the Yukawa Lagrangian:

$$\mathcal{L} = [i\hbar c\bar{\psi}\gamma^\mu \partial_\mu\psi - m_1c^2\bar{\psi}\psi] + \left[\tfrac{1}{2}(\partial_\mu\phi)(\partial^\mu\phi) - \tfrac{1}{2}\left(\frac{m_2c}{\hbar}\right)^2\phi^2\right] + \alpha_Y\bar{\psi}\psi\phi.$$

What are the spins and masses of the particles? What are their propagators? Draw the Feynman diagram for their interaction, and determine the vertex factor.

11.22. Derive equation (11.120).

11.23. Suppose we took

$$\psi_1 \equiv (\eta + \xi)/\sqrt{2}, \qquad \psi_2 \equiv (\eta - \xi)/\sqrt{2}$$

as the fundamental fields, instead of definition (11.119). Express the Lagrangian (11.120) in terms of ψ_1 and ψ_2. [*Comment:* Offhand, it looks as though we have two *massive* fields here, and thus escape Goldstone's theorem. Unfortunately, there is *also* a term of the form $-\mu^2\psi_1\psi_2$. If you interpret this as an *interaction*, it converts ψ_1 into ψ_2, and vice versa, but that means neither one exists as an independent free particle. Rather, such an expression should be interpreted as an off-diagonal term in the mass matrix (11.45), indicating that we have incorrectly identified the fundamental fields in the theory. The *physical* fields are those for which M is *diagonal*, and for which no direct transitions from one to the other can occur. We have encountered this situation once before, in Section 4.8 of Chapter 4: We found that $K^0 \leftrightarrow \bar{K}^0$, and hence that these are not the physical particle states; instead, the linear combinations K_1 and K_2, in terms of which the mass matrix is diagonal, are the "true" particles.]

11.24. Generalize the argument following equation (11.115) to *three* fields (ϕ_1, ϕ_2, ϕ_3). What are the masses of the three particles? How many Goldstone bosons are there in this case?

11.25. Starting from expressions (11.129) and (11.130), derive equation (11.131).

11.26. Draw the primitive vertices for all the interactions in curly brackets in equation (11.131). Circle the ones that survive in equation (11.136).

Appendix A

The Dirac Delta Function

The Dirac delta function, $\delta(x)$, is an infinitely high, infinitesimally narrow spike at the origin, with area 1 (Fig. A.1). Specifically

$$\delta(x) = \begin{cases} 0, & \text{if } x \neq 0 \\ \infty, & \text{if } x = 0 \end{cases} \quad \text{and} \quad \int_{-\infty}^{\infty} \delta(x)dx = 1 \tag{A.1}$$

Technically, it's not a function at all, since its value is not finite at $x = 0$. In the mathematical literature it is known as a *generalized* function, or *distribution.* It is, if you like, the *limit* of a *sequence* of functions, such as rectangles of height n and width $1/n$, or isosceles triangles of height n and base $2/n$ (Fig. A.2), or any other shape you might wish to use.

If $f(x)$ is some "ordinary" function (that is, *not* another delta function—in fact, just to be on the safe side let's say that $f(x)$ is *continuous*—then the *product* $f(x)\delta(x)$ is zero everywhere except at $x = 0$. It follows that

$$f(x)\delta(x) = f(0)\delta(x) \tag{A.2}$$

(This is the most important fact about the delta function, so make sure you understand why it is true. The point is that since the product is zero anyway *except* at $x = 0$, we may as well replace $f(x)$ by the value it assumes at the origin.) In particular

$$\int_{-\infty}^{\infty} f(x)\delta(x)dx = f(0)\int_{-\infty}^{\infty} \delta(x)dx = f(0) \tag{A.3}$$

Under an integral, the delta function "picks out" the value of $f(x)$ at $x = 0$. (Here and below, the integral need not run from $-\infty$ to $+\infty$; it is sufficient that the domain extend across the delta function, and $-\epsilon$ to $+\epsilon$ would do just as well.)

Of course, we can move the spike from $x = 0$ to some other point, $x = a$:

$$\delta(x - a) = \begin{cases} 0, & \text{if } x \neq a \\ \infty, & \text{if } x = a \end{cases} \quad \text{and} \quad \int_{-\infty}^{\infty} \delta(x - a)dx = 1 \tag{A.4}$$

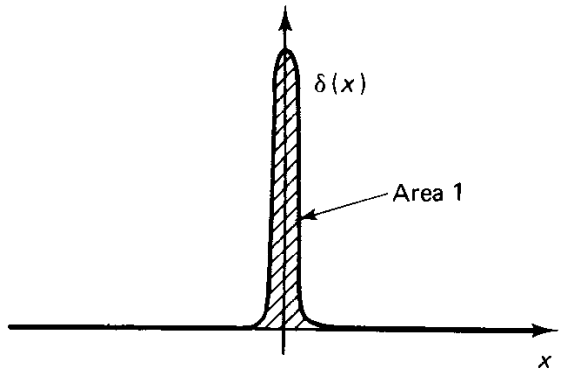


Figure A.1 The Dirac delta function (you must imagine, however, that the curve is infinitely high and infinitesimally narrow).

(see Fig. A.3). Equation (A.2) generalizes to

$$f(x)\delta(x-a) = f(a)\delta(x-a) \tag{A.5}$$

and equation (A.3) generalizes to

$$\int_{-\infty}^{\infty} f(x)\delta(x-a)dx = f(a) \tag{A.6}$$

Now, how would we interpret the expression $\delta(kx)$, if k is some nonzero (real) number? Suppose we multiply by an "ordinary" function $f(x)$ and integrate:

$$\int_{-\infty}^{\infty} f(x)\delta(kx)dx$$

We may change variables, letting $y \equiv kx$, so that $x = y/k$, and $dx = 1/k\, dy$. If k is positive, the integration still runs from $-\infty$ to $+\infty$, but if k is *negative*, then $x = \infty$ implies $y = -\infty$, and *vice versa*, so the order of the limits is reversed. Restoring the "proper" order costs a minus sign. Thus

$$\int_{-\infty}^{\infty} f(x)\delta(kx)dx = \pm\int_{-\infty}^{\infty} f(y/k)\delta(y)\frac{dy}{k} = \pm\frac{1}{k}f(0) = \frac{1}{|k|}f(0) \tag{A.7}$$

(The lower signs apply when k is negative, and we can account for this neatly by putting absolute value bars around the k, as indicated.) In this context, then, $\delta(kx)$ serves the same purpose as $(1/|k|)\delta(x)$:

$$\int_{-\infty}^{\infty} f(x)\delta(kx)dx = \int_{-\infty}^{\infty} f(x)\left[\frac{1}{|k|}\delta(x)\right]dx \tag{A.8}$$



Figure A.2 Two sequences of functions whose limit is $\delta(x)$.

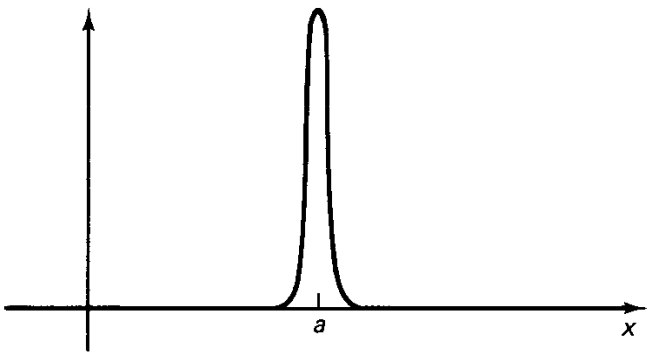


Figure A.3 "Graph" of $\delta(x - a)$.

Because this holds for *any* $f(x)$, it follows that the delta function expressions are equal:*

$$\delta(kx) = \frac{1}{|k|}\,\delta(x) \tag{A.9}$$

What we have just analyzed is really a special case of the general form $\delta(g(x))$, where $g(x)$ is some function of x. $\delta(g(x))$ has spikes at the zeros, $x_1, x_2, x_3, \ldots$, of $g(x)$:

$$g(x_i) = 0 \qquad (i = 1, 2, 3, \ldots, n) \tag{A.10}$$

In the neighborhood of the ith zero, we may expand $g(x)$ as a Taylor series:

$$g(x) = g(x_i) + (x - x_i)g'(x_i) + \tfrac{1}{2}(x - x_i)^2 g''(x_i) + \cdots \cong (x - x_i)g'(x_i) \tag{A.11}$$

In view of equation (A.9), the spike at x_i has the form

$$\delta(g(x)) = \frac{1}{|g'(x_i)|}\,\delta(x - x_i) \qquad (x \cong x_i) \tag{A.12}$$

The factor $|g'(x_i)|^{-1}$ tells us the "strength" of the delta function at x_i. Putting this together with the spikes at the other zeros, we conclude

$$\delta(g(x)) = \sum_{i=1}^{n} \frac{1}{|g'(x_i)|}\,\delta(x - x_i) \tag{A.13}$$

Thus any expression of the form $\delta(g(x))$ can be reduced to a sum of simple delta functions.†

EXAMPLE A.1

Simplify the expression $\delta(x^2 + x - 2)$.

Solution. Here $g(x) = x^2 + x - 2 = (x - 1)(x + 2)$; there are two zeros, at $x_1 = 1$ and $x_2 = -2$. Differentiating, $g'(x) = 2x + 1$, so $g'(x_1) = 3$ and $g'(x_2) = -3$. Thus

$$\delta(x^2 + x - 2) = \tfrac{1}{3}\delta(x - 1) + \tfrac{1}{3}\delta(x + 2)$$

* You ought to ponder that last step for a moment. Ordinarily, the equality of two integrals certainly does *not* imply equality of the integrands. The crucial point here is that the integrals are equal for *any* $f(x)$. Suppose the delta function expressions $\delta(kx)$ and $(1/|k|)\delta(x)$ actually *differed,* say, in the neighborhood of the point $x = 17$. Then I would pick a function $f(x)$ that was sharply peaked about $x = 17$, and the integrals would not be equal. Since, on the contrary, the integrals *must* be equal, it follows that the delta function expressions are themselves equal. [Well, technically they might still differ at isolated points, provided these contribute nothing to the integral. But we can silence this objection by noting that both sides of equation (A.9) are clearly zero except at $x = 0$.]

† Equation (A.13) is *exact,* notwithstanding the truncated Taylor series (A.11) I used in its derivation. At x_i, the "extra" terms are zero, since they contain powers of $(x - x_i)$.

It is an easy matter to generalize to three (or more) dimensions:

$$\delta^3(\mathbf{r}) = \delta(x)\delta(y)\delta(z) \tag{A.14}$$

This three-dimensional delta function is zero everywhere except at the origin, where it blows up. The triple integral over $\delta^3(\mathbf{r})$ is 1:

$$\int \delta^3(\mathbf{r})d^3r = \int \delta(x)\delta(y)\delta(z)dx\,dy\,dz = 1 \tag{A.15}$$

and

$$\int f(\mathbf{r})\delta^3(\mathbf{r} - \mathbf{r}_0)d^3r = f(\mathbf{r}_0) \tag{A.16}$$

For example, the charge density (charge per unit volume) of a point charge q located at the point $\mathbf{r}_0$ can be written

$$\rho(\mathbf{r}) = q\,\delta^3(\mathbf{r} - \mathbf{r}_0)$$

PROBLEMS

A.1. **(a)** $\int_0^3 (2x^2 + 7x + 3)\delta(x - 1)dx = ?$

(b) $\int_0^3 \ln(1 + x)\delta(\pi - x)dx = ?$

A.2. Use equation (A.13) to simplify the expression $\delta(\sqrt{x^2 + 1} - x - 1)$.

A.3. Use equation (A.13) to simplify the expression $\delta(\sin x)$. Sketch this function.

A.4. Let $f(y) = \int_0^2 \delta(y - x(2 - x))dx$. Find $f(y)$, and plot it from $y = -2$ to $y = +2$.

A.5. $\int_{-1}^5 x^4\left[\frac{d^2}{dx^2}\delta(x - 3)\right]dx = ?$ [*Hint:* Integrate by parts.]

A.6. Evaluate $\int \mathbf{r}\cdot(\mathbf{a} - \mathbf{r})\delta^3(\mathbf{r} - \mathbf{b})d^3r$, if $\mathbf{a} = (1, 2, 3)$, $\mathbf{b} = (3, 2, 1)$, and the integration is over a sphere of radius 1.5 centered at (2, 2, 2).

Decay Rates and Cross Sections

B.1 Decays

Suppose particle 1 decays into particles 2, 3, 4, . . . , n:

$$1 \longrightarrow 2 + 3 + 4 + \cdots + n$$

The decay rate is given by the formula

$$d\Gamma = |\mathcal{M}|^2 \frac{S}{2\hbar m_1} \left\{ \left[\frac{c\, d^3\mathbf{p}_2}{(2\pi)^3 2E_2} \right] \left[\frac{c\, d^3\mathbf{p}_3}{(2\pi)^3 2E_3} \right] \cdots \left[\frac{c\, d^3\mathbf{p}_n}{(2\pi)^3 2E_n} \right] \right\} \times (2\pi)^4\, \delta^4(p_1 - p_2 - p_3 - \cdots - p_n) \tag{B.1}$$

where $p_i = (E_i/c,\ \mathbf{p}_i)$ is the 4-momentum of the ith particle (which carries mass m_i, so that $E_i^2 - \mathbf{p}_i^2 c^2 = m_i^2 c^4$). The decaying particle is presumed to be at rest: $p_1 = (m_1 c,\ \mathbf{0})$. S is a product of statistical factors: $1/j!$ for each group of j identical particles in the final state.

Two-Body Decays If there are just *two* particles in the final state, the integrals can be performed explicitly. The total decay rate is

$$\Gamma = \frac{S|\mathbf{p}|}{8\pi\hbar m_1^2 c} |\mathcal{M}|^2 \tag{B.2}$$

Where $|\mathbf{p}|$ is the magnitude of either outgoing momentum:

$$|\mathbf{p}| = \frac{c}{2m_1} \sqrt{m_1^4 + m_2^4 + m_3^4 - 2m_1^2 m_2^2 - 2m_1^2 m_3^2 - 2m_2^2 m_3^2} \tag{B.3}$$

In particular, if the outgoing particles are *massless*, then $|\mathbf{p}| = m_1 c/2$, and

$$\Gamma = \frac{S}{16\pi\hbar m_1} |\mathcal{M}|^2 \tag{B.4}$$

B.2 Cross Sections

Suppose particles 1 and 2 collide, producing particles 3, 4, . . . , n:

$$1 + 2 \longrightarrow 3 + 4 + \cdots + n$$

The cross section is given by the formula

$$d\sigma = |\mathcal{M}|^2 \frac{\hbar^2 S}{4\sqrt{(p_1 \cdot p_2)^2 - (m_1 m_2 c^2)^2}} \left\{\left[\frac{c\, d^3\mathbf{p}_3}{(2\pi)^3 2E_3}\right]\left[\frac{c\, d^3\mathbf{p}_4}{(2\pi)^3 2E_4}\right] \cdots \left[\frac{c\, d^3\mathbf{p}_n}{(2\pi)^3 2E_n}\right]\right\} \times (2\pi)^4\, \delta^4(p_1 + p_2 - p_3 - p_4 - \cdots - p_n) \tag{B.5}$$

where (as before) $p_i = (E_i/c, \mathbf{p}_i)$ is the 4-momentum of particle i (mass m_i), $E_i = c\sqrt{m_i^2 c^2 + \mathbf{p}_i^2}$, and S is a statistical factor ($1/j!$ for each group of j identical particles in the final state).

Two-Body Scattering If there are just *two* particles in the final state, the integrals can be performed explicitly.

(a) *In the center-of-momentum frame.* Here

$$\sqrt{(p_1 \cdot p_2)^2 - (m_1 m_2 c^2)^2} = (E_1 + E_2)|\mathbf{p}_1|/c \tag{B.6}$$

and

$$\frac{d\sigma}{d\Omega} = \left(\frac{\hbar c}{8\pi}\right)^2 \frac{S|\mathcal{M}|^2}{(E_1 + E_2)^2} \frac{|\mathbf{p}_f|}{|\mathbf{p}_i|} \tag{B.7}$$

where $|\mathbf{p}_i|$ is the magnitude of either incoming momentum, and $|\mathbf{p}_f|$ is the magnitude of either outgoing momentum. In particular, for *elastic* scattering ($A + B \longrightarrow A + B$), $|\mathbf{p}_i| = |\mathbf{p}_f|$, so [letting $E \equiv (E_1 + E_2)/2$]:

$$\frac{d\sigma}{d\Omega} = \left(\frac{\hbar c}{16\pi}\right)^2 \frac{S|\mathcal{M}|^2}{E^2} \tag{B.8}$$

(b) *In the lab frame (particle 2 at rest).* Here

$$\sqrt{(p_1 \cdot p_2)^2 - (m_1 m_2 c^2)^2} = m_2 c|\mathbf{p}_1| \tag{B.9}$$

In the case of *elastic* scattering ($A + B \longrightarrow A + B$), the differential cross section is

$$\frac{d\sigma}{d\Omega} = \left(\frac{\hbar}{8\pi}\right)^2 \frac{\mathbf{p}_3^2 S|\mathcal{M}|^2}{m_2|\mathbf{p}_1|\,|\,|\mathbf{p}_3|(E_1 + m_2 c^2) - |\mathbf{p}_1|E_3 \cos\theta|} \tag{B.10}$$

If, in particular, the incident particle is *massless* ($m_1 = 0$), this reduces to

$$\frac{d\sigma}{d\Omega} = \left(\frac{\hbar E_3}{8\pi m_2 c E_1}\right)^2 S|\mathcal{M}|^2 \tag{B.11}$$

If the target recoil is negligible ($m_2 c^2 \gg E_1$), then (B.10) reduces to

$$\frac{d\sigma}{d\Omega} = \left(\frac{\hbar}{8\pi m_2 c}\right)^2 |\mathcal{M}|^2 \tag{B.12}$$

If the outgoing particles are massless ($m_3 = m_4 = 0$), (B.5) yields

$$\frac{d\sigma}{d\Omega} = \left(\frac{\hbar}{8\pi}\right)^2 \frac{S|\mathcal{M}|^2|\mathbf{p}_3|}{m_2|\mathbf{p}_1|(E_1 + m_2 c^2 - |\mathbf{p}_1|c\cos\theta)} \tag{B.13}$$

Appendix C

Pauli and Dirac Matrices

C.1 Pauli Matrices

These are three Hermitian, unitary, traceless 2×2 matrices:

$$\sigma_x \equiv \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \qquad \sigma_y \equiv \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \qquad \sigma_z \equiv \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \tag{C.1}$$

(Often we use *numerical* indices: $\sigma_1 = \sigma_x$, $\sigma_2 = \sigma_y$, $\sigma_3 = \sigma_z$; σ is not part of a 4-vector, and we do not distinguish upper and lower indices: $\sigma_1 = \sigma^1$, $\sigma_2 = \sigma^2$, $\sigma_3 = \sigma^3$.)

(a) *Product Rules.*

$$\sigma_i \sigma_j = \delta_{ij} + i\epsilon_{ijk}\sigma_k \tag{C.2}$$

(A 2×2 unit matrix is implied in the first term, and summation over k in the second). Thus, in particular:

$$\sigma_x^2 = \sigma_y^2 = \sigma_z^2 = 1 \tag{C.3}$$

$$\sigma_x\sigma_y = i\sigma_z, \qquad \sigma_y\sigma_z = i\sigma_x, \qquad \sigma_z\sigma_x = i\sigma_y \tag{C.4}$$

$$[\sigma_i, \sigma_j] = 2i\epsilon_{ijk}\sigma_k \qquad \text{(commutator)} \tag{C.5}$$

$$\{\sigma_i, \sigma_j\} = 2\delta_{ij} \qquad \text{(anticommutator)} \tag{C.6}$$

and for any two vectors **a** and **b**,

$$(\mathbf{a}\cdot\sigma)(\mathbf{b}\cdot\sigma) = \mathbf{a}\cdot\mathbf{b} + i\sigma\cdot(\mathbf{a}\times\mathbf{b}) \tag{C.7}$$

(b) *Exponentials.*

$$e^{i\theta\cdot\sigma} = \cos\theta + i\hat{\theta}\cdot\sigma \sin\theta \tag{C.8}$$

C.2 Dirac Matrices

These are four unitary traceless 4×4 matrices:

$$\gamma^0 \equiv \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}; \qquad \gamma^i = \begin{pmatrix} 0 & \sigma^i \\ -\sigma^i & 0 \end{pmatrix} \tag{C.9}$$

(Here 1 is the 2×2 unit matrix, and 0 is the 2×2 matrix of zeros; σ^i are the Pauli matrices. Lowering indices changes the sign of the "spatial" components: $\gamma_0 = \gamma^0$, $\gamma_i = -\gamma^i$.) We introduce as well the auxiliary matrices

$$\gamma^5 \equiv i\gamma^0\gamma^1\gamma^2\gamma^3 \tag{C.10}$$

$$\mathbf{\Sigma} \equiv \begin{pmatrix} \boldsymbol{\sigma} & 0 \\ 0 & \boldsymbol{\sigma} \end{pmatrix} \tag{C.11}$$

$$\sigma^{\mu\nu} = \frac{i}{2}(\gamma^\mu\gamma^\nu - \gamma^\nu\gamma^\mu) \tag{C.12}$$

For any 4-vector a^μ, we define the 4×4 matrix $\not{a}$ as follows:

$$\not{a} \equiv a_\mu\gamma^\mu \tag{C.13}$$

(a) *Product Rules.* In terms of the metric

$$g^{\mu\nu} \equiv \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix} \tag{C.14}$$

(note that $g^{\mu\nu}g_{\mu\nu} = 4$), we have:

$$\gamma^\mu\gamma^\nu + \gamma^\nu\gamma^\mu = 2g^{\mu\nu}, \qquad \not{a}\not{b} + \not{b}\not{a} = 2a\cdot b \tag{C.15}$$

$$\gamma_\mu\gamma^\mu = 4 \tag{C.16}$$

$$\gamma_\mu\gamma^\nu\gamma^\mu = -2\gamma^\nu, \qquad \gamma_\mu\not{a}\gamma^\mu = -2\not{a} \tag{C.17}$$

$$\gamma_\mu\gamma^\nu\gamma^\lambda\gamma^\mu = 4g^{\nu\lambda}, \qquad \gamma_\mu\not{a}\not{b}\gamma^\mu = 4a\cdot b \tag{C.18}$$

$$\gamma_\mu\gamma^\nu\gamma^\lambda\gamma^\sigma\gamma^\mu = -2\gamma^\sigma\gamma^\lambda\gamma^\nu, \qquad \gamma_\mu\not{a}\not{b}\not{c}\gamma^\mu = -2\not{c}\not{b}\not{a} \tag{C.19}$$

(b) *Trace Theorems.* The trace of the product of an *odd* number of gamma matrices is *zero.*

$$Tr(1) = 4 \tag{C.20}$$

$$Tr(\gamma^\mu\gamma^\nu) = 4g^{\mu\nu}, \qquad Tr(\not{a}\not{b}) = 4a\cdot b \tag{C.21}$$

$$Tr(\gamma^\mu\gamma^\nu\gamma^\lambda\gamma^\sigma) = 4(g^{\mu\nu}g^{\lambda\sigma} - g^{\mu\lambda}g^{\nu\sigma} + g^{\mu\sigma}g^{\nu\lambda}),$$
$$Tr(\not{a}\not{b}\not{c}\not{d}) = 4(a\cdot b\, c\cdot d - a\cdot c\, b\cdot d + a\cdot d\, b\cdot c) \tag{C.22}$$

Since γ^5 is the product of an *even* number of γ matrices, it follows that $Tr(\gamma^5\gamma^\mu) = 0$ and $Tr(\gamma^5\gamma^\mu\gamma^\nu\gamma^\lambda) = 0$. When γ^5 is multiplied by an *even* number of γ's, we find

$$Tr(\gamma^5) = 0 \tag{C.23}$$

$$Tr(\gamma^5\gamma^\mu\gamma^\nu) = 0, \qquad Tr(\gamma^5\not{a}\not{b}) = 0 \tag{C.24}$$

$$Tr(\gamma^5\gamma^\mu\gamma^\nu\gamma^\lambda\gamma^\sigma) = 4i\epsilon^{\mu\nu\lambda\sigma}, \qquad Tr(\gamma^5\not{a}\not{b}\not{c}\not{d}) = 4i\epsilon^{\mu\nu\lambda\sigma}a_\mu b_\nu c_\lambda d_\sigma \tag{C.25}$$

where $\epsilon^{\mu\nu\lambda\sigma} = -1$, if $\mu\nu\lambda\sigma$ is an even permutation of 0123, $+1$ for an odd permutation, and 0 if any two indices are the same. Note that

$$\epsilon^{\mu\nu\lambda\sigma}\epsilon_{\mu\nu\kappa\tau} = -2(\delta^\lambda_\kappa\delta^\sigma_\tau - \delta^\lambda_\tau\delta^\sigma_\kappa) \tag{C.26}$$

Feynman Rules (Tree Level)

D.1 External Lines

Spin 0: (nothing)

Spin $\frac{1}{2}$:
- Incoming particle: u
- Incoming antiparticle: $\bar{v}$
- Outgoing particle: $\bar{u}$
- Outgoing antiparticle: v

Spin 1:
- incoming: ϵ^{μ}
- outgoing: $\epsilon^{\mu *}$

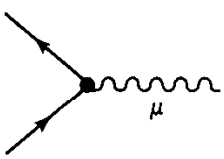
D.2 Propagators

Spin 0: $$\frac{i}{q^2 - (mc)^2}$$

Spin $\frac{1}{2}$: $$\frac{i(\not{q} + mc)}{q^2 - (mc)^2}$$

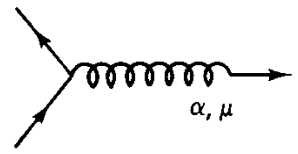
Spin 1:
- Massless: $\dfrac{-ig_{\mu\nu}}{q^2}$
- Massive: $\dfrac{-i[g_{\mu\nu} - q_\mu q_\nu/(mc)^2]}{q^2 - (mc)^2}$

D.3 Vertex Factors

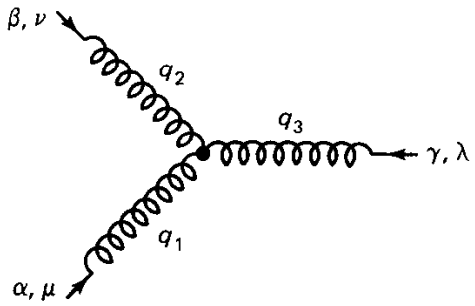
QED: 

$ig_e\gamma^{\mu}$ $\quad (g_e = \sqrt{4\pi\alpha})$

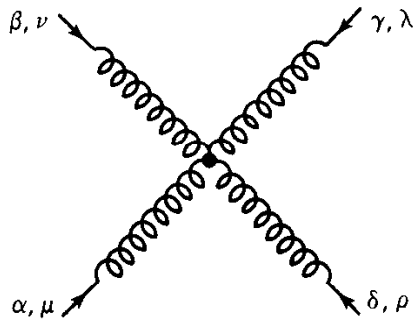
QCD:



$$\frac{-ig_s}{2}\lambda^{\alpha}\gamma^{\mu}$$

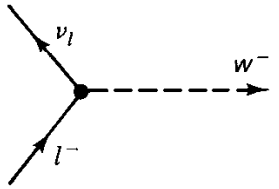


$$-g_s f^{\alpha\beta\gamma}[g_{\mu\nu}(q_1 - q_2)_\lambda + g_{\nu\lambda}(q_2 - q_3)_\mu + g_{\lambda\mu}(q_3 - q_1)_\nu]$$



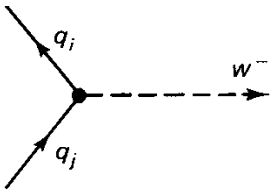
$$-ig_s^2[f^{\alpha\beta\eta}f^{\gamma\delta\eta}(g_{\mu\lambda}g_{\nu\rho} - g_{\mu\rho}g_{\nu\lambda}) + f^{\alpha\delta\eta}f^{\beta\gamma\eta}(g_{\mu\nu}g_{\lambda\rho} - g_{\mu\lambda}g_{\nu\rho}) + f^{\alpha\gamma\eta}f^{\delta\beta\eta}(g_{\mu\rho}g_{\nu\lambda} - g_{\mu\nu}g_{\lambda\rho})]$$

GWS:



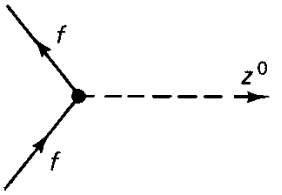
$$\frac{-ig_w}{2\sqrt{2}}\gamma^{\mu}(1 - \gamma^5)$$

(Here l is any lepton, and ν_l the corresponding neutrino.)



$$\frac{-ig_w}{2\sqrt{2}}\gamma^{\mu}(1 - \gamma^5)U_{ij}$$

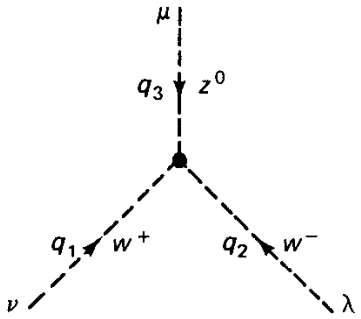
(Here $i = u$, c, or t, and $j = d$, s, or b; U is the Kobayashi–Maskawa matrix.)



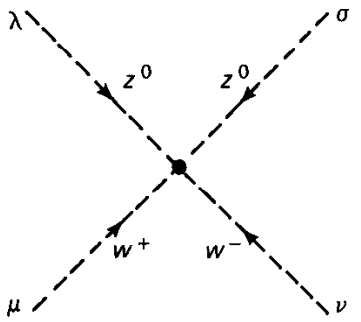
$$\frac{-ig_z}{2}\gamma^{\mu}(c_V^f - c_A^f\gamma^5)$$

(Here f is any quark or lepton.)

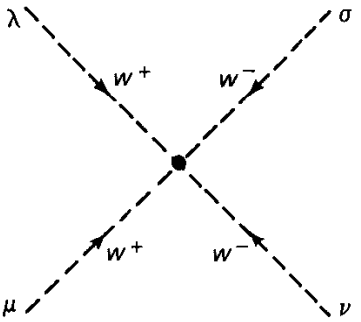
f	c_V	c_A
ν_e, ν_μ, ν_τ	$\frac{1}{2}$	$\frac{1}{2}$
e^-, μ^-, τ^-	$-\frac{1}{2} + 2\sin^2\theta_w$	$-\frac{1}{2}$
u, c, t	$\frac{1}{2} - \frac{4}{3}\sin^2\theta_w$	$\frac{1}{2}$
d, s, b	$-\frac{1}{2} + \frac{2}{3}\sin^2\theta_w$	$-\frac{1}{2}$



$$ig_w \cos\theta_w[g_{\nu\lambda}(q_1 - q_2)_\mu + g_{\lambda\mu}(q_2 - q_3)_\nu + g_{\mu\nu}(q_3 - q_1)_\lambda]$$



$$-ig_w^2 \cos^2\theta_w(2g_{\mu\nu}g_{\lambda\sigma} - g_{\mu\lambda}g_{\nu\sigma} - g_{\mu\sigma}g_{\nu\lambda})$$

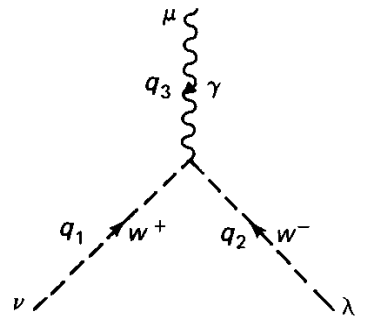


$$ig_w^2(2g_{\mu\lambda}g_{\nu\sigma} - g_{\mu\nu}g_{\lambda\sigma} - g_{\mu\sigma}g_{\nu\lambda})$$

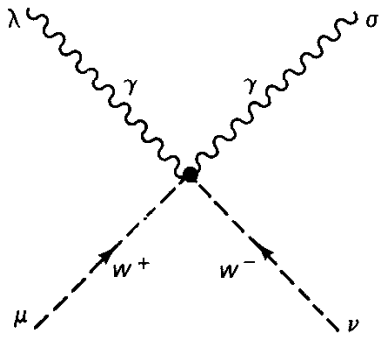
The weak coupling constants are related to the electromagnetic coupling constant:

$$g_w = \frac{g_e}{\sin\theta_w}; \qquad g_z = \frac{g_e}{\sin\theta_w \cos\theta_w}.$$

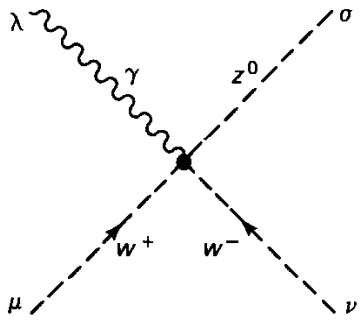
There are also "mixed" couplings of the photon to the W and Z:



$$ig_e[g_{\nu\lambda}(q_1 - q_2)_\mu + g_{\lambda\mu}(q_2 - q_3)_\nu + g_{\mu\nu}(q_3 - q_1)_\lambda]$$



$$-ig_e^2(2g_{\mu\nu}g_{\lambda\sigma} - g_{\mu\lambda}g_{\nu\sigma} - g_{\mu\sigma}g_{\nu\lambda})$$



$$-ig_e g_w \cos\theta_w (2g_{\mu\nu}g_{\lambda\sigma} - g_{\mu\lambda}g_{\nu\sigma} - g_{\mu\sigma}g_{\nu\lambda})$$

Index

A: *see* Conservation laws; Baryon number
ABC theory, 201–213
Abelian, 106, 354
  See also Gauge invariance; Groups
Accelerator, 4
Adjoint, 223, 228
Allowed energies, 148
α: *see* Fine structure constant
α_s, 61, 77, 165, 279, 294
α_w, 76, 302, 309
α particle: *see* Particles
Amplitude, 119, 189, 194–195, 200–201, 229, 231
Anderson, C. D., 18, 20, 29
Angular momentum, 103
  addition, 109–113
  eigenfunctions: *see* Spherical harmonics
  eigenvalues, 146
  matrices: *see* Spin matrices
  orbital, 107–109, 153, 252
  spin, 103, 107–109, 116, 252
Annihilation: *see* Pair annihilation
Anomalous magnetic moment
  electron, 17, 153, 156, 232
  proton, 156–157
Anticommutator, 139, 216
Antielectron: *see* Positron
Antineutrino, 24, 26–28, 124–125
Antineutron, 21
Antiparticle, 3, 18–22, 36–37, 135, 217, 220–221
Antiproton, 21, 30–31, 96–97
Antisymmetric state, 112, 118, 174–180
Antisymmetric tensor, 100, 224
Antisymmetrization, 231
Associated Laguerre polynomial, 149
Associated Legendre function, 147
Associated production, 32–33
Asymptotic freedom 62–64, 164–165, 209, 279, 292–295
Axial vector, 126–127, 224, 302–303, 324

B: *see* Beauty
B meson, 44, 79, 167
b quark, 44, 122
Bare: *see* Charge; Coupling constant; Mass
Barn, 200
Baryon, 17, 29, 34, 36
  decuplet, 34–36, 38, 111, 179
  magnetic moment, 180–182
  mass, 182–184
  number: *see* Conservation laws
  octet, 33, 39, 111, 179–180
Beautiful baryon, 44
Beautiful meson, 44
Beauty, 44, 47
Beta decay, 22–24, 27, 46, 52, 56, 301, 309–314
β particle: *see* Electron
Bethe, H. A., 155, 159, 208
Bevatron, 21, 96
Bhabha scattering, 57–59, 78, 232, 234
Bilinear covariants, 222–225
Bispinor: *see* Dirac spinor
Bjorken, J. D., 42, 269
Bjorken scaling, 269–273, 294–295
Bohr, N., 13, 23, 149, 151
Bohr energies, 148, 186
Bohr magneton, 153

Bohr model, 13
Bohr radius, 149
Bohr theory, 149–151
Boson, 109, 175
Bottom: *see* Beauty
Bottomonium, 143–144, 167–168
  See also Υ meson
Bound state, 2, 39, 42, 52, 143–188
Branching ratio, 72, 190, 316
Brookhaven, 32, 42, 132
Breit frame, 102
Brick wall frame: *see* Breit frame
Broken symmetry, 336
  See also Gauge invariance
Bubble chamber, 35

C: *see* Charm; Charge, conjugation
c quark, 42–44, 122
Cabibbo, N., 70, 317
Cabibbo angle, 301, 314, 317–321
Cabibbo theory, 317–322
Callan–Gross relation, 270–273
Casimir's trick, 236–238, 253
Center of momentum, 96–98
Central potential, 146
Centrifugal barrier, 147, 159
CERN, 6, 40, 44, 46, 67, 323–325, 327
Chadwick, J., 14
Charge
  bare, 63, 249
  conjugation, 103, 128–130, 134–135, 162, 222, 252
  conservation: *see* Conservation laws
  effective, 63, 249
  electric, 47, 60, 118
  independence, 117
  renormalization, 246–250
  weak, 65
  See also Coupling constant
Charged weak interactions, 301–304
Charm, 42–44, 47, 165
Charmed baryons, 43–45
Charmed mesons, 43–44
Charmonium, 44, 143–144, 162–169
  See also ψ meson
Chiral states, 330–333
Chromodynamics, 55–56, 355–357
  See also Quantum chromodynamics
Clebsch–Gordan coefficients, 111–112
Cloud chamber, 7, 20
CM: *see* Center of momentum
Coleman–Glashow formula, 53
Colliding beams, 6, 46, 98–99
Collisions, 91–93
Color, 41–42, 60–61, 64, 74, 261–262, 279
  factor, 164, 285–289, 291
  octet, 280, 285–286
  sextet, 288
  singlet, 178, 188, 280–281, 284–286, 289, 291, 298
  $SU(3)$, 178, 281, 355–356
  triplet, 287–288
Colorless particle, 41, 64–65, 280–281
  See also Color, singlet
Commutator, 139
Completeness, 217, 221, 229, 237
Compton, A. H., 15
Compton
  scattering, 15–16, 58, 78, 102, 232, 235–236
  wavelength, 15, 63
Confinement, 40, 42, 64–65, 286, 289, 295–296
Conservation laws, 72–76, 79, 103, 105
  angular momentum, 105
  baryon number, 29–31, 72–74, 77, 118
  charge, 32, 72–73, 105, 128, 226, 313
  color, 74
  electron number, 27, 47, 74, 317
  energy, 51–52, 60, 91–92, 105, 203, 230
  flavor, 67, 74–75
  isospin, 117–120
  lepton number, 26–27, 31
  mass, 91–92
  momentum, 91–92, 105, 203, 230
  muon number, 27, 74, 317
  parity, 125–128
  quark number, 74
  strangeness, 32–34, 74–75, 118
  tau number, 47, 74, 317
Conserved current, 313
Continuity equation, 226, 347
Contraction of indices, 87
Contraction theorems, 239
Contravariant four-vector, 85, 214
Cosmic rays, 4, 18–20, 29, 100
Coulomb force, 57
Coulomb gauge, 227–229, 281, 302
Coulomb potential, 148, 164, 194, 285, 287
Coupling constant, 61, 202
  bare, 249
  dimensions, 230
  effective, 63, 209–210
  electromagnetic, 62–63, 230, 336, 359
  renormalized, 247–248, 292
  running, 62, 77, 209–210, 249, 292, 294–295
  strong, 61, 77, 279, 359
  weak, 77, 302, 309, 324, 335–337
  See also Charge
Covariant derivative, 349, 352, 356, 365
Covariant four-vector, 85–86, 214, 251
Cowan, C. L., 26
CP violation, 130–134, 321
Cronin, J. W., 132–133, 135
Crossing symmetry, 21–22, 58
Cross section, 119, 189–194, 378
See also Golden rule, for scattering
  $A + A \rightarrow B + B$, 204–206
  hard sphere, 193
  Mott, 241
  nucleon–nucleon, 119

Cross section (*Continued*)
pair annihilation (QCD), 292
pair annihilation (QED), 164, 187, 245, 261
pion–nucleon, 119–120, 140
Rutherford, 194, 241
two-body, 199–201
Current
charged weak, 332
color, 357
conserved, 226, 313
electromagnetic, 225–227, 333, 335–336, 349
weak hypercharge, 334–335
weak isospin, 334–335
Yang–Mills, 354
Cutoff, 208, 247
CVC hypothesis, 313
Cyclotron formula, 7

D meson, 44, 76, 79, 166
D_s meson: *see F* meson
d quark, 37, 122
D'Alembertian, 227
Davis, R., 26
De Broglie wavelength, 6
Decay rate, 189–190, 195–198, 377
kaon, 318
muon, 304–309
neutron, 311–312
pion, 316
positronium, 164, 245–246
quarkonium, 292
two-body, 197–198
Decays, 2, 3, 72–76, 93–94, 241
Decuplet: *see* Baryon, decuplet
Deep inelastic scattering, 40, 48, 257–258, 266–273
Degeneracy, 149
Delbruck scattering, 78
Δ, 31, 34, 69, 120
Delta function: *see* Dirac delta function
Density of states, 194
See also Phase space
DESY, 67
Detailed balance, 21–22, 134
Detectors, 7
Deuteron, 118, 143
Differential cross section, 192, 200, 245
Differential decay rate, 195
Dimensions
amplitudes, 200–201
coupling constants, 202, 230
fields, 347
Dipole function, 267
Dipole moments: *see* Electric dipole moment; Magnetic moment
Diquarks, 289
Dirac, P. A. M., 18, 208–209, 214–215
Dirac delta function, 157, 195–196, 203, 230–231,373–376
Dirac equation, 18, 213–222
momentum space, 218, 228
Dirac matrices, 215, 380
See also Gamma matrices
Dirac sea, 18–21, 217
Dirac spinor, 216, 226
Disconnected diagram, 207
Discrete symmetry, 103
Dot product, 86
Down quark: *see d* quark
Downness, 47
Drell–Yan process, 295
Dresden, M., 201

Effective charge: *see* Charge
Effective mass: *see* Mass
Eigenfunction, 146
Eigenvalue, 114, 146
Eigenvector, 114
Eight-baryon problem, 121
Eightfold Way, 33–39, 107, 121–122
Einstein, A., 14–15, 76
Einstein summation convention, 64
Elastic collision, 91–92
Electric dipole moment, 135
Electric form factor, 267
Electrodynamics, 55–56, 225–228
See also Quantum electrodynamics
Electromagnetic current: *see* Current
Electromagnetic decay, 241
Electromagnetic field, 225
Electromagnetic force, 55–56
Electromagnetic potential, 226
Electron, 4, 11–12, 230
Electron number: *see* Conservation laws
Electron–deuteron scattering, 278
Electron–electron scattering: *see* Møller scattering
Electron–muon scattering, 232–233, 238–240
Electron–neutron scattering, 276, 278
Electron–positron scattering
elastic: *see* Bhabha scattering
inelastic, 257–262, 327–330
Electron–positron annihilation: *see* Pair annihilation
Electron–proton scattering
elastic, 262–267, 269
inelastic, 266–269
Electroweak force, 3, 46, 56, 322, 330–337
See also GWS theory
Electroweak interference, 326, 330
Elementary particles, 1
See also Particles
Energy
conservation: *see* Conservation laws
kinetic, 90–92, 152
operator, 145; *see also* Hamiltonian
relativistic, 87–91, 152
rest, 90
Energy-momentum four-vector, 89
η meson, 38–39, 170

η' meson, 38–39, 170, 172, 188
Euler–Lagrange equation, 344–345
Exchange of particles, 16, 57, 61
  See also Mediator
Exclusion principle: *see* Pauli
Exotic atom, 159
Exotic particle, 53
Expectation value, 151–152, 186
External line, 58–60, 229–231, 282–283

F meson, 44 (now called D_s)
Faddeev–Popov ghost, 282
Family: *see* Generation
Fermi, E., 23, 56, 120
Fermi constant (G_F), 307–309
Fermi's Golden Rule: *see* Golden Rule
Fermi theory of beta decay, 23–24, 44, 46, 56, 307–308
Fermion, 109, 175
Feynman, R. P., 3, 21, 27, 56, 155, 203, 208
Feynman calculus, 59, 189–212, 361–362
Feynman diagram, 55, 57, 59, 189, 194, 201, 203
Feynman rules, 3, 59, 194, 213, 357–360, 381–384
  ABC theory, 201–204
  GWS, 336–337
  QCD, 279–284
  QED, 213, 228–231, 255
  weak interactions, 302, 317–318, 322
Feynman–Stückelberg interpretation, 21
Field strength tensor, 225
Fifth force, 296
Fine structure, 151–155
Fine structure constant, 9, 59, 148, 152, 230, 249, 294, 309
Flavor, 41–44, 47, 53, 64, 116–122, 279, 280
Flavordynamics, 55–56, 67
Form factor, 266–269, 312, 315
Four-momentum, 89
Four-vector, 84–87
Four-velocity: *see* Proper velocity
ft-value, 312
Fundamental representation, 121, 171

G: *see* G-parity
G_F: *see* Fermi constant
Gamma matrices, 216, 224, 238–239, 380
Gamma rays: *see* Photon
Gauge fields, 348–349, 352, 356
Gauge invariance, 253, 298
  abelian, 354
  broken, 362–365
  global, 348, 350, 352, 354
  local 348–350, 352, 354, 365–367
  nonabelian, 354–355
Gauge theory, 210, 343–371
Gauge transformation, 105, 226–227, 348–349
Gell-Mann, M., 32–34, 36–37, 56, 131–132
Gell-Mann matrices, 282, 355
Gell-Mann–Nishijima formula, 118, 140, 334
Gell-Mann–Okubo mass formula, 52
Generation, 47–48, 317, 321–322
Ghost particle, 228, 282, 367
GIM mechanism, 44, 70–71, 301, 319–320, 322
Glashow, S. L., 42–44, 56, 70, 322–323, 330
Global gauge transformation, 348, 352, 354
Glueball, 48, 61
Gluon, 48, 55–56, 60–61, 260, 275, 279–281
  octet, 280
Gluon–gluon coupling, 61, 282–284, 291, 299, 359–360
Golden Rule, 189, 194–201, 305
  for decays, 195, 305
  for scattering, 198–199
Goldstone boson, 364, 366–367
Goldstone's theorem, 364, 371
G-parity, 129–30
Grand Unification, 31, 76–77
Gravitational force, 55
Graviton, 16, 48, 55–56
Greenberg, O. W., 41
Ground state
  baryon, 176
  hydrogen, 186
  See also Vacuum
Group theory, 103, 106
Groups, 106
  abelian, 106
  continuous, 106
  finite, 106
  infinite, 106
  $O(n)$, 106, 137
  $SO(3)$, 107
  $SO(n)$, 106–107, 137
  $SU(2)$, 107, 121, 352
    See also Isospin
  $SU(2) \times U(1)$, 334–337, 355
  $SU(3)$, 103, 107, 121
    See also Eightfold Way; Color, $SU(3)$
  $SU(6)$, 103, 121
  $SU(n)$, 106, 137
  $U(n)$, 106
GUTs; *see* Grand Unification
GWS (Glashow/Weinberg/Salam) theory, 3, 46, 48, 56, 66, 76, 210, 301, 322–325, 330–337
Gyromagnetic ratio, 153, 156

Hadron, 28, 33, 143
  See also Baryon; Meson
Hadron production, 257–262
Half-life, 72, 211, 312
Hamiltonian, 146, 151–153, 158, 160, 252
Hard-sphere scattering, 191–193
Heaviside–Lorentz units, 9, 230, 249, 347
Heavy lepton, 44
Heavy quark, 165, 172
Helicity, 27, 124, 221, 330–333, 339

Heisenberg, W., 116
See also Uncertainty principle
Hermitian conjugate, 220, 223
Hermitian matrix, 351, 355
Hidden symmetry, 364
Higgs mechanism, 76, 330, 343, 354, 360, 365–368
Higgs particle, 48, 367
Higher-order processes, 58, 63, 202, 206–210, 246–250, 281, 292–294
Hole theory, 20–21
t'Hooft, G., 210, 323, 360
Hydrogen, 143, 148–159
Hypercharge, 34, 334–335
See also Weak hypercharge
Hyperfine splitting
in baryons, 182–184
in hydrogen, 156–159
in mesons, 172
in positronium, 161

Identical particles, 1, 41, 175–176
Impact parameter, 191
Inclusive cross section, 267–269
Indistinguishable particles, 1
Inertial frame, 81
Infinite momentum frame, 272
Intermediate vector boson, 46, 56, 335
See also W; Z
Internal line, 58–60, 202, 229
Internal momenta, 203, 231
Internal quantum number, 128
Internal symmetries, 103, 105, 116–120
Intrinsic angular momentum: *see* Angular momentum, spin
Intrinsic parity: *see* Parity
Intrinsic strength of weak force, 76, 309
Intersecting storage rings, 6
Invariance, 105
See also Symmetry
Invariant, 85, 88–89, 94, 97, 281
Inverse beta decay, 26, 303
Inversion, 125–126
Ionization, 7
Irreducible representation, 107
Isospin, 103, 116–120, 335
See also Weak isospin

J/ψ: *see* ψ meson
Jets, 48, 258–260

K meson, 29–38, 128, 131–133, 318
Kaon: *see* K meson
Ket, 3, 109
Klein–Gordon equation, 213–215, 227
KM matrix: *see* Kobayashi–Maskawa matrix
Kobayashi–Maskawa matrix, 70–71, 133, 301, 321–322, 325
Kronecker delta, 139

Lagrangian, 343–347
classical, 343–344
Dirac, 345–346
Higgs, 367
Klein–Gordon, 345
Maxwell, 347
Proca, 346–347
QCD, 357
QED, 350
Yang–Mills, 354
Yukawa, 371
Lagrangian density, 344, 347
See also Lagrangian
Lamb, W. E., 32, 154
Lamb shift, 17, 154–156, 161, 209, 249, 294
Λ (scale parameter in QCD), 295
Λ baryon, 29, 32–35, 70
λ matrices: *see* Gell-Mann matrices
Laplacian, 146
Lederman, L. M., 28, 132
Lee, T. D., 56, 122, 128
Left-handed doublet, 334–335
Left-handed state, 27, 124, 330–334
Lepton number, 26–27
See also Conservation laws
Leptonic decay, 318
Leptons, 17–18, 28, 65–67
families, 47
table, 28, 47
weak interactions, 301–304, 317
See also Electron; Muon; Neutrino; Tau
Levi-Civita symbol, 139, 239, 254, 303
Lifetime, 42, 52, 72–73, 75–76, 132, 189–190
A particle, 204
muon, 307
neutron, 312–314
pion, 316
positronium, 164, 187, 246
See also Decay rate
Light quark baryons, 172–184
Light quark mesons, 143, 168–172
Lightlike four-vector, 86
Linear-plus-coulomb potential, 165–167
Local gauge invariance, 3, 343, 348–350, 352, 360
Local gauge transformations, 348, 365
Logarithmic divergence, 208, 247
Longitudinal polarization, 301
Loop diagram, 207–210, 231, 246, 255, 281–282, 292–294
Lorentz condition, 226–229, 281, 301–302
Lorentz contraction, 83
Lorentz invariance, 85–86, 97
Lorentz transformations, 81–84
Lowering operator, 140
Luminosity, 194, 245–246

Magnetic form factor, 267
Magnetic moment, 153
anomalous: *see* Anomalous magnetic moment

baryons, 180–182
electron, 17, 153, 156
proton, 156–157
Mandelstam variables, 102
Marshak, R. E., 18
Mass formulas
baryons, 182–184
Coleman–Glashow, 53
Gell-Mann–Okubo, 52
mesons, 172
pion, 51
quarkonium, 165
W and Z, 325
Mass
bare, 122
constituent, 122
current, 122
effective, 121–122, 209–210
matrix, 351, 371
neutral kaon, 132, 135
origin, 368
physical, 209
relativistic, 90, 92
renormalized, 209, 247
running, 209
shell, 60
term, 360–362
Massless particle, 90
Matrix element, 194
See also Amplitude
Matter–antimatter asymmetry, 22, 134
Maximal parity violation, 123, 133, 303
Maxwell, J. C., 56, 76
Maxwell Lagrangian, 347
Maxwell's equations, 225–228
Mechanics, 2
Mediator, 16, 47–48, 55–57, 61, 301, 308–309
See also Gluon; Graviton; Intermediate vector boson; Photon; Pion; W; Z
Meson, 17–18, 29, 31, 33–34, 36, 38–39, 128, 168–72
mass, 171–172
nonet, 36, 38–39, 128, 169
octet, 34
Metric, 85, 216
Millikan, R. A., 15, 39
Minkowski metric: *see* Metric
Minimal coupling, 350
Mixing
Cabibbo, 70–71, 74, 317–322
Kobayashi-Maskawa, 70–71, 321–322, 325
neutral kaons, 130–134
neutral mesons, 170–171
Møller scattering, 57–58, 232–233
Momentum
conservation: *see* Conservation laws
four-vector, 89
operator, 144, 214
relativistic, 87–91
space, 218, 228–229, 358
Mott scattering, 232–233, 240–241, 265, 270, 309, 313
Multiplets, 117–118
Multiplicative quantum number, 127, 129
Muon, 4, 18–19, 24–25, 27, 66, 100, 304–309
Muon number, 27, 47
See also Conservation laws
Muonium, 159–160, 168

N: *see* Nucleon
n: *see* Neutron
Neddermeyer, S., 18
Ne'eman, Y., 33
Negative energy states, 18, 217–218, 221
Neutral weak interactions, 65–67, 322–324, 326
Neutrino, 14, 19, 22–28, 124–125
Neutrino–electron scattering, 303, 323–327
Neutrino–nucleon scattering, 323–324
Neutron, 14, 24, 27, 68, 116, 135, 309–315
Ninth gluon, 280–281, 296, 356
Noether's theorem, 103, 105, 117, 370
Nonabelian gauge, 354–355
Nonet: *see* Meson, nonet
Nonleptonic decay, 318–319
Normalization
Dirac spinor, 218, 220–221, 225, 229
Pauli spinor, 113
polarization vector, 229
wave function, 145
November Revolution, 41–45, 166
Nucleon, 116
See also Neutron; Proton
Nucleon–nucleon scattering, 118–119
Nucleus, 12–14

$O(n)$: *see* Groups
Octet, 35
See also Baryon, octet; Color, octet; Gluon, octet; Meson, octet
Ω^-, 34–36, 52–53, 70
ω, 170, 172
Orthogonal matrix, 107
Orthogonal polarization vectors, 229
Orthogonal spinors, 228, 251
OZI rule, 75–76, 79, 166–167, 294

P: *see* Parity, operator
Pair annihilation, 58, 161–162, 215, 232, 241–245, 257–262, 289–292
Pair production, 58, 215, 232
Pais, A., 32, 131–132
Parity, 103, 122–128, 130–131, 134, 223–224
baryon, 127
conservation, 125, 127–128
intrinsic, 127
invariance, 122–123
meson, 127–128
operator, 125–127, 224
particle and antiparticle, 127, 252
photon, 127

Parity (*Continued*)
quarks, 127
violation, 123, 131
Particle Data Booklet, 50, 111
Particles
α, 12–13, 141
antineutrino, 24, 26–28, 124–125
antineutron, 21
antiproton, 21, 30–31, 96–97
B meson, 44, 79, 167
b (beauty or bottom) quark, 44, 122
β: *see* Electron
c (charm) quark, 42–44, 122
D meson, 44, 76, 79, 166
d (down) quark, 37, 122
Δ, 31, 34, 69, 120
deuteron, 118, 143
electron, 4, 11–12, 230
η, 38–39, 170
η', 38–39, 170, 172, 188
F meson, 44 (now called D_s)
γ: *see* Photon
gluon, 48, 55–56, 60–61, 260, 275, 279–281
graviton, 16, 48, 55–56
J or J/ψ meson: *see* ψ meson
K meson
charged, 29, 31–38, 128, 318
neutral, 29–30, 32–38, 131–133
kaon: *see* K meson
Λ, 29, 32–35, 70
Λ_c, 44–45
muon, 4, 18–19, 24–25, 27, 66, 100, 304–309
neutrino, 14, 19, 22–28, 124–125
neutron, 14, 24, 27, 68, 116, 135, 309–314
nucleon, 116
Ω^-, 34–36, 52–53, 70
ω meson, 170, 172
ϕ meson, 75, 170, 172, 262–263
photon, 14–17, 47, 55–57, 225–230
π meson: *see* Pion
pion, 18–19, 24–25, 27, 51, 68, 124–125, 211, 314–317
positron, 14, 20–21, 217, 230
proton, 4, 13, 29–34, 39–40, 77, 116, 262–275
ψ meson, 42, 44, 75, 144, 262–263; *see also* Charmonium
ρ meson, 170, 172
s (strange) quark, 37, 122
Σ, 31–33, 182–184
t (top or truth) quark, 44, 47, 122
τ lepton, 44, 47, 261
u (up) quark, 37, 122
Υ meson, 44, 79, 144, 167–169, 262–263
W, 44–47, 55–56, 301
Ξ, 31, 33, 35, 78, 182–184
Z, 6, 46–47, 55–56, 301, 340
Parton model, 269–273
Partons, 42, 270
See also Gluon; Quark
Pauli, W., 23, 125, 208, 215
Pauli matrices, 115, 139, 216, 334, 351, 379
Pauli principle, 1, 3, 18, 41, 174–176
PCAC hypothesis, 313, 319
Perturbation theory, 59
nonrelativistic, 143, 151
relativistic: *see* Feynman calculus
Perturbative QCD, 279, 295
Phase transformation: *see* Gauge transformation
Phase space, 194–195
ϕ meson, 75, 170, 172, 262–263
Photoelectric effect, 15
Photon, 14–17, 47, 55–57, 225–230
Pion, 18–19, 24–25, 27, 51, 68, 124–125, 211, 314–317
Pion decay constant, 315–316
Planck, M., 14
Planck formula, 14, 90, 149
Planck's constant, 14, 108
Plane waves, 218, 227
Polar vector, 126–127
Polarization, 62–63, 227–229
Polarization vector, 227–229
Positron, 14, 20–21, 217, 230
Positronium, 143, 159–64, 245–246
Potential
Coulomb, 148, 164, 194, 285, 287
four-vector, 226–228
interquark, 165–167, 279, 284–289
linear-plus-coulomb, 165–167
scalar, 226
vector, 226–228
Potential energy term, 361, 363
Powell, C. F., 18–19, 24–25, 29, 31
Primitive vertex: *see* Vertex
Proca equation, 213, 346
Projection matrix, 339
Propagator, 203, 357
electron, 230
gluon, 283
modified, 247
photon, 230
quark, 283
spin zero, 203, 358
spin one-half, 203, 358
spin one, massive, 302, 358
spin one, massless, 230, 359
unstable particle, 329
W and Z, 302, 325, 329
Proper time, 87
Proper velocity, 88
Proton, 4, 13, 29–34, 39–40, 77, 116, 262–275
Pseudoscalar, 126–127, 223–224
Pseudoscalar mesons, 34, 110, 169–172
Pseudovector: *see* Axial vector
ψ meson, 42, 44, 75, 144, 262, 263
See also Charmonium

Q: *see* Charge, electric
Quantum chromodynamics (QCD), 3, 55–56, 60–65, 164, 279–299, 355–357

Quantum electrodynamics (QED), 3, 18, 55–60, 63, 153, 155, 213–255
Quantum field theory, 2, 3, 10, 16, 21, 135, 175, 189, 213, 343, 357
Quantum mechanics, 2, 3, 143–148, 189
Quark
  confinement: *see* Confinement
  distribution functions, 273–276
  masses, 53, 122
  model, 37–44
  number, 74
  search, 39–40
  table, 47, 122
  weak interactions, 317–322
  See also particles
Quark–quark interaction, 284–289
Quarkonium, 160, 162–169
  See also Bottomonium; Charmonium
Quasi-bound state, 162–163, 166

R, 261–263
Rabi, I. I., 29, 149
Radial equation, 147
Radiative corrections, 155
Raising operator, 140
Range, 17, 51–52
Reactor, 4, 26
Real particle, 59–60
Reduced mass, 160
Reflection, 125–126
Regularization, 208
Relativistic correction, 151–152
Relativistic mechanics, 2, 87–99
Relativity, 3, 81–82
Reines, F., 26
Renormalization, 156, 209–210, 246–250, 360
Representation, 107, 116
Resonance, 120, 140, 191, 262–263
ρ meson, 170, 172
Richter, B., 42
Right-handed state, 27, 124, 331–333
Rochester, G. D., 29–30
Rosenbluth formula, 266, 268–269
Rotation, 105–106
Rotation group, 107
  See also Groups
Rotation matrix, 115, 137, 139
Rubbia, C., 46, 327
Running coupling constant: *see* Coupling constant
Rutherford, E., 12–13, 249
Rutherford scattering, 12–13, 40, 193–194, 232–233, 240–241, 309, 313
Rydberg formula, 151

S: *see* Strangeness
s (strange) quark, 37, 122
Salam, A., 56, 323, 330
Scalar, 87, 224
  See also Invariant
Scalar product, 86
Scaling: *see* Bjorken scaling
Scattering, 2
  See also Bhabha scattering; Collisions: Compton; Cross sections; Electron–muon scattering; Electron–positron scattering; Electron–proton scattering; Golden Rule; Møller scattering; Mott scattering; Neutrino–electron scattering; Rutherford scattering
Scattering amplitude, 119
  See also Amplitude
Scattering angle, 191–192
Schrodinger equation, 143–148, 213–214, 217
Schwinger, J. S., 56, 153, 155, 208
Screening, 62–64, 249, 294
Sea: *see* Dirac sea
Sea quarks, 275–276
Semileptonic decay, 67, 318–319
Separation of variables, 145–146
Sextet, 288
Σ, 31–33, 182–184
Simultaneity, 82
Singlet, 112, 118, 158–159
  See also Color, singlet
SLAC, 5, 8, 40, 258, 273, 327
Slash notation, 235–236, 238
$SO(n)$: *see* Groups
Solar neutrinos, 26
Spacelike four-vector, 86
Spectator quark, 68, 70
Spectrum, 148, 150–151, 155, 162–163, 167
Spherical harmonics, 147–148, 186
Spin: *see* Angular momentum
Spin $\frac{1}{2}$, 113–116
Spin matrices, 113–115, 139, 221
Spin and statistics, 175
Spin-averaged amplitude, 236
Spin–orbit coupling, 151, 153–154, 157
Spin–spin coupling: *see* Hyperfine structure
Spinor, 113–115
  See also Dirac spinor
Spontaneous symmetry-breaking, 323, 343, 360, 362–365
Stable particle, 72–73
Standard Model, 3, 46–48, 121–122, 322, 325, 350, 367–368
State, 3
Statistical factor, 195
Stevenson, E. C., 18
Storage rings, 5
Strange particles, 28–33
Strangeness, 32–34, 37–39, 47, 118
Street, J. C., 18
Strength, 55, 76, 78
Strong force, 17, 32–33, 55–56, 279
Structure constants, 282, 297
Structure functions, 264–276
  See also Quark, distribution functions; Form factor
Stückelberg, E. C. G., 21, 29
Subparticles, 48, 53

Sum rules, 276, 278
Supermultiplet, 36, 43, 121, 171–80
  See also Baryon, decuplet; Baryon, octet; Meson, nonet
Su(n): *see* Groups
Symmetric state, 112, 118, 174
Symmetric tensor, 100
Symmetry, 103–141
  See also Gauge invariance; Groups; Invariance

T: *see* Time reversal
t quark, 44, 47, 122
τ lepton, 44, 47, 261
τ number, 47
  See also Conservation laws
τ–θ puzzle, 128
TCP theorem, 3, 135
Tensor, 86–87
θ_C: *see* Cabibbo angle
θ_w: *see* Weak mixing angle
Thomas precession, 153
Thomson, J. J., 11–12
Three-jet event, 258–260
Threshold, 22, 96–97, 101, 261–263
Time dilation, 83
Time reversal, 103, 134–135
Timelike four-vector, 86
Ting, C. C., 42
Top, 44, 47
Top quark: *see t* quark
Toponium, 167
Total cross section, 193
Trace, 237–239
Trace theorems, 238–239, 380
Transformation
  Dirac spinor, 222–223
  four-vector: *see* Lorentz transformations
  tensor, 86–87
  See also Charge, conjugation; Parity; Time reversal
Transition probability, 3
Translation, 105
Transverse gauge: *see* Coulomb gauge
Transverse polarization, 228
Tree-level diagram, 206, 249, 282
Triangle function, 101
Triplet, 112, 118, 158–159
Truth, 44, 47
Two-jet event, 258–259
Two-neutrino hypothesis, 27–28

u quark, 37, 122
U(n): *see* Groups
Uncertainty principle, 51–52, 73
Unification, 31, 76–78, 330–337
  See also: Electroweak force; Grand Unification; GWS theory
Unitary matrix, 106, 351
Units, 8, 9, 200–201, 230, 347
Up quark: *see u* quark
Upness, 47
Υ meson, 44, 79, 144, 167–169, 262–263

V–A interaction, 303, 324, 330
V-events, 29
Vacuum, 362–364
Vacuum polarization, 63, 156, 246, 293
Valence quarks, 275–276
Vector, 224
Vector interaction, 302–303, 309, 313, 324, 330, 332
Vector meson, 110, 169–172
Vector potential, 226–228
Velocity addition rule, 83, 88
Vertex, 73
  ABC, 201
  QCD, 60–61, 280, 282–283, 359
  QED, 56–57, 230, 359
  weak, 65–67, 69–71, 302, 317, 322, 336–337
Vertex factor: *see* Feynman rules; Vertex
Virial theorem, 144
Virtual particle, 58–60, 78, 257–258

W boson, 46–47, 48, 55–56, 301
Ward identity, 250
Wave function, 3, 145, 176, 214, 347
Weak force, 32, 55–56, 123, 301–341
Weak hypercharge, 333–335
Weak interaction, 65–72
  charged, 65, 67, 301–304, 317–322
  neutral, 66, 69, 322–330
Weak isospin, 333–335
Weak mixing angle, 324, 336
Weinberg, S., 56, 323, 330
Weinberg angle: *see* Weak mixing angle
Weyl, H., 125, 350, 354, 356
Work function, 15
Wu, C. S., 123

Ξ, 31, 33, 35, 78, 182–184

Yang, C. N., 56, 122, 128, 350
Yang–Mills theory, 350–355
Yukawa, H., 17–18, 47, 56
Yukawa coupling, 368
Yukawa meson, 14, 17–18, 47, 51, 65
  See also Pion

Z boson, 6, 46–47, 55–56, 301, 340
Zweig, G., 37

PAULI MATRICES

$$\sigma_1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}; \qquad \sigma_2 = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}; \qquad \sigma_3 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$$

$$\sigma_i\sigma_j = \delta_{ij} + i\epsilon_{ijk}\sigma_k; \qquad [\sigma_i, \sigma_j] = 2i\epsilon_{ijk}\sigma_k; \qquad \{\sigma_i, \sigma_j\} = 2\delta_{ij}$$

$$(\mathbf{a}\cdot\boldsymbol{\sigma})(\mathbf{b}\cdot\boldsymbol{\sigma}) = \mathbf{a}\cdot\mathbf{b} + i\boldsymbol{\sigma}\cdot(\mathbf{a}\times\mathbf{b}); \qquad e^{i\boldsymbol{\theta}\cdot\boldsymbol{\sigma}} = \cos\theta + i\hat{\boldsymbol{\theta}}\cdot\boldsymbol{\sigma}\sin\theta$$

DIRAC MATRICES

$$\gamma^0 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}; \qquad \gamma^i = \begin{pmatrix} 0 & \sigma_i \\ -\sigma_i & 0 \end{pmatrix}; \qquad \gamma^5 = i\gamma^0\gamma^1\gamma^2\gamma^3 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$$

$$\{\gamma^\mu, \gamma^\nu\} = 2g^{\mu\nu}; \qquad g^{\mu\nu} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix}$$

Product rules and trace theorems: see Appendix C.

DIRAC EQUATION

$$i\hbar\gamma^\mu\partial_\mu\psi - mc\psi = 0$$

$$(\not{p} - mc)u = 0; \qquad (\not{p} + mc)v = 0$$

$$\bar{u}(\not{p} - mc) = 0; \qquad \bar{v}(\not{p} + mc) = 0$$

$$\bar{\psi} = \psi^\dagger\gamma^0; \qquad \bar{\Gamma} = \gamma^0\Gamma^\dagger\gamma^0; \qquad \gamma^0\gamma^\mu\gamma^0 = \gamma^{\mu\dagger}$$

FEYNMAN RULES

EXTERNAL LINES

Spin 0: (Nothing)

Spin $\frac{1}{2}$:
- Incoming particle: u
- Incoming antiparticle: $\bar{v}$
- Outgoing particle: $\bar{u}$
- Outgoing antiparticle: v

Spin 1:
- Incoming: ϵ^μ
- Outgoing: $\epsilon^{\mu*}$

PROPAGATORS

Spin 0: $$\frac{i}{q^2 - (mc)^2}$$

Spin $\frac{1}{2}$: $$\frac{i(\not{q} + mc)}{q^2 - (mc)^2}$$

Spin 1:
- Massless: $\dfrac{-ig_{\mu\nu}}{q^2}$
- Massive: $\dfrac{-i[g_{\mu\nu} - q_\mu q_\nu/(mc)^2]}{q^2 - (mc)^2}$

Vertex factors: see Appendix D.

PHYSICAL CONSTANTS AND CONVERSION FACTORS

Speed of light	$c = 2.9979 \times 10^{8}$ m/s
Planck's constant	$\hbar = h/2\pi = 1.0546 \times 10^{-34}$ J s $= 6.5822 \times 10^{-22}$ MeV s
Electron charge (magnitude)	$e = 1.6022 \times 10^{-19}$ C $= 4.8032 \times 10^{-10}$ esu
Electron mass	$m_e = 0.511003$ MeV/$c^2 = 9.10953 \times 10^{-31}$ kg
Proton mass	$m_p = 938.280$ MeV/$c^2 = 1.67265 \times 10^{-27}$ kg
Fine structure constant	$\alpha = e^2/\hbar c = 1/137.036$
Classical electron radius	$r_e = e^2/m_e c^2 = 2.8179 \times 10^{-15}$ m
Electron compton wavelength	$\lambda_e = h/m_e c = 2.4263 \times 10^{-12}$ m
Bohr radius	$a = \hbar^2/e^2 m_e = 5.29177 \times 10^{-11}$ m
Bohr energies	$E_n = -m_e e^4/(2\hbar^2 n^2) = -\alpha^2 m_e c^2/2n^2 = -13.6$ eV/n^2

1 Å $= 10^{-10}$ m
1 fm $= 10^{-15}$ m
1 barn $= 10^{-28}$ m^2
1 eV $= 1.60219 \times 10^{-19}$ J
1 MeV/$c^2 = 1.782676 \times 10^{-30}$ kg
1 C $= 2.997925 \times 10^{9}$ esu

PARTICLE DATA

(Mass in MeV/c^2; Lifetime in Seconds; Charge in Units of Proton Charge.)

QUARKS (Spin $\frac{1}{2}$)

	Flavor	Charge	Mass (speculative)		
			Bare	Effective	
				In baryons	In mesons
First generation	d	$-\frac{1}{3}$	7.5	363	310
	u	$+\frac{2}{3}$	4.2		
Second generation	s	$-\frac{1}{3}$	150	538	483
	c	$+\frac{2}{3}$	1100	1500	
Third generation	b	$-\frac{1}{3}$	4200	4700	
	t	$+\frac{2}{3}$	>23,000		

LEPTONS (Spin $\frac{1}{2}$)

	Lepton	Charge	Mass	Lifetime	Principal decays
First generation	e	-1	0.511003	∞	—
	ν_e	0	0	∞	—
Second generation	μ	-1	105.659	2.197×10^{-6}	$e\nu_\mu\bar{\nu}_e$
	ν_μ	0	0	∞	—
Third generation	τ	-1	1784	3.3×10^{-13}	$\mu\nu_\tau\bar{\nu}_\mu$, $e\nu_\tau\bar{\nu}_e$, $\rho\nu_\tau$
	ν_τ	0	0	∞	—

MEDIATORS (Spin 1)

Mediator	Charge	Mass	Lifetime	Force	
gluon	0	0	∞	strong	
photon (γ)	0	0	∞	electromagnetic	electroweak
$W^\pm$	± 1	81,800	unknown	(charged) weak	electroweak
Z^0	0	92,600	unknown	(neutral) weak	electroweak

BARYONS (Spin $\frac{1}{2}$)

Baryon	Quark content	Charge	Mass	Lifetime	Principal decays
$N\begin{cases} p \\ n \end{cases}$	uud udd	$+1$ 0	938.280 939.573	∞ 900	— $pe\bar{\nu}_e$
Λ	uds	0	1115.6	2.63×10^{-10}	$p\pi^-$, $n\pi^0$
Σ^+	uus	$+1$	1189.4	0.80×10^{-10}	$p\pi^0$, $n\pi^+$
Σ^0	uds	0	1192.5	6×10^{-20}	$\Lambda\gamma$
Σ^-	dds	-1	1197.3	1.48×10^{-10}	$n\pi^-$
Ξ^0	uss	0	1314.9	2.90×10^{-10}	$\Lambda\pi^0$
Ξ^-	dss	-1	1321.3	1.64×10^{-10}	$\Lambda\pi^-$
Λ_c^+	udc	$+1$	2281	2×10^{-13}	not established

BARYONS (Spin $\frac{3}{2}$)

Baryon	Quark content	Charge	Mass	Lifetime	Principal decays
Δ	uuu, uud, udd, ddd	$+2, +1, 0, -1$	1232	0.6×10^{-23}	$N\pi$
Σ^*	uus, uds, dds	$+1, 0, -1$	1385	2×10^{-23}	$\Lambda\pi$, $\Sigma\pi$
Ξ^*	uss, dss	$0, -1$	1533	7×10^{-23}	$\Xi\pi$
Ω^-	sss	-1	1672	0.82×10^{-10}	ΛK^-, $\Xi^0\pi^-$, $\Xi^-\pi^0$

PSEUDOSCALAR MESONS (Spin 0)

Meson	Quark content	Charge	Mass	Lifetime	Principal decays
$\pi^\pm$	$u\bar{d}$, $d\bar{u}$	$+1, -1$	139.569	2.60×10^{-8}	$\mu\nu_\mu$
π^0	$(u\bar{u}-d\bar{d})/\sqrt{2}$	0	134.964	8.7×10^{-17}	$\gamma\gamma$
$K^\pm$	$u\bar{s}$, $s\bar{u}$	$+1, -1$	493.67	1.24×10^{-8}	$\mu\nu_\mu$, $\pi^\pm\pi^0$, $\pi^\pm\pi^\pm\pi^\mp$
K^0, $\bar{K}^0$	$d\bar{s}$, $s\bar{d}$	$0, 0$	497.72	K_S^0 0.892×10^{-10} K_L^0 5.18×10^{-8}	$\pi^+\pi^-$, $\pi^0\pi^0$ $\pi e\nu_e$, $\pi\mu\nu_\mu$, $\pi\pi\pi$
η	$(u\bar{u}+d\bar{d}-2s\bar{s})/\sqrt{6}$	0	548.8	7×10^{-19}	$\gamma\gamma$, $\pi^0\pi^0\pi^0$, $\pi^+\pi^-\pi^0$
η'	$(u\bar{u}+d\bar{d}+s\bar{s})/\sqrt{3}$	0	957.6	3×10^{-21}	$\eta\pi\pi$, $\rho^0\gamma$
$D^\pm$	$c\bar{d}$, $d\bar{c}$	$+1, -1$	1869	9×10^{-13}	$K\pi\pi$
D^0, $\bar{D}^0$	$c\bar{u}$, $u\bar{c}$	$0, 0$	1865	4×10^{-13}	$K\pi\pi$
$F^\pm$ (now $D_s^\pm$)	$c\bar{s}$, $s\bar{c}$	$+1, -1$	1971	3×10^{-13}	not established
$B^\pm$	$u\bar{b}$, $b\bar{u}$	$+1, -1$	5271	14×10^{-13}	$D+?$
B^0, $\bar{B}^0$	$d\bar{b}$, $b\bar{d}$	$0, 0$	5275		
η_c	$c\bar{c}$	0	2981	6×10^{-23}	$KK\pi$, $\eta\pi\pi$, $\eta'\pi\pi$

VECTOR MESONS (Spin 1)

Meson	Quark content	Charge	Mass	Lifetime	Principal decays
ρ	$u\bar{d}$, $d\bar{u}$, $(u\bar{u}-d\bar{d})/\sqrt{2}$	$+1, -1, 0$	770	0.4×10^{-23}	$\pi\pi$
K^*	$u\bar{s}$, $s\bar{u}$, $d\bar{s}$, $s\bar{d}$	$+1, -1, 0, 0$	892	1×10^{-23}	$K\pi$
ω	$(u\bar{u}+d\bar{d})/\sqrt{2}$	0	783	7×10^{-23}	$\pi^+\pi^-\pi^0$, $\pi^0\gamma$
ϕ	$s\bar{s}$	0	1020	20×10^{-23}	K^+K^-, $K^0\bar{K}^0$
J/ψ	$c\bar{c}$	0	3097	1×10^{-20}	e^+e^-, $\mu^+\mu^-$, 5π, 7π
D^*	$c\bar{d}$, $d\bar{c}$, $c\bar{u}$, $u\bar{c}$	$+1, -1, 0, 0$	2010	$>1\times10^{-22}$	$D\pi$, $D\gamma$
Υ	$b\bar{b}$	0	9460	2×10^{-20}	$\tau^+\tau^-$, $\mu^+\mu^-$, e^+e^-