

Group Theory Videos by doctorphys

Very Nice Short Lectures on Group Theory

Math for physics videos

doctorphys is a physics professor

<http://www.youtube.com/watch?v=yeK2N7FbkVI>

<http://www.youtube.com/watch?v=KyTdDssSoNw>

http://www.youtube.com/watch?v=K_2n7IiGzn0

<http://www.youtube.com/watch?v=k7mhjythy8g>

<http://www.youtube.com/watch?v=geO7SME8SmM>

Unitary Groups

Groups

Military Group

Matrix Groups

Lie Algebra

Four Lectures on Group Theory by Robert de Mello Koch

Very Nice One Hour Lectures for Physicists

<http://www.youtube.com/watch?v=3wNPrSwbtQ8>

<http://www.youtube.com/watch?v=Zflfm-gU4yA>

http://www.youtube.com/watch?v=5VDZtk_3TiU

<http://www.youtube.com/watch?v=Abw4nP9xEck>

Visualizing Group Theory by Nathan Carter

Fantastic!!! But focused on discrete groups.

"It's one of the best introductions to group theory—or to any branch of higher math—I've ever read."

"I had two students email me and tell me that they picked up your book to thumb through after it came in the mail and suddenly found themselves actually reading the book. Both students indicated that this caught them by surprise because they didn't intend to start reading it and that they had never really read a math textbook before."

<http://www.youtube.com/watch?v=E3hg3nk7gTQ>

<http://www.youtube.com/watch?v=MnRcc9eINaQ>

<http://web.bentley.edu/empl/c/ncarter/vgt/TalkWithoutWords-320x240.mov>

<http://www.youtube.com/watch?v=F2YXeEXV4WY>

<http://web.bentley.edu/empl/c/ncarter/vgt/index.html>

<http://web.bentley.edu/empl/c/ncarter/vgt/media.html>

Download the free software from sourceforge

<http://groupeexplorer.sourceforge.net/>

Evariste Galois and Galois Groups

The famous and tragic story of Evariste Galois

<http://www.youtube.com/watch?v=WRhFW6p832o>

<http://www.youtube.com/watch?v=iikdqV5-H2k>

<http://www.youtube.com/watch?v=J6dsanpnpt0>

<http://www.youtube.com/watch?v=xfJ4vwQ3vpo>

Group Theory for Chemistry

Point Groups and Space Groups for Molecules

Focused on discrete groups.

http://www.youtube.com/watch?v=hRI_4WTLuuc

<http://www.youtube.com/watch?v=Mean9EHUwAc>

<http://www.youtube.com/watch?v=B5VRXbOz-ho>

<http://www.youtube.com/watch?v=W5CCHQWvIdg>

<http://www.youtube.com/watch?v=Uzt5OyF3wNc>

<http://www.youtube.com/watch?v=P7BbHi462lQ>

<http://www.youtube.com/watch?v=oyGFXP8Lxz8>

http://www.youtube.com/watch?v=_FNgo_8GSH4

<http://www.youtube.com/watch?v=nkCudb-wbEU>

<http://www.youtube.com/watch?v=QnnAInt4mLM>

More Math Videos by doctorphys

More math for physics videos

doctorphys is a physics professor

<http://www.youtube.com/watch?v=JQIywFbCeVQ>

<http://www.youtube.com/watch?v=rtsu3CUB6LM>

<http://www.youtube.com/watch?v=rnhmrBo2bF0>

<http://www.youtube.com/watch?v=qApKLj4Y6AA>

<http://www.youtube.com/watch?v=O4IwSBshOkI>

<http://www.youtube.com/watch?v=rwFpqYWRJ7g>

http://www.youtube.com/watch?v=BN_WFnUWr1Q

<http://www.youtube.com/watch?v=QtDQLteKZpA>

http://www.youtube.com/watch?v=_hio6ilv8gw

Fun with Matrices

The Green's Function

Green's Theorem

Poles

The Residue

The Residue Theorem

Complex Integration Example 1

Complex Integration Example 2

Complex Integration Example 3

Group Theory Videos by MathDoctorBob

Math style lectures

<http://www.youtube.com/watch?v=b2Ozprvt8cw>
<http://www.youtube.com/watch?v=LgvmvCMMEg>
<http://www.youtube.com/watch?v=aWxznOOMPsk>
<http://www.youtube.com/watch?v=vwgk49wLMS4>
<http://www.youtube.com/watch?v=hgbnua35tE4>
<http://www.youtube.com/watch?v=le66R31omSU>
<http://www.youtube.com/watch?v=zgjIZ5fZD2A>
<http://www.youtube.com/watch?v=Tvtvz32EpRU>
<http://www.youtube.com/watch?v=nj8jIzWE-Gk>
http://www.youtube.com/watch?v=-_TlxtG1AMQ
<http://www.youtube.com/watch?v=RAPxnZKIWA0>
<http://www.youtube.com/watch?v=yAIVhUZuQRO>
<http://www.youtube.com/watch?v=-fG2xHpaEos>
<http://www.youtube.com/watch?v=tqpBmtfVwVg>
http://www.youtube.com/watch?v=4psS2Gys_SU
<http://www.youtube.com/watch?v=c-VlZK2GLJg>
<http://www.youtube.com/watch?v=6rOVG2ps41Y>
<http://www.youtube.com/watch?v=WaKzx7CSL5c>
<http://www.youtube.com/watch?v=RTEX87QNz00>
<http://www.youtube.com/watch?v=ETx2Yp3PRrE>
http://www.youtube.com/watch?v=bSj8_iilM0o
<http://www.youtube.com/watch?v=wqgEpvjEuLo>
<http://www.youtube.com/watch?v=VYF4uMOvBWI>
<http://www.youtube.com/watch?v=VYF4uMOvBWI>
http://www.youtube.com/watch?v=U_618kB6P1Q
<http://www.youtube.com/watch?v=h9LLlyKA3S4>
<http://www.youtube.com/watch?v=Mix8G-Ye9uI>
<http://www.youtube.com/watch?v=BvZXPFBqUw0>
<http://www.youtube.com/watch?v=qQ4NcrV6u24>

GT1 Definition of a group
GT1 Example of Group Inverse
GT2 Definitions of Subgroup
GT3 Coset's and Lagrange's Theorem
GT4 Normal Groups and Quotient Groups
GT5 Index 2 theorem and Dihedral Groups
GT6 Centralizers, Normalizers, and Direct Products
GT7 The Commutator Subgroup
GT8 Group Homomorphisms
GT9 Group Isomorphisms
GT9 Example of a Group Isomorphism
GT10 Example of Non-Isomorphic Groups
GT11 Group Automorphisms
GT11 Automorphisms of A_4
GT12 $\text{Aut}(Z/n)$ and Fermat's Little Theorem
GT12 Automorphisms of Dihedral Groups
GT13 Groups of Order 8
GT14 Semidirect Products
GT15 Group Actions
GT16 Cayley's Theorem
GT16 Examples of Cayley's Theorem
GT17 Symmetric and Alternating Groups
GT18 Conjugacy and The Class Equation
GT18 Class Equation for Dihedral Groups
GT18 An is Simple
GT19 Cayley's Theorem
GT20 Overview of Sylow Theory
GT20 Sylow Theorems--Proofs
GT20 Sylow Theory for Simple 60

Representation Theory Videos by MathDoctorBob

Math style lectures

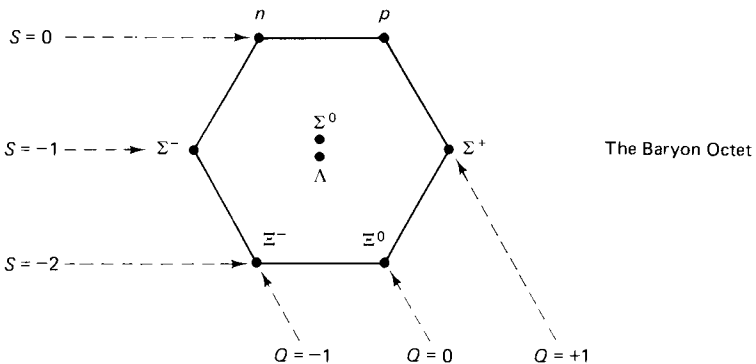
<http://www.youtube.com/watch?v=9rDzaKASMTM>
http://www.youtube.com/watch?v=eJKgN8j_C1Y
<http://www.youtube.com/watch?v=WGmbslwC114>
<http://www.youtube.com/watch?v=6StZcQI0CdE>
<http://www.youtube.com/watch?v=pHH2Z9ZbWJg>
<http://www.youtube.com/watch?v=Uw9QbZ9Fxxw>
<http://www.youtube.com/watch?v=JP038CZQeqM>
<http://www.youtube.com/watch?v=KtmC1hTq6PI>
<http://www.youtube.com/watch?v=LQfwtwPx5w8>

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)

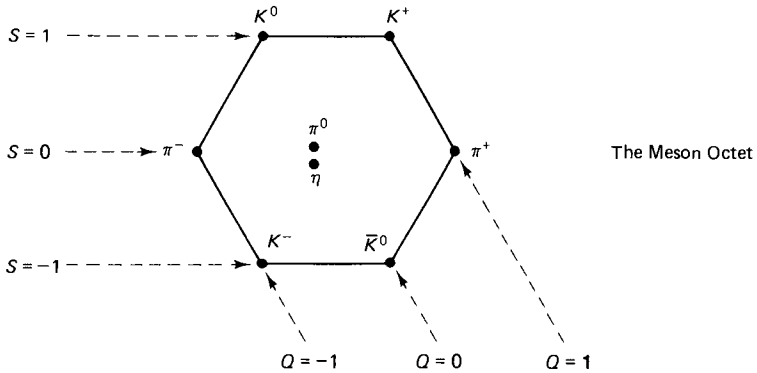
The Mendeleev of elementary particle physics was Murray Gell-Mann, who introduced the so-called *Eightfold Way* in 1961.²¹ (Essentially the same scheme was proposed independently by Ne'eman.) The Eightfold Way arranged the baryons and mesons into weird geometrical patterns, according to their charge and strangeness. The eight lightest baryons fit into a hexagonal array, with two particles at the center:



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

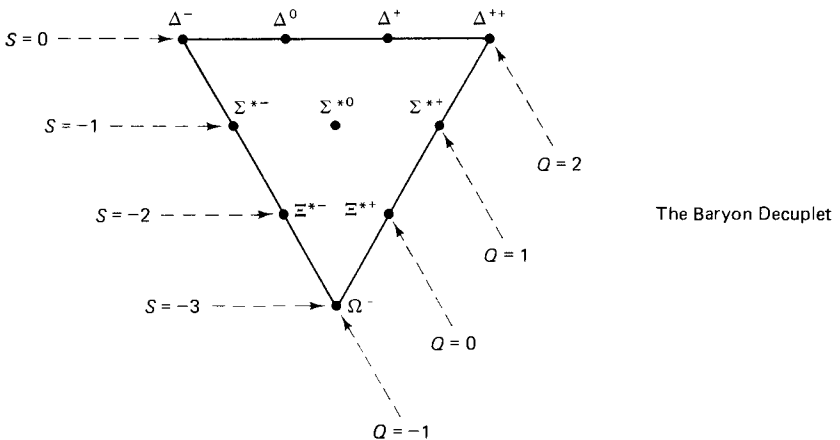
charge) for the proton and the Σ^+ ; $Q = 0$ for the neutron, the lambda, the Σ^0 , and the Ξ^0 ; $Q = -1$ for the Σ^- and the Ξ^- . *Horizontal* lines associate particles of like *strangeness*: $S = 0$ for the proton and neutron, $S = -1$ for the middle line and $S = -2$ for the two Ξ 's.

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



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*:



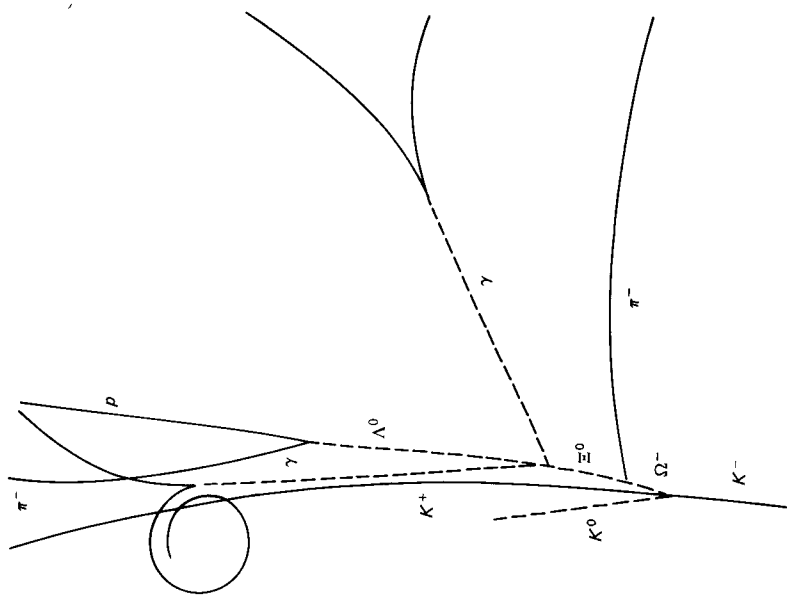
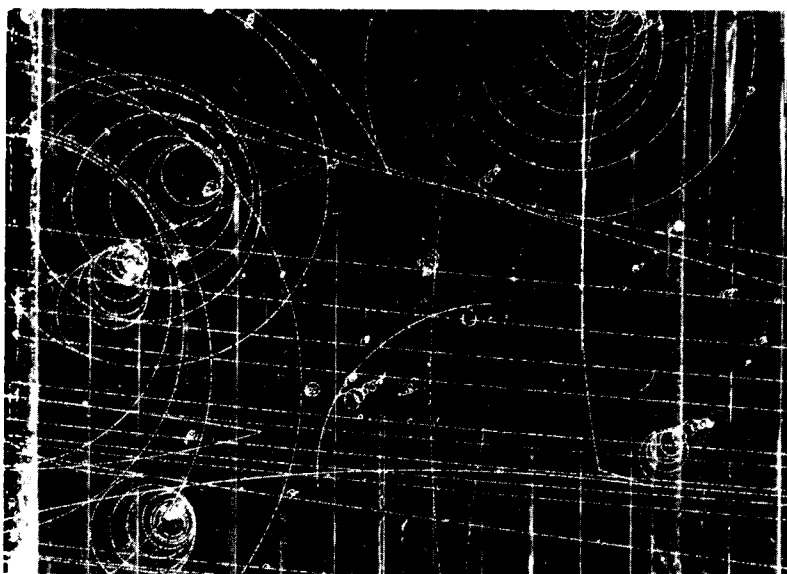


Figure 1.10 The discovery of the Ω^- . The actual bubble chamber photograph is shown on the left; a line diagram of the relevant tracks on the right. (Photo courtesy Brookhaven National Laboratory.)

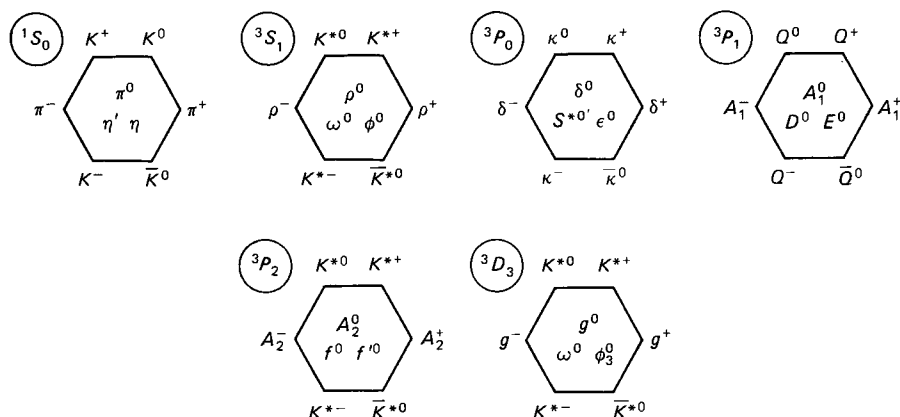


Figure 1.11 Established meson nonets. Obviously, we are running out of letters. It is customary to distinguish different particles represented by the same letter by indicating the mass parenthetically (in MeV/ c^2), thus $K^*(892)$, $K^*(1430)$, $K^*(1650)$, and so on. In this figure the supermultiplets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

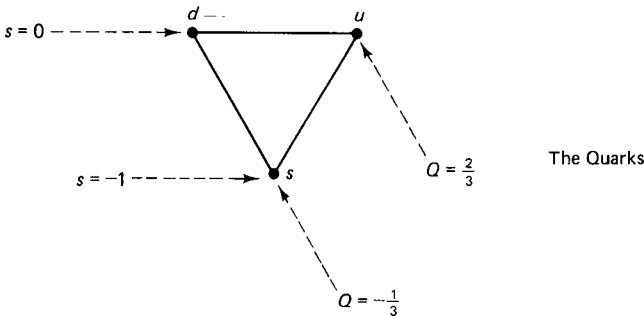
* A similar thing happened in the case of the Periodic Table. There were three famous “holes” (missing elements) on Mendeleev’s chart, and he predicted that new elements would be discovered to fill in the gaps. Like Gell-Mann, he confidently described their properties, and within 20 years all three—gallium,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

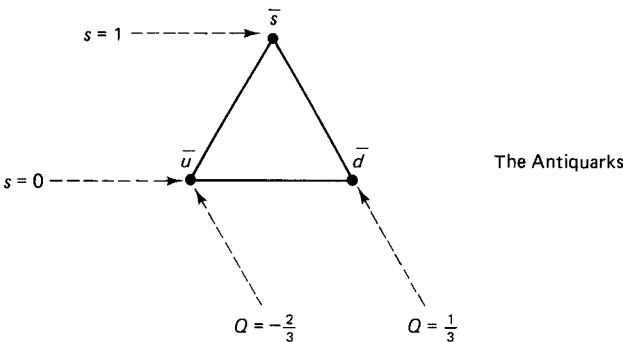
Classification is the first stage in the development of any science. The Eightfold Way did more than merely classify the hadrons, but its real importance lies in the organizational structure it provided. I think it's fair to say that the Eightfold Way initiated the modern era in particle physics.

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 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 *antiquark* (\bar{q}), with the opposite charge and strangeness:



The quark model asserts that

1. Every baryon is composed of three quarks (and every *antibaryon* is composed of three *antiquarks*).
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 *antimesons* 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,

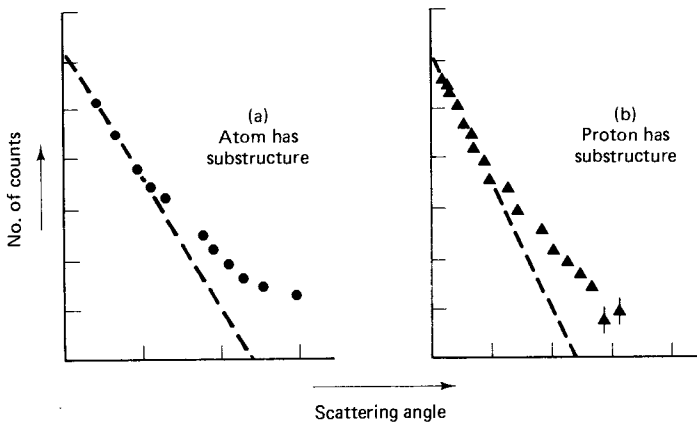


Figure 1.12 (a) In Rutherford scattering the number of particles deflected through large angles indicates that the atom has internal structure (a nucleus). (b) In deep inelastic scattering the number of particles deflected through large angles indicates that the proton has internal structure (quarks). 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

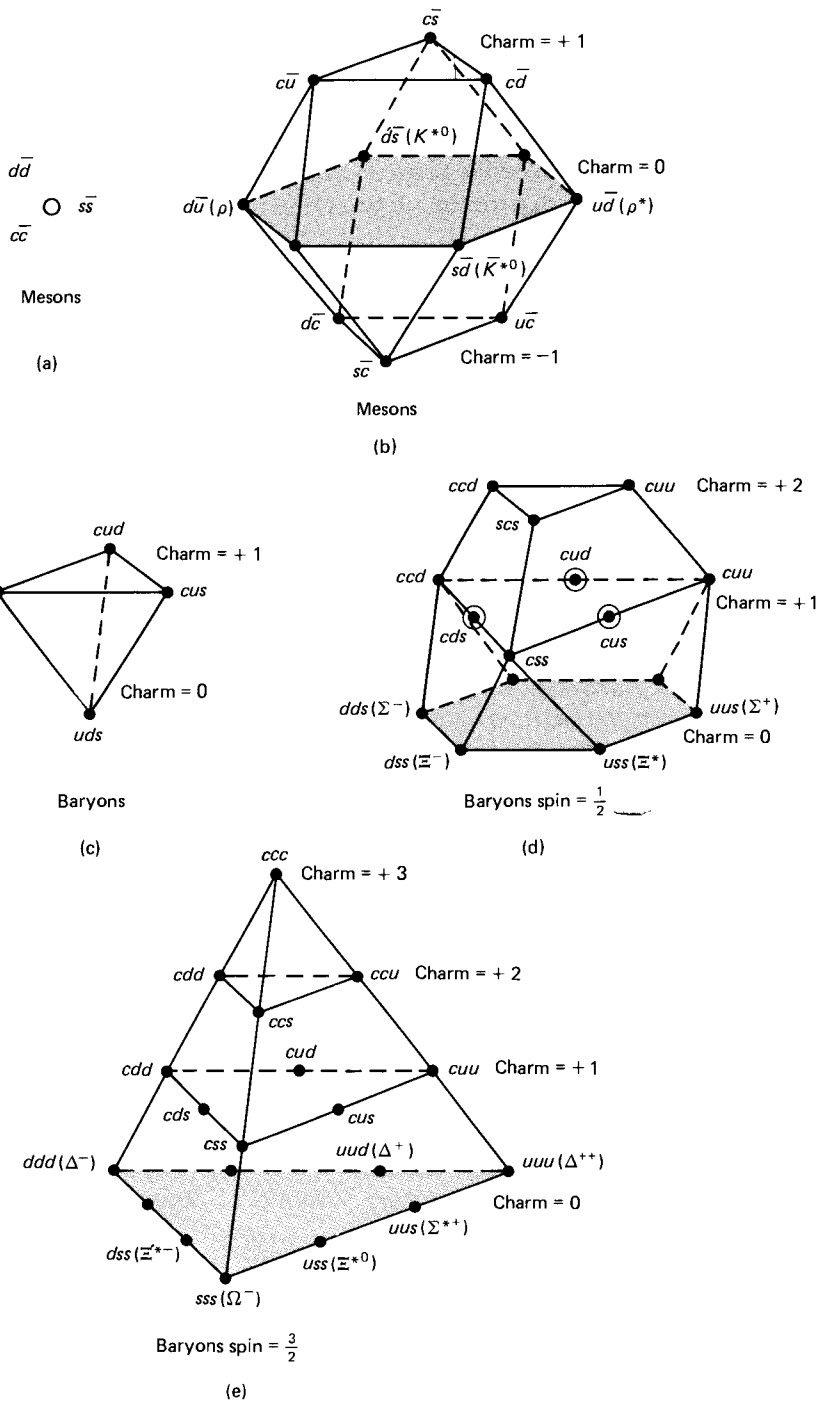


Figure 1.13 Supermultiplets constructed with four quarks. (From "Quarks with Color and Flavor," by S. Glashow. Copyright © Oct. 1975 by Scientific American, Inc. All rights reserved.)

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 Maiani³² 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 despair, 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 1981³⁹ (the claim is hotly contested⁴⁰); 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

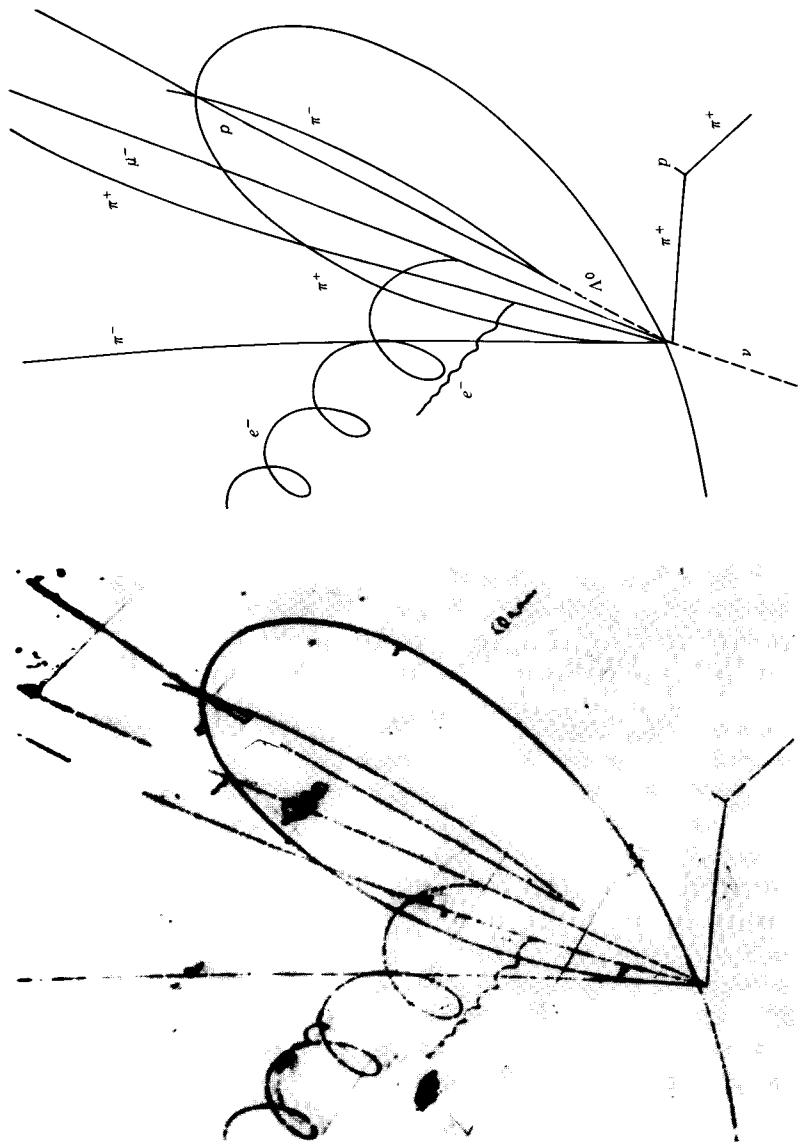


Figure 1.14 The charmed baryon. The probable interpretation of this event is $\nu_\mu + p \rightarrow \Lambda_c^+ + \mu^- + \pi^-$. The charmed baryon decays ($\Lambda_c^+ \rightarrow \Lambda + \pi^+$) too soon to leave a track, but the subsequent decay of the Λ is clearly visible. (Photo courtesy of 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 be⁴²

$$M_W = 82 \pm 2 \text{ GeV}/c^2, \quad M_Z = 92 \pm 2 \text{ GeV}/c^2 \quad (1.30)$$

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 $0.94 \text{ GeV}/c^2$, so we're talking about something nearly 100 times as heavy). In January 1983 the discovery of the W (at $81 \pm 5 \text{ GeV}/c^2$) was reported by Carlo Rubbia's group,⁴³ and five months later the same team announced discovery of the Z (at $95 \pm 3 \text{ GeV}/c^2$).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-

according to their charge (Q), electron number (L_e), muon number (L_μ), and tau number (L_τ). They fall naturally into three *families* (or *generations*):

LEPTON CLASSIFICATION

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

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

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

QUARK CLASSIFICATION

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

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

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

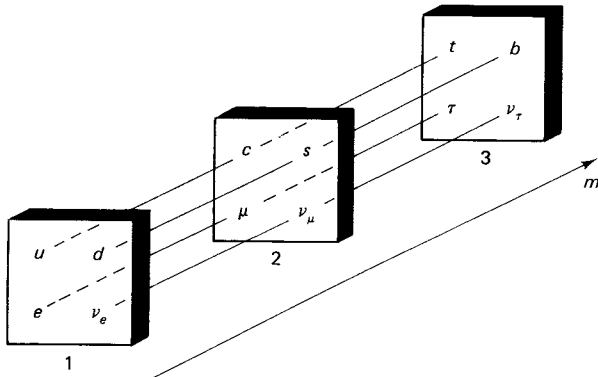


Figure 1.15 The three generations of quarks and leptons, in order of increasing mass.

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

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

REFERENCES AND NOTES

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

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

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

40. D. Drijard et al., *Phys. Lett.* **108B**, 361 (1982).
41. S. Behrends et al., *Phys. Rev. Lett.* **50**, 881 (1983).
42. The formula for the W and Z masses was first obtained by S. Weinberg, *Phys. Rev. Lett.* **19**, 1264, (1967). It involved a parameter θ_w whose value was unknown at that time, and all Weinberg could say for sure was that $M_W \geq 37 \text{ GeV}/c^2$ and $M_Z \geq 75 \text{ GeV}/c^2$. 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 \mathbf{E} and \mathbf{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, $\Delta x \Delta p \geq \hbar$, to estimate the minimum momentum of an electron confined to a nucleus (radius 10^{-13} cm). From the relativistic energy-momentum relation, $E^2 - \mathbf{p}^2 c^2 = m^2 c^4$, 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 $\Sigma \rightarrow \pi$, $\Lambda \rightarrow \eta$, 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, $\Delta^{++} \rightarrow p^+ + \pi^+$. 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.)

10

SU(3) and the quark model: classification and dynamic probes

10.1 Introduction

The discovery of such a wealth of apparently ‘elementary’ particles stimulated new activity in the search for a pattern amongst them, as a first step towards the understanding of their nature. The discovery of such a pattern is analogous to, for instance, the discovery of the Rydberg formula in atomic spectroscopy. The Bohr atom finally provided an explanation of the formula, and we shall see that the quark model provides an explanation of the symmetry pattern of the elementary particles.

We have already become familiar with the limited symmetry of isotopic spin multiplets. In that case we grouped together particles which were the same except for properties associated with the electric charge. The degeneracy of the multiplet is removed by the symmetry-breaking Coulomb interaction. Alternatively, we can regard the members of the multiplet as states linked by rotations in isotopic spin space and we can define a group of rotation operators which enable us to step from one state to another.

The Coulomb interaction is not strong compared with the so called ‘strong’ interactions, and the symmetry breaking to which it gives rise is small. For instance, the masses of particles in the same isotopic spin multiplet differ only by at most a few per cent. In order to extend the

symmetry, to group larger numbers of particles together, we must recognise the existence of much stronger symmetry breaking forces since the mass differences between, say, I -spin multiplets, are substantial, even compared with the particle masses themselves. Nevertheless it turns out to be true that the symmetry, though broken, remains in many ways very useful.

10.2 Baryon and meson multiplets

In attempting to group together different I -spin multiplets we may seek to group particles having the same baryon number, spin and parity, but allow the strangeness (or equivalently the hypercharge) to vary within the multiplet. This hypothesis has the merit of success over others, such as allowing spin and parity to vary within a multiplet and demanding the same strangeness, which might *a priori* appear equally reasonable.

For I -spin multiplets, we may represent the members of the multiplet as points spaced at unit intervals on the I_3 -axis. The raising and lowering operators allow steps to the right and left along the axis. For instance, for the $\Delta(1232)$ multiplet we have the configuration shown in fig. 10.1. In order to extend the classification to include other I -spin multiplets of the same J^P and B we need to move from a one- to a two-dimensional diagram, where the axes are I_3 and Y . We first take the $J^P = \frac{1}{2}^+$ baryons, of which eight were known when the classification was first proposed. The original $\frac{1}{2}^+$ octet consists of p, n, Σ^+ , Σ^0 , Σ^- , Λ^0 , Ξ^0 and Ξ^- ; two I -spin doublets, a triplet and a singlet. The diagram for this octet is shown in fig. 10.2. The octet forms a hexagonal pattern on the Y - I_3 diagram and it is clear that in order to make transitions between states we need operators which effect diagonal steps as well as the step operators within the I -spin multiplet.

Many other particles also readily fall into similar patterns. In particular, for the pseudo-scalar, vector and tensor mesons we find nonets which are shown on the Y - I_3 plot in fig. 10.3. In the meson nonets, we notice that, unlike the baryon case, particle and antiparticle appear in the same multiplet, since for both particle and antiparticle B is the same.

10.3 Symmetry groups

Having noted the existence of these multiplets we may seek a classification to describe them. The SU(3) group was proposed for this purpose in 1961 by Gell-Mann and independently by Ne'mann.

Some of the formal properties of groups are summarised below, though the reader wishing to pursue this aspect of the subject more deeply is recommended to consult one of the several excellent texts such as *Lie Groups for Pedestrians* (Lipkin, 1965). In this section we use some of these properties to bring out the principal physical ideas. The idea of examining

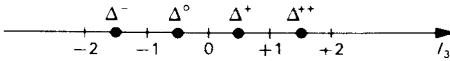


Fig. 10.1. The members of the $\Delta(1232)$ isotopic spin multiplet plotted on the I_3 axis.

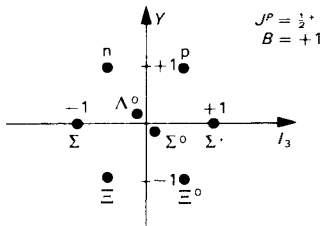


Fig. 10.2. The $J^P = \frac{1}{2}^+$, $B = +1$ octet plotted on the $Y-I_3$ plane.

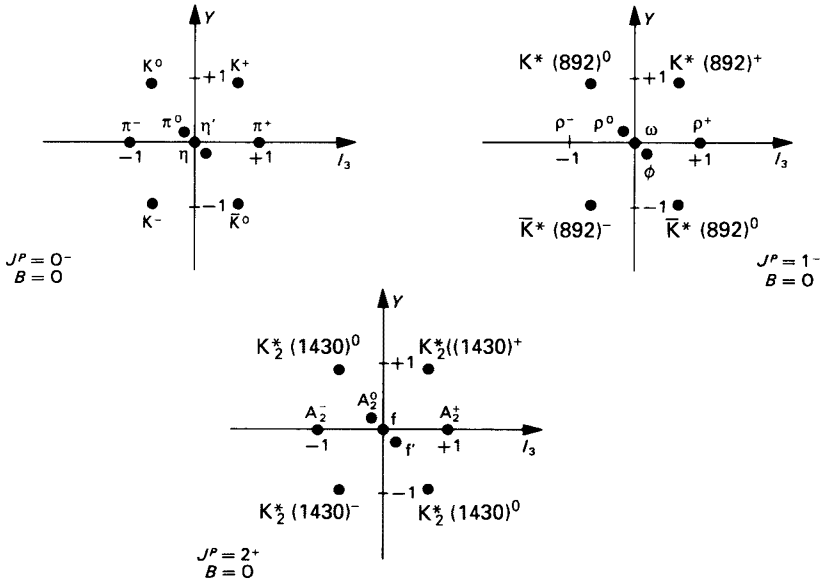


Fig. 10.3. The pseudo-scalar, vector and tensor meson nonets plotted on the $Y-I_3$ plane.

the symmetry of any system by ‘rotating’ it in the appropriate space is a familiar one. We accomplish this by means of ‘rotation’ operators, in the most general sense of the term ‘rotation’. The set of rotation operators is said to form a symmetry group. For instance, on functions of geometric coordinates we have the translation, space-rotation and inversion groups. These operators act on the wave functions.

The most familiar example to compare with the operators we shall require is the space-rotation group. These operators act on the angular coordinates of the wave function $\psi(\theta_1, \theta_2, \theta_3) = \psi(\theta)$. They have the form

$$R(\theta) = e^{i\theta \cdot \mathbf{J}}.$$

Here, the θ are the rotation angles and \mathbf{J} the angular momentum. These operators form an infinite group.

Now, if we have a limited space of states such that under the operators of the group each state acted on by an operator of the group transforms into another state in this space, then we have an invariant subspace of states. For such a set of states, an operator of the group cannot connect a state within the set to a state outside. For instance, the set of spherical harmonic functions $Y_{0,0}, Y_{1,1}, Y_{1,0}, Y_{1,-1}$ form a four dimensional invariant subspace of the rotation group. The subspace is said to be irreducible if it contains no smaller subspace. Thus, pursuing the above example, we see that:

$Y_{0,0}$ is an irreducible singlet,
 $Y_{1,1}, Y_{1,0}, Y_{1,-1}$ is an irreducible triplet.

The conventional way of expressing this division is to write

$$(Y_{0,0}, Y_{1,1}, Y_{1,0}, Y_{1,-1}) = Y_{0,0} \oplus (Y_{1,1}, Y_{1,0}, Y_{1,-1}).$$

An irreducible invariant subspace is known as a multiplet of the operator group.

A general theorem can be proved which states that if the Hamiltonian for the system commutes with the operators of the symmetry group then the multiplets are sets of degenerate eigenstates of the system. We know that for the elementary particles the states are not degenerate, so that the multiplet is split. This is due to additional terms in the Hamiltonian which do not commute with the group operators. An example of terms which cause splitting for the rotation group is, for instance, the $L \cdot S$ coupling which splits the levels in atomic spectra.

In the space rotation group there is an infinite number of operators, but we can deal with this infinite group in a simple way because all the

operators can be expressed in terms of only three basic operators, J_1 , J_2 and J_3 . If $R(\theta)$ depends analytically on θ , at least near $\theta = 0$ and thus near $R(\theta) = 1$, then we can relate the J s to R by means of the relation

$$iJ_k = \left[\frac{\delta R(\theta)}{\delta \theta_k} \right]_{\theta=0}.$$

A Lie group is a continuous group composed of the operators $U(\alpha) = U(\alpha_1, \dots, \alpha_n)$ which depends analytically on all the n parameters and for which $U(0) = 1$. As in the particular case above, we can extract basic operators

$$iG_i = \left. \frac{\delta U}{\delta \alpha_i} \right|_{\alpha=0}$$

which are called the generators of the group but which are not themselves members of the group. The dimension of the group is given by the number of the generators.

We have already seen that the larger multiplets for particles contain the I -spin multiplets. Suppose, therefore, that we enlarge a Lie group by adding more coordinates for the group to transform. In this case we have to add more operators to the group.

An idea of fundamental importance in the present application of group theory is that of the basic or elementary multiplet. For a Lie group of rank l there exist l such elementary multiplets, and from these can be built up more complicated multiplets by taking repeated 'products' between the basic multiplets. A simple example from the rotation group is the triplet deuterium ground state consisting of spin $\frac{1}{2}$ neutron and proton. The spin $\frac{1}{2}$ particles are members of two-member multiplets or doublets of the rotational group for angular momentum $\frac{1}{2}$ (the generators of which are the Pauli spin matrices). The product of two such doublets then yields triplet and singlet multiplets according to $2 \otimes 2 = 3 \oplus 1$. This kind of decomposition of the product of two multiplets is a process which is familiar under another guise in angular momentum theory as the Clebsch–Gordan expansion

$$Y_{l_1 m_1}(\theta, \phi) Y_{l_2 m_2}(\theta, \phi) = \sum_{L=|l_1-l_2|}^{l_1+l_2} \sum_{M=-L}^{+L} C_{m_1 m_2 M l_1 l_2 L} Y_{LM}(\theta, \phi).$$

The form of the decomposition or the values of the Clebsch–Gordan coefficients must be worked out from the commutation relations of the generators.

Thus for the Simple Unitary (Lie) group of order 3, SU(3), the elementary multiplet is a triplet, and the multiplet products can be shown to decompose as, for instance,

$$3 \otimes 3 \otimes 3 = 10 \oplus 8 \oplus 8 \oplus 1.$$

10.4 The SU(3) classification and the quark model

For SU(2) there are only two basic states, that is, the basic multiplet is a doublet such as the p, n *I*-spin multiplet. In order to include states with non-zero hypercharge, it is necessary to move to a larger group of order 3, such as the group SU(3) for which the basic multiplet is a triplet. The state to be added to the SU(2) doublet is then a state with non-zero hypercharge and *I* = 0. The strangeness and isotopic spin quantum numbers of the triplet are then the same as p, n, Λ .

An early model in which it was indeed proposed that all the known particles were built from the physical proton, neutron and lambda particles was proposed by Sakata in 1956. However, this model quite rapidly ran into some problems, such as the parity of the Σ , which was on this model composed of $N\bar{N}\Lambda$ (e.g. $\Sigma^+ = p\bar{n}\Lambda$) in mutual S-states, so that $J^P = \frac{1}{2}^-$ and not $\frac{1}{2}^+$ as is experimentally found to be the case. The basic SU(3) multiplet is now referred to as the SU(2) doublet u, d plus the non-zero hypercharge state s. These states are now identified with the *quarks* from which all the particles are built. It has often been stressed that the SU(3) classification does not depend on the actual existence of particles corresponding to the basic multiplet. We shall return to the question of existence of the quarks below. For the present we merely use a basic triplet of states, equivalent under the strong interaction, to construct the pattern of the observed particles.

In order to find the patterns we require the set of operators forming the group SU(3) which will transform the basic triplet states. For SU(3) the operators may be represented by a group of unimodular, unitary 3×3 matrices in a similar way to the Pauli spin matrices for SU(2). There are $3 \times 3 - 1$ such independent traceless matrices. However, the proper treatment of these eight operators requires a more detailed knowledge of group theory than can be presented here.

We rather follow the treatment of Lipkin (1965) to derive the operators in terms of bilinear products of the creation and annihilation operators for u, d and s. These operators are denoted $a_u^\dagger, a_d^\dagger, a_s^\dagger$ and a_u, a_d, a_s respectively.

The technique is best illustrated by application first to the simplest case of the *I*-spin doublet u, d. Only bilinear products which do not change the

baryon number can be permitted, so that in this case there are four such products:

- $a_u^\dagger a_d$ – changes d to u \equiv step up operator = τ_+ ;
- $a_d^\dagger a_u$ – changes u to d \equiv step down operator = τ_- ;
- $a_u^\dagger a_u$ – counts ‘ups’ (i.e. the number of protons is given by the eigenvalue);
- $a_d^\dagger a_d$ – counts ‘downs’.

In order to include states with $S \neq 0$ we include the a_s^\dagger, a_s , obtaining:

- $a_s^\dagger a_u$ – changes u to s = C_- ;
- $a_s^\dagger a_d$ – changes d to s = C_+ ;
- $a_u^\dagger a_s$ – changes s to u = B_+ ;
- $a_d^\dagger a_s$ – changes s to d = B_- .

Together with the first two we thus have six step operators:

- τ_+ and τ_- leave Y unchanged and change I_3 by ± 1 ;
- B_+ and B_- change Y by $+1$ and I_3 by $\pm \frac{1}{2}$;
- C_+ and C_- change Y by -1 and I_3 by $\pm \frac{1}{2}$.

The operations are illustrated in fig. 10.4. Such a set of step operators satisfies the rule that operation on any state transforms it into another state of the multiplet or annihilates it. These operators require a hexagonal type symmetry for all the multiplet diagrams. This is due to the fact that instead of the single I_3 symmetry axis in SU(2) there are now three symmetry axes at 120° to each other.

Along the ‘ I -spin’ axis the hypercharge is constant but the charge varies. The step operators for the basic multiplet change $d \leftrightarrow u$. Along the B_-C_+ axis the charge Q remains constant, Y varies and the step operators change $d \leftrightarrow s$. This is known as the ‘ U -spin’ axis. If invariance under I_3

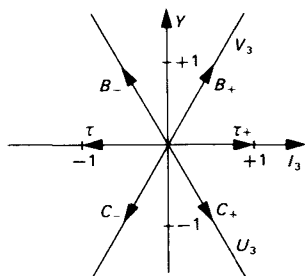


Fig. 10.4. Step operators in the $Y-I_3$ plane.

changes was exact then the properties of a system would be independent of its charge. If invariance under U -spin was true the properties would be independent of the hypercharge. The third axis, corresponding to $u \leftrightarrow s$ transformations, has no such simple interpretation. It is sometimes known as the V -spin axis.

As has already been mentioned, in $SU(3)$ there are eight independent operators of which we have discussed the six step operators formed from bilinear combinations of the creation and annihilation operators for u, d, s . It is convenient to choose the other two operators as

$$\begin{aligned} \tau_0 &= \frac{1}{2}(a_u^\dagger a_u - a_d^\dagger a_d) \\ N &= \frac{1}{3}(a_u^\dagger a_u + a_d^\dagger a_d - 2a_s^\dagger a_s). \end{aligned}$$

By writing the charge, strangeness and baryon number operators which count these quantities as Q, S and B we see that

$$\begin{aligned} Q &= \frac{1}{3}(2a_u^\dagger a_u - a_d^\dagger a_d - a_s^\dagger a_s) \\ S &= -a_s^\dagger a_s \\ B &= \frac{1}{3}(a_u^\dagger a_u + a_d^\dagger a_d + a_s^\dagger a_s) \end{aligned}$$

so that using $Q = I_3 + \frac{1}{2}(B + S)$ we have that operation with τ_0 gives the I_3 eigenvalue. Also

$$N = B + S = Y.$$

None of these operators changes B .

For these eight operators the basic triplet and antitriplet are shown in fig. 10.5. The operators transform the basic states into each other. By combining these basic triplets nine states are obtained, which can be generated from the vacuum by the products of creation and annihilation operators. This is shown in fig. 10.6, where we have omitted the states at the centre generated by the remaining three products or by operators such as τ_0, B and N . The nine states split into an octet plus a singlet. The singlet has the quantum numbers of the vacuum, while the octet also has two states with $I_3 = Y = 0$. As far as the $SU(3)$ algebra is concerned these two

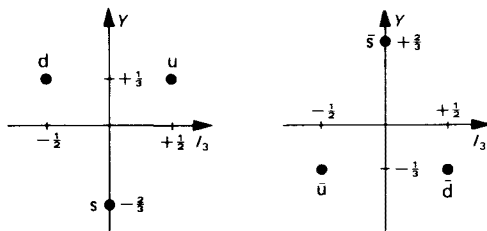


Fig. 10.5. The basic $SU(3)$ triplet (u, d, s) and antitriplet ($\bar{u}, \bar{d}, \bar{s}$).

octet states are degenerate, but we note that the four $Y = 0$ states must include the $I_3 = 0$ member of the I -spin triplet. Thus the two states at the origin have $I = 1$ and $I = 0$.

From this analysis we see that in the simplest model all the mesons ($B = 0$) can be constructed from quark-antiquark ($q\bar{q}$) pairs.

In a similar way the baryons and antibaryons may be constructed from three quark (qqq) or three antiquark ($\bar{q}\bar{q}\bar{q}$) combinations. We shall return to this aspect of the model in section 10.7.

A consequence of the nature of the step operators is that the multiplets are all hexagonal lattices in the Y - I_3 plane, having the following general properties:

(a) All diagrams are symmetrical about the Y -axis ($+I_3 \leftrightarrow -I_3$) corresponding to such transformations as $d \leftrightarrow u$ and also about axes corresponding to $u \leftrightarrow s$ and $d \leftrightarrow s$ transformations.

(b) The multiplicity of points at any coordinate increases by one at each 'ring' as one moves in from the boundary, until one arrives at a point, or triangle, inside which triangle it remains constant.

(c) Charge conjugation changes the sign of Y and I_3 . Thus the charge-conjugate multiplet is the original reflected through the origin.

SU(3) gives, in addition to the multiplets we have already discussed, others, such as those formed by combining two octets or three triplets:

$$8 \otimes 8 = 27 \oplus 10 \oplus \bar{10} \oplus 8 \oplus 8 \oplus 1$$

$$3 \otimes 3 \otimes 3 = 10 \oplus 8 \oplus 8 \oplus 1.$$

On the Y - I_3 diagram the decuplet and 27-plet have the forms shown in fig. 10.7, where the quark content of the states in the decuplet is also indicated.

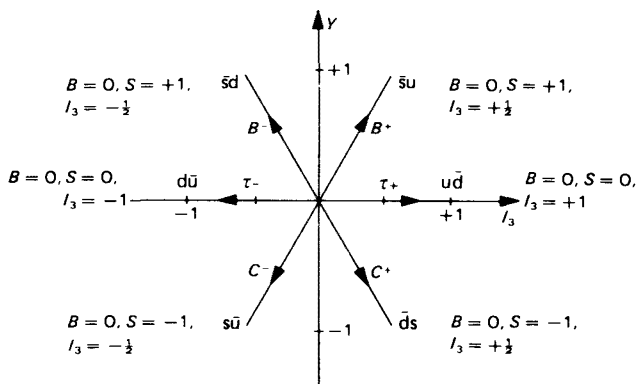


Fig. 10.6. States obtained by combining the basic triplet and antitriplet.

10.5 **Prediction of the Ω^-**

We might expect the u- and d-quarks to be very similar in mass since they are members of the same *I*-spin doublet and also since we know that the neutron and proton built from different combinations of u- and d-quarks are very close in mass. The symmetry breaking which is evident in the SU(3) multiplet may, however, be manifest in a difference in mass between the u- and d-quarks, on the one hand, and the s-quark. We shall see that other effects arising from quark-quark interactions can also give rise to mass differences in their combinations but for the $J = \frac{3}{2}$ decuplet we may note that the situation is particularly simple in that

(a) no space on the diagram is occupied by more than one particle, so that possible ‘mixing’ problems are avoided;

(b) as we move down from one *I*-spin multiplet to another the number of s-quarks increases at each step.

If in this case the mass differences between *I*-spin multiplets are simply due to the difference in mass between the u-, d- and s-quarks then we expect an *equal mass spacing rule*.

In 1962 there were known only nine members of the $J^P = \frac{3}{2}^+$ -multiplet with the space at the bottom vacant (fig. 10.8). The Ω^- proposed by Gell-Mann (and independently by Ne’eman) as the missing member has the quantum numbers:

$$B = +1, \quad J^P = \frac{3}{2}^+, \quad Y = -2, \quad S = -3, \quad I = 0, \quad I_3 = 0, \\ Q = -1.$$

Applying the equal mass spacing rule we have:

$$m_{\Omega} - m_{\Xi(1530)} = m_{\Xi(1530)} - m_{\Sigma(1385)} \\ = m_{\Sigma(1385)} - m_{\Delta(1232)}.$$

The best values for the latter two differences were

$$m_{\Xi(1530)} - m_{\Sigma(1385)} = 147 \pm 1.5 \text{ MeV}/c^2 \\ m_{\Sigma(1385)} - m_{\Delta(1232)} = 146 \pm 1.4 \text{ MeV}/c^2.$$

Thus the mass of the Ω^- was predicted to be

$$m_{\Xi(1530)} + 146 = 1675 \text{ MeV}/c^2.$$

With such properties the only $\Delta S = 1$ decays which are kinematically allowed are

$$\Omega^- \rightarrow \begin{cases} \Lambda K^- \\ \Xi^- \pi^0 \\ \Xi^0 \pi^- \end{cases}$$

The discovery of the Ω^- , its decay by the above modes and the measurement of its mass, the best present value of which is

$1672.43 \pm 0.32 \text{ MeV}/c^2$ have already been described in section 5.10 and represented a triumph for the SU(3) scheme. The Ω^- should have spin $\frac{3}{2}$ and this, too, has been confirmed in measurements in recent experiments. In one of these, some five million photographs of $8.25 \text{ GeV}/c$ K^- -mesons were taken in the CERN 2 m hydrogen bubble chamber to produce observed Ω^- particles. In another remarkable experiment at CERN a beam of hyperons was created by collisions of very high energy protons from the SPS with a primary target. Even short lived particles like Ξ^- and Ω^- may survive passage down a long beam line to the detection apparatus if their energy is high so that the lifetime is relativistically dilated. In both cases the decay angular distribution of the Ω^- is characteristic of a spin $\frac{3}{2}$ particle.

10.6 Mass formulae and mixing

The masses of the states within a multiplet are (approximately) related by the Gell-Mann–Okubo mass formula

$$m = m_0 + m_1 Y + m_2 (I(I+1) - \frac{1}{4} Y^2). \quad 10.1$$

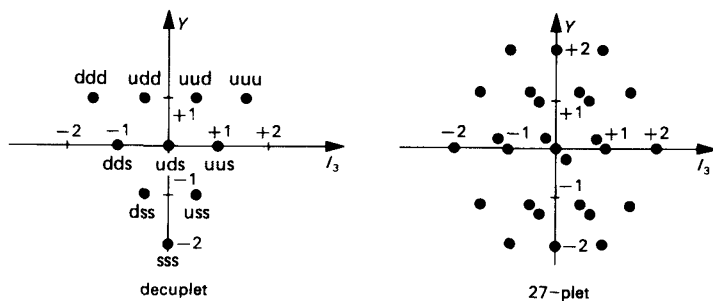


Fig. 10.7. SU(3) decuplet and 27-plet states on the Y - I_3 plane.

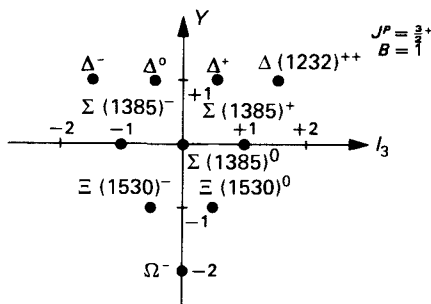


Fig. 10.8. The $J^P = \frac{3}{2}^+$, $B = +1$ decuplet on the Y - I_3 plane.

This formula was originally based on a speculative argument concerning the nature of the forces responsible for the symmetry breaking. We prefer here to treat the formula as an empirical rule which will in due course be accounted for by quantum chromodynamics (see chapter 13). In the baryon decuplet

$$Y = B + S = 2(I - 1)$$

so that **10.1** gives

$$m = (m_0 + 2m_2) + Y(m_1 + \frac{3}{2}m_2)$$

giving equal mass spacing. In the $\frac{1}{2}^+$ octet **10.1** gives

$$2m_n + 2m_{\Sigma^0} = m_{\Sigma^+} + 3m_{\Lambda}$$

while the measured masses yield 4515 and 4539 MeV/ c^2 for the left- and right-hand sides of this equation so that the relation is seen to be moderately well satisfied.

We might expect that the Gell-Mann–Okubo formula could be applied to the meson octets. For mesons the field equations always involve the squares of the masses, so that it is plausible that the formula here should be written with mass-squared instead of mass as in the baryon case. A test of whether mass or mass-squared is the correct quantity is complicated by the phenomenon of mixing, which we shall discuss below.

If we write the Gell-Mann–Okubo formula **10.1** in a general form for mesons where the subscripts $\frac{1}{2}$ and 1 refer to the I -spin and where m_8 is the mass of the I -spin zero member of the octet, then

$$m_{\frac{1}{2}}^2 = \frac{1}{4}m_1^2 + \frac{3}{4}m_8^2, \tag{10.2}$$

since for the mesons particle and antiparticle appear reflected about the I_3 axis.

We have already noted that the mesons appear to be grouped in nonets rather than octets. In fact, the $I = 0$ member of the SU(3) octet has quantum numbers identical to those of the SU(3) singlet, so that if there is SU(3) breaking we might expect mixing between these particles. We will denote the true SU(3), $I = 0$, $Y = 0$, octet and singlet particles as α_8 and α_0 respectively. However, the observed physical particles may be (specific) mixtures of the α_8 and α_0 and the multiplet containing these mixtures will have nine members. The idea of mixing was proposed by Sakurai (1962), who proposed also a mixing parameter in the form of the ‘mixing angle’. This parameterisation makes it easy to preserve the normalisation and has other formally satisfying features. The situation is very similar to that of the $K^0\bar{K}^0$ particles.

Formally both the $K^0\text{--}\bar{K}^0$ and octet–singlet mixing can be treated rather

similarly by means of the 'mass matrix'. We shall not attempt here to develop the treatment of the mass matrix fully or rigorously, but merely introduce the idea and use it to obtain an expression for the mixing angle.

First note that we expect to obtain the mass m_8 of the pure octet state from the expression 10.2. For instance, in the vector meson nonet we should have

$$m_8^2 = \frac{4}{3}m_{K^*(892)}^2 - \frac{1}{3}m_\rho^2 = 0.863 \quad (\text{GeV}/c^2)^2.$$

In this nonet the physical singlet particles are the ω and the ϕ for which the mass-squared values are 0.610 and 1.035 $(\text{GeV}/c^2)^2$ respectively.

Now we write the physical particle wave functions β and β' , say, in terms of the Sakurai mixing parameter

$$\begin{aligned} |\beta\rangle &= |\alpha_0\rangle \cos \theta + |\alpha_8\rangle \sin \theta \\ |\beta'\rangle &= -|\alpha_0\rangle \sin \theta + |\alpha_8\rangle \cos \theta. \end{aligned} \quad 10.3$$

For a stable stationary state, the solutions of the wave equation contain a factor e^{im_0t} , and the mass of the particle is the eigenvalue of the total Hamiltonian for the interaction. Thus

$$\begin{aligned} \langle \beta | H_0 | \beta \rangle &= m_\beta^2 \quad (\text{e.g. } m_\omega^2) \\ \langle \beta' | H_0 | \beta' \rangle &= m_{\beta'}^2. \end{aligned} \quad 10.4$$

If there is no mixing then

$$\langle \beta | H_0 | \beta' \rangle = \langle \beta' | H_0 | \beta \rangle = 0,$$

but if some part of H_0 mixes the basic states this will no longer be true.

For the no-mixing case we could write a two-component time-dependent Schrödinger equation of the form

$$i \frac{d}{dt} \begin{bmatrix} \beta \\ \beta' \end{bmatrix} = \begin{bmatrix} m_\beta^2 & 0 \\ 0 & m_{\beta'}^2 \end{bmatrix} \begin{bmatrix} \beta \\ \beta' \end{bmatrix}$$

where in this case the mass matrix is diagonal.

If there is mixing, the off-diagonal terms are no longer zero and we have

$$i \frac{d}{dt} \begin{bmatrix} \alpha_0 \\ \alpha_8 \end{bmatrix} = \begin{bmatrix} m_0^2 & m_{0,8}^2 \\ m_{8,0}^2 & m_8^2 \end{bmatrix} \begin{bmatrix} \alpha_0 \\ \alpha_8 \end{bmatrix} \quad 10.5$$

where m_0 and m_8 are the masses of the I -spin zero members of the SU(3) singlet and octet multiplets and $m_{0,8}$ and $m_{8,0}$ are due to mixing. The mass matrices corresponding to the representation of the states in terms of the physical particles and the pure SU(3) states must satisfy the conditions that their traces and determinants are equal:

$$\begin{aligned} m_0^2 + m_8^2 &= m_\beta^2 + m_{\beta'}^2 \\ m_0^2 m_8^2 - m_{0,8}^4 &= m_\beta^2 m_{\beta'}^2 \end{aligned}$$

($m_{0,8}$ has been taken equal to $m_{8,0}$).

The two representations must be linked by a rotation which is the operator

$$R = \begin{bmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{bmatrix}.$$

Thus θ is the angle of rotation which will diagonalise the mass matrix

$$M = \begin{bmatrix} m_0^2 & m_{0,8}^2 \\ m_{8,0}^2 & m_8^2 \end{bmatrix}$$

of equation **10.5**. Imposing this condition in the form

$$R^{-1}MR = \begin{bmatrix} m_\beta^2 & 0 \\ 0 & m_\beta^2 \end{bmatrix}, \tag{10.6}$$

straightforward algebra shows that the condition that the off-diagonal elements in RMR^{-1} be zero is

$$\tan 2\theta = \frac{m_{0,8}^2 + m_{8,0}^2}{m_8^2 - m_0^2}. \tag{10.7}$$

10.3 and **10.4** or consideration of the diagonal elements of **10.6** yield

$$\begin{aligned} m_\beta^2 &= m_0^2 \cos^2 \theta + m_8^2 \sin^2 \theta - (m_{0,8}^2 + m_{8,0}^2) \sin \theta \cos \theta, \\ m_\beta^2 &= m_0^2 \sin^2 \theta + m_8^2 \cos^2 \theta + (m_{0,8}^2 + m_{8,0}^2) \sin \theta \cos \theta. \end{aligned} \tag{10.8}$$

We may eliminate $(m_{0,8}^2 + m_{8,0}^2)$ and m_0^2 from **10.7** and **10.8** to obtain an expression for the mixing angle in terms of m_8^2 (which can be calculated from **10.2**). This expression can be written

$$\sin^2 \theta = \frac{m_\beta^2 - m_8^2}{m_\beta^2 - m_0^2}.$$

For the pseudo-scalar mesons we find $\theta = -11.1^\circ$ (mixing η and η'). For the vector mesons $\theta = +38.6^\circ$ (ω, ϕ) and for the tensor mesons $\theta = +28^\circ$ (f_2, f_2'). Thus for the pseudo-scalar mesons the mixing is small and the physical η nearly obeys the Gell-Mann–Okubo formula, but for both the vector and tensor mesons the mixing is large and the physical particle masses do not obey the GMO relation. The vector and tensor nonets exhibit close to ‘ideal mixing’ ($\theta \approx 35^\circ$) for which ϕ and f_2' are pure $s\bar{s}$ states (see below).

Note that by introducing the new free parameter θ , the mass relation ceases to be a test of the theory or to be a predictive tool. There are other results involving θ which are open to independent test, such as the ratios of certain decay modes of the $I = 0$ mesons within the nonet and, on the basis of a particular quark interaction model, the ratios of production of, say, η and η' in certain reactions. Such results, although in general

agreement with the mixing angles obtained from the masses, usually have large errors for θ while the model used in the production reaction analysis is also not well proved.

10.7 Mesons and baryons constructed from quarks

The quarks as building blocks for the mesons and baryons were introduced in section 10.4. From fig. 10.5 and the relations between Y, B, S, Q and I_3 , we can deduce the properties of the quarks, which are summarised in table 10.1.

The simplest assumption is that all the quarks have spin $\frac{1}{2}$ (see section 10.11). Thus, for instance, a K^+ -meson will be constructed from a $u\bar{s}$ quark combination and a proton from uud .

We have already seen that $SU(3)$ allows singlets, octets, decuplets, antidecuplets, 27-plets and higher multiplets. However, the quark model as described above is more restrictive. For mesons we expect only

$$3 \otimes \bar{3} = 8 \oplus 1,$$

i.e. singlets and octets, while for baryons we expect only

$$3 \otimes 3 \otimes 3 = 1 \oplus 8 \oplus 8 \oplus 10,$$

i.e. no antidecuplet or 27-plet or higher multiplet.

For instance, consider the meson 27-plet shown in fig. 10.9. Some possible decay modes are shown for the 'far-out' states such as a doubly-charged $S = +2$ meson (K^+K^+), a doubly-charged ($I = \frac{3}{2}$) $S = +1$ meson (doubly-charged $K^* \rightarrow K^+\pi^+$) and a doubly-charged $S = 0$ meson (decaying to $\pi^+\pi^+$ or $\pi^0\pi^+\pi^+$). In order to construct such an $S = +2, I = +1$ meson we require four quarks, $\bar{s}suu$. Similarly, the student may check that a strangeness $S = +1$ and charge -2 baryon must belong to a baryon 27-plet and be constructed from five quarks ($\bar{s}uuud$), while a $B = +1, S = +1$ singly-charged particle must belong to an antidecuplet.

Table 10.1. *Properties of the u, d and s quarks*

	B	I	I_3	Q	Y	S
u	$\frac{1}{3}$	$\frac{1}{2}$	$+\frac{1}{2}$	$\frac{2}{3}$	$+\frac{1}{3}$	0
d	$\frac{1}{3}$	$\frac{1}{2}$	$-\frac{1}{2}$	$-\frac{1}{3}$	$+\frac{1}{3}$	0
s	$\frac{1}{3}$	0	0	$-\frac{1}{3}$	$-\frac{2}{3}$	-1
\bar{u}	$-\frac{1}{3}$	$\frac{1}{2}$	$-\frac{1}{2}$	$-\frac{2}{3}$	$-\frac{1}{3}$	0
\bar{d}	$-\frac{1}{3}$	$\frac{1}{2}$	$+\frac{1}{2}$	$+\frac{1}{3}$	$-\frac{1}{3}$	0
s	$-\frac{1}{3}$	0	0	$+\frac{1}{3}$	$+\frac{2}{3}$	+1

Some evidence of such states has been reported in formation experiments in the form of peaks in the total cross-section for K^+p scattering (pure $I = 1$) and in the total $I = 0$, K -nucleon scattering cross-section as deduced from the K^+p and $K^+n(K^+d)$ cross-sections. However, the peak in the K^+p cross-section may probably be explained by the opening up of inelastic pion-production channels, while other possible explanations of the $I = 0$ peak have also been proposed. In addition, a number of double-charged peaks have been found in production experiments. These peaks have, in general, either found explanations as reflections of other resonances or have not been substantiated. Thus, at the present time, all well established meson resonances fit into octets or singlets (nonets) and all the baryon resonances into singlets, octets or decuplets. It seems clear that 'far-out' or 'exotic' resonances are produced only weakly, if at all, and this result must be counted as evidence for the correctness of the quark model.

10.8 The orbital-excitation model for quarks

If the mesons are taken to consist of $\bar{q}q$ pairs, the q and \bar{q} may rotate about each other with relative orbital angular momentum lh . We may then have a series of states of increasing values of l . Such a model has been proposed by Dalitz.

Mesons built in this way must have some specific properties. The parity is given by $P = (-1)^{l+1}$ and the charge-conjugation parity in the strangeness-zero case will be $C = (-1)^{l+s}$, where s is the spin of the $\bar{q}q$ pair. Thus, for instance, all natural parity states must have $C = P$.

The lowest $\bar{q}q$ states are then the 3S_1 and 1S_0 states, which have the appropriate J^{PC} for the vector and scalar nonets respectively. From the

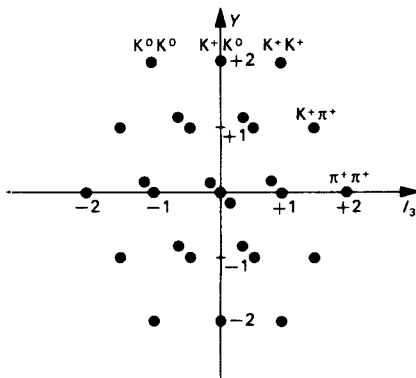


Fig. 10.9. Possible 27-plet meson states.

basic 3S_1 and 1S_0 nonets there can be generated two series of rotational levels of increasing l . The triplet 3S_1 yields P-states 3P_0 , 3P_1 and 3P_2 , for which $P = +1$, and for the neutral states of which $C = +1$. The singlet P-state is 1P_1 , with again positive parity, but negative under charge conjugation.

The four P-states may be rendered non-degenerate by spin-orbit SU(3)-breaking forces proportional to $\mathbf{I} \cdot \mathbf{s}$. But since

$$\langle \mathbf{I} \cdot \mathbf{s} \rangle = \frac{j(j+1) - l(l+1) - s(s+1)}{2}$$

we have equal spacing in mass-squared for the P-levels

$$\begin{array}{cccc} {}^3P_{L-1} & {}^3P_L & {}^1P_L & {}^3P_{L+1} \\ \leftarrow 2L \rightarrow & \leftarrow 2L \rightarrow & \leftarrow 2L \rightarrow & \leftarrow 2L \rightarrow \end{array}$$

For D-states the inner states are twice as close as the other spacings.

It is possible to assign some of the higher-mass mesons to such orbital-excitation multiplets. For the baryon three-quark system the orbital-excitation model is, of course, considerably more complicated.

10.9 Developments of the quark model: magnetic moments

The SU(3) classification has been enlarged to encompass also the spin quantum number. The appropriate group is SU(6), which should be a good symmetry if the interactions are invariant with respect to rotations of the spin and if spin-orbit couplings are not important. This limitation means that SU(6) can be valid only in the non-relativistic region, unlike the SU(3) symmetry which was independent of space-time.

In considering the calculation of magnetic moments we shall need to construct the appropriate states for proton and neutron from the spinning quarks. With certain simple assumptions SU(3) itself can give certain relations, such as, for instance, $\mu(\Sigma^+) = \mu(p)$ (measured values 2.5 ± 0.5 and 2.79 respectively). (This is immediately obvious if it is assumed that the electromagnetic interaction is a U -spin scalar.) The SU(6) quark model, however, while including these relations, yields additional ones and in particular relates the very precisely known proton and neutron magnetic moments.

We give below a naïve calculation of the proton and neutron magnetic moments which yields results in remarkably good agreement with the experimentally measured values. It has lately however become clear from muon scattering experiments that the gluons (see chapter 13) which bind the quarks, and also virtual pairs of $s\bar{s}$ quarks, may make important

contributions to the spin. Although we should bear in mind that the agreement of the naïve result with experiment may be an accident the calculation is nevertheless a useful exercise of its kind.

We take the proton to be constructed from the quark combination (uud). We now use the Clebsch–Gordan coefficients to construct the spin $\frac{1}{2}$ quark states. We shall take $j_z = +\frac{1}{2}$ so that the $(\frac{1}{2}, +\frac{1}{2})$ physical proton will be built from the three states (A), (B), (C) below, where we make the simplest assumption that the quarks are all in relative S-states. (This implies that the overall wave function is symmetric whereas the generalised Pauli principle (3.7) would require (Fermion) spin $\frac{1}{2}$ quarks to have an antisymmetric wave function. The explanation of this paradox lies in the idea of colour – see section 13.1.)

	u	u	d
(A)	$(\frac{1}{2}, +\frac{1}{2})$	$(\frac{1}{2}, +\frac{1}{2})$	$(\frac{1}{2}, -\frac{1}{2})$
(B)	$(\frac{1}{2}, +\frac{1}{2})$	$(\frac{1}{2}, -\frac{1}{2})$	$(\frac{1}{2}, +\frac{1}{2})$
(C)	$(\frac{1}{2}, -\frac{1}{2})$	$(\frac{1}{2}, +\frac{1}{2})$	$(\frac{1}{2}, +\frac{1}{2})$

States in which the up and down quarks are arranged in a different order, e.g. udu, are all the same, the only point of consequence for this problem being the spin orientations.

Grouping now the two up quarks we see that

$$(A) \equiv (1, +1)(\frac{1}{2}, -\frac{1}{2}),$$

$$(B) \text{ and } (C) \equiv (1, 0)(\frac{1}{2}, +\frac{1}{2}).$$

We can write

$$(1, 0)(\frac{1}{2}, +\frac{1}{2}) = \frac{1}{2}(B) + \frac{1}{2}(C).$$

Thus, referring to the table of Clebsch–Gordan coefficients, we have

$$\begin{aligned} p_{\frac{1}{2}} &= \sqrt{\frac{2}{3}}(A) - \sqrt{\frac{1}{3}}\sqrt{\frac{1}{2}}\{(B) + (C)\} \\ &= \sqrt{\frac{1}{6}}(2A - B - C) \\ &= \sqrt{\frac{1}{6}}(2u\uparrow u\uparrow d\downarrow - u\uparrow u\downarrow d\uparrow - u\downarrow u\uparrow d\uparrow), \end{aligned} \tag{10.9}$$

where the arrows indicate the spin directions.

Similarly, the physical neutron may be written in terms of spinning quarks as

$$n_{\frac{1}{2}} = \sqrt{\frac{1}{6}}(2d\uparrow d\uparrow u\downarrow - d\uparrow d\downarrow u\uparrow - d\downarrow d\uparrow u\uparrow), \tag{10.10}$$

If we write a basic ‘quark magneton’ in the usual way as

$$\mu_q = \frac{Q\hbar}{2m_q c}$$

where Q is the charge and m_q the mass of the quark (here assumed approximately equal for the u and d quarks), then the up and down

quarks will have magnetic moments expressed in terms of m_q and the z -component of the spin σ_3 as

$$\begin{aligned}\mu_u &= \frac{2}{3}\sigma_{3,u}\mu_q \\ \mu_d &= -\frac{1}{3}\sigma_{3,d}\mu_q\end{aligned}$$

We now calculate the proton and neutron magnetic moments as the sums of the moments arising from the quark states of **10.9** and **10.10**.

$$\begin{aligned}\mu_p &= \frac{1}{6}[4(\frac{2}{3} + \frac{2}{3} + \frac{1}{3}) + (\frac{2}{3} - \frac{2}{3} - \frac{1}{3}) + (-\frac{2}{3} + \frac{2}{3} - \frac{1}{3})]\mu_q = \mu_q \\ \mu_n &= -\frac{2}{3}\mu_q\end{aligned}$$

so that $\mu_n/\mu_p = -\frac{2}{3}$. The experimental value is $-1.913/2.792 = -0.68$. The agreement represents a further striking success for the quark model.

10.10 Nucleon structure from scattering experiments

It has long been realised that the nucleons are not point particles but are of finite size $\sim 10^{-13}$ cm while the evidence from symmetries discussed in earlier sections has suggested that they are built from quarks. Probing the nucleons dynamically by means of deep inelastic scattering of leptons (electrons, muons and neutrinos) has led to an understanding of nucleon structure, to the confirmation of the quarks as the elements of the nucleon substructure and to measurement of the quark properties. The leptons, having no strong interaction, probe the nucleon structure by means of the electromagnetic and weak interactions.

10.10.1 *Elements of scattering theory: elastic scattering*

We start from the familiar Rutherford scattering formula which gives the cross-section for elastic scattering of two charged, spinless, point particles as a result of the electromagnetic interaction

$$\left(\frac{d\sigma}{d\Omega}\right)_R = \frac{(Ze)^2 E^2}{4p^4 \sin^4(\theta/2)}.$$

Ze is the charge on the target (assumed infinitely massive so that there is no recoil), e is the incident charge, p and E the incident momentum and energy, θ the scattering angle and α the fine-structure constant.

For scattering of high energy electrons on protons, several factors modify the simple Rutherford formula:

- (a) the particles are not spinless and the proton magnetic moment contributes to the interaction;
- (b) the proton is not infinitely massive and recoils;
- (c) the proton is not a point charge but a charge distribution $e\rho(x)$.

In general, the cross-section for scattering by an extended target can be

written in terms of that for a point target by inclusion of a ‘form factor’ $F(\mathbf{q})$

$$\left(\frac{d\sigma}{d\Omega}\right) = \left(\frac{d\sigma}{d\Omega}\right)_{\text{Point}} |F(\mathbf{q})|^2$$

where \mathbf{q} is the momentum transfer from the projectile to the target, $\mathbf{q} = \mathbf{p}_i - \mathbf{p}_f$, where \mathbf{p}_i and \mathbf{p}_f are the initial and final momenta. The form factor in such a case can be shown to be the Fourier transform of the charge distribution

$$F(\mathbf{q}) = \int \rho(\mathbf{x}) e^{i\mathbf{q} \cdot \mathbf{x}} d^3x.$$

When all the above factors (a), (b), (c) are taken into account, the cross-section formula can be written

$$\left(\frac{d\sigma}{d\Omega}\right)_{\text{lab}} = \frac{\alpha^2}{4E^2 \sin^4(\theta/2)} \frac{E'}{E} \left[\frac{G_E^2 + \tau G_M^2}{1 + \tau} \cos^2\left(\frac{\theta}{2}\right) + 2\tau G_M^2 \sin^2\left(\frac{\theta}{2}\right) \right]$$

10.11

where we have written $\tau = -(q^2/4m^2)$, where m is the proton mass, and where E' is the new energy of the projectile after scattering. The G_E and G_M are form factors related to the charge and magnetic moment distributions in the proton (but, due to the complications introduced by the recoil, not simply as the Fourier transforms of these distributions). They are functions of q^2 .

10.10.2 *Elements of scattering theory: inelastic scattering*

If we wish to probe to dimensions of order x then the projectile wavelength must, in general, be of this order or less. For scattering as a result of the electromagnetic interaction the process takes place as a result of photon exchange as in fig. 10.10. Thus short-distance structure studies require high energy photon exchange or *large values of* $-q^2$. Such large q^2

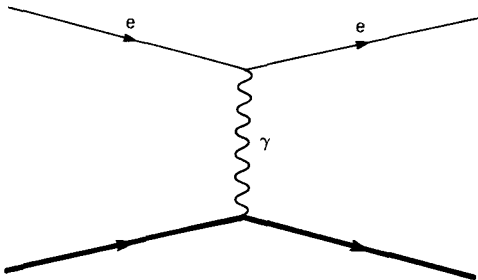


Fig. 10.10. Photon exchange in charged-particle scattering.

collisions, where the interaction is with the individual nucleon constituents, will in general result in break up of the nucleon, resulting at high energies in jets including mesons and nucleons, so that the initial particle identities are no longer preserved and the collision becomes *inelastic*. The effective mass W of the final hadronic state is no longer constant as in elastic scattering so that we require *two independent variables rather than one* in order to describe the scattering.

It is convenient at this stage to move to a formalism which is independent of the reference frame (above formulae refer to the laboratory frame) and to this end we take q^2 to be the *four-momentum transfer* (see appendix A1). Since the energy-momentum four-vectors $(\mathbf{p}, i\varepsilon)$ are Lorentz vectors the square of the four-momentum transfer q^2 is an invariant quantity (fig. 10.11):

$$\begin{aligned} q^2 &= (\mathbf{p}_i - \mathbf{p}_f)^2 - (\varepsilon_i - \varepsilon_f)^2 \\ &= -2m^2 - 2p_i p_f \cos \theta + 2\varepsilon_i \varepsilon_f \\ &\simeq 2p_i p_f (1 - \cos \theta) = 4p_i p_f \sin^2 \theta / 2, \end{aligned}$$

the approximation being good if $m^2 \ll q^2$. In this form q^2 is seen to be positive and (referred to the four-momentum components) is known as 'spacelike'. For real rather than virtual, particle exchange q^2 is negative or 'timelike'. Considering the struck proton vertex

$$q^2 = p'^2 - (E' - M)^2.$$

The energy $(E' - M)$ acquired by the hadronic system is equal to the energy transfer from the virtual photon, i.e. the timelike component of the four-vector $(\mathbf{p}, i\varepsilon)$. In this application ε is normally referred to by the symbol ν , so that

$$\nu = E' - M.$$

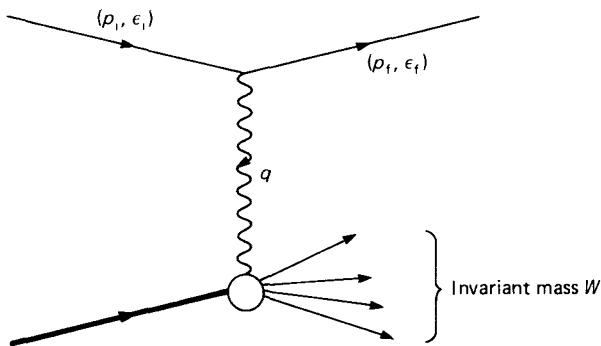


Fig. 10.11. Four-momentum transfer q in scattering processes.

If W is the effective mass of the hadronic system, then

$$W^2 = E'^2 - p'^2$$

with

$$q^2 = M^2 + 2\nu M - W^2.$$

In *elastic* scattering $W = M$ and $q^2 = 2\nu M$ so that ν and q^2 are not independent variables.

The cross-section in terms of variables q^2 and ν can be written

$$\frac{d^2\sigma}{dq^2 d\nu} = \frac{4\pi\alpha^2 E'}{q^4} \frac{1}{E M} \left[W_2(q^2, \nu) \cos^2\left(\frac{\theta}{2}\right) + 2W_1(q^2, \nu) \sin^2\left(\frac{\theta}{2}\right) \right]. \quad \mathbf{10.12}$$

The $W_1(q^2, \nu)$ and $W_2(q^2, \nu)$ are form factors related to the transverse and longitudinal polarisation states of the exchanged virtual photon (*virtual* photons can have longitudinal polarisation).

A useful variable in this area is

$$x = q^2/2M\nu$$

where it is clear that $0 \leq x \leq 1$; $x = 1$ corresponding to the elastic, and $x = 0$ to the totally inelastic, situations.

10.10.3 *Bjorken scaling and partons*

If the proton contains point-like scattering centres then, for sufficiently small wavelength photon probes, the cross-section should be expressible in the point-like form. In order to compare more directly the forms of equations **10.12** and **10.11** we write form factors

$$F_1(q^2, \nu) = W_1(q^2, \nu)$$

$$F_2(q^2, \nu) = \frac{\nu}{M} W_2(q^2, \nu)$$

so that **10.12** becomes

$$\frac{d^2\sigma}{dq^2 d\nu} = \frac{4\pi\alpha^2 E'}{q^4} \frac{1}{E \nu} \left[F_2(q^2, \nu) \cos^2\left(\frac{\theta}{2}\right) + \frac{2\nu}{M} F_1(q^2, \nu) \sin^2\left(\frac{\theta}{2}\right) \right]. \quad \mathbf{10.13}$$

which compares directly with the elastic case **10.11** for which $\nu/M = 2\tau$. Thus for point-like scattering the functions $F(q^2, \nu)$ can not depend on two variables such as q^2 and ν but should be expressible in terms of a single parameter such as x . This property is known as the Bjorken scaling hypothesis (Bjorken, 1967), and states that in the limit where q and ν tend to infinity the form factors depend only on the *ratio* of these quantities determined by the parameter x . The fact that x is dimensionless and has no associated scale leads to the description of the phenomenon as 'scale invariance'. The verification of the independence of form factors with

respect to q^2 is illustrated in fig. 10.12. Although scale invariance is not perfect for all values of x the dependence on q^2 is always weak compared with the very strong dependence on q^2 of the elastic form factor for electron–proton scattering where the interaction is with the proton as a whole and where the form factor falls by $\sim 10^6$ as q^2 increases from 1–25 GeV². Reasons for some deviation from scale invariance are discussed in chapter 13, section 13.3, but the main features remain unaffected.

The point constituents of the nucleon originally named ‘partons’ by Feynman (1969) are now naturally identified with the quarks and gluons (see 13.2). Beyond the scaling features discussed here inelastic scattering can be used as a probe to investigate the nature and momentum distributions of the constituents. In addition, the ratio of the form factors in neutrino– and electron–nucleon scattering provides a test of the quark charge assignments which is in good agreement with the $\frac{2}{3}e$ and $-\frac{1}{3}e$ values for u and d quarks discussed in earlier sections. The development of this subject in more detail is beyond the scope of this text and we confine our treatment to just one further important aspect: the parton spins.

10.10.4 Parton spins

In our comparison of equations 10.11 and 10.13 we implicitly assumed scattering off a spin $\frac{1}{2}$ target (the proton). We can express this element of the comparison more explicitly in terms of the form factors

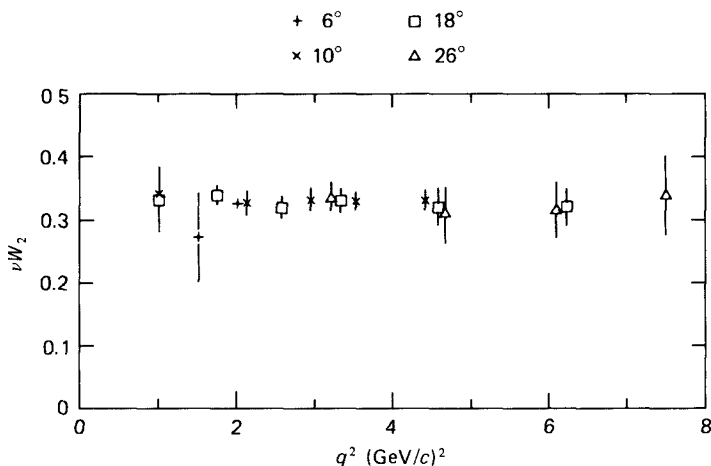


Fig. 10.12. An example of scaling behaviour for the structure function W_2 for electron–proton scattering which is seen to be independent of q^2 . (Friedman and Kendall, 1972).

$F_1(x)$ and $F_2(x)$. If we take the point charge version of **10.11** ($G_M = G_E = 1$) and use the relations for q in terms of θ for an elastic collision:

$$q^2 = 2E^2(1 - \cos \theta) \quad \text{and} \quad dq^2 = \frac{E^2 d\Omega}{\pi} \quad (E \gg m, E = P_i)$$

then **10.11** has the form

$$\frac{d^2\sigma}{dq^2} = \frac{4\pi\alpha^2 Z^2}{q^4} \left(\frac{E'}{E}\right) \left(\cos^2\left(\frac{\theta}{2}\right) + \frac{q^2}{2m^2} \sin^2\left(\frac{\theta}{2}\right)\right) \tag{10.14}$$

where we have written Ze for the parton charge and m for the parton mass.

We require to compare this equation with **10.13** or its analogue expressed in terms of the variable x , but we note that a proper comparison requires us to replace M in **10.13** by the parton mass.

We can interpret the physical significance of x and relate it to the quark and proton masses by an argument using the ‘infinite momentum’ frame in which the proton has a momentum so large that we can neglect its rest mass and write its four-momentum P as $(p, 0, 0, ip)$. The proton thus consists of a bunch of partons moving along the x -axis with momentum such that the parton masses and the transverse components of their momenta can also be neglected. We suppose that each parton carries a fraction f of the total momentum p ($0 < f < 1$) so that a parton momentum is fp . In our model one parton scatters, picking up the momentum transfer q . Thus

$$(fP + q)^2 = -m^2 \simeq 0$$

where m is the parton mass. Then

$$f^2 P^2 + q^2 + 2fP \cdot q \simeq 0.$$

But $|f^2 P^2| = f^2 M^2 \simeq 0$ (cf. q^2) so that

$$f = -\frac{q^2}{2P \cdot q} = \frac{q^2}{2Mv}$$

since we can choose to evaluate the invariant scalar product of the four-vectors in the laboratory where $P = M$ (nucleon at rest) and the time component of $q = v$. Thus we see that, in fact, f is equal to our previously defined parameter x and that x represents the fraction of the proton momentum carried by the struck quark. We may also write the elastic relation $q^2 = 2mv$ to get (in this approximation)

$$x = \frac{m}{M}.$$

We may now substitute this relation in equation **10.13** to obtain

$$\frac{d^2\sigma}{dq^2 dx} = \frac{4\pi\alpha^2}{q^4} \frac{E'}{E} \frac{1}{x} \left[F_2(x) \cos^2\left(\frac{\theta}{2}\right) + 2xF_1(x) \frac{q^2}{2m^2} \sin^2\left(\frac{\theta}{2}\right) \right]. \quad \mathbf{10.15}$$

Comparing the coefficients in **10.15** and **10.14** we see that for the scattering to behave as that from point-like constituents of spin $\frac{1}{2}$ and normal Dirac magnetic moments (assumed in **10.11**) we must have

$$\frac{2xF_1(x)}{F_2(x)} = 1.$$

This is known as the Callan–Gross relation (1968). It is well satisfied by the experimental measurements establishing the quarks as spin $\frac{1}{2}$ particles.

10.11 The search for free quarks

In the preceding sections we have discussed the successes of the quark model in the classification of resonances, in the calculation of some magnetic moments and in accounting for the features of lepton–nucleon scattering.

Although, as we shall see in chapter 13, it may well be the case that the force between quarks is such that free quarks can never be observed, nevertheless this is not presently an inevitable consequence of quantum chromodynamic theory and the search for free quarks is still vigorously pursued.

The most distinctive observable property of the quarks is their fractional charge. The ionisation produced by $\frac{1}{3}e$ and $\frac{2}{3}e$ charged quarks will be $\frac{1}{9}$ and $\frac{4}{9}$ that of singly-charged particles travelling at the same speed. Extensive studies of particles from accelerator targets and in the cosmic radiation have not produced convincing evidence for the existence of fractionally-charged particles. Early results on cloud-chamber tracks of cosmic rays have not been confirmed by subsequent experiments and were probably instrumental artefacts.

The only positive results on the existence of fractionally-charged particles come from a search for fractionally-charged particles in matter (LaRue *et al.*, 1977, 1981). The measurement is a sophisticated version of the classical Millikan oil drop experiment but using very small diamagnetic niobium balls (mass $\sim 9 \times 10^{-5}$ g) levitated between two horizontal plates. The results of the experiment indicated clearly that some of the spheres carried residual charges of $\pm\frac{1}{3}$ of the electron charge.

Such measurements have however been repeated by P. F. Smith *et al.* (1987) using a similar experimental arrangement, and in an extensive set

of data these workers find no evidence for fractional charges with a limit of $< 10^{-21}$ quarks/nucleon.

In another long running search for quarks in matter, Marinelli and Morpurgo (1980) levitated small iron cylinders in a magnetic field. An electric field was used to apply a force to the particle and thus to measure the charge. No fractional charge particles have been detected with a limit for free quarks in iron of $\leq 3 \times 10^{-21}$ quarks/nucleon.

On balance the evidence to date is against the existence of free quarks in matter; a result which may be accounted for by the nature of the quark–quark interaction discussed in section 13.3.

14

Higher symmetries

14.1 Grand unification

We first recall some facts which emerge from the studies and analysis described in the preceding chapters concerning the properties of *quarks* and *leptons*. Both quarks and leptons

- (i) have spin $\frac{1}{2}$;
- (ii) obey the Pauli principle;
- (iii) exhibit no internal structure (at the present limit $\sim 10^{-18}$ m), i.e. they act as point particles;
- (iv) are left-handed in respect of weak processes;
- (v) fall into the same kind of doublets for weak processes;
- (vi) obey similar gauge theories and interact by exchange of spin 1 bosons.

The only difference we recognise between quarks and leptons is that the quarks carry colour while the leptons do not.

These similarities and the successful unification of the weak and electromagnetic interactions naturally lead to intensive efforts to unify the electro-weak with the strong interactions. Such unification implies the existence of a larger group encompassing

$$SU(2) \times U(1) \times SU(3)$$

In such a group, the couplings for the different processes would not be

independent, so that, for instance, the Weinberg angle θ_w linking the weak and electromagnetic couplings would be predicted by the theory. All the interactions are then described in terms of a single coupling to which the other couplings can be related once the grand unified group is recognised. Such grand unified theories are frequently referred to by the abbreviation 'GUTS'.

We recall that the strong coupling decreases as a result of the logarithmic screening term (section 13.3) while the electromagnetic coupling increases due to the screening effect, as a function of momentum transfer. If we write α_3 , α_2 and α_1 for the couplings corresponding to the SU(3) of colour, the SU(2) and U(1) subgroups, then we can calculate their variation as a function of momentum Q (or impact parameter $1/Q$) and, using the measured values at accessible energies, we can determine the momentum at which the couplings become equal. The corresponding energy is known as the unification mass M_x and its value is found to be $\sim 10^{15} \text{ GeV}/c^2$ ($\sim 10^{-10} \text{ gm}$). The behaviour of the couplings is illustrated in fig. 14.1. The unification mass is so large that there is no chance that we can approach it with current accelerator techniques. Nevertheless, grand unification can still have small effects even at presently attainable energies.

In order to illustrate such possible effects we return to consider the unification group required to encompass $SU(2) \times U(1) \times SU(3)$. Georgi and Glashow (1974) have shown that the smallest such group is SU(5). An SU(N) gauge theory has $(N^2 - 1)$ gauge bosons (cf. SU(3) with eight coloured gluons), so that for SU(5) we expect 24. The 24 break down into multiplets

$$24 = (8, 1) \oplus (1, 3) \oplus (1, 1) \oplus (3, 2) \oplus (\bar{3}, 2)$$

where the brackets represent an $(SU(3)_{\text{colour}}, SU(2)_L)$ decomposition. The (8, 1) corresponds to the gluons, the $(1, 3) \oplus (1, 1)$ to the W, Z and γ while the $(3, 2) \oplus (\bar{3}, 2)$ describes a new *weak doublet* of 12 *coloured* bosons responsible for the unification. These superheavy bosons, usually called X and Y, will have mass $\sim 10^{15} \text{ GeV}/c^2$ and are responsible for transforming quarks to leptons by processes such as

$$u + d \rightarrow Y \rightarrow e^+ + \bar{u}$$

or

$$u \rightarrow Y + \bar{d}.$$

We may compare this process with the corresponding electro-weak and strong processes

$$d \rightarrow u + W^-$$

and

$$u_R \rightarrow u_B + g(r\bar{b}).$$

Such processes fix the charges of the X and Y as $\frac{4}{3}e$ and $\frac{1}{3}e$ or $-\frac{4}{3}e$ and $-\frac{1}{3}e$; for example,

$$\begin{aligned} u &\rightarrow Y + \bar{d} \\ +\frac{2}{3} &+ \frac{1}{3} + \frac{1}{3} \\ u &\rightarrow X + \bar{u} \\ +\frac{2}{3} &+ \frac{4}{3} - \frac{2}{3}. \end{aligned}$$

We return to the transformation of quarks to leptons in the next section.

A fascinating aspect of the grand unification group is that it accounts for the relationship between the quark charges and the fact that the electron and proton charges are *identically equal*. The argument depends on the fact that since the photon is one of the gauge bosons of SU(5) it follows (not proved here) that the charge operator is a generator of the group. In addition, it is a property of simple groups such as SU(3) and SU(5) that the trace of each generator vanishes. In order to exploit this

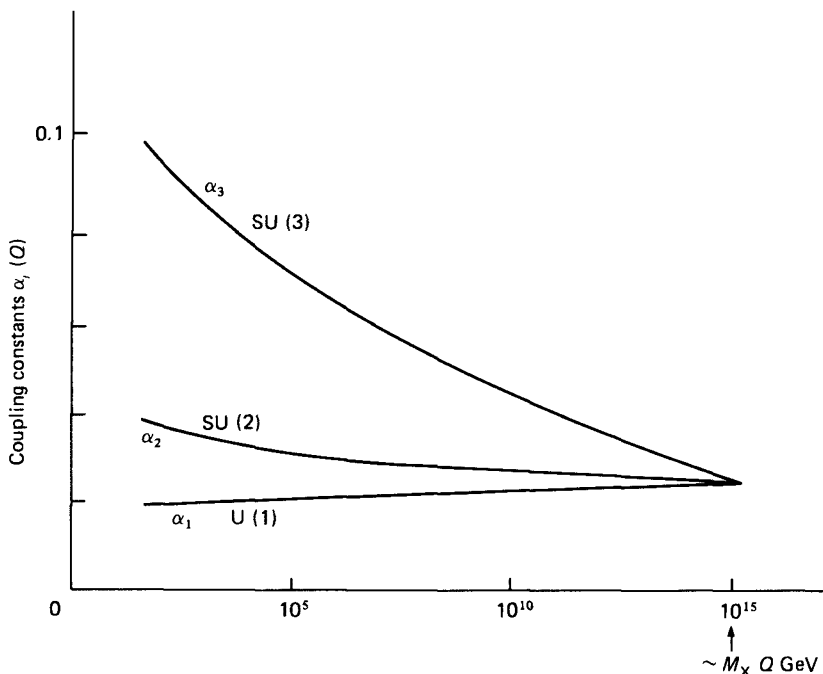


Fig. 14.1. Variation of the running coupling constants for the electro-weak and strong interactions as a function of Q to the (speculative) 'grand unification' mass at $\sim 10^{15}$ GeV.

property we have to identify multiplets of the overall gauge group. For the SU(5) group the multiplets comprise a $\bar{5}$ and a 10 (for *each* generation) where the $\bar{5}$ is usually taken to consist of

$$\bar{5} \equiv \left\{ \begin{array}{c} \nu_e \\ e^- \\ \bar{d}_r \\ \bar{d}_b \\ \bar{d}_g \end{array} \right\}_{\text{LH}}.$$

On the basis of the argument just presented the sum of the charges in the $\bar{5}$ should equal zero. Thus

$$3Q_d + Q_\nu + Q_e = 0$$

so that $Q_d = \frac{1}{3}Q_e$.

For the 10 the same argument yields

$$Q_u = -2Q_d.$$

It follows that

$$Q_p = -Q_e.$$

Thus

- (a) charge *must* be quantised;
- (b) because quarks and leptons appear in the same multiplet their charges are related;
- (c) because there are three colours while the electron is colourless the quarks must have $\frac{1}{3}e$ and $\frac{2}{3}e$ fractional charges;
- (d) the electron and proton charges are equal and opposite.

The $\bar{10}$ is the antisymmetric combination of two $\bar{5}$ s (and the 10 of two fives) and we can see that such a $\bar{10}$ plus the $\bar{5}$ yield all the required states: the u and d quarks in three colours and two helicities, the electron in two helicities and the neutrino in one helicity only (table 14.1).

A satisfying feature of the SU(5) group is that the predicted value of $\sin^2\theta_w$ is ~ 0.20 to be compared with the measured value of 0.2276 ± 0.0004 .

SU(5) is the *simplest* satisfactory group required to account for the structure of the electro-weak and QCD subgroups but it is not the only possible group.

14.2 Proton decay

14.2.1 Lifetime and decay modes

We have mentioned that grand unification can have small effects at low energies. The most dramatic such effect is the decay of the proton.

Such a possibility is immediately obvious from the discussion of section 14.1 above. Figure 14.2 gives diagrams corresponding to processes which we have already discussed but with the addition of a 'spectator' quark. The diagrams correspond to the process

$$p \rightarrow \pi^0 + e^+$$

From a slightly different point of view it is clear that such transitions must follow from inclusion of quarks and leptons in the same multiplet.

The decay rate for a diagram of the kind shown in fig. 14.2 can be estimated in the same way, as for instance, for muon decay by the weak interaction via a W -propagator. The result has the form

$$\tau_p = \frac{4\pi A M_x^4}{g^2 m_p^5}$$

where g is the grand unified $SU(5)$ coupling, the factor M_x^4 is introduced

Table 14.1.

$\bar{5}$ state combination Left-handed	$1\bar{0}$ Right-handed
$\nu_e \bar{e}^-$	e^-
$\nu_e \bar{d}_r$	\bar{d}_r
$\nu_e \bar{d}_b$	\bar{d}_b
$\nu_e \bar{d}_g$	\bar{d}_g
$e^- \bar{d}_r$	\bar{u}_r
$e^- \bar{d}_b$	\bar{u}_b
$e^- \bar{d}_g$	\bar{u}_g
$\bar{d}_r \bar{d}_b$	u_g
$\bar{d}_r \bar{d}_g$	u_b
$\bar{d}_b \bar{d}_g$	u_r

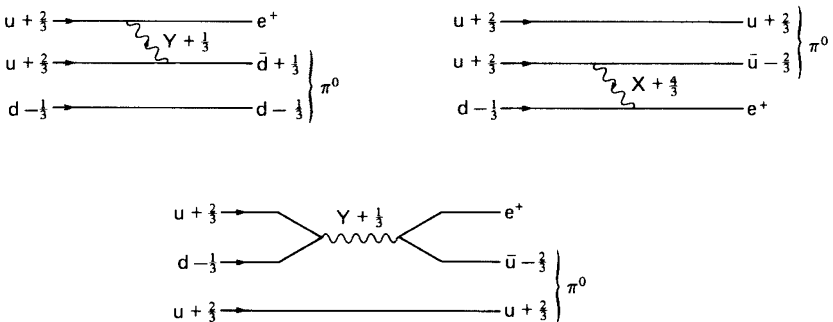


Fig. 14.2. Some mechanisms for proton decay to $e\pi^0$.

by the X-propagator, m_p is the proton mass and the factor A depends on the hadronic matrix elements. We have seen (fig. 14.1) that $g/4\pi \sim \frac{1}{40}$ with $M_x \sim 10^{15}$ GeV while $A \sim 1$, so that we get $\tau_p \sim 5 \times 10^{31}$ years (cf. lifetime of the Universe $\sim 10^{10}$ years). The lifetime is clearly highly sensitive to the unification mass M_x . More sophisticated calculations based on SU(5) (see Langacker, 1981) yield values $\tau_p = 3 \times 10^{29 \pm 13}$ years.

The principal decay channels predicted under SU(5) are:

$p \rightarrow e^+\pi^0$	30 %	$n \rightarrow e^+\pi^-$	54 %
$e^+\rho^0$	14 %	$e^+\rho^-$	23 %.
$e^+\omega$	30 %		
$\bar{\nu}_e\pi^+$	11 %		

We may note that most GUTS other than SU(5) also require proton decay though the unification mass and the lifetimes are, in general, different. Another potentially important difference for the nature of detectors designed to study this phenomenon is that other decay modes such as $p \rightarrow \bar{\nu}_\tau K^+$ are more favoured in some models.

We shall return to consider the cosmological relevance of GUTs in chapter 15.

14.2.2. Measurements

For studies aiming to detect proton decay with lifetimes as long as, say, 10^{32} years, a very large quantity of detector material is required. At such a lifetime, for instance, we require ~ 1000 tons of iron to yield one $e^+\pi^0$ -decay per year.

A critical factor in the design of such experiments is the background which could simulate proton-decay events. The dominant background is due to atmospheric neutrinos, the interaction rate for which is *also* proportional to the mass of the detector material, so that the decay and neutrino interaction rates become equal for a lifetime $\sim 10^{31}$ years. Distinction between background and real events and identification of decay modes other than $e^+\pi^0$ demand sophisticated detector systems.

Two main varieties of detector have been developed to attack the problem. The simpler system consists of a very large tank of water monitored by photomultipliers surrounding or even within the tank. The multipliers detect Cerenkov light from relativistic particles, in particular from e^+ and e^- generated in decays such as $p \rightarrow e^+\pi^0$, $\pi^0 \rightarrow \gamma\gamma$, $\gamma \rightarrow e^+e^-$. Pure water is transparent to the Cerenkov light over a large proportion of the spectrum, allowing the use of very large water volumes. The electromagnetic showers from e^+ and the π^0 -decay γ s will appear back-to-back

so that the Cerenkov light detected by opposite groups of multipliers can be used to measure the decay product momenta. In addition, muons can be detected via their electron decay as pulses with a delay $\sim 2 \mu\text{s}$.

One of the most ambitious detectors of this kind to date is the IMB detector (Bionta *et al.*, 1983) which consists of an approximately cubical mass of 7000 tons of water ($18 \times 17 \times 22.5 \text{ m}$) monitored by 2408 5-in photomultiplier tubes distributed over the surface of the tank on a 1 m grid (fig. 14.3). The detector is situated at a depth equivalent to 1570 m of water underground in the Morton Salt Mine in Ohio in order to minimise the cosmic-ray background. Nevertheless some 2×10^5 muons per day cross the detector to give triggers (> 12 PMs firing within 50 ns or > 3 PMs in any two of 32 groups of 64 firing within 150 ns). The muons serve to calibrate the detector and are distinguished from decay events by an energy selection plus a cut which requires the reconstructed vertex to lie within a fiducial region everywhere $> 2 \text{ m}$ within the detector surface. This cut reduces the effective mass of material to 3300 tons. After elimination of the muon events there remained from a run of 132 days 112 events attributable to neutrino interactions. The absolute number, the energy distribution and the up-down ratio were all consistent with the expectation for neutrino interactions. In addition, the topology of the energy distribution for 109 of the events was such that most of the energy was concentrated in one hemisphere ('one-track hypothesis') inconsistent with $p \rightarrow e^+\pi^0$ or $\mu^+\pi^0$. The remaining three events are also found to be inconsistent with $p \rightarrow e^+\pi^0$ on detailed examination and are attributed to inelastic neutrino absorptions $\nu N \rightarrow \mu\pi N$ or $\nu N \rightarrow e\pi N$.

Although detectors like the IMB experiment have given the best limits to date on the decay lifetime, it is of interest to look briefly at the other main type of detector which can yield better track information. Although such detectors to date have not approached the large mass of the IMB system, some very massive devices of this kind are presently under construction.

Examples of such detectors are the KGF calorimeter at a depth of 7600 m water-equivalent in the Kolar Gold Mines in India (which claimed several examples of proton decay not, however, generally accepted as convincing decay events) which was the first such operating detector, and the NUSEX detector, operating at a depth of 5000 m water-equivalent in the Mont Blanc tunnel. This latter detector consists of a sandwich construction of 1 cm-thick iron plates interleaved with $1 \text{ cm} \times 1 \text{ cm}$ streamer tubes. The tubes are of plastic with a central high-voltage anode wire. The cathodes consist of two orthogonal sets of pick-up strips above

and below the tubes, so that each tube signal gives a three-dimensional coordinate. The mass of the detector is 150 tonnes. Compared with the water Cerenkov, detector devices like NUSEX have a high vertex-reconstruction precision (~ 1 cm compared with ~ 1 m), much better track definition and better containment of events due to the high stopping power of the material. Thus a larger fraction of the detector is useful as a fiducial volume.

Limits on decay have been measured for about 30 modes. In particular the large detectors yield

$$\tau(p \rightarrow e^+\pi^0) > 6.8 \times 10^{32} \text{ y}$$

$$\tau(p \rightarrow \nu K^+) > 9 \times 10^{31} \text{ y}$$

In summary, we can say that there is at present no totally convincing evidence for nucleon decay. In addition, the large detector results appear

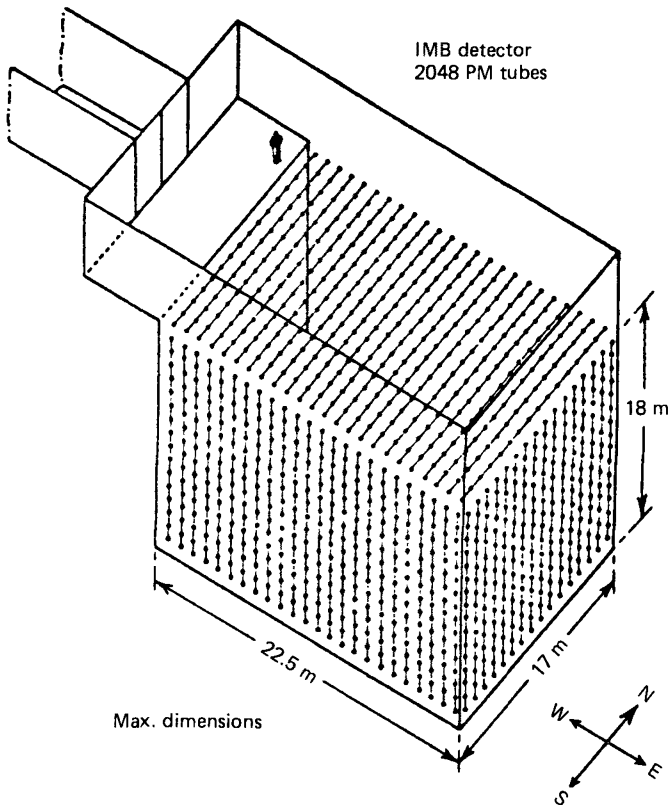


Fig. 14.3. Schematic of the IMB proton-decay detector.

to set a limit for the $p \rightarrow \pi^0 e^+$ decay which is inconsistent with the SU(5) prediction.

Although the great proton decay detectors have so far not found any events we shall however see in the next chapter that they are proving to be very powerful detectors of neutrinos from extraterrestrial sources.

14.3 Magnetic monopoles

Another prediction of grand unified theories is that *magnetic monopoles* should exist as stable particles appearing at the breaking of the grand unified group such as SU(5) into the subgroups which include U(1) (t'Hooft, 1974, and Polyakov, 1974). The magnetic charge is given by the Dirac (1931) relation between magnetic and electric charge

$$g_{\text{Dirac}} = \frac{\hbar c}{2e} \cdot n = \frac{nc}{2\alpha}$$

where α is the fine-structure constant ($\alpha = e^2/\hbar c = \frac{1}{137}$) and the integer n in the Dirac theory is unity. Thus

$$g_{\text{D}} = 68.5e = 3.29 \times 10^{-8} \text{ cgs units.}$$

The monopole mass M_{M} is related to the mass of the X-boson mass M_{X} by

$$M_{\text{M}} \gtrsim \frac{M_{\text{X}}}{g} \sim 10^{16} \text{ GeV} = 0.02 \mu\text{g}$$

(g = strong-interaction coupling strength). This mass is so enormous that there is no prospect of creating monopoles at any accelerator and the only possible source of such particles is as a residue from the Big Bang where grand unification held until the temperature dropped below about 10^{15} GeV at around 10^{-35} s after time zero (see chapter 15). Simple cosmological models, in fact, predict a flux of monopoles at the Earth 10^7 times larger than the upper bound from existing measurements while modifications involving an inflationary Universe scenario yield a vanishingly small flux. The cosmological question must be regarded as open, so that experimental searches for monopoles are of high interest both from the GUT and cosmological points of view. We return to this topic in section 15.1.7.

The dimensionless coupling constant is

$$g_{\text{D}}^2/\hbar c = 34.25$$

to be compared with $e^2/\hbar c = \frac{1}{137}$ so that perturbative calculation techniques are inapplicable.

A number of experimental techniques have been used to search for monopoles: track-etch detectors; ionisation detectors (ionisation is $(g_D/e)^2 \sim 4700$ times the ionisation for an electron charge with the same velocity); bulk matter searches; acoustic wave detection; superconducting induction devices and, indirectly, by searching for monopoles acting as a catalyst for proton decay. We discuss further only the last two of these in that there is *one* observation of an event in an induction device and in view of the interesting cross connection with the proton decay. If a monopole passes through a superconducting coil it produces a step change in the magnetic flux through the coil which generates a step in the electric current in the coil. The method, although technically sophisticated in that it is important to provide very effective magnetic shielding of the coil, is nevertheless particularly attractive in that the signal will be independent of the monopole velocity, mass and electric charge. A group at Stanford (Cabrera, 1982) has used this technique and in 1982 recorded a single current jump the size of which corresponds to a magnetic charge g_D passing through their 5 cm-diameter coil. Although the sensitivity of the Stanford experiment and other independent searches is now at least three orders of magnitude greater than that at the time of the original observation no further events have been observed. Many other experiments using a variety of techniques have set limits on the cosmic ray monopole flux as $\leq 10^{-14} \text{ cm}^{-2} \text{ s}^{-1} \text{ sr}^{-1}$.

Finally on this topic we mention the possibility that monopoles may catalyse proton decay with a significant cross-section (Rubakov, 1981, 1982, and Callan, 1982, 1983) via processes of the kind



The cross-section σ_c is expected to be proportional to $1/\beta$ as in exothermic capture processes in nuclei at low velocities where β is the relative velocity of monopole and proton. Thus

$$\sigma_c = \sigma_0/\beta$$

and σ_0 is expected to be ~ 0.1 mb while for slow monopoles $\beta \sim 0.1$ due to the Fermi motion of the protons in the nucleus. The mean free path is then

$$\lambda(\text{cm}) = \frac{1}{N\sigma_c} = \frac{\beta}{N\sigma_0} \sim \begin{cases} 2-3 \text{ m for iron} \\ 16 \text{ m for water} \end{cases} \quad (N = \text{nucleons cm}^{-3})$$

so that in the larger proton-decay detectors multiple proton decays could be catalysed along the track of a single monopole. Detection of such

multiple decays will depend on appropriately short dead time following the first event or a sufficiently long sensitive time to catch all such events in the same time window. For single detected events it may be impossible to distinguish such catalysed decays from neutrino interactions. No multiple events have so far been observed.

14.4 Supersymmetry

SU(5), as already mentioned, is not the only candidate theory for grand unification and at a stage where GUTs are still far from established as a correct description of nature it may seem over ambitious to attempt to formulate even more general theories. There are however good reasons why theoreticians have devoted immense effort to such studies over the past two decades. First we recognise that the Standard Model and even GUTs are incomplete, leaving a large number of apparently arbitrary features unaccounted for such as the number of generations of quarks and leptons and many of their properties which have to be 'put in by hand'. The number of such constants differs in different models but is always ~ 20 . A feature of the Standard Model which is, in the broad picture, disturbing is the distinction between the particles (fermions) and the quanta of the forces which act between them (bosons).

There are also more technical problems with ordinary GUTs. For instance higher order corrections involving X and Y exchange give rise to corrections to the W and Z masses $\sim g^2 M_X^2$ ($\sim 10^{14}$ GeV/ c^2) which must somehow be got rid of to achieve the real measured masses ~ 90 GeV/ c^2 .

On the other hand quantum gravity theory is recognised to be an *unrenormalisable* field theory. A theory which offers an escape from this difficulty at the same time as achieving grander unification is a synthesis of general relativity with *supersymmetry* (SUSY) in what has become known as *supergravity*.

14.4.1. *General features of supersymmetry.*

In supersymmetry every 'normal' fermion (quarks and leptons) has a supersymmetric spin zero partner (squarks and sleptons) while every normal boson (photon, gluons, W, Z, Higgs) has a spin $\frac{1}{2}$ partner. The quantum of the gravitational force, the graviton, which has spin 2, has a supersymmetric partner, the gravitino, which has spin $\frac{3}{2}$. The list of supersymmetric particles with the usual nomenclature is given in table 14.2. Two doublets of Higgs fields are needed to give masses to $+\frac{2}{3}$ and $-\frac{1}{3}$ charge quarks and to cancel certain anomalies. In the minimal version of

supersymmetry (see below) these give rise to five physical boson states: two charged scalar bosons heavier than the W; two neutral scalar bosons, one heavier than the Z^0 and one lighter than the Z^0 , and one neutral pseudoscalar boson heavier than the lightest scalar boson.

Thus we have

$$Q^i|F\rangle = B \quad \text{and} \quad Q^i|B\rangle = F$$

where Q^i carries no internal quantum numbers so all particles and their supersymmetric partners have identical quantum numbers other than the spin. We see that no presently known particle can be the SUSY partner of another. The index $i = 1 \dots N$ designates the complexity of the particle multiplets in different versions of supersymmetry and turns out to be equal to the number of gravitinos. For instance in 'minimal supergravity' $N = 1$ and there is one gravitino while in 'extended supergravity' theories N can range up to 8 with 8 gravitinos.

In this picture the 'superspin' direction in the appropriate superspace defines the particle as a boson or a fermion. The theory satisfies local gauge invariance and remarkably it turns out that transformation of bosons to fermions also shifts the particles slightly in ordinary space-time so that the Pauli principle which otherwise would be violated, is preserved intact.

The infinities in fermion exchange have the opposite sign from those in boson exchange so in a supersymmetric theory cancellations are possible which render supergravity finite up to two loops of exchanged particles. Unfortunately this is not true for all higher orders so the problem is only partly solved. (Compare for instance the situation in weak interactions where the introduction of the W was the *first stage* in solving the infinity problem). One important gain is the cancellation of the higher order

Table 14.2

Particle	Spin	Supersymmetric partner	Spin
quark	$\frac{1}{2}$	squark	\tilde{q} 0
lepton	$\frac{1}{2}$	slepton	\tilde{l} 0
photon	1	photino	$\tilde{\gamma}$ $\frac{1}{2}$
gluon	1	gluino	\tilde{g} $\frac{1}{2}$
W	1	wino	\tilde{W} $\frac{1}{2}$
Z	1	zino	\tilde{Z} $\frac{1}{2}$
H^\pm	0	Higgsino	\tilde{H}^\pm $\frac{1}{2}$
H^0	0		\tilde{H}^0 $\frac{1}{2}$
graviton	2	gravitino	\tilde{g} $\frac{3}{2}$

corrections mentioned above, to the masses M_w and M_z . Cancellation of infinities also goes some way to explaining the existence of the *two* mass scales of 100 GeV and 10^{15} GeV.

It is of course apparent that the supersymmetry is a *broken* symmetry (like I-spin symmetry, electro-weak symmetry...) since it is clear from experiment that since the SUSY particles have not been observed they can not have the same masses as their ordinary partners and are probably heavy. The SUSY couplings always involve pairs of supersymmetric particles so these particles should be produced in associated pairs and each sparticle must have an odd number of sparticles among its decay products. Thus there must be at least one stable sparticle, being the lightest such particle. Arguments mainly based on the failure to observe stable sparticles in ordinary matter suggest that the lightest is probably the photino, though there is no reason why the photino should be massless. The photino is expected to be weakly interacting with ordinary matter so will not be easy to detect. The \tilde{W}^\pm and \tilde{H}^\pm are expected to mix to give mass eigenstates different from pure winos or Higgsinos usually referred to as charginos and designated χ .

14.4.2. *The search for SUSY particles*

A priori it is possible that the masses of some or all of the sparticles are sufficiently low that they could be produced by existing accelerators and searches for such particles have been vigorously pursued. Bearing in mind the features of sparticles discussed in the previous section we can see how such particles might be expected to be produced and decay.

As an example let us consider the production of sleptons at an e^+e^- collider of sufficient energy. We may expect two effective diagrams for production: (a) via the same process involving an intermediate virtual Z^0 or γ as is effective for normal leptons or quarks and (b) via photino exchange (fig. 14.4).

The sleptons should commonly decay to a photino and an ordinary lepton, e.g. the selectron should decay to an electron and a photino. The photinos being weakly interacting will generally escape without detection carrying off substantial energy and transverse momentum. Such interactions will thus appear as dileptonic events which are acoplanar i.e. where the beam line is not in the plane of the two secondary particles. Background to such events will come from QED processes and in particular from $l^+l^-\gamma$ and $l^+l^+l'^-l'^-$ final states. Such background processes are however well understood and calculable so that it is possible

to determine whether these backgrounds account for the data or whether there are events which could arise from slepton production.

There is presently no convincing evidence for the production of SUSY particles in the processes described above or in any of the other processes in which we might expect to find them. Lower mass limits have been obtained for the sparticles which generally depend on the assumption that the photino is the lightest sparticle. Recent data from ALEPH at LEP for example, give lower limits for the masses of sleptons \tilde{e} , $\tilde{\mu}$, $\tilde{\tau}$ as 43.5, 42.6, and 40.4 GeV/ c^2 respectively while a pure wino of less than 45.5 GeV/ c^2 is excluded as is a pure Higgsino of mass less than 44.5 GeV/ c^2 . These limits are roughly independent of the photino mass (assumed lightest) up to $m_\gamma \sim 28$ GeV/ c^2 .

Supersymmetry must thus for the present be considered as an attractive *hypothesis* awaiting more data and possibly higher energy accelerators for its establishment or its rejection into the limbo of interesting ideas which do not apply to our universe as it is.

14.5 Superstrings

During the past decade many theorists have devoted themselves to the development of an alternative and very ambitious 'theory of everything' known as *superstring* theory. These ideas originated in work with a somewhat different orientation by Veneziano in the late 1960s but owe their development and present popularity primarily to the work of John Schwarz and Michael Green over the past two decades.

The basic idea is that *point* particles are replaced by *strings*, that is

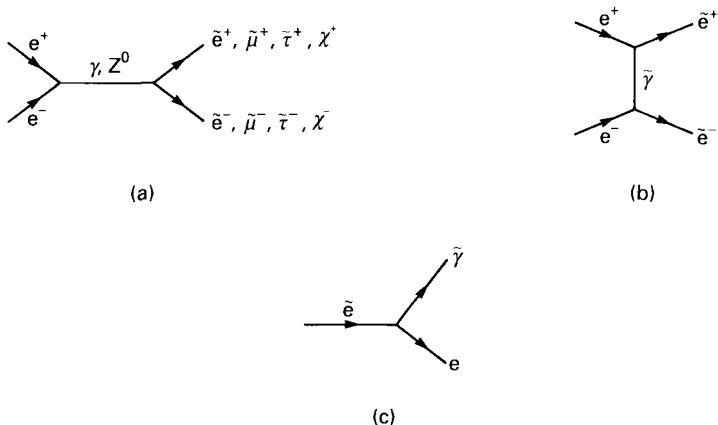


Fig. 14.4. Production (a), (b), and decay (c), processes for supersymmetric particles in e^+e^- collisions.

objects having a linear dimension. Originally both open and closed (loop) strings figured in such theories but latterly the closed loop theories have appeared to be the most promising. Thus instead of particles generating 'world lines' in space-time the loops as they move generate tube-like 'world sheets'. Different types of particle correspond to different oscillations of the loop although the oscillations may be in a space-time of 10 dimensions!

One of the early successes of string theory was the way in which a massless spin 2 particle, identified with the graviton, fell out of the mathematics, encouraging Schwarz and his then collaborator Scherk to see the theory as including gravity (and indeed also the other fundamental forces) in addition to the strong nuclear forces which they had set out to study. In fact the superstring theory has impressive success in eliminating infinities in quantum gravity – possibly to all orders – a unique theoretical achievement.

The natural scale to use in a theory including gravity is the Planck scale where the Planck mass is given by ($G_N =$ gravitational constant $= 6.67 \times 10^{-11} \text{ m}^3 \text{ kg}^{-1} \text{ s}^{-2}$)

$$M_p = \left(\frac{\hbar c}{G_N} \right)^{\frac{1}{2}} = 2.1 \times 10^{-8} \text{ kg} \quad E_p = M_p c^2 = 1.2 \times 10^{19} \text{ GeV}$$

and the corresponding Compton wavelength

$$l_p = \frac{\hbar}{M_p c} = 1.6 \times 10^{-35} \text{ m}$$

The corresponding time and temperature which we shall refer to in chapter 15 are then

$$t_p = \left(\frac{\hbar G_N}{c^5} \right)^{\frac{1}{2}} = 5.4 \times 10^{-44} \text{ s} \quad \text{and} \quad T_p = \frac{1}{k} \left(\frac{\hbar c^5}{G_N} \right)^{\frac{1}{2}} = 1.4 \times 10^{32} \text{ K}$$

The string length is thus expected to be $\sim 10^{-35}$ m which is clearly beyond the reach of investigation of any conceivable particle accelerator although possibly relevant in the first instants following the Big Bang. For this reason it will not be possible to achieve *direct* tests of superstring theory. This does not preclude indirect tests at attainable energies although in view of the energy scale of 10^{19} GeV it seems likely that only a very complete formulation of the theory will allow predictions within the range of foreseeable accelerators. At present the theory is far from such a developed state and there are no such predictions.

Superstring theory differs from most physical theories in that it is a 'top down' approach where the guiding principles are the universality and the

3. A LITTLE ABOUT GROUP THEORY

3.1 Preliminaries

It is an apparent fact that nature exhibits many symmetries, both exact and approximate. A symmetry is an invariance property of a system under a set of transformations. For example, our faces have approximate reflection symmetry, because we look approximately the same in a photograph as in a mirror. As another example, a sphere has rotational symmetry because it looks the same no matter how it is rotated.

Symmetry transformations of physical systems have properties analogous to those of a mathematical *group*. These properties are: If we successively perform two symmetry transformations we obtain a unique symmetry transformation; the transformations are associative; and inverse and identity transformations exist.

We have already mentioned in chapter 1 a theorem, called *Noether's theorem*, which relates symmetry principles to conservation laws. Noether's theorem says that if a physical system can be described by a classical Lagrangian which is invariant under a continuous group of transformations, then the system has a conserved current. If a classical field is quantized, the resulting quantum field theory usually has the same symmetry. However, the quantized theory may have an *anomaly*, which breaks the classical symmetry. We briefly discuss anomalies in Section 4.8.

As two examples of symmetry, we note that the fundamental interactions of nature are apparently invariant under the group of translations and the group of rotations in three dimensions. Noether's theorem relates symmetry under translations to the law of conservation of momentum, and the symmetry under rotations to the law of conservation of angular momentum. The translation and rotation groups are examples of *Lie* groups, which we define in Section 3.3.

The Lagrangian of the standard model is invariant under the group of gauge transformations $SU(3) \times SU(2) \times U(1)$. (We define gauge transformations in chapter 4.) The standard model is also invariant under the proper Poincaré group, which includes translations in space and time, rotations, and proper Lorentz transformations. (Improper Lorentz transformations include space and time reflections.) The Poincaré group is a Lie group. The groups $SU(3)$, $SU(2)$, and $U(1)$ are *special unitary* groups, which are also Lie groups. A unitary group is a group of unitary matrices, and a special unitary group is a group of unitary matrices with determinants equal to unity. In order to understand the standard model, we have to have some familiarity with the Lie groups and their Lie algebras, especially unitary groups. (We discuss unitary groups further in Section 3.5.)

In addition to the space-time symmetries of the proper Poincaré group, some theories have additional symmetries under finite transformation groups. Of these, we single out space reflection or parity P , time inversion T , and charge conjugation C . We do not give much discussion of these groups, but they play an important role in the standard model. As we have already mentioned, none of the groups P , C , and T is an exact symmetry of the standard model, but the combined symmetry CPT , taken in any order, is exact.

In this chapter we briefly discuss groups in general, then Lie groups and their algebras, and the unitary groups and their algebras. We concentrate on group representations (which we define in the Section 3.2), especially irreducible unitary representations. We can have unitary representations of many different groups, not only of unitary groups. The treatment in this chapter may seem to be a little condensed for those who only know a little about group theory. More details can be found in many places, for example, in a book on unitary symmetry (Lichtenberg, 1978). We do not discuss the Poincaré group in any detail.

A group G is a set of elements which satisfy four postulates:

- 1) A law of combination, often called a product, is defined so that if a and b belong to G , the product ab is a unique element of G .
- 2) Multiplication is *associative*, i.e., $a(bc) = (ab)c$.
- 3) An *identity* e exists such that $ea = ae = a$.
- 4) An *inverse* a^{-1} exists to any element a such that $a^{-1}a = aa^{-1} = e$.

The number of elements of a group may be finite, in which case the group is called a finite group, or infinite. If all the elements of a group commute with one another, the group is said to be *abelian*. Otherwise the group is *nonabelian*. A subgroup of a group is a subset of elements which is itself a group under the same multiplication law. Every group has at least two subgroups: itself and the group consisting only of the identity. These are called *improper* subgroups; any others are called *proper* subgroups. A group H is *homomorphic* to a group G if there is a mapping of the elements of G onto the elements of H . The groups are *isomorphic* if the mapping is one-to-one.

An element a belonging to G is said to be *conjugate* to an element b in G if there exists an element u in G such that $a = ubu^{-1}$. Let H be a subgroup of G , and let h be in H and g be in G . Form the product elements $h' = ghg^{-1}$ for all h . Then the h' form a group H' which is isomorphic to H . If, for all g in G , the elements of H and H' are identical, then H is called an *invariant* or self-conjugate subgroup of G .

The group G is said to be the *direct product* of two groups H and H' if every h in H commutes with every h' in H' and if every g in G can be written uniquely as a product of an element in H and an element in H' . The direct product is written in the form $G = H \times H'$.

3.2 Group representations

A *representation* of a group is a homomorphism between the group and a group of linear operators which operate on a vector space. We can think of the vectors in this space as being the states (wave functions) of a quantum mechanical system. A finite-dimensional *matrix representation* of a group is a homomorphism between the group and a group of matrices. We often simply use the word “representation” to mean a matrix representation. If a representation is isomorphic to the group, it is said to be *faithful*. We shall consider only representations by square matrices. If G is a group with elements g , then we often denote the corresponding element of the representation by $D(g)$.

The matrices of a representation are a special case of linear operators which act on a vector space. If the matrices are n -by- n , the vectors (wave functions) are column matrices

with n entries, and their hermitian conjugates are row matrices. The vectors are members of an n -dimensional vector space, and therefore the matrices are said to be n -dimensional.

A similarity transformation is a transformation by means of a matrix S which leaves unaltered the algebra of the transformed system. A similarity transformation acts differently on a representation D and on a vector V , namely

$$D' = SDS^{-1}, \quad V' = SV. \quad (3.1)$$

If a representation can be brought into the following form by a similarity transformation:

$$D(g) = \begin{pmatrix} D_1(g) & X(g) \\ 0 & D_2(g) \end{pmatrix} \quad (3.2)$$

for all g , then the representation is called *reducible*. If not, it is *irreducible*. If $X(g) = 0$, the representation is fully reducible. We shall restrict our considerations to cases in which reducible representations are fully reducible, and we shall omit the word “fully.” A theorem (Schur’s lemma) states: A matrix which commutes with all matrices of an irreducible representation is a multiple of the unit matrix.

We next discuss the importance of irreducible unitary representations of groups within the Hamiltonian formalism, as using this formalism is somewhat easier than using the Lagrangian formalism. Let us consider an n -dimensional irreducible unitary representation of a group G . The unitary matrices act on a set of n linearly-independent vectors, which can be chosen to be orthonormal. The members of this orthonormal set (or basis) constitute a *multiplet*.

Let a unitary representation of a symmetry group (that is, a group of transformations which leaves the physical system invariant) be denoted by U_a , where a stands for all the parameters which specify individual group elements. If we have any transition matrix (ϕ, ψ) , where ϕ and ψ are state vectors (or wave functions) describing physical states, then the transformed states $\phi' = U_a\phi$ and $\psi' = U_a\psi$ satisfy the condition

$$(\phi', \psi') = (U_a\phi, U_a\psi) = (U_a^{-1}U_a\phi, \psi) = (\phi, \psi). \quad (3.3)$$

Thus, unitary transformations are important in quantum mechanics because they leave transition matrixes invariant.

If the Hamiltonian H of a physical system is invariant under a symmetry group G , then all members of a multiplet belonging to an irreducible unitary representation have the same energy, as we now show. Now consider the eigenvalue equation

$$H\psi_n = E_n\psi_n. \quad (3.4)$$

If we operate on this equation with U_a , we get

$$U_aH\psi_n = U_aHU_a^{-1}U_a\psi_n = E_nU_a\psi_n. \quad (3.5)$$

Now if H' and ψ'_n are defined as

$$H' = U_aHU_a^{-1}, \quad \psi'_n = U_a\psi_n, \quad (3.6)$$

our equation becomes

$$H'\psi'_n = E_n\psi'_n. \quad (3.7)$$

But because U_a is a symmetry group, by definition it leaves the Hamiltonian H invariant, so that $H' = H$. This implies that U_a commutes with the Hamiltonian:

$$HU_a = U_aH, \text{ or } [H, U_a] = 0, \quad (3.8)$$

where $[H, U_a] = HU_a - U_aH$ is called the *commutator* of H and U_a . Then Eq. (3.7) becomes simply

$$H\psi'_n = E_n\psi'_n, \quad (3.9)$$

so that the transformed wave functions ψ'_n are also eigenfunctions of the Hamiltonian with the same energy eigenvalue. But the transformed wave functions are in general linear combinations of all members of the original multiplet. Therefore, in order for Eq. (3.9) to be true, all members of the multiplet must have the same energy eigenvalue. We mention that if the representation is reducible, the new wave functions are not in general linear combinations of *all* the wave functions belonging to the representation, so that all the wave functions do not need to have the same energy.

It should be clear from the above arguments that if any operator A commutes with the U_a , then all members of a multiplet have the same eigenvalue of the operator A . Thus, for example, let us consider the rotation group $R(3)$. Not only is the Hamiltonian invariant under rotations, so that all members of a multiplet have the same energy, but U_a also commutes with the operator J^2 , so that all members of a multiplet have the same eigenvalue of J^2 , namely, $J(J+1)$.

3.3 Lie groups

We have noted that a group may have a finite or infinite number of elements. A *Lie* group has a continuously infinite number of elements characterized by a finite number of parameters which can vary continuously. Furthermore, if an element of a Lie group is characterized by a set of r parameters collectively denoted by a ($a = a_1, a_2, \dots, a_r$), and another element is characterized by a set of parameters b , then the product element is characterized by a set of parameters c which are *analytic functions* of a and b .

As an example of a Lie group, consider the rotations in two dimensions. These are characterized by a parameter θ . The transformation is

$$\begin{aligned} x'_1 &= x_1 \cos \theta - x_2 \sin \theta, \\ x'_2 &= x_1 \sin \theta + x_2 \cos \theta. \end{aligned} \quad (3.10)$$

The transformation can be written in matrix form as

$$x' = R(\theta)x, \quad (3.11)$$

where

$$x = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}, \quad R = \begin{pmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{pmatrix}. \quad (3.12)$$

The rotation matrix $R(\theta)$ is the group element characterized by the single parameter θ . Rotations in two dimensions constitute an abelian group, but in more dimensions the rotation group is nonabelian. Note that the groups $R(n)$ are faithful representations of themselves in n dimensions.

The group multiplication law for rotations in 2 dimensions can be stated as follows:
If

$$R(\theta) = R(\theta_2)R(\theta_1), \quad (3.13)$$

then

$$\theta = \theta_2 + \theta_1. \quad (3.14)$$

The rotation groups are *compact*. This means that the parameters vary over a *finite, closed* region. For example, the parameter θ of the two-dimensional rotation group varies over the interval $0 \leq \theta \leq 2\pi$.

On the other hand, the translation groups are not compact because the parameters are unbounded. For example, a translation in 1 dimension,

$$x' = x + a,$$

is characterized by a parameter a which can vary from $-\infty$ to ∞ . Likewise, the group of Lorentz transformations is not compact because the group is characterized by a parameter v (the velocity) which varies in the interval $0 \leq v < c$, which is open at one end. Rotations and Lorentz transformations are both subgroups of the Lorentz group.

The concepts of simple and semisimple Lie groups are important but somewhat complicated. An oversimplified definition, which is adequate for our purposes, is that a Lie group is *simple* if it is nonabelian and has no proper invariant Lie subgroups. It is *semisimple* if it is nonabelian and has no abelian invariant Lie subgroups. Clearly, a simple group is also semisimple. If a group is the direct product of two or more groups H, H', \dots , then the subgroups H, H', \dots are invariant. The direct product of simple and/or semisimple Lie groups is semisimple.

Recall that the local gauge group of the standard model is $SU(3) \times SU(2) \times U(1)$. This group is not semisimple because it has an abelian invariant subgroup $U(1)$. However, the group $SU(3) \times SU(2)$ is semisimple. The groups $SU(3)$ and $SU(2)$ are simple.

3.4 Lie algebras

Let us consider a Lie group of transformations. We obtain the *Lie algebra* of the group by considering group elements which differ only infinitesimally from the identity. From these elements we can construct operators called *generators* which allow us to obtain a unitary representation of the group. More precisely, we obtain all the elements of the group which can be generated by continuous transformations from the identity. There is one generator for each parameter of the group. Methods for obtaining the generators of a Lie group have been discussed in many places (see, e.g., Lichtenberg, 1978).

Let the generators of a Lie group be $X_i, i = 1, 2, \dots, r$, where the group is characterized by r real parameters a_i . If the generators are Hermitian, a unitary representation of an arbitrary group element U_a is given by

$$U_a = e^{-i \sum a_i X_i}.$$

It can be shown that the X_i form a Lie algebra, which means that they satisfy the algebraic equations

$$[X_i, X_j] = i \sum_{k=1}^r c_{ij}^k X_k. \quad (3.15)$$

Here the *commutator* $[A, B] = AB - BA$ and the c_{ij}^k are constants called the *structure constants* of the group. (Some people call other constants $b_{ij}^k = ic_{ij}^k$, the group structure constants.) There is no significance to the fact that we write c_{ij}^k with both lower and upper indices in Eq. (3.15). We do this because in the future we shall use the *summation convention* of omitting the summation sign and summing over a repeated upper and lower index (in any order). The structure constants of a Lie algebra can differ with different choices of generators.

As we see from Eq. (3.15), a Lie algebra has the property that the commutator of any two members of the algebra (generators of the Lie group) is a linear combination of the members of the Lie algebra. We also see that the algebra is in general neither commutative nor associative. A *representation* of a Lie algebra is a set of matrices which obey the commutation relations of the algebra.

If a Lie group is abelian, all the commutators of its Lie algebra vanish, i.e. all its structure constants are zero. The maximum number of commuting generators of a Lie group is called the *rank* of the group. Since any generator commutes with itself, every Lie group is at least rank one. The k commuting generators of a rank k Lie group can be simultaneously diagonalized in a matrix representation.

If a Lie group of rank k is semisimple and compact, then one can construct from the members of its Lie algebra k nonlinear *invariant* operators, called *Casimir* operators, which commute with every member of the algebra.

As an example, let us consider the Lie algebra and Casimir operators of the familiar rotation group in three dimensions $R(3)$. This group is characterized by 3 parameters (for example, the Euler angles). Therefore, it has three generators, which can be taken to be the familiar angular momentum operators J_x , J_y , and J_z . They satisfy the Lie algebra

$$[J_x, J_y] = iJ_z, \quad \hbar = 1, \quad (3.16)$$

and cyclic permutations. This group is rank one because none of the J_i commutes with any other. It is also semisimple (actually, simple), so that it has one Casimir operator J^2 given by

$$J^2 = J_x^2 + J_y^2 + J_z^2. \quad (3.17)$$

A representation of the Casimir operator in n dimensions commutes with all the members of an irreducible representation in n dimensions. Therefore, by Schur's lemma, a representation of the Casimir operator is a multiple of the unit matrix.

It is convenient to denote the generators of $R(3)$ by J_1 , J_2 , and J_3 , and write

$$[J_i, J_j] = i\epsilon_{ij}^k J_k, \quad i, j, k = 1, 2, 3, \quad (3.18)$$

where $\epsilon_{ij}^k = \epsilon_{ijk}$ is completely antisymmetric in its indices and $\epsilon_{123} = 1$. Then we see that the structure constants of $R(3)$ are given by ϵ_{ijk} or $i\epsilon_{ijk}$, depending on whether the

structure constants are defined with or without the i . It is easy to show from the definition of rotations that the number of parameters of $R(n)$ is $(n^2 - n)/2$.

3.5 Unitary groups and algebras

The unitary group in n dimensions $U(n)$ is the group of $n \times n$ matrices U_a satisfying

$$U_a^\dagger = U_a^{-1}, \quad (3.19)$$

where a stands for the parameters of the group, the dagger denotes the Hermitian conjugate matrix, and the superscript -1 denotes the inverse. By definition, for any matrix A , we have $(A^\dagger)_{ij} = A_{ji}^*$, with the asterisk denoting the complex conjugate.

A complex matrix in n dimensions is specified by $2n^2$ real numbers. If the matrix is unitary, there are n^2 relations among these numbers, so that $U(n)$ is characterized by n^2 parameters. The group $U(1)$ is one-dimensional and is characterized by only one parameter. Each element of $U(1)$ is a phase $e^{i\theta}$.

The special unitary groups $SU(n)$ have matrices with determinants equal to unity. This provides another relation so that $SU(n)$ is characterized by $n^2 - 1$ parameters. The rank of $SU(n)$ is $n - 1$. The $SU(n)$ groups are semisimple and compact, so that $SU(n)$ has $n - 1$ Casimir operators.

Like $R(3)$, $SU(2)$ has 3 parameters and is of rank 1. In fact, the generators of $SU(2)$ satisfy the same Lie algebra as the generators of $R(3)$. This implies that the two groups are locally isomorphic (i.e., the mapping of a neighborhood of one onto a neighborhood of the other is one-to-one) and globally homomorphic. In fact, the homomorphism is two-to-one from $SU(2)$ onto $R(3)$.

The groups $U(n)$ and $SU(n)$ are matrix groups, and so are faithful representations of themselves. In discussing representations of the unitary groups, we usually confine ourselves to $SU(n)$. The reason is that the algebra of $U(n)$ is the same as the algebra of $SU(n) \times U(1)$, and all the representations of $U(1)$ are one dimensional. The group $SU(n)$ has $n - 1$ so-called fundamental representations. Of these, two are n -dimensional if $n > 2$. There is only one fundamental (two-dimensional) representation if $n = 2$.

The group $SU(n)$ also has a representation of $n^2 - 1$ dimensions, the same number as the number of generators of the group. This representation is called the *adjoint* representation.

We can construct n -dimensional representations of the algebra of $SU(n)$. For $n = 2$, we can chose these matrices to be the familiar Pauli spin matrices σ_1 , σ_2 , and σ_3 , given by

$$\sigma_1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad \sigma_2 = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \quad \sigma_3 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}. \quad (3.20)$$

The Pauli matrices satisfy the commutation relations

$$[\sigma_i, \sigma_j] = 2i\epsilon_{ij}^k \sigma_k. \quad (3.21)$$

Note the factor 2 difference between the structure constants when expressed in terms of the σ 's rather than in terms of the J 's, given in Eq. (3.10). This follows because $\sigma_i = 2J_i$,

and is an example of the fact that the structure constants depend on the representation of the Lie algebra. Still another representation of the Lie algebra of $SU(2)$ or $R(3)$ is by the two-dimensional matrices σ_+ , σ_- , and σ_3 , where

$$\sigma_+ = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}, \quad \sigma_- = \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}. \quad (3.22)$$

The matrix σ_+ is called a raising operator and σ_- is called a lowering operator because of their action on the eigenvectors of the operator σ_3 (see Section 3.6). The matrices σ_+ and σ_- can be written in terms of σ_1 and σ_2 as follows:

$$\sigma_+ = (\sigma_1 + i\sigma_2)/2; \quad \sigma_- = (\sigma_1 - i\sigma_2)/2.$$

The Casimir operator of $SU(2)$ is the same as that of $R(3)$. We can write the two-dimensional Casimir operator in terms of the Pauli matrices:

$$J^2 = \frac{1}{4}\sigma^2,$$

where

$$\sigma^2 = \sum_{i=1}^3 \sigma_i^2 = 2(\sigma_+\sigma_- + \sigma_-\sigma_+) + \sigma_3^2. \quad (3.23)$$

We now turn to $SU(3)$. The generalization of the Pauli matrices are the so-called Gell-Mann matrices λ_i ($i = 1, 2, \dots, 8$), which are given by

$$\begin{aligned} \lambda_1 &= \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, & \lambda_2 &= \begin{pmatrix} 0 & -i & 0 \\ i & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, & \lambda_3 &= \begin{pmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \\ \lambda_4 &= \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix}, & \lambda_5 &= \begin{pmatrix} 0 & 0 & -i \\ 0 & 0 & 0 \\ i & 0 & 0 \end{pmatrix}, & & \\ \lambda_6 &= \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}, & \lambda_7 &= \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & -i \\ 0 & i & 0 \end{pmatrix}, & \lambda_8 &= \frac{1}{\sqrt{3}} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -2 \end{pmatrix}. \end{aligned} \quad (3.24)$$

The λ_i satisfy the commutation relations

$$[\lambda_i, \lambda_j] = 2if_{ij}^k \lambda_k, \quad (3.25)$$

with summation implied. The f_{ijk} are themselves often called the structure constants of the group. They are given in Table 3.1.

There are two Casimir operators of $SU(3)$, one quadratic and the other cubic in the generators. We shall have occasion to use only the quadratic Casimir operator F^2 , which is given by

$$F^2 = \frac{1}{4}\lambda^2 = \frac{1}{4} \sum_{i=1}^8 \lambda_i^2. \quad (3.26)$$

Table 3.1. Nonvanishing structure constants of $SU(3)$. The f_{ijk} are antisymmetric under permutation of any two indices.

ijk	f_{ijk}	ijk	f_{ijk}
123	1	345	1/2
147	1/2	367	-1/2
156	-1/2	458	$\sqrt{3}/2$
246	1/2	678	$\sqrt{3}/2$
257	1/2		

We now introduce a different representation for the generators, which is suitable to generalization to any $SU(n)$. We introduce the notation H_a , ($a = 1, 2, \dots, n-1$) for the mutually commuting generators (which can be simultaneously diagonalized) and the notation E_{ab} for the $n^2 - n$ nondiagonal generators. These are $n \times n$ matrices with matrix elements given by

$$(H_a)_{jk} = \delta_{jk} \left[\sum_{l=1}^a \delta_{jl} - a\delta_{j,a+1} \right], \quad (3.27)$$

$$(E_{ab})_{jk} = \delta_{aj} \delta_{bk}, \quad a \neq b, \quad (3.28)$$

where δ_{ab} are elements of the unit matrix. If $a > b$, E_{ab} is a lowering operator; if $a < b$, it is a raising operator. Also, $E_{ab} = E_{ba}^\dagger$. In $SU(2)$,

$$H_a = \sigma_3, \quad E_{12} = \sigma_+, \quad E_{21} = \sigma_-. \quad (3.29)$$

We see from Eqs. (3.27) and (3.28) that the $SU(3)$ generators are

$$\begin{aligned}
 H_1 &= \begin{pmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 0 \end{pmatrix}, & H_2 &= \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -2 \end{pmatrix}, & E_{12} &= \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \\
 E_{21} &= \begin{pmatrix} 0 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, & E_{13} &= \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, & E_{31} &= \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix}, \\
 E_{23} &= \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix}, & E_{32} &= \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}.
 \end{aligned} \quad (3.30)$$

It should be clear from the examples we have given for $SU(2)$ and $SU(3)$ that it is straightforward to write down the matrix generators of any $SU(n)$ in both the λ_i and (H_a, E_{ab}) representations (up to normalization constants).

We can use the lowering operators $E_{a+1,a}$ and their hermitian conjugates to obtain the Clebsch-Gordan coefficients of any $SU(n)$. We show in the next section how this is done.

The trace of a matrix is the sum of its diagonal elements. The $n^2 - 1$ generators of $SU(n)$ are traceless matrices in n dimensions. Any real $n \times n$ traceless matrix can be written as a linear combination of them, and any real $n \times n$ matrix can be written as a linear combination of them and the unit matrix in n dimensions.

3.6 Multiplets of unitary groups

The n -dimensional generators of $SU(n)$ operate on n -dimensional column vectors. Clearly, there are n linearly independent vectors, which we may denote by u_a , ($a = 1, 2 \dots n$). A convenient representation for these vectors is that the j -th row of u_a is equal to δ_{aj} . In $SU(2)$ the u_a are

$$u_1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \quad u_2 = \begin{pmatrix} 0 \\ 1 \end{pmatrix}. \quad (3.31)$$

In $SU(3)$ they are

$$u_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \quad u_2 = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, \quad u_3 = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}. \quad (3.32)$$

We can order the vectors from *highest* to *lowest*, such that u_a is *higher* than u_b if $a < b$. The vectors u_a of $SU(n)$ are said to belong to the first *fundamental* representation. Altogether, $SU(n)$ has $n - 1$ inequivalent so-called fundamental representations, two of which have n dimensions, except for $SU(2)$, which has only one fundamental representation. All the multiplets of $SU(n)$ can be built up from the vectors of the first fundamental representation using only the raising and lowering matrices of the n -dimensional representation of the Lie algebra. Sometimes, however, it is convenient to use more than one fundamental representation in building the multiplets.

The eigenvalues of the H_a operating on a state vector is called the *weight* \mathbf{m} of the vector. For example, in $SU(3)$, we see from Eqs. (3.30) and (3.32) that $(H_1, H_2)u_2 = (-1, 1)u_2$. therefore, the weight \mathbf{m} of u_2 is $\mathbf{m} = (m_1, m_2) = (-1, 1)$.

We see from the definitions of E_{ab} and the u_a , that

$$E_{ab}u_c = u_a\delta_{bc}. \quad (3.33)$$

With our ordering of the vectors such that u_a is higher than u_b if $a < b$, we can see that the operators E_{ab} are indeed raising or lowering operators for $a < b$ or $a > b$ respectively. We shall restrict ourselves to the the lowering operators $E_{a+1,a}$ and their hermitian conjugates, which are raising operators.

Let us use the u_a in $SU(2)$ and $SU(3)$ to build up some other multiplets of these groups. We start with $SU(2)$, which is a case which should be familiar. To make things more concrete, u_1 and u_2 can be the spin-up and spin-down wave functions (state vectors) of a particle of spin 1/2. As another possibility, they can be the isospin wave functions of a u and d quark respectively. If we wish to consider N -particle wave functions, we start

with the wave function $u_1(1)u_1(2)\dots u_1(N)$, where the numbers in parentheses stand for particle 1, particle 2, etc. A simpler notation is to omit the numbers in parentheses, and by convention write the wave function of particle 1 first, etc. We shall adopt this simpler notation in the following. We also introduce the notation that the lowering operator E_{ab} operating on an N -particle state is given by

$$E_{ab} = \sum_{i=1}^N E_{ab}(i). \quad (3.34)$$

It is best to begin with only two particles. Let $\chi_1 = u_1u_1$ and operate on χ_1 with the lowering operator

$$E_{21} = E_{21}(1) + E_{21}(2) = \sigma_-(1) + \sigma_-(2). \quad (3.35)$$

We get

$$E_{21}\chi_1 = u_1u_2 + u_2u_1 = \sqrt{2}\chi_2, \quad (3.36)$$

where we have defined χ_2 to be a normalized state. Repeating the operation, we get

$$E_{21}\chi_2 = \sqrt{2}u_2u_2 = \sqrt{2}\chi_3. \quad (3.37)$$

If we operate on χ_3 we get 0. Thus, starting from two doublets of $SU(2)$, we have obtained a triplet state χ_i , corresponding to spin or isospin 1. We next construct a state ϕ_1 orthogonal to χ_2 . We see that ϕ_1 must be given by

$$\phi_1 = (u_1u_2 - u_2u_1)/\sqrt{2}. \quad (3.38)$$

If we operate on ϕ_1 we get 0, as we expect, so that this state is a singlet, corresponding to spin or isospin 0. We can obtain the eigenvalues of the diagonal operators J_3 and J^2 by directly operating on the χ_i and ϕ_1 .

The coefficients multiplying the product wave functions u_iu_j in the expressions for χ_i and ϕ_i are known as Clebsch-Gordan coefficients. In the case we have considered, these Clebsch-Gordan coefficients are unique, but in the case of the product of three or more wave functions, the Clebsch-Gordan coefficients can depend on somewhat arbitrary definitions of wave functions. We can see this as follows: If we start with the product $u_1u_1u_1$, we can use the lowering operator E_{21} to construct all the symmetric wave functions belonging to the same multiplet as $u_1u_1u_1$. The problem arises when we want to construct other multiplets. For example, consider the (unnormalized) wave function ξ given by

$$\xi = u_1u_1u_2 + u_1u_2u_1 + u_2u_1u_1. \quad (3.39)$$

There are *two independent* wave functions orthogonal to ξ , and the Clebsch-Gordan coefficients depend on what linear combination of these wave functions we choose. The choice in some instances is given by convention, but real questions of physics can influence what choice is convenient.

Let us now generalize to $SU(3)$. The χ_i and ϕ_1 from $SU(2)$ are also wave functions of $SU(3)$, but they are not complete multiplets. We can operate on χ_2 with E_{32} to get

$$\chi_4 = (u_1u_3 + u_3u_1)/\sqrt{2}. \quad (3.40)$$

Operating on χ_4 with E_{21} , we get

$$\chi_5 = (u_2u_3 + u_3u_2)/\sqrt{2}, \quad (3.41)$$

and operating on χ_5 with E_{32} , we get after normalizing

$$\chi_6 = u_3u_3. \quad (3.42)$$

Thus, the χ_i are a sextet of $SU(3)$. Likewise, from ϕ_1 , we can obtain

$$\phi_2 = u_1u_3 - u_3u_1, \quad \phi_3 = u_2u_3 - u_3u_2, \quad (3.43)$$

so that the ϕ_i are a triplet of $SU(3)$.

We now define a new triplet of vectors \bar{u}_a given by

$$\bar{u}_1 = \phi_3, \quad \bar{u}_2 = -\phi_2, \quad \bar{u}_3 = \phi_1. \quad (3.44)$$

By explicit construction we find

$$E_{ab}\bar{u}_c = -\bar{u}_b\delta_{ac}. \quad (3.45)$$

This differs from the action of E_{ab} on u_c . The \bar{u}_a are the vectors belonging to the second fundamental representation of $SU(3)$.

3.7 Young tableaux

A *Young tableau* or *Young diagram* describes the symmetry of a collection of an integer ν identical particles, each of which can be in one of several available states. We shall confine our considerations to the description of Young diagrams in the case that the symmetry group is $U(n)$ or $SU(n)$ and the particles belong to the first fundamental representation of the group. Then the number of possible states of a particle is also n . An example is a collection of five electrons, each of which can be in one of two spin states (spin up or spin down). In this case $\nu = 5$, $n = 2$.

A Young tableau is a collection of boxes, one for each particle, arranged in rows and columns to represent the symmetry of the state: symmetric in rows, antisymmetric in columns. A *proper* tableau is one in which each row is no longer than the one above it and each column is no longer than the one to the left of it. When we refer to a Young tableau, we shall mean a proper tableau unless otherwise specified. An example of a proper tableau is the following:

$$\begin{array}{|c|c|c|c|}
 \hline
 & & & \\
 \hline
 & & & \\
 \hline
 & & & \\
 \hline
 & & & \\
 \hline
 \end{array} \quad (3.46)$$

We begin by using Young tableaux in connection with $SU(2)$. We denote the basis vectors of the fundamental doublet of $SU(2)$ by u_1 and u_2 . These vectors may, for example, denote the two states of a particle with spin $1/2$. Another notation for these one-particle states is by means of a Young tableau with a single box. We make the identification

$$u_1 = \boxed{1}, \quad u_2 = \boxed{2} \quad (3.47)$$

The single box without a number stands for both members of the doublet. The same considerations hold for $U(2)$.

Now suppose we have a two-particle state. If it is a symmetric ψ_s , we denote it by a row, and if it is an antisymmetric state ψ_a , by a column:

$$\psi_s = \boxed{\quad\quad}, \quad \psi_a = \begin{array}{|c|} \hline \square \\ \hline \square \\ \hline \end{array} \quad (3.48)$$

These tableaux represent multiplets which are different than the two-dimensional fundamental multiplet. Consider first the symmetric state. If both particles are in the state u_1 or both are in the state u_2 , the corresponding tableaux are

$$\boxed{1\ 1}, \quad \boxed{2\ 2}.$$

There is one symmetric state with one particle having the state vector u_1 and the other having the vector u_2 , namely $(u_1u_2 + u_2u_1)/\sqrt{2}$. (We adopt the convention of writing the state vector of the first particle at the left, the vector of the second particle next, etc. This convention saves us the trouble of writing $[u_1(1)u_2(2) + u_2(1)u_1(2)]/\sqrt{2}$.) This symmetric state is represented by the tableau

$$\boxed{1\ 2}.$$

The arrangement

$$\boxed{2\ 1}$$

is obviously the same as the previous arrangement, and must not be counted. It is called a *nonstandard* arrangement. Thus, the symmetric state is a triplet. There is only one antisymmetric two-particle state $(u_1u_2 - u_2u_1)/\sqrt{2}$, corresponding to the arrangement

$$\begin{array}{|c|} \hline 1 \\ \hline 2 \\ \hline \end{array}.$$

The other arrangement

$$\begin{array}{|c|} \hline 2 \\ \hline 1 \\ \hline \end{array}$$

is nonstandard and must not be counted.

The above considerations for $U(2)$ or $SU(2)$ can be generalized to any $U(n)$ or $SU(n)$ and lead us to the following definition:

A *standard arrangement* of a tableau of $U(n)$ or $SU(n)$ is a proper tableau containing a positive integer i in each box ($1 \leq i \leq n$) such that the integers increase in going from top to bottom in a column and do not decrease in going left to right in a row. Hereafter, unless we explicitly state otherwise, an arrangement will mean a standard arrangement.

An important theorem which we do not prove is that the number N of standard arrangements of a Young tableau with positive integers no greater than n is equal to the dimension of an irreducible representation of $U(n)$ or $SU(n)$.

We see that a Young tableau for $U(n)$ or $SU(n)$ consists of ν boxes in no more than n rows. The tableaux are limited to n rows because one cannot antisymmetrize a configuration of more than n particles when each particle has only n available states.

As an example, consider a collection of five electrons, each of which can be in one of two spin states (spin “up” u_1 or spin “down” u_2). In this case $\nu = 5$, $n = 2$, and the symmetry group of the spins is $SU(2)$. If we include the lepton number in our description (an additive quantum number), then the symmetry group is $U(2)$ or $SU(2) \times U(1)$. (We do not distinguish between these last two groups, as we use only the Lie algebra, which is the same for both.) The dimensionality is the same whether the lepton number is included in the description. Electrons must obey Fermi statistics, that is, their state vectors must be antisymmetric under the interchange of *all* the coordinates of any two electrons. However, the symmetry under the interchange of only the spins is given by any standard Young tableau.

Consider a three-electron state belonging to the Young tableau



This state has the same multiplicity as the one-electron state belonging to the tableau



but the lepton number is 3 in the first case and 1 in the second. If we are interested only in the $SU(2)$ of the spins and not in the lepton number, the two diagrams are equivalent in that they correspond to the same quantum numbers.

To avoid the complication of the quantum number associated with the $U(1)$ subgroup of $U(n)$, we often restrict ourselves to $SU(n)$. Then all columns with n boxes may be removed from a given tableau, as there is only one way to antisymmetrize a state of n particles, each of which has n degrees of freedom. The number of states N_n is equal to the result obtained by counting the number of ways one can put positive integers $\leq n$ in the remaining boxes, consistent with the rules that numbers must not decrease going from left to right in any row and must increase going from top to bottom in any column.

We see that in the case of $SU(n)$, a tableau with ν boxes can also denote states of the corresponding symmetry containing a different number of particles. We can therefore divorce the concept of the symmetry of a state corresponding to a Young tableau from the concept of particles belonging to the first fundamental representation.

A tableau with no more than $n - 1$ rows can be specified by a set of $n - 1$ integers p_i , which is the number of boxes in row i minus the number of boxes in row $i + 1$. The multiplicity N_n of any diagram is a function of the p_i . It is a complicated combinatorial problem to find the number of standard arrangements of a tableau with a given \mathbf{p} . The formula for $N_n(p_1, p_2 \dots p_{n-1}) = N_n(\mathbf{p})$ is known for any n but we write it down just for $n = 2$ and $n = 3$. For $SU(2)$ we have

$$N_2(p) = p + 1. \quad (3.49)$$

The number of states of a given angular momentum j is $2j + 1$. Then, using Eq. (3.40), we can make the identification

$$p = 2j. \quad (3.50)$$

For $SU(3)$ the formula is

$$N_3(\mathbf{p}) = (p_1 + 1)(p_2 + 1)(p_1 + p_2 + 2)/2. \quad (3.51)$$

These formulas give the number of states in a multiplet belonging to an irreducible representation of the group with the symmetry specified by the Young tableau \mathbf{p} .

We see that for $SU(2)$ the number N_2 can be any positive integer. However, for $SU(n)$ with $n > 2$, the numbers N_n include only a proper subset of the positive integers. For example, in $SU(3)$, the numbers N_3 have the values 1, 3, 6, 8, 10, 15, etc, as determined either by counting the standard arrangements of Young tableaux or from Eq. (3.51), substituting non-negative integers for p_1 and p_2 .

All the formulas $N_n(\mathbf{p})$ are symmetric under the interchange

$$p_i \leftrightarrow p_{n-i},$$

that is,

$$N_n(p_1, p_2 \dots p_{n-1}) = N_n(p_{n-1} \dots p_1). \quad (3.52)$$

Two Young tableaux which transform into each other under this transformation are called *conjugate tableaux*, and the irreducible representations which act on them are called *conjugate representations*. The first fundamental representation of $SU(3)$ is characterized by $(p_1, p_2) = (1, 0)$; the second fundamental representation is the conjugate representation, and is characterized by $(p_1, p_2) = (0, 1)$. However, it is common to characterize these (and other representations) by a single number which gives their multiplicity: in the case of the first and second representations, we use 3 and $\bar{3}$ respectively. Conjugate representations have the same dimensionality, but conjugate tableaux do not necessarily have the same number of boxes. If a tableau is unchanged under the transformation of interchanging p_i and p_{n-i} , it is self-conjugate, and likewise for the corresponding representation. Since all representations of $SU(2)$ consist of a single row, all representations of $SU(2)$ are self-conjugate.

We now show how to build all irreducible representations of $SU(n)$ starting from the first fundamental one. First we note that a product of two or more representations is in

general reducible. We can see this by considering the basis vectors on which they act. It is simplest to begin with $SU(2)$. The product states are written using Young tableaux as

$$\square \times \square$$

These stand for the four product states

$$u_1 u_1, \quad u_1 u_2, \quad u_2 u_1, \quad u_2 u_2.$$

But we know that to obtain the basis vectors of irreducible representations we must take the linear combinations which correspond to the symmetric and antisymmetric Young tableaux. This result is true in any $SU(n)$. We write

$$\square \times \square = \begin{array}{|c|c|} \hline \square & \square \\ \hline \end{array} + \begin{array}{|c|} \hline \square \\ \hline \square \\ \hline \end{array} \quad (3.53)$$

In $SU(2)$, the multiplicities are

$$SU(2): \quad 2 \times 2 = 3 + 1. \quad (3.54)$$

In $SU(3)$, we have

$$SU(3): \quad 3 \times 3 = 6 + \bar{3}, \quad (3.55)$$

and in any $SU(n)$ we have

$$SU(n): \quad n \times n = n(n+1)/2 + n(n-1)/2. \quad (3.56)$$

In the above examples we have found the irreducible representations contained in the product of two irreducible representations. This decomposition is called the *Clebsch–Gordan series*. If the decomposition contains no representation more than once, the product is called *simply reducible*. If $n > 2$, the decomposition of $n \times \bar{n}$ is different from the decomposition of $n \times n$. We have

$$SU(n): \quad n \times \bar{n} = (n^2 - 1) + 1, \quad (3.57)$$

which is different from the decomposition given in Eq. (3.57). In particular, in $SU(3)$ we have

$$3 \times \bar{3} = 8 + 1. \quad (3.58)$$

We now tell how to find the Clebsch–Gordan series for the product of any two representations of $SU(n)$. We do not give the proof, as it is quite complicated (even the recipe is complicated).

Recipe. We draw the Young tableaux of the two representations, marking each box of the second with the number of the row to which it belongs. We then attach the boxes of

the second tableau in all possible ways to the first tableau, subject to the following rules for the combined tableaux:

- 1) Each tableau should be proper.
- 2) No tableau should have a column with more than n boxes, and we can remove all columns with n boxes.
- 3) The numbers must not decrease from left to right in a row.
- 4) The numbers must increase from top to bottom in a column.
- 5) We can make a path by moving along each row from the right, starting at the top. At each point of the path, the number of boxes encountered with the number i must be less or equal to the number of boxes with $i - 1$.

As an example, if we follow the rules, we find the irreducible representations contained in 8×8 of $SU(3)$ to be

$$8 \times 8 = 27 + 10 + \bar{10} + 8 + 8 + 1. \quad (3.59)$$

We see that the Clebsch–Gordan series contains two equivalent representations, namely, the two 8’s. This means that the product of 8×8 is not simply reducible.

The product of any two representations of $SU(2)$ is simply reducible, but this result does not hold for any $n > 2$. Even for $SU(2)$, the product of three representations is not necessarily simply reducible. For example, in $SU(2)$, we have

$$2 \times 2 \times 2 = 4 + 2 + 2. \quad (3.60)$$

If two or more equivalent representations appear in the reduction of a product, the use of group theory alone is not sufficient to enable us to label the states. We must know something of the physics of the problem in order to obtain the most useful labeling. The Clebsch–Gordan coefficients are also not determined uniquely without additional input.

For example, the reduction given in Eq. (3.60) can come about from obtaining the irreducible multiplets from three electron spins. One of the two doublets arises by combining the spins of the first two electrons to give a triplet and then combining the third spin to give a doublet. The other doublet arises from combining the first two spins to form a singlet and then combining the third spin to give a doublet. Group theory says nothing about how the first two spins in fact combine, as that depends on the forces involved. As another example, in $SU(3)$ we have

$$3 \times 3 \times 3 = 10 + 8 + 8 + 1. \quad (3.61)$$

The way the two 8’s are chosen in either the decomposition (3.59) or (3.61) depends on the physics.

3.8 Evaluating quadratic Casimir operators

The group $SU(n)$ has $n - 1$ Casimir operators, one of which is quadratic in the generators. The quadratic Casimir of $SU(n)$ may be evaluated by making use of the raising and lowering operators. For $SU(2)$ this operator is J^2 and is given by

$$J^2 = \frac{1}{4}(2\sigma_+\sigma_- + 2\sigma_-\sigma_+ + \sigma_z^2). \quad (3.62)$$

We take advantage of the fact that J^2 is an invariant operator to operate on the particular state for which σ_+ vanishes. We eliminate the term $\sigma_+\sigma_-$ by making use of the commutation relation

$$[\sigma_+, \sigma_-] = \sigma_z. \quad (3.63)$$

This enables us to write J^2 (when operating on our particular state) only in terms of the diagonal operator σ_z . Since σ_z is an additive quantum number, it is straightforward to evaluate J^2 for any representation. Details are given in books on quantum mechanics. The answer is $J^2 = j(j+1)$, or, since $p = 2j$,

$$J^2 = \frac{1}{2}p\left(\frac{1}{2}p + 1\right). \quad (3.64)$$

The same method works for any $SU(n)$ except that the algebra is more complicated. In particular, for $SU(3)$ we get

$$F^2 = p_1 + p_2 + \frac{1}{3}(p_1^2 + p_2^2 + p_1p_2). \quad (3.65).$$

1. SOME BASIC IDEAS

In this chapter we introduce briefly and with minimal explanation a large number of our present ideas about elementary particle physics. Put simply, we describe the subject, not necessarily the way it is, but the way we now think it is. In later chapters we expand upon and clarify the more complicated of the ideas we outline here.

1.1 What is an elementary particle?

Our understanding of bulk matter is that it is composed of microscopic entities called molecules. However, we know that molecules are not elementary particles, but are composed of atoms. The atoms in turn are not elementary, but are made of atomic nuclei and electrons. The nuclei are made of nucleons (protons and neutrons), and the nucleons are made of quarks. Perhaps the quarks and electrons are elementary, and perhaps not.

There are three properties that we associate with an elementary particle. First, the particle should not be a composite of other particles. Second, the particle should not have a detectable size. Third, there should be no way to distinguish between two examples of an elementary particle, i.e, they should be identical or indistinguishable in their properties, such as charge or mass. (The orientation of the spin may vary, but not the magnitude of the spin.)

One can see from the way these criteria are stated that as more experimental evidence is accumulated, a particle may lose its status of being elementary. There are also times when a theorist might like to treat a body as if it were elementary, for example, as a point mass, when it clearly is not an elementary particle. We know from observation that the earth is a composite object, made up of many simpler things. But if we are considering the motion of the earth in the gravitational field of the sun, then it is a good first approximation to treat the earth as an elementary particle, although obviously it is not. To a much better approximation, if we want to calculate the motion of stars moving in a galaxy, we can treat the stars as pointlike elementary particles. Of course, if we are interested in understanding the light coming from the stars, we must treat them as composite. Likewise, if we are interested in the motion of the tides on earth, we must regard the earth as composite.

In looking at many aspects of the kinetic theory of gases, we may consider molecules to be elementary. However, the failure of real gases at high density to obey the ideal gas equation is most easily explained if molecules have sizes greater than zero. A particle with a size greater than zero is a natural candidate for a composite object, although a greater-than-zero size does not in itself prove that a particle is composite.

On the scale of atomic physics, to a good approximation the nucleus may be treated as a point. However, in the most accurate calculations of atomic properties, the fact that the nucleus has a size greater than zero must be taken into account. At the present time, we have good evidence that nuclei are composite particles made primarily of protons and neutrons.

High-energy electron-proton elastic scattering experiments show that protons and neutrons in turn are not pointlike. Now we believe that nucleons are composite and are made primarily of quarks. At our present level of knowledge, we treat quarks as elementary particles. However, we do not know whether future developments, either in theory or experiment, will make it useful to regard quarks as composite particles.

According to our present understanding, a necessary, but not sufficient, condition for a particle to be elementary is that two of the same kind must be truly indistinguishable. This means we believe that no experiment can be designed which can find an intrinsic property of one which is different from that same property in the other. With sufficiently precise experiments, two dust particles, say, can be distinguished, so dust particles can't really be elementary particles. However, to the best of our knowledge, in no experiment can two electrons be distinguished, so the electron is a candidate for an elementary particle. In fact, at the present time, physicists believe that the electron is elementary. On the other hand, just as with two electrons, we cannot distinguish between two hydrogen atoms, but we now know that a hydrogen atom is a composite particle, made of a proton (itself a composite particle) and an electron.

To sum up, we depend on both theory and experiment to help us decide whether a particle should be regarded as elementary or composite. It may be convenient to treat even a composite particle as elementary in some contexts. The classification of a particle as elementary is always provisional. If a particle has a size which is measured to be greater than zero, we may suspect that the particle is composite. However, in the absence of a successful theory which gives the composite structure of a particle in terms of simpler entities, we may usefully continue to treat the particle as elementary.

1.2 Interactions

Since the time of Isaac Newton, the acceleration of particles has been ascribed to forces acting upon them. We can generalize the notion of a force to an *interaction*, which can create and annihilate particles and cause them to decay into other particles, as well as causing them to accelerate.

According to our present ideas, to every kind of elementary particle that exists in nature, there is also an antiparticle. We do not know for sure whether in our universe there exist equal numbers of particles and antiparticles, or whether there is an excess of particles. Because particles and antiparticles can annihilate, there will exist very few antiparticles in a region with an excess of particles. We have no evidence that any region in the visible universe is made up primarily of

antiparticles, and so most physicists assume provisionally that the universe now contains an overwhelming excess of particles. Whether that state of affairs also existed in the early universe is by no means clear.

On a phenomenological level, we have firm evidence for the existence of four different interactions, known as electromagnetic, gravitational, weak, and strong. In the 19th century, after Maxwell's theory of electromagnetism, physicists knew of only two forces, the gravitational and the electromagnetic. The weak and the strong interactions were just discovered during the 20th century. We should not necessarily think that we live at a privileged time in which it is given to us to know of all of nature's forces; there may exist other interactions of which we are still unaware. On the other hand, it is now known that the weak and electromagnetic interactions, although phenomenologically quite different from each other, are actually related. It is an attractive idea that all interactions are different manifestations of one fundamental interaction, and there have been many attempts, so far without marked success, to unify the known interactions.

The gravitational interaction has the property that the strength of the force between two particles is proportional to the product of their energies. At energies presently accessible in the laboratory, the gravitational interactions between elementary particles are very small compared to the strength of their other interactions, and so can usually be safely neglected. For this reason, we emphasize the remaining three interactions, strong, electromagnetic, and weak, all of which play important roles in the behavior of elementary particles.

Each of the interactions is characterized by its strength and range. The electromagnetic and gravitational interactions are long range, while the strong and the weak interactions are short range. Of the four interactions at laboratory energies, the strong interaction has the greatest strength, the electromagnetic and weak interactions have comparable strengths, and the gravitational interaction is by far the weakest. Phenomenologically, the weak interaction appears to be much weaker than the electromagnetic owing to the short range of the former and the long range of the latter.

According to our present ideas, the interactions between particles (matter) are carried by fields. The strength of an interaction is measured by the magnitude of the coupling of matter to the field that transmits the interaction. The effective coupling strengths of the strong, electromagnetic, and weak interactions are not constants even though they are often called coupling constants, but depend on the distance between the particles. If an interaction between particles falls off approximately exponentially with distance, it is said to be short range, and its range is the characteristic distance appearing in the exponent. If an interaction falls off approximately as the inverse of the distance, it is called a long-range interaction.

The ideas of strength and range which we have given here are considerably oversimplified, as they do not take into account the notion of forces so strong that they confine certain particles into the interior of others. We discuss these concepts

more fully in later chapters.

1.3 Symmetry and conservation laws

The universe gives the impression of unceasing change; some of the changes are apparently haphazard, like certain fluctuations in the weather, while others are regular, like the rotation of the earth. However, it is remarkable that, amid all the change in nature, we are able to identify certain properties which remain constant as time goes on—the conservation laws. Among these laws are the conservation of energy, momentum, and angular momentum. Others are the conservation of certain charges, the most familiar of which is electric charge. These conservation laws are, as far as we know, exact, while others, like the conservation of parity, are only approximate.

Another remarkable fact is the symmetry of the interactions of nature under certain transformations. By the symmetry of a physical system under a transformation we mean that the system is invariant under the transformation. Examples of symmetries which, as far as we know, are exact when gravity can be neglected, are invariances under translation, rotation, and time translation. Other symmetries, like space reflection and time reversal, are only approximate in nature.

The transformations carried out on a physical system have the following properties:

- (1) A given transformation (such as a rotation) followed by another transformation, is a unique transformation.
- (2) If we make three transformations, we get the same result whether we combine the first two and follow it with the third or do the first one and follow it by the combination of the second and third. (This is an associative property.)
- (3) We can leave the system untransformed. (This is an identity transformation.)
- (4) After making a transformation, we can transform back again to the original state. (This is an inverse transformation.)

These properties of transformations are just the properties of a mathematical group, and so the study of group theory can help us shed light on the properties of physical systems which are invariant under certain groups of transformations. A discrete symmetry, like reflection, corresponds to a finite group, while a continuous symmetry, like a rotation through any angle, corresponds to a continuous group. We discuss group theory in more detail in Chapter 3.

There is a close connection between symmetries and conservation laws, expressed by a beautiful theorem due to Emmy Noether. Noether's theorem says that if the Lagrangian of a classical system is invariant under a continuous group of transformations, then the system has a conserved quantity (sometimes called a "current").

Some classical theories have been generalized to quantum mechanics and quantum field theory. Then under normal circumstances, if the classical form of the theory has a symmetry and a corresponding conservation law, so does the

quantum mechanical generalization. However, this is not always true, and in those cases in which it is false, the quantum version of the theory is said to have an *anomaly*.

If a symmetry and its conservation law are only approximate, we say the symmetry is broken. We mention three ways in which a symmetry may be broken:

- (1) The symmetry is broken by an anomaly.
- (2) The Lagrangian of the system contains a large term, which respects the symmetry, and a small term, which does not.
- (3) Although the Lagrangian of a system respects the symmetry, the lowest-energy state does not. In this case, the symmetry is said to be “hidden” or “spontaneously broken.”

In the theory of elementary particles, all three kinds of symmetry breaking occur. Chapter 4.

1.4 The standard model of elementary particles

We treat particle physics within the framework of the so-called *standard model* of elementary particles. In other areas of knowledge, there are other standard models, like the standard solar model and the standard big-bang model. In this book, by “the standard model” we mean the standard model of elementary particles. The mathematical tools necessary (but, unfortunately, not sufficient) for a deep understanding of the standard model are quantum field theory and group theory. In this section, we briefly discuss quantum field theory and some of the invariance groups of the standard model. In later chapters we amplify this discussion and also define and treat the standard model in considerable detail.

The standard model is a *relativistic quantum field theory* which describes the strong, electromagnetic, and weak interactions of elementary particles. It is a *realistic* field theory, by which we mean that it is described by a Lagrangian which exists in four space-time dimensions, satisfies the postulates of special relativity, is local and causal, and has a lowest-energy state, called the *ground state* or *vacuum*. By a local theory, we mean one in which the quanta of the theory are pointlike particles which interact at a point in space-time. Particles separated from one another interact via the field whose quanta carry the force in such a way as to satisfy causality: namely, no signal can propagate faster than the speed of light.

The Lagrangian of the standard model, like that of any realistic field theory, is invariant under certain continuous groups of transformations: translations, rotations, time translations, and proper Lorentz transformations. These transformations taken together constitute the proper Poincaré group. A realistic field theory allows for the creation and annihilation of particles and antiparticles, in agreement with our observation that such processes occur.

The standard model is invariant under several groups of transformations in addition to the transformations of the Poincaré group. Among these transforma-

tion groups are certain groups of local *gauge transformations*. A theory which is invariant under a group of local gauge transformations is called a *gauge theory*.

We now briefly describe what a gauge theory is. The Lagrangian describing a field theory is a function of fields which are defined at each space–time point (see Chapter 4). Each field has a phase, which may be different at each point. The Lagrangian of a gauge theory has two kinds of fields, matter fields and gauge fields. The phases of the matter fields can be varied locally, that is, by an amount which is different at each point in space–time. In a gauge theory this variation does not change the Lagrangian because the gauge fields can be transformed in a way to cancel any new terms arising from the transformation of the matter fields. A combined transformation of the matter fields and the gauge fields which differs from point to point is called a local gauge transformation.

We shall see in Chapter 4 that local gauge invariance is a very powerful principle which leads to the specification of the form of the interactions of the theory.

In order to describe the gauge group of the standard model, we need first to define unitary and special unitary groups. A unitary matrix in n dimensions is an $n \times n$ matrix having the property that its inverse is equal to its Hermitian conjugate. A special unitary matrix is one with determinant unity. It can be proved that the unitary matrices in a given number of dimensions form a group, and likewise so do the special unitary matrices. The unitary group $U(n)$ is the group of unitary matrices in n dimensions, and the special unitary group $SU(n)$ is the group of unitary matrices with determinant unity. We shall describe these groups in more detail in Chapter 3.

The gauge group of the standard model is the *product* (see Chapter 3) of three groups of unitary matrices, namely, $SU(3) \times SU(2) \times U(1)$. The gauge group $SU(3)$ is the symmetry group of the strong interactions, described by a theory called *quantum chromodynamics* (QCD), while the group $SU(2) \times U(1)$ is the group of the partially unified electromagnetic and weak, or *electroweak*, interactions. The electromagnetic part of the electroweak interaction is *quantum electrodynamics* (QED), which at the present time is our most successful quantum field theory. Agreement between the predictions of QED and experimental measurements are sometimes better than 1 part in a billion.

According to the standard model, the electromagnetic and weak interactions are related because the groups $SU(2)$ and $U(1)$ are mixed by mechanism known as *spontaneous symmetry breaking*, which we shall describe in Chapter 4. The combined electroweak field theory is due to Sheldon Glashow, Steve Weinberg, Abdus Salam, and others. The theory of the strong interactions, QCD, is the work of many physicists, including Y. Nambu, Murray Gell-Mann, H. Fritzsch, H. Leutwyler, Weinberg, David Politzer, David Gross, and Frank Wilczek.

In addition to having continuous symmetries, the standard model has a discrete symmetry known as *CPT*. In order to define *CPT* symmetry, we first define the operations of charge conjugation C , parity P , and time reversal T . Charge

conjugation replaces the charges carried by a particle by those of the antiparticle. In addition, C replaces certain internal additive quantum numbers by their negatives (see Section 1.9). Parity P , or space reflection, replaces the spatial coordinates of the fields by their negatives. Finally, T replaces the time coordinate of the field by its negative. The Lagrangian of the standard model is not invariant under the separate operations C , P , and T , but it is invariant under the combined operations of all three, taken in any order. This combined invariance is known as CPT symmetry.

There is a theorem, known as the CPT theorem, that the Lagrangian of any local field theory which is invariant under proper Poincaré transformations and has a vacuum state is also invariant under CPT . So far, we have not observed any violation of CPT invariance in nature. If we do in the future, the most probable explanation would be that nature is not described by a local field theory.

A realistic field theory has as a consequence a connection between spin and statistics, called the spin-statistics theorem. So far, all particles observed in nature are either bosons (which satisfy Bose–Einstein statistics) or fermions (which satisfy Fermi–Dirac statistics). Hypothetical particles obeying more complicated statistics have been discussed in the literature, but, so far, there is no evidence for their existence. The kind of statistics obeyed by a collection of identical particles is a consequence of the symmetry of its wave function. The wave function of a collection of identical bosons is symmetric under the interchange of all the coordinates of any two of them. In contrast, the wave function of a collection of identical fermions is antisymmetric under the interchange of all the coordinates of any two. As a consequence, identical fermions obey the Pauli exclusion principle.

The spin-statistics theorem says that if particles are restricted to be either bosons or fermions, then particles of integral spin are bosons and particles of half-integral spin are fermions. This theorem is also in accord with observation.

At the present time, nobody knows how to solve a realistic quantum field theory exactly. Nevertheless, we are able to solve such a theory approximately in a perturbation expansion in powers of the interaction strength (or coupling constant). It turns out that infinite terms arise in the expansion of many field theories, but it is possible to circumvent this problem by a process known as *renormalization*. The renormalization procedure involves the introduction of a number of parameters whose values must be determined by a comparison of the theory with experiment. One then obtains a unique finite answer for each observable quantity which is calculated. For realistic field theories, renormalization is always possible with the introduction of a finite number of parameters, and such theories are said to be renormalizable.

The perturbation expansion of a field theory can be schematically pictured by means of *Feynman diagrams* or *graphs*. A Feynman graph is a picture in space-time describing the interactions of particles. In order to be able to draw a graph in a plane, one shows only one of the three spatial dimensions. A graph contains lines (which describe the motion (or propagation) of free particles and vertices

which describe interactions at points in space–time. We shall describe Feynman graphs in more detail in Chapter 7.

Although the standard model has much predictive power, many physicists regard it as an approximation to a more fundamental theory which is yet to be discovered. One reason for this belief is the fact that the standard model requires at least 18 parameters for its specification. Another reason is that the model is not as beautiful as we like a fundamental theory to be.

1.5 Elementary particles of the standard model

The principal elementary particles of the standard model are *quarks* and *leptons*, which are fermions of spin $\frac{1}{2}$, and *gauge bosons*, of spin 1. Also, a *Higgs boson*, of spin 0, exists in the model. The quarks and leptons are quanta of so-called *spinor* (spin 1/2) fields, the gauge bosons are quanta of *vector* (spin 1) gauge fields, and Higgs bosons are quanta of a *scalar* (spin 0) field. The Higgs boson and one of the quarks (the top quark) have not yet been observed.

Although quarks, leptons, gauge bosons, and the Higgs particle are the elementary particles of the standard model, they may be composite in another theory. Furthermore, in the future they may be discovered to have complicated structure. So far, however, all experiments to measure the size of the elementary particles of the standard model have just obtained upper limits. The upper limits derived from experiment are somewhat model dependent, but for the quarks and leptons are in the region 10^{-16} to 10^{-18} cm.

In the standard model, the two kinds of elementary spin- $\frac{1}{2}$ fermions, leptons and quarks, are supposed to be point-like. These elementary fermions carry one or more kinds of *charge*, which are measures of the strength of the interaction of the fermions with the gauge fields. These charges are not only the ordinary electric charge, but also the weak and strong charge, the last being called *color*. The spin-1 bosons of the standard model, which are the quanta of the gauge fields, are the photon γ , which interacts with electrically charged particles, the weak bosons W^+ , W^- , and Z^0 , which interact with particles which carry weak charge, and eight colored gluons G_i , which interact with particles which carry strong charge, or color. Just as a particle without electric charge is said to be electrically neutral, a particle without strong charge is said to be color-neutral (or colorless).

A major distinction between the leptons and quarks is that the quarks carry color, whereas the leptons do not. This is a way of saying that quarks, but not leptons, have strong interactions. All quarks and all leptons carry weak charge, and all quarks and half the leptons carry electric charge.

Usually, it is stated that there are two kinds of electric charge, positive and negative, but we prefer to take the point of view that there is only one kind of electric charge, plus its anticharge, which is opposite in sign to the charge. This point of view is more easily generalized to other kinds of charges, like strong and weak charges, which occur in more than one variety. Unfortunately, the generalization from electric charge to the charge in more complicated theories is

not completely trivial. There are at least three ways various authors use the word charge and sometimes one has to be careful about the usage. Once we get to the discussions of group theory and gauge theory any confusion should be alleviated. Often, charge is used to refer to the coupling constant of the theory. In that case, there is one for each component of the gauge theory. $SU(3) \times SU(2) \times U(1)$ has three separate components and there are three coupling constants, one for the strong group $SU(3)$ and two for the electroweak theory $SU(2) \times U(1)$.

A second way of looking at a charge is as the integral of a current *a la* Emmy Nother. In this case, there is one charge for each parameter that describes the gauge group. The color group $SU(3)$ is an 8 parameter group, so there are eight kinds of charges. The *weak isospin* group $SU(2)$ is a three parameter group, and the *weak hypercharge* $U(1)$ has one more charge. These four charges are the sources for the W^+ , W^- , Z and photon. In the literature, you are likely to see references to a *charged* current and a *neutral* current, which have different strengths because of spontaneous symmetry breaking. The terms “charged” and “neutral” here refer to the electric charge carried by the weak currents. There are actually two charged currents of charge ± 1 which couple to the W^+ and W^- . The neutral current couples to the Z^0 . The fourth current is also neutral is the familiar electromagnetic current to which the photon couples.

Returning to $SU(3)$, we find the third way in which the word charge is used. The quarks are fundamental objects that are part of the source for the color fields. (The gluons themselves are also sources of the color fields.) The quarks have a label that can take three values. So, we often say that the quarks have three kinds of strong charge or color. These colors are conventionally called “red,” “green,” and “blue,” although sometimes other colors are chosen. There are also three anticolors. Particles carrying color come in color multiplets, with a multiplicity greater than unity, whereas color-neutral particles come in color singlets. This is a somewhat different usage of the word charge, and perhaps we should just say that there are three color states, not three color charges, but this usage is common in the literature and I hope you will not find yourself confused.

We next consider what charges are carried by the gauge bosons. The photon is not only electrically neutral, but also does not carry either weak or strong charge. The eight gluons are electrically neutral and do not carry weak charge, but they do carry color and therefore interact strongly with one another, as well as with quarks. The weak bosons W^+ and W^- carry electric charge and so interact with photons, but the Z^0 is electrically neutral. None of the weak bosons has color, but all three carry weak charge. We amplify this in Chapter 5.

Not only do the quarks and leptons carry charges, but they also have another property, called *flavor*, which distinguishes them. At the present time, there is good experimental evidence for six lepton flavors and five quark flavors. There is indirect evidence for a sixth quark flavor.

The evidence for a sixth quark flavor is theoretical, coming from the standard model, which requires the sixth quark to exist. Within the standard model,

certain calculations agree rather well with experiment if a sixth quark exists, but disagree with experiment in the absence of a sixth quark. In fact, experimental measurements of certain quantities have led to theoretical estimates, based on the standard model, of the mass of the sixth quark. Other calculations are insensitive to whether a sixth quark exists.

The quarks have been given whimsical names: up (u), down (d), strange (s), charmed (c), bottom or beauty (b), and top or truth (t). The three charged leptons are the electron (e), the muon (μ), and the tauon (τ). Each has a neutral companion called a neutrino (ν). Their names are the electron neutrino (ν_e), the muon neutrino (ν_μ), and the tauon neutrino (ν_τ).

In principle, each flavor of quark and lepton can have a different mass. However, in the simplest version of the standard model, the three neutral leptons, the neutrinos ν_e , ν_μ , and ν_τ , all have zero mass. Nevertheless, there is no particular reason within the standard model for the neutrinos to have zero mass. Therefore, if it should turn out that any of the neutrinos has mass greater than zero, as some preliminary evidence indicates, the model can easily be modified to accommodate such a mass. In fact, the mass of each fermion is simply a free parameter in the standard model. The fact that the model has so many free parameters is an advantage in terms of flexibility but a disadvantage in terms of beauty and predictive power.

1.6 Families

According to the standard model, the fermions come in *families* or *generations*, each family containing two flavors of quarks and two flavors of leptons. Thus, the standard model requires that the number of quark flavors be even and equal to the number of lepton flavors. Ordinary matter is composed only of particles of the first family.

The standard model, by itself, says nothing about how many families should exist. However, present experimental evidence and the standard model together imply that there exist at least three families of quarks and leptons. If a fourth family should be discovered, then existing experiments and the standard model together require that the neutrino belonging to the new family have a large mass, as discussed in Chapter 2. But in the present version of the standard model, all neutrinos have zero mass. Thus, the existence of a fourth family would necessitate at least some modification of the standard model. The elementary fermions belonging to the three known families are shown in Table 1.1.

In Table 1.1 we have omitted color. Because each quark can be red, green, or blue, each family actually contains six quarks and two leptons. Each of the charged fermions has an antiparticle with the opposite charge. According to the standard model, the antiparticles of the neutrinos are also different from the particles.

1.7 Hadrons

Table 1.1 Elementary fermions belonging to the three known families.

First family	Second family	Third family
ν_e (e -neutrino)	ν_μ (μ -neutrino)	ν_τ (τ -neutrino)
e (electron)	μ (muon)	τ (tau lepton)
u (up quark)	c (charmed quark)	t (top quark)
d (down quark)	s (strange quark)	b (bottom quark)

In addition to the elementary particles of the standard model, there exist many chargeless (electrically neutral) and colorless (color-neutral) bound states. We mention, for example, the helium atom, the uranium nucleus, and the water molecule. Of particular importance in elementary particle physics are the *hadrons*, which are colorless bound states of quarks, antiquarks and gluons. Some are electrically charged and others are electrically neutral.

A *baryon* is a bound state of three so-called *valence* quarks, plus a *sea* of quark-antiquark pairs and gluons, and a *meson* is a bound state of a valence quark and an antiquark, plus particles of the sea. The valence particles are those which give a hadron its quantum numbers, while the sum of the sea particles have the quantum numbers of the vacuum.

The hadrons are treated as part of elementary particle physics not only for historical reasons (they were formerly thought to be elementary), but for the fundamental reason that, because we have not observed free quarks and gluons, all our information about those elementary particles comes from the study of hadrons.

But why have we not observed free quarks and gluons, or, for that matter, any colored bound states of quarks and gluons? No one knows for sure, but it is conjectured that QCD requires that colored particles cannot exist as free particles, but must be confined to the interior of hadrons. So far, no one has been able to prove that QCD requires the confinement of color, and this is one of the most important unsolved problems of the theory. The difficulty of proving confinement from QCD probably stems from the fact that perturbation theory is inadequate for this purpose, but we must remember the alternative possibility that QCD does not confine color. The conjecture that QCD confines color is bolstered by nonperturbative calculations done in an approximation in which the continuum field theory is replaced by space-time points on a lattice. Lattice calculations indicate that QCD is confining.

In the simplest approximation, an ordinary meson is composed just of a quark-antiquark pair (or, as we sometimes loosely say, of two quarks), and an ordinary baryon is composed of three quarks. These *valence* quarks of hadrons determine their flavor quantum numbers. In a better approximation, hadrons also contain particles of the sea. The quark-antiquark pairs of the sea carry no manifest flavor,

but might carry hidden flavor.

In ordinary hadrons, the wave function must contain amplitudes for configurations which are more complicated than we have just described. For example, the wave function of an ordinary baryon has a large amplitude for a term containing just three valence quarks plus the sea; but it has a smaller amplitude for a term containing four valence quarks, a valence antiquark, plus the sea; a still smaller amplitude for a term with five valence quarks, two valence antiquarks, plus the sea, etc. In addition, there are terms containing one or more valence gluons plus other particles. For many purposes, all these complicated terms in the wave function can be neglected in ordinary hadrons. In a still more drastic approximation, even the particles of the sea are neglected.

If the wave function of a hadron does not contain the first term in the expansion discussed above (three valence quarks or a valence quark-antiquark pair), the hadron is said to be *exotic*. Thus, the leading term in an exotic baryon may contain either four valence quarks and a valence antiquark (plus the sea) or three valence quarks and a valence gluon (plus the sea). Similarly, the leading term in an exotic meson may contain two valence quarks and two valence antiquarks (plus the sea), a valence quark-antiquark pair and a valence gluon (plus the sea), or even no valence quark-antiquark pairs but just two or more valence gluons (plus the sea). The concept of a valence gluon is a fuzzy one. Perhaps it is better to think of such a gluon as having additional energy and perhaps additional angular momentum compared to a gluon of the sea.

If a hadron has quantum numbers which forbid it to be made strictly of three quarks or a quark-antiquark pair, then it is said to be *manifestly* exotic. Otherwise, its exotic nature is hidden and is difficult to determine from experiment. There is only weak evidence for the existence of exotic hadrons.

If a meson contains only valence gluons and no valence quark-antiquark pairs, it is called a *glueball*. The simplest glueballs are composites of two valence gluons, but this is an approximation, as glueballs also contain a sea of gluons and quark-antiquark pairs. Because gluons do not carry flavor quantum numbers, glueballs must be flavorless, whereas some, but not all, mesons, have manifest flavor quantum numbers. Other exotic hadrons, which contain both valence quarks (and or antiquarks) and valence gluons, are called *hybrids*. In nature, pure ordinary mesons, pure glueballs, and pure hybrids almost surely do not exist, but, depending on quantum numbers, are mixed with each other and with ordinary and other exotic mesons. For example, a meson with only hidden flavor, such as the η , can mix with a predominantly glueball state. For this reason, it is difficult to identify hadrons which are predominantly glueballs.

Many of the properties of hadrons (as well as leptons and gauge bosons) are given in the *Review of Particle Properties* of the Particle Data Group. Traditionally, that review appears in even-numbered years. Also, an updated pocket condensation of the review, called the *Particle Properties Data Booklet*, appears every two years. It is available from the Particle Data Group, Lawrence Berkeley

Lab, Berkeley, CA 94720, USA, and from CERN Science Information Service, CH-1211 Geneva 23, Switzerland. The Data Booklet is a valuable handy reference. The latest review as of the time of writing is by Hikasa et al. (1992).

1.8 Quark and lepton masses

The mass of a free charged lepton traveling at high speed can be obtained by measuring its momentum p and energy E , and determining its mass from the equation $m^2 = E^2 - p^2$. By a “free” particle, we mean a particle which is not bound to any other particle. According to this definition, an electron in a hydrogen atom is not free, but an ionized electron is free. We determine the momentum of a charged particle by measuring its curvature in a homogeneous magnetic field of known strength. We determine the energy of a charged particle by measuring the ionization it produces when traveling in a material medium. The mass of a neutrino is measured indirectly by measuring the energies and momenta of charged particles participating in an interaction which produces the neutrino and then applying energy and momentum conservation laws. Attempted measurements of the masses of the neutrinos by this method have so far yielded only upper limits. However, there may be a far more sensitive way to measure neutrino masses, or at least neutrino mass differences. If different kinds of neutrinos can convert to one another, in a process called “neutrino oscillations,” then a measurement of such oscillations gives information about neutrino mass differences.

Because free quarks have not been observed, their masses have not been measured in the usual way. In the absence of direct measurements of the energy and momentum of free quarks, the quark masses can be inferred only indirectly from measurements of properties of hadrons. In this situation, different kinds of measurements of hadrons yield different results for the masses of bound quarks. The problem is that considerable interpretation is necessary to extract a value of the quark mass from a measured hadron property. The very meaning of mass is somewhat ambiguous for a particle that can exist only in a bound state.

Two different masses are usually associated with a quark of a given flavor: a “current” mass and a “constituent” mass. The current mass is the smaller of the two, and is usually associated with processes involving large momentum transfers. The constituent mass is usually more convenient for dealing with static properties of hadrons. But the quark mass is a function of the momentum transfer at which the mass is measured, albeit indirectly, and some measurements yield quark masses which are intermediate between the current and constituent masses.

A constituent quark has a higher mass than a current quark because it includes some of the mass associated with the sea. If a quark is struck gently by a probe (a process with low momentum transfer), then, as the quark moves it drags along particles of the sea, thereby gaining additional inertia. On the other hand, if a quark is struck violently by a probe (a process with high momentum transfer), the particles of the sea cannot follow the struck quark’s motion, and so the quark appears less massive. The difference in mass between a constituent and current

quark is of order $300 \text{ MeV}/c^2$

In elementary particle physics, particle masses are often given in energy units divided by c^2 , where c is the speed of light. This is because, according to Einstein's famous equation, the rest energy E_0 of a particle is related to its mass m by $E_0 = mc^2$. The unit of energy is commonly taken to be the eV (electron volt), the energy acquired by an electron when accelerated through a potential difference of one volt. Also used are meV (10^{-3} eV), keV (10^3 eV), MeV (10^6 eV), GeV (10^9 eV), and TeV (10^{12} eV).

In Table 1.2 are given the lepton masses and the approximate values of the current and constituent quark masses in MeV/c^2 . The constituent mass of a quark of a given flavor is approximately half the mass of a meson containing that flavor. (In applying this rule, we exclude the pion, which is an anomalously light meson.) Often, the u , d , and s quarks are called light quarks, while the c , b , and t are called heavy quarks.

Table 1.2 Lepton and quark masses in MeV/c^2 . The d - u mass difference is about $4 \text{ MeV}/c^2$ for both current and constituent quark masses. The lepton masses are taken from the tables of the Particle Data Group (Hikasa et al., 1992). The quark masses are rough estimates based on the work of Prof. Lichtenberg except for the mass of the t quark, which is a lower limit based on an experimental search by the DO collaboration (D. Zieminska, private communication).

Fermion	Mass	Current mass	Constituent mass
e	$0.51099906 \pm 0.00000015$		
μ	105.658387 ± 0.000034		
τ	1784.1 ± 3.6		
ν_e	$< 17 \times 10^{-6}$		
ν_μ	< 0.27		
ν_τ	< 35		
u		~ 5	~ 300
d		~ 9	~ 300
s		~ 200	~ 500
c		~ 1500	~ 1800
b		~ 4800	~ 5100
t		$> 131,000^*$	$> 131,000^*$

*A theoretical estimate of the mass of the t quark is $140,000 \pm 40,000 \text{ MeV}/c^2$.

From here on, we usually let $\hbar = c = 1$ (\hbar is Planck's constant divided by 2π and c is the speed of light), and write particle masses in energy units.

1.9 Quantum numbers

In the interactions of the elementary fermions, several *number conservation laws* are observed to hold, at least to a good approximation. We can assign a lepton number to all leptons (1 for a lepton and -1 for an antilepton) and a lepton family number to each of the three lepton families. To a good approximation, at least, the lepton number and the three lepton family numbers are conserved quantities. The three lepton family number conservation laws are examples of additive number conservation laws. For example, if the electron family conservation law holds, then the number of electrons plus electron neutrinos (minus their antiparticles) is a constant.

So far, we have not observed any violation of lepton number conservation. However, the question of whether we have observed lepton family number violation is at present uncertain. The uncertainty comes from the so-called solar neutrino puzzle. Apparently, fewer electron neutrinos from the sun are detected on the earth than the number calculated within the framework of the standard solar model. If the experiments are correct, as seems very likely, then there is something wrong with the theory. At present, we do not know whether the standard solar model is wrong or whether some electron neutrinos are converted to other kinds of neutrinos either in the sun or between the sun and the earth (neutrino oscillations). If neutrinos from one family can indeed convert to those of another family, this means that not only is lepton family number violated but also that at least one neutrino has mass different from zero. Fortunately, with only minor modifications, the standard model of elementary particles can accommodate both neutrino masses and neutrino oscillations.

Because quark number (the number of quarks minus the number of antiquarks) is observed to be conserved, we assign a quark number 1 to every quark and a number -1 to every antiquark. In the strong and electromagnetic interactions, quark flavor numbers are observed to be conserved. Therefore, we can assign a quark flavor number to each quark. However, in weak interactions, not only are these flavor number laws violated, but quark family numbers are also violated. However, so far we have not observed any violation of quark number; for example, we have not seen any process in which a quark has turned into a lepton. Such transitions are possible in some theories, including some so-called *grand unified theories*, which are different from the standard model. Even the standard model allows violation of quark and lepton flavor numbers because of an anomaly in the theory. However, the level of violation is predicted to be negligible in experiments that have been done so far.

All quarks have baryon number $A = 1/3$. (Baryon number is often given the symbol B , but we use B both for “bottomness” (defined in later in this section) and for B mesons (a B meson contains a b quark). The value $A = 1/3$ is chosen for quarks so that a baryon has $A = 1$. Baryon number conservation is a consequence of the conservation of quark number. If baryon number is exactly conserved, then the lightest baryon (the proton) must be absolutely stable.

We call *charge number* Q the value of the electric charge in units of the proton charge e . (The charge on the electron is $-e$.) The u , c , and t quarks have charge number $Q = 2/3$, and the d , s , and b quarks have $Q = -1/3$. Charge number is exactly conserved in the standard model. Charge number is conserved as a consequence of charge conservation, which follows from gauge invariance of QED. We would be astonished to find a violation. Not only does gauge invariance require that charge number be conserved, but also the magnitude of the charge. If charge number is exactly conserved, the lightest charged particle (the electron) must be absolutely stable.

On the other hand, we have no gauge principle which requires the conservation of baryon number. Although the Lagrangian of the standard model has a form which apparently conserves quark number—and therefore baryon number—the presence of an anomaly in the theory means that baryon number, like lepton number, holds only approximately. However, the quantum number $A - L$ (baryon number minus lepton number) is exactly conserved in the standard model.

The amount of violation of baryon number and lepton number can be estimated in the standard model, and turns out to be negligible and outside the range of experimental observation (except at extremely high temperatures). Therefore, if the standard model is correct, experiments at normal temperature will be unable to detect any proton decay. However, a number of models which go beyond the standard model predict that the amount of baryon number violation is large enough to be observable. Thus far, no experiment has detected any violation of A or L , in accordance with the standard model.

We have already noted that quark flavors do not change in strong or electromagnetic interactions, but may change in weak interactions. Because quark flavor is approximately conserved, it is useful to introduce flavor quantum numbers U (upness), D (downness), S (strangeness), C (charm), B (bottomness), and T (topness). These quantum numbers are defined so that a quark of charge $2/3$ has appropriate flavor quantum number (U , C , or T) equal to 1, and a quark of charge $-1/3$ has flavor quantum number (D , S , or B) -1 . An antiquark has opposite values of all these quantum numbers. Then the (electric) charge number of any quark is given by

$$Q = \frac{1}{2}(A + U + D + S + C + B + T). \quad (1.1)$$

In turn, if more quark flavors are discovered which fit into the pattern of the standard model, the generalization of Eq. (1.1) will be obvious. The quantum numbers in Eq. (1.1) of an antiquark are opposite in sign to those of a quark.

The charge Q and the flavor quantum numbers are examples of *internal* quantum numbers, i.e., quantum numbers of a particle having nothing to do (as far as we know) with ordinary space or time. They are *additive* quantum numbers, which means that one can obtain the values of these quantum numbers for a collection of particles by adding algebraically the values for the individual particles.

Therefore, Eq. (1.1) holds, not only for quarks and antiquarks, but for any hadron which is a bound state of quarks and/or antiquarks and/or gluons. Gluons have the value zero for all the quantum numbers entering Eq. (1.1).

The charge conjugation operator C replaces all the charges by anticharges and all the internal additive quantum numbers by their negatives. This means that the state obtained by operating with C on a particle state yields a state with the charges and additive quantum numbers of the antiparticle. However, because C is not in general an exact symmetry of field theory and is not an exact symmetry of the standard model, we do not use C to define an antiparticle state.

Because the d quark is only about 4 MeV heavier than the u quark, a value which is much smaller than the constituent masses of these quarks, an approximate symmetry, called (strong) isospin symmetry, holds for the interactions of the u and d . Then another quantum number, the isospin I , can be defined for these quarks, and also a z component I_z . Conventionally, the quantum numbers I and I_z are used to describe the u and d quarks instead of the flavor quantum numbers U and D .

In terms of I_z , Eq. (1.1) is

$$Q = I_z + \frac{1}{2}(A + S + C + B + T). \quad (1.2)$$

This equation is a generalization of the Gell-Mann–Nishijima relation for u , d , and s quarks. Equation (1.2) is much more commonly seen than Eq. (1.1). We exhibit Eq. (1.1) to point out that all the quarks can be treated on the same footing. It is because the mass difference between the u and d quark is small that it is more convenient to drop the quantum numbers U and D and instead use the quantum numbers I and I_z . Isospin symmetry is an approximate $SU(2)$ flavor symmetry of the strong interactions. The z direction in isospin space is related to the charge, as can be seen from Eq. (1.2). However, the rules for addition of isospin are the same as the rules for addition of ordinary angular momentum.

As far as we know from experiment, in Eqs. (1.1) and (1.2), the conservation laws corresponding to the quantum numbers Q and A of Eqs. (1.1) and (1.2) are exact and the others are approximate. However, we have a gauge principle which says that Q is exact, whereas in the standard model A is violated by an anomaly. The extent of the violation is negligible at present-day temperatures, but might have been important in the early hot universe shortly after the time of the big bang.

Still another quantum number, (strong) hypercharge Y is sometimes given. This is not an independent quantum number. We define it only for the light u , d , and s quarks, in which cases it is given by

$$Y = A + S \quad (1.3)$$

so that for these quarks

$$Q = I_z + Y/2. \quad (1.4)$$

Hypercharge is used for the u , d , and s quarks because I_z and Y are quantum numbers associated with an approximate $SU(3)$ flavor symmetry of these quarks. There seems little reason to introduce hypercharge for the heavy quarks, although some people have done so. (Later, we introduce weak hypercharge and weak isospin for all fermions.) All the quantum numbers of Eqs. (1.1)–(1.4) are internal additive quantum numbers.

In addition to the additive quantum numbers, we commonly use two other kinds, *vector* and *multiplicative*. Examples of vector quantum numbers are angular momentum J and isospin I . Each of these quantum numbers is associated with invariance under rotations in a three-dimensional space $R(3)$, or, essentially equivalently, rotations in a complex two-dimensional unitary space $SU(2)$. We shall see in a later chapter on group theory in what sense these two groups are equivalent. For the groups $R(3)$ and $SU(2)$, the resultant J of two angular momenta J_1 and J_2 has the possible values

$$J = J_1 + J_2, J_1 + J_2 - 1, \dots, |J_1 - J_2|.$$

The same law of “addition” (actually, of combination) holds for isospin because the underlying group is the same. However, we shall learn in the chapter on group theory that different laws of combination hold for vector quantum numbers of other symmetry groups like $SU(3)$.

Examples of multiplicative conservation laws are parity P (associated with invariance under space reflection) and charge conjugation C (associated with replacing the charges of a particle by their anticharges and replacing internal additive quantum numbers by their negatives). As we have said, these are only approximate conservation laws of the standard model. Time reversal T is also an approximate symmetry of the standard model, but, because of the nature of the operator T , particles do not have a corresponding quantum number.

A definite parity can be associated with the relative orbital angular momentum L of two particles. This parity is given by $P = (-1)^L$. Most particles can also be assigned intrinsic parities. If a particle can be produced in a parity-conserving reaction in which no other particle is created or destroyed, then the intrinsic parity of the particle can be measured. An example is the π^0 meson, which can be produced in the reaction

$$p + p \rightarrow p + p + \pi^0.$$

The π^0 intrinsic parity has been measured to be negative. On the other hand, a charged particle cannot be produced without the creation or destruction of another charged particle. Therefore, the parity of a charged particle can be measured only relative to the parity of some other particle. For example, the π^+ meson can be produced in the reaction

$$p + p \rightarrow p + n + \pi^+.$$

From this reaction, we can measure that the product of the intrinsic parities of the p , n and π^+ is negative. It is customary to *define* the intrinsic parity of the p

and n both to be positive. With this convention, the parity of the π^+ is negative. In the same convention, the parity of the π^- is negative.

The product of the intrinsic parities of a boson and its antiparticle is positive (from measurement and from field theory). Therefore, we conventionally define the intrinsic parity of an antiparticle of integral spin to be the same as the intrinsic parity of the particle. The parity of a boson-antiboson pair in a state of orbital angular momentum L is given by $P = (-1)^L$. On the other hand, the product of the intrinsic parities of a fermion-antifermion pair is negative (again, both from measurement and field theory). As a consequence, we often define the parity of a fermion to be positive and its antifermion to be negative. The parity of a fermion-antifermion pair with orbital angular momentum L is $P = -(-1)^L$.

If a particle is indistinguishable from its antiparticle, it is an eigenstate of the charge conjugation operator C . A photon has negative C parity. A π^0 can decay electromagnetically into two photons, which shows it has positive C parity. A particle-antiparticle pair (either boson or fermion) in a state with orbital angular momentum L and total spin S is in an eigenstate of C with the value $C = (-1)^{L+S}$.

Some of the quantum numbers of the quarks are given in Table 1.3. Except for the quantum number I , all the quantum numbers in Table 1.3 are additive. In addition, all quarks have spin $1/2$ and baryon number $A = 1/3$, and all can be defined to have positive parity. None is an eigenstate of C .

Table 1.3 Quantum numbers of the quarks. All quarks have baryon number $A = 1/3$. The quantum numbers U and D are rarely used.

Flavor	Q	I	I_z	U	D	S	C	B	T	Y
u	$2/3$	$1/2$	$1/2$	1	0	0	0	0	0	$1/3$
d	$-1/3$	$1/2$	$-1/2$	0	-1	0	0	0	0	$1/3$
s	$-1/3$	0	0	0	0	-1	0	0	0	$-2/3$
c	$2/3$	0	0	0	0	0	1	0	0	$-$
b	$-1/3$	0	0	0	0	0	0	-1	0	$-$
t	$2/3$	0	0	0	0	0	0	0	1	$-$

The leptons also have flavor quantum numbers. However, it is more useful to discuss the three lepton family quantum numbers. We call them electron family number L_e , muon family number L_μ , and tau family number L_τ . Both e and ν_e have $L_e = 1$, both μ and ν_μ have $L_\mu = 1$, and both τ and ν_τ have $L_\tau = 1$. The antiparticles have opposite values of these quantum numbers. According to the standard model with zero neutrino masses, the lepton family numbers are conserved, except for the negligible violation arising from the anomaly, but in

some other models observable violations can occur. As we have said earlier, if the neutrinos turn out to have masses greater than zero, the standard model can be revised to accommodate the nonconservation of lepton family numbers. It would require a more profound change in the standard model to violate the conservation of lepton number. Such a modification would be necessary if, for example, a reaction were observed in which a quark turned into a lepton, except possibly at extremely high energy.

The leptons and quarks have weak isospin i and weak hypercharge y , which, in analogy to Eq. (1.4), are related to electric charge number by

$$Q = i_z + y/2. \tag{1.5}$$

Because the weak interactions do not conserve parity, fermions with negative helicity (left-handed fermions) do not interact in the same way as fermions with positive helicity (right-handed fermions). Note: the helicity of a particle is the component of its spin in the direction of its motion.

According to the standard model, fermions interact weakly as left-handed doublets and right-handed singlets. This has as a consequence that left-handed fermions have weak isospin $i = 1/2$, while right-handed fermions have $i = 0$, with corresponding differences in hypercharge because of Eq. (1.5). Weak isospin and hypercharge cannot be conserved for fermions with masses greater than zero because the helicity of such particles can be changed by Lorentz transformations.

We see this as follows: Consider a massive particle with positive helicity traveling in some direction. Transform to a different Lorentz frame in which the particle is traveling in the opposite direction. This transformation does not change the direction of spin of the particle, and therefore reverses its helicity. In order to reverse the direction of the particle in the new Lorentz frame, the transformation must involve a speed greater than the speed of the particle in the original Lorentz frame. This is always possible for a particle with mass, but is impossible for a massless particle, which travels with the speed of light in any Lorentz frame.

In the present version of the standard model, the neutrinos are supposed to be massless. Therefore, helicity is a good quantum number for them. Furthermore, right-handed neutrinos and left-handed antineutrinos do not exist in the model.

1.10 Boson masses

According to the standard model, after the spontaneous breaking of $SU(2) \times U(1)$ symmetry (discussed later) there remain the manifestly unbroken symmetries $SU(3)$ of QCD and $U(1)$ of QED. Now the gauge bosons associated with a manifestly gauge-invariant theory must be massless. Therefore, the gluons and the photon have zero mass. The weak bosons, on the other hand, are associated with a spontaneously broken gauge symmetry, and do acquire masses in the theory. Because gluons, like quarks, are confined to the interior of hadrons, gluon masses cannot be directly measured. It is sometimes useful to define a constituent mass

for gluons which is different from zero. The constituent gluon can be considered to be a valence gluon plus a sea of gluons and quark-antiquark pairs. There is some indirect evidence that the gluon constituent mass is about 700 MeV. In Table 2.4 we give the masses and electric charges of the gauge bosons.

Table 1.4 Masses and charges of the gauge bosons. The masses of the W and Z are from the Particle Properties Data Booklet (June, 1992). All these bosons have spin 1 and baryon and lepton number zero.

Name	mass (GeV)	charge
γ (photon)	0	0
W^\pm (charged weak bosons)	80.22 ± 0.26	± 1
Z^0 (neutral weak boson)	91.173 ± 0.020	0
G_i ($i = 1, 2 \dots 8$ gluons)	0^*	0

*Theoretical value. The gluon constituent mass may be approximately 0.7 GeV.

The Higgs scalar boson is also an elementary particle of the standard model, but it has not been observed. It is the particle that spontaneously breaks the $SU(2) \times U(1)$ symmetry of the model by the Higgs mechanism, which we discuss in a later chapter. As we shall see later, according to the model the Higgs boson is electrically neutral and has a mass greater than zero. Unfortunately, the model does not predict what the mass of the Higgs should be or even relate the mass to experiments which can be done at low energy. Therefore, experimental physicists do not know where to hunt for the Higgs. Experiments apparently rule out a Higgs with mass less than about 24 GeV. Calculations within the framework of the standard model do not seem self consistent if the Higgs boson has a mass much greater than about 700 GeV. Aside from the fact that the Higgs boson has not been observed, some theoretical physicists believe that an elementary Higgs particle is the weakest link of the standard model, and that there might be another mechanism to break the symmetry spontaneously.

2. THE DISCOVERY OF ELEMENTARY PARTICLES AND HADRONS

In this chapter we discuss some of the major discoveries in the development of the subject of elementary particle physics. We do this, not only to give a perspective on a subject which is basically a twentieth century area of study, but also to discuss some relevant experimental methods and apparatus as well as theoretical advances. We shall see that the discovery of an elementary particle requires not only a good experiment but a theoretical interpretation of the results.

2.1 The electron

The first of what we now call elementary particles to be discovered, and the only one recognized before the twentieth century, is the electron. This particle was first observed by Sir Joseph J. Thomson in 1897, and he received the Nobel prize for the discovery. (Note: in this chapter we discuss a number of Nobel prize winners, nearly all in physics. Usually, we do not distinguish between those who were sole winners and those who shared the prize.)

Prior to Thomson's work, it was known that so-called *cathode rays* are emitted by a hot negatively-charged filament and attracted to a positively-charged electrode in a glass tube, called a cathode ray tube. (A typical tube might be 2 cm in diameter and 20 cm long.) The cathode rays are not observed directly. Rather, they excite the atoms of gas along their path in the tube, and we see the light emitted by the atoms when they spontaneously return to their ground state.

At the time Thomson did his experiment, it was known that cathode rays could be deflected by a magnetic field in such a direction that they must have negative charge. Thomson suggested that the rays were in fact a beam of charged particles, later called electrons.

But it was not enough simply to conjecture that the cathode rays were particles. Thomson went further and found the charge-to-mass ratio of the electrons. His procedure was to measure the radius of curvature of the electron beam in a magnetic field of known strength, and then to apply an electric field of sufficient strength to restore the beam to its undeflected path. Using the known laws of classical electromagnetism, it was then a simple exercise for him to compute the charge-to-mass ratio of the particles in the beam.

Note that Thomson did not see electrons directly, but only light from atoms along their paths. It turns out that essentially all the observations of elementary particles (and of many composite particles, such as hadrons) are indirect, so that a discovery of an elementary particle not only requires observation of certain phenomena but also a proper interpretation of the observations.

Thomson and others made crude measurements of the electric charge of the electron, observing how charged water droplets moved in an electric and gravitational field. Later, in 1909, Robert Millikan measured the charge of the electron

in an oil drop experiment. Milikan discovered that tiny oil droplets can be made to pick up electrical charges. By measuring the constant rate of fall of drops in air, Milikan was able to estimate their masses using Stokes's law for the speed of fall of a small sphere in a viscous medium. If an electric field is applied, there is a different rate of fall, which depends on the electric charge carried by the drop. Milikan made measurements of the charges of many oil drops and found that they were all integral multiples of a constant charge, the charge on the electron $-e$. Milikan received the Nobel prize for this and other work.

Taken together with Thomson's measurement of the charge-to-mass ratio of the electron, the electron's mass could also be determined. The present values of the charge $-e$ and mass m of the electron are

$$-e = -1.602 \times 10^{-19} \text{ C}, \quad m = 0.511 \text{ MeV} = 9.109 \times 10^{-31} \text{ kg}. \quad (2.1)$$

These numbers are known to considerably more precision than given here. For more precise values of these and other fundamental constants given in this chapter, see the latest issue of the *Particle Properties Data Booklet* (June, 1992). At the present time, we have no generally accepted theory which predicts either the charge or the mass of the electron. It is unfortunate that the same symbol e is used for the electron and for the magnitude of its charge. Sometimes e^- is used as a symbol for the electron.

The electron is the least massive charged particle known. Also, as far as is known, the electron is absolutely stable against decay. If indeed there exists no charged particle lighter than the electron, then the law of charge conservation, which as far as is known, is exact, forbids the electron to decay.

The year 1925 was a watershed year for physics, with the invention of quantum mechanics by Erwin Schrödinger and Werner Heisenberg. (Both are Nobel prize winners.) In the same year, G. E. Uhlenbeck and Sam Goudsmit interpreted a number of puzzling experiments involving electrons bound in atoms. For example, a beam of silver atoms was found to be split into two by a suitable inhomogeneous magnetic field. Uhlenbeck and Goudsmit's idea was that an electron has a spin \mathbf{S} , whose z component can take on only two possible values, $S_z = \pm 1/2$ (times \hbar), and a magnetic moment μ_e given by

$$\mu_e = -e\hbar/(2m) \quad (2.2)$$

(times \hbar/c). The unit $e\hbar/(2m)$ is called a *Bohr magneton*.

Hereafter, we say for simplicity that the spin of a particle is the integer or half integer S which gives its maximum component along an arbitrary z axis in units with $\hbar = 1$. The actual magnitude of S in these units is $\sqrt{S(S+1)}$. We use the same simple notation for orbital angular momentum L (which is always an integer) and for total angular momentum J (which may be either an integer or half integer).

There is a magnetic moment μ_L associated with the orbital angular momentum L of an electron, given by $\mu_L = -eL/(2m)$. However, the magnetic moment

of the electron associated with its spin is $\mu_e = 2eS/(2m)$, twice as large as was expected. Note that the energy E_μ , associated with the magnetic moment μ of a particle in a magnetic field \mathbf{B} is given by $E_\mu = -\mu \cdot \mathbf{B}$. (The same symbol μ is used for the muon, to be discussed in Section 2.7.)

In 1928, not very long after the discovery of the magnetic moment of the electron's magnetic moment, Paul A. M. Dirac invented a relativistic wave equation to describe the electron. The Dirac equation accounted for the spin and magnetic moment of the electron in a natural way. Dirac received the Nobel prize for his achievement.

Later measurements of the magnetic moment of the electron showed a small (about 0.1%) deviation from the Dirac value, and this so-called “anomalous” moment has been accounted for by the theory of quantum electrodynamics, invented in the 1940's independently by Richard Feynman, Julian Schwinger, and Shinichiro Tomonaga. These physicists shared the Nobel prize for their work.

The theory of quantum electrodynamics (QED), now incorporated into the standard model, is our best theory of particle interactions. Agreement between the theory of quantum electrodynamics and experiment is very impressive. In some instances, very precise experiments have agreed with theory to better than one part in a billion (10^9).

2.2 The photon

Quantum theory began in the last year of the nineteenth century, when, in 1900, Max Planck used a quantum hypothesis to explain black-body radiation. Planck's break with classical theory came from his postulate that the energy of oscillators in matter is quantized in unit multiples of a fundamental constant h , known as Planck's constant. Planck received the Nobel prize for his work.

Planck's constant divided by 2π has the value

$$\hbar = 1.055 \times 10^{-34} \text{ J s} = 6.582 \times 10^{-22} \text{ MeV s}. \quad (2.3)$$

Combining this value with the speed of light

$$c = 2.998 \times 10^8 \text{ m/s}, \quad (2.4)$$

we obtain

$$\hbar c = 197.32 \text{ MeV fm}, \quad (2.5)$$

where $1 \text{ fm} = 10^{-15} \text{ m}$. Equations (2.3)–(2.5) are useful in converting numbers representing physical quantities to conventional units from units in which $\hbar = c = 1$. Note that from Eq. (2.3), we obtain $1 \text{ MeV} = 1.603 \times 10^{-13} \text{ J}$.

Even after Planck's work, physicists still believed that light behaves as a wave, owing to Young's diffraction and interference experiments in the 19th century. However, in 1905, Albert Einstein postulated that light of frequency ν is quantized into particles, later called photons, each having energy $E = h\nu$. Einstein proposed

this idea in order to explain the photoelectric effect. The photoelectric effect can be demonstrated as a table-top experiment. If light of sufficiently high frequency ν is incident on a metal, the metal can emit electrons with a maximum kinetic energy T given by Einstein's equation

$$T = h\nu - \phi, \tag{2.6}$$

where ϕ is a constant characteristic of the particular metal.

Light near the blue end of the spectrum has sufficient frequency to knock electrons out of many metals, while red light, no matter how intense the beam, does not. Einstein proposed Eq. (2.6) and interpreted it to mean that a *photon*, a particle of light, can knock an electron out of a metal provided the energy of the photon is greater than the energy ϕ with which the electron is bound in the metal. Subsequent experiments with many metals verified the correctness of Eq. (2.6). Einstein received the Nobel prize, not for his work in special or general relativity, but principally for his law of the photoelectric effect.

Nobody understood the significance of photoelectric experiments until Einstein proposed his law. Afterwards, some physicists, although willing to accept the law, remained unconvinced of the particle interpretation. Then in 1923 Arthur Compton scattered a beam of monochromatic light by electrons and found that the scattered light beam had lower frequency than the incident beam. Compton showed that the frequency of the scattered light depends on the angle of scattering in such a way that the energy loss of a photon when scattered by an electron is equal to the recoil energy gained by the electron. Compton's experiments and interpretation finally convinced most physicists that light has particle as well as wave properties. Compton received the Nobel prize for his work.

Our present picture of the photon is that it is a quantum of the electromagnetic field. As such, it is the carrier of the force between charged particles, as described by the theory of QED. Quantum electrodynamics, like other "realistic" field theories, cannot be solved exactly, but must be solved in a perturbation expansion. As we have said, QED was invented independently by Tomonaga, Schwinger, and Feynman. Today, most physicists use the Feynman formulation for calculational purposes. In the Feynman approach, the terms in the perturbation theory expansion are pictorialized by Feynman diagrams or Feynman graphs.

The mass of the photon is zero within experimental error. According to our present ideas, if the photon had a mass, the force between two charged particles would not go inversely as the square of the distance between them. The theory of special relativity is built on the principle that the speed of light is a constant in all inertial frames, and that only massless particles can travel at that speed. Also, the principle of gauge invariance, which we discuss in a later chapter, forbids the photon to have a mass.

2.3 The atomic nucleus and the proton

Radioactivity was first observed in 1896 by Henri Becquerel on film accidentally exposed to uranium in a desk drawer. When Becquerel developed the film, he observed the tracks of particles emitted by radioactive uranium atoms. Becquerel shared the Nobel prize with Pierre and Marie Curie, the latter two receiving it for their work with radioactive atoms. (Incidentally, Marie Curie also received the Nobel prize in chemistry for the discovery and isolation of radioactive materials, including radium.)

Experiments revealed that radioactive atoms may emit α rays (now known to be ${}^4\text{He}$ nuclei), β rays (now known to be electrons), and γ rays (now known to be photons). In fact, this radiation is emitted by atomic nuclei, but the existence of nuclei was not known when Becquerel made his discovery. The energy of an emitted particle depends on the emitting nucleus and on the nature of the process, and may be as much as a few MeV, about a million times as energetic as the energy of a photon of visible light.

In 1910, H. Geiger and E. Marsden, two assistants in Ernest Rutherford's laboratory, measured the scattering by a gold foil of α particles emitted by a radioactive source. The detector was a tube with a high voltage across it. A charged particle entering the tube causes ionization of some of the atoms of gas in the tube, and the high voltage accelerates the free electrons and ions, which in turn cause further ionization. The result is a discharge of electricity, which is recorded as a "count." Geiger and Marsden found that sometimes the α particles were scattered through very large angles, a fact which could not be explained by the then existing "plum-pudding" model of the atom in which electrons (the plums) were thought to be embedded in a positively-charged medium (the pudding).

Rutherford interpreted the experiments in 1911 with a model in which an atom consists of a tiny positively-charged nucleus at the center, which carries nearly all the mass of the atom, plus electrons going around the nucleus in orbits, much like planets go around the sun (except for a vast difference in scale). Rutherford was able to show that Coulomb scattering from the nucleus can give rise to the observed large-angle scattering of the α particles. (Coulomb scattering is the scattering by a $1/r$ Coulomb potential). Strangely, Rutherford received the Nobel prize, not in physics, but in chemistry for work with radioactive substances.

Rutherford was lucky. He calculated the scattering using classical (non quantum) physics. For some potentials he would have gotten the wrong answer (that is, an answer in disagreement with experiment), but a Coulomb potential has the special property that the scattering cross section is the same when calculated classically and quantum mechanically, and, furthermore, is in agreement with experiment. Although it is now known that nuclei are not point particles, the experiments of Geiger and Marsden were not sufficiently sensitive to give a measure of the nuclear size.

Some physicists of the time did not at first believe in Rutherford's model of the atom because, classically, such a model is unstable. The argument is as follows: Electrons in orbits around a nucleus undergo acceleration. But acceler-

ating charged particles emit electromagnetic radiation, losing energy. Therefore, electrons orbiting a nucleus should lose energy and spiral into nucleus.

Then in 1913 Bohr solved the problem by simply postulating that electrons move around the nucleus in certain *stationary* orbits without emitting radiation. According to Bohr's model, electrons emit radiation as photons only when making a transition from one stationary orbit with a higher energy to another stationary orbit with a lower energy. According to the model, an electron in the lowest-energy stationary orbit is stable. The model was taken seriously, because, with it, Bohr was able to predict the wavelengths of the observed spectral lines emitted by hydrogen atoms. The relation between wavelength λ , frequency ν , and photon energy E is $\lambda = c/\nu = hc/E$, where we have included h and c explicitly. Bohr received the Nobel prize for his work.

After Rutherford's explanation of the scattering of α particles by gold, Geiger and Marsden performed further experiments, scattering α particles by a wide variety of elements. Because the scattering cross section is proportional to the nuclear charge, Geiger and Marsden deduced that the number of "elementary" positively-charged particles in a nucleus is approximately equal to half the atomic weight (in atomic units). The "elementary" charged nuclear particles are now known as protons.

The lightest atomic nucleus is the nucleus of ordinary (light) hydrogen, consisting of a single proton. We shall not discuss the experiments that established many of the properties of the proton. Within experimental error, the charge of the proton is equal in magnitude (although opposite in sign) to the charge of the electron. Hereafter, we shall *assume* that the proton charge is e , but we should not forget that this is an experimental question. The spin of the proton is $1/2$.

The mass and magnetic moment of the proton have been measured very precisely. The proton mass M is

$$M = 938.3 \text{ MeV} = 1.673 \times 10^{-27} \text{ kg}. \quad (2.7)$$

The first measurement of the proton's magnetic moment showed that the proton is not a Dirac particle. The magnetic moment of the proton μ_p is

$$\mu_p = 2.793 e/(2M). \quad (2.8)$$

The anomalous magnetic moment of the proton has the value $1.793 e/(2M)$ and is larger than its Dirac moment $e/(2M)$. The unit $e/(2M)$ is called a *nuclear magneton*. The large value of the proton's anomalous magnetic moment is a clue that the proton is not an elementary particle.

As far as is known, the proton is the lightest baryon and is stable against decay. This stability led to the law of conservation of baryon number. We do not know of any principle, such as the principle of gauge invariance, which is responsible for conservation of baryon number. As we have already remarked, an anomaly in the standard model violates conservation of baryon number, but the predicted

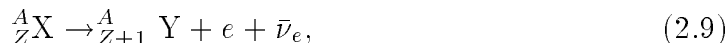
proton lifetime is far too long to be observable. (Physicists, using the standard model, have estimated that in the early universe, when the temperature was very high, baryon-number nonconservation was appreciable.) Some other theories do predict observable proton decay in the present universe, and at least one such theory has been ruled out by the existing experimental lower limit on the lifetime of the proton.

2.4 The neutrino

We have already mentioned that radioactive nuclei commonly emit α , β , or γ rays, now known to be ${}^4_2\text{He}$ nuclei, electrons, and energetic photons respectively. The α and γ particles were observed to be monoenergetic, but the β particles were observed to be emitted with a broad spectrum of kinetic energies, ranging from as close to zero as could be detected up to a well-defined maximum for a given decay process.

Until 1930, it was not understood why in β decay electrons emerge from a nucleus with a spectrum of energies. Beta decay experiments show that the total energy of the daughter nucleus plus emitted electron is less than the rest energy of the parent nucleus, except when the electron was emitted with its maximum energy. Furthermore, the sum of the momenta of the the electron and daughter nucleus is measured to be different from zero, whereas the parent nucleus is at rest. Also, the total angular momentum of the electron and daughter nucleus is different from the angular momentum of the parent nucleus.

In 1930 Wolfgang Pauli suggested that an undetected particle is emitted in β decay, carrying off energy, momentum, and angular momentum. In order to account for the observed characteristics of β decay, the new particle must be electrically neutral, have little or no mass, and have half-integral spin. Furthermore, its interactions with matter must be so weak as to enable it to have escaped detection in the early experiments. Fermi named the particle the “neutrino,” which means “little neutral one.” It is denoted by the symbol ν . Now we know that the particle participating in β decay is actually an antiparticle, the electron antineutrino $\bar{\nu}_e$. (Antiparticles are discussed in more detail in Section 2.5.) A general β decay process is



where X is the parent nucleus and Y is the daughter. The superscript A denotes the baryon number, in this case the number of protons plus neutrons in the nucleus, and the subscript Z denotes the charge number, in this case the number of protons in the nucleus. (See Section 2.6 for further discussion of the neutron.)

Pauli received the Nobel prize, not for postulating the existence of the neutrino, but for proposing the Pauli exclusion principle. (The generalization of this principle is the statement that the wave function of a collection of identical fermions must be antisymmetric under the interchange of the coordinates of any two of them.)

In 1934 Fermi invented a theory of beta decay, in which four fermions, two charged and two neutral, interact at a point. Fermi's theory describes β decay of the neutron (see Section 2.6 for more about the neutron):

$$n \rightarrow p + e + \bar{\nu}_e, \quad (2.10)$$

where the symbols n and p stand for neutron and proton respectively. Note that when a neutron decays, an antineutrino, not a neutrino, is emitted. Fermi received the Nobel prize, not for his theory of beta decay, but for experiments bombarding nuclei with neutrons.

Fermi knew that his theory could not be correct at high energy, and so was probably a low-energy approximation to a better theory. Fermi's theory provides a good description of many properties of β decay, but had to be modified to take into account subsequent experiments which showed that parity is not conserved in weak interactions. Also, the pointlike nature of the Fermi interaction is inadequate at high energy. The standard model includes a generalization of Fermi's theory to take into account these effects plus the existence of more than one family of fermions.

In the early 1950's, Fred Reines and C. Cowen carried out an experiment in which antineutrinos emitted from a nuclear reactor were detected by an inverse β process

$$\bar{\nu}_e + p \rightarrow n + e^+, \quad (2.11)$$

where e^+ is an anti-electron, or positron (see Section 2.5).

In the late 1950's, Madame Wu and her co-workers discovered in a β decay experiment that parity is not conserved. Following a suggestion by T. D. Lee and C. N. Yang, Wu and collaborators measured the angular distribution of β rays from a polarized radioactive source (^{60}Co), and found a decay asymmetry that established that parity is not conserved. Soon afterward, several physicists showed that Fermi's theory could be altered to include parity violation.

Subsequent developments included the discovery of a second neutrino ν_μ and then a third ν_τ . Then, in the late 1960's Fermi's theory was superseded by the electroweak theory of Sheldon Glashow, Steven Weinberg, and Abdus Salam, now a part of the standard model.

2.5 The positron

In 1925, quantum mechanics was invented independently in two equivalent forms: the matrix mechanics of Heisenberg and the wave mechanics of Schrödinger. Schrödinger not only was the first to write down the nonrelativistic wave equation that bears his name but also was the first to write a relativistic wave equation to describe a particle. Schrödinger's relativistic equation is now usually called the Klein–Gordon equation.

It was soon discovered that if one tries to interpret the Klein–Gordon equation as the equation for a single particle, one runs into difficulties. An important

example is that one cannot define a probability function which is positive under all circumstances. The difficulties led physicists to give the Klein–Gordon equation a different interpretation, namely, that the equation describes a *field*, with an infinite number of degrees of freedom. We shall introduce field theory in a later chapter.

Three years after Schrödinger invented the Klein–Gordon equation, Dirac wrote down a different relativistic wave equation, which provides an excellent description of the electron. The Dirac equation also runs into difficulties if it is interpreted as the equation of a single particle. A major problem arises because the Dirac equation has solutions with negative energy. There is no obvious mechanism to prevent an electron at rest from making a transition to a state with negative energy, radiating energy in the process. Thus, the electron apparently does not have a ground state. In order to forbid the unwanted transitions, Dirac postulated that all the negative energy levels are filled with a “sea” of electrons. The Pauli principle, which forbids more than one electron to be in the same quantum state, then prevents electrons from making transitions to the negative-energy states. In Dirac’s picture, the vacuum or ground state of the system consists of a filled infinite sea of negative energy electrons.

Suppose an electron from the negative-energy sea gets a positive energy by absorbing an energetic photon. Dirac interpreted the resulting “hole” in the sea as a particle with positive charge and positive energy. Dirac at first hoped that such a particle could be interpreted as a proton, but it was soon realized that the hole must have the same mass as the electron. Thus, the Dirac equation predicts that, a positively charged particle like the electron should exist. This particle is the *antiparticle* of the electron.

Today we believe that every particle has an antiparticle. The antiparticle is distinguished from the particle by having opposite additive quantum numbers. If the particle and antiparticle have no distinguishing quantum numbers, they are identical. All antiparticles of known fermions are distinguishable from their particles. However, some bosons, including the photon, are the same as their antiparticles.

In 1932, just a few years after Dirac’s work, Carl. D. Anderson discovered the positron, a particle with the same mass as the electron and with opposite charge. Anderson used as a detector a *cloud chamber*, which is a chamber containing a supersaturated gas. When a charged particle traverses the chamber, it ionizes atoms along its path. The ions serve as nuclei for the condensation of liquid drops, which form along the path taken by the charged particle. These drops are observed (usually on film). After a photograph is taken, the cloud chamber is expanded, and the drops evaporate. Then the chamber is compressed again, and is ready to detect the next charged particle passing through it. The cloud chamber was invented by Charles Wilson, for which he received the Nobel prize.

Anderson exposed his cloud chamber to so-called cosmic radiation, which we now briefly describe. Cosmic rays are particles which impinge on the earth’s atmosphere from outer space. They were discovered in the period 1911–1914 by

Victor Hess and W. Kohlörster in balloon observations at high altitude. Additional particles are created by the collision of cosmic rays with atoms in the upper atmosphere. The cosmic rays vary in energy from moderate to very high. The energy spectrum shows that the flux of incident particles falls steeply with increasing energy. This feature has as a consequence that very high energy collisions of cosmic rays are rare. In fact, one of the principal reasons for building accelerators is to obtain beams of particles much more intense than the natural beams of cosmic rays. Another reason is that in an accelerator the type of particles in the beam, their intensity, and their energy are all controlled. Hess and Anderson shared the Nobel prize, Hess for cosmic rays, and Anderson for the positron.

Anderson's cloud chamber contained a magnetic field and a lead plate. Among the particles Anderson saw in the chamber (actually, he saw only their tracks) were particles of approximately the mass of the electron but positively charged. Anderson determined the sign of the charge by the direction of motion of the particle and its direction of curvature in the magnetic field. He estimated the magnitude of the charge by the energy loss in the lead plate and estimated the mass by the the amount of ionization and curvature in the chamber.

It is clear from the above description that, in order to discover the positron, Anderson had to go far beyond the mere observation of curved tracks in a cloud chamber. He also had to make detailed measurements of curvature, ionization, etc, and successfully interpret his measurements so as to obtain the charge and mass of the particles making the tracks.

A positron is the *antiparticle* of an electron. A photon can turn into positron-electron pair (near another particle, which can take up momentum and so satisfy the conservation of momentum). Dirac's interpretation of pair creation is that an electron from the negative-energy sea can make a transition to a state of positive energy by absorbing a photon of sufficient energy, leaving a hole (a positron) behind. Conversely, a positron and an electron can annihilate, converting into a photon (again near another particle to take up momentum). In Dirac's interpretation, electron-positron annihilation corresponds to a positive-energy electron radiating a photon and dropping into a hole in the negative energy sea.

With the invention of field theory (also pioneered by Dirac), we have a different interpretation of electrons and positrons. The Dirac equation is interpreted as the equation of a field, whose quanta are electrons and positrons. In this many-body interpretation of the Dirac equation (which will be discussed later), we do not need the concept of an infinite negative-energy sea.

Within experimental error, the positron is observed to have the same mass as the electron and a charge and magnetic moment which are equal in magnitude opposite in sign to those of the electron. These observations agree with the predictions of the *CPT* theorem, which follows from quantum field theory.

2.6 The neutron

In 1919 Rutherford was able to produce protons from nuclei bombarded by α

particles, demonstrating the transmutation of one element into another. About a year later, Rutherford proposed the existence of a neutral subnuclear particle, the neutron, having about the same mass as that of the proton.

About a dozen years later, in 1932, Irene Curie and F. Joliot bombarded a beryllium target with α particles. They observed charged particles emerging from the target and estimated their energy. They concluded from energy conservation that electrically neutral particles must also come from the target. These particles were in fact neutrons, but, unfortunately, Curie and Joliot did not realize that they had evidence for this particle.

Later in 1932 John Chadwick improved the experiments of Curie and Joliot and correctly made the deduction that the neutral particles were neutrons, emitted in the reaction



The neutron n is sometimes given the symbol 1_0n , so as to bring out that it has baryon number $A = 1$ and charge $Z = 0$. From the energy of the observed recoil carbon nucleus, Chadwick was able to calculate the mass of the neutron to about 0.1% accuracy, finding it to be very nearly equal to the mass of the proton. The mass M_n of the neutron is

$$M_n = 939.57 \text{ MeV}, \quad (2.13)$$

which is 1.29 MeV more than the mass of the proton. Chadwick received the Nobel prize for his discovery of the neutron.

The discovery of the neutron led to a picture of atomic nuclei made of protons and neutrons, or *nucleons*. The atomic mass number A of a nucleus is the number of nucleons it contains, while its atomic number Z is the number of protons it contains.

Later experiments showed that the neutron has spin $1/2$, the same as the proton. However, the magnetic moment μ_n was observed to be different from zero and equal to

$$\mu_n = -1.913 e/(2M). \quad (2.14)$$

This is entirely an anomalous magnetic moment, because the Dirac magnetic moment of a particle is proportional to its charge and so is zero for the neutron. As we shall explain in a later chapter, the quark model gives us a good qualitative understanding of the magnetic moment of the neutron.

The neutron is unstable, decaying exponentially in time with a mean life τ_n given by

$$\tau_n = 889 \pm 2 \text{ s}. \quad (2.15)$$

Its principal decay mode is given by Eq. (2.10).

The mean life (or lifetime) τ of an unstable particle appears in the exponent of the formula

$$f = e^{-t/\tau}, \quad (2.16)$$

where f is the fraction of the initially produced particles remaining after a time t . We do not use the “half life,” which is the time required for half the particles to decay.

In the 1960’s electron scattering experiments showed that the neutron has a size similar to that of the proton and a charge distribution different from zero. Although the neutron is electrically neutral, it is positively charged at the center and negatively charged at its edges.

In the early 1930’s the nature of the nuclear force was not understood, except that it was measured to have a short range (about 1.4 fm). There followed a program in which physicists tried to deduce the properties of the force between two nucleons by an array of proton-proton and neutron-proton scattering experiments. Physicists succeeded in learning some of the properties of the force, especially that it is complicated.

One simple feature of the nuclear force is that it is approximately the same between two protons as between two neutrons, or between a proton and a neutron. Heisenberg introduced the notion of isospin to describe the proton and neutron in terms of a single particle, the nucleon. In this formalism, the nucleon has isospin $1/2$, just as the electron has spin $1/2$, and the proton and neutron correspond to the two orientations of the nucleon in an abstract isospin space, analogous to the two possible orientations of the electron spin in ordinary space. Subsequently, Wigner elaborated on the formalism of isospin, assuming it is conserved in strong interactions. In the formalism, the number of charge states N_I available to a particle is related to its isospin I by the equation

$$N_I = 2I + 1. \tag{2.17}$$

The fact that the neutron and proton have slightly different masses implies that isospin is only approximately conserved.

In 1932, after the discovery of the neutron, the list of elementary particles was very short, but seemed sufficient to account for the structure of atoms and atomic nuclei, including radioactivity. Only five known particles were thought to be elementary: the proton p , the neutron n , the electron e , the photon γ , and the neutrino ν .

The antiparticles of the five particles were also believed to exist, but the only distinct antiparticle observed at the time was the positron. (The photon is the same as its antiparticle.) The \bar{p} and \bar{n} were not yet discovered, and only indirect evidence for both the ν and the $\bar{\nu}$ existed. A particle is distinguished from its antiparticle by its additive quantum numbers, charge, and magnetic moment, all of which are opposite in sign for the antiparticle. Because the neutron has a magnetic moment which is different from zero, the \bar{n} must have a magnetic moment of opposite sign. Also, the n has baryon number $A = 1$, but the \bar{n} has $A = -1$. Thus, the \bar{n} is distinguishable from the n . The neutrino is distinguished from its antiparticle by lepton number.

It was more than a quarter century later that the antiproton and antineutron were discovered at the Bevatron, a synchrotron at Berkeley built to provide beams of protons with enough energy to create a nucleon-antinucleon pair in a collision. At the time these antinucleons were discovered, many additional particles had been discovered, and the view of the world as containing only five particles had been abandoned.

2.7 The pion and muon

The simple picture of a universe made of only five elementary particles lasted just a few years. The first attack on the picture was a theoretical one. In 1935 H. Yukawa proposed the existence of a *meson* (this is the modern name) to carry the nuclear force, just as a photon was believed to be the carrier of the electromagnetic force. The photon is a massless particle and leads to a Coulomb force of infinite range. (We shall discuss in a later chapter how this arises in QED.) Yukawa reasoned by analogy that a nuclear force of finite range should arise from a carrier with a mass greater than zero. Yukawa received the Nobel prize for his proposal.

The potential V between two charged particles, each of charge e , arising from the exchange of a massless photon, is

$$V = e^2/r. \quad (2.18)$$

In analogy, Yukawa proposed that the strong potential between two nucleons arising from the exchange of a meson of mass m is

$$V = -g^2 e^{-mr}/r, \quad (2.19)$$

where g is the strong-interaction coupling strength, analogous to the charge e . The range R of this interaction is $1/m$, or, in more conventional units, $R = \hbar/(mc)$. Using a range of 1.4 fm for the nuclear force, we obtain a meson mass $m \simeq 140$ MeV. Yukawa proposed that the meson have spin zero, in order for the potential be attractive. (A particle of spin one leads to a repulsive potential between like charges.) Also, to explain the similarity of the nuclear force between two protons, two neutrons, and a proton and neutron, three such mesons should exist, with charges e , 0 , and $-e$. In obtaining a meson mass of 140 MeV, we use Eq. (2.5) as follows: since $R = \hbar/(mc)$, we have $Rmc^2 = \hbar c = 197.3$ MeV fm. Then since $R = 1.4$ fm, we get $mc^2 = 197.3/1.4 = 141$ MeV.

The next important development came in 1937, when Anderson and S. H. Neddemyer discovered the muon μ in cosmic radiation. This particle, which has mass 105.7 MeV, was first thought to be the meson predicted by Yukawa. But the interactions of muons in matter were observed to be far too weak to explain the strong nuclear force. The muon was called a “ μ meson” in the early years after its discovery, but, in our present terminology, muon is a lepton, not a meson.

The muon was observed in two charge states, with charges e and $-e$ (within experimental error), both of which have the same mass (again within experimental

error). We now regard the μ^- as the particle and the μ^+ as the antiparticle. The muon, like the electron, has spin $1/2$, and a magnetic moment consisting of a Dirac part and a small anomalous part which can be calculated in the theory of QED. In its electromagnetic interactions, the muon behaves like a heavy electron.

However, the μ is unstable, decaying with a mean life

$$\tau_\mu = 2.197 \times 10^{-6} \text{ s.} \quad (2.20)$$

A direct measurement of the lifetime of muons at rest was not made until the early 1940's. On the average, high-energy muons travel much further through the atmosphere before decaying than one would expect from a simple nonrelativistic calculation using their mean lifetime. In fact, experiments show that the mean life of high-energy muons is lengthened in accordance with time dilation of special relativity.

When a muon decays, an electron is emitted. The observation of a spectrum of energies of emitted electrons shows that at least two neutral particles are emitted in the decays of muons. It was later established that the principal decay mode of the muon (almost 100% of the time) is

$$\mu^- \rightarrow e^- + \bar{\nu}_e + \nu_\mu. \quad (2.21)$$

The muon neutrino ν_μ is distinguishable from ν_e . This fact was determined in a measurement about 30 years later by a team of physicists led by Jack Steinberger, Leon Lederman, and Mel Schwartz, who shared the Nobel prize for their work. The team let a beam of high-energy muon neutrinos collide with matter and found that the neutrino interactions caused muons but not electrons, to be created. This is in contrast with high energy electron neutrinos, which interact with matter to produce electrons but not muons.

After the muon was shown not to be Yukawa's meson, nobody understood what its place was in nature's scheme. The physicist Isidor Rabi asked, "Who ordered that?" Now we know that the muon was the first particle to be discovered belonging to the second family of quarks and leptons. If Rabi could ask his question today, it might be, "Who ordered the second and third families, and are there any more?"

Because the muon does not interact strongly with nuclei, Robert Marshak suggested in 1947 that the muon was not Yukawa's meson, and that if Yukawa's theory was correct, a different meson remained to be discovered. Actually, unknown to Marshak a new meson, called the π meson, or pion, was discovered by Cecil F. Powell and his collaborators about a month earlier in a photographic plate exposed to cosmic rays. A special kind of emulsion containing a large percentage of silver compounds is used to detect the tracks of charged particles. The particles ionize the material along their path, and when the emulsion is developed, particle tracks clearly stand out. Powell received the Nobel prize for his work.

In the early experiments, only positively and negatively charged pions were observed. (Later, neutral pions were seen.) The masses of the π^+ and π^- are the same within experimental error. The charged pion mass is

$$m_\pi = 139.57 \text{ MeV}. \quad (2.22)$$

Charged pions were observed to decay with a lifetime τ_{pi} given by

$$\tau_\pi = (2.603 \pm 0.002) \times 10^{-8} \text{ s}. \quad (2.23)$$

The principal decay modes are

$$\pi^+ \rightarrow \mu^+ + \nu_\mu, \quad \pi^- \rightarrow \mu^- + \bar{\nu}_\mu. \quad (2.24)$$

It was known quite early that the decay was into two particles because the muon is emitted with a unique energy in the c.m. frame of the pion.

A proton accelerator at the University of California, Berkeley, produced a beam of sufficient energy (340 MeV) so that when atomic nuclei were struck by the beam, pions were produced. At first, only charged pions were seen, but in 1950 neutral pions were observed as well.

The Berkeley accelerator was a cyclotron, a device to accelerate charged particles. The cyclotron was invented in 1932 by Ernest O. Lawrence and M. S. Livingston. Lawrence received the Nobel prize for this invention. A cyclotron is a circular accelerator in which charged particles move, guided by a fixed magnetic field. At each revolution, the particles are accelerated by a high-frequency voltage. Because the magnetic field does not change with time, as the particles gain energy, they move in orbits of larger radii, and so spiral outward until their most energetic orbit is reached.

The mass of the neutral pion is

$$m_{\pi^0} = 134.97 \text{ MeV}. \quad (2.25)$$

This mass is similar to that of the charge pion and close to the mass originally predicted by Yukawa. Furthermore, pions were observed to interact strongly with nucleons, an observation which is consistent with the idea that the pion is Yukawa's meson. Because the pion exists in three charge states, it has isospin $I = 1$, in accordance with Eq. (2.17). Not only do the π^+ and π^- belong to the same isospin triplet, but they are antiparticles of each other. The π^0 is its own antiparticle.

The lifetime of the neutral pion is

$$\tau_{\pi^0} = (8.4 \pm 0.6) \times 10^{-17} \text{ s}. \quad (2.26)$$

The error is rather large because the lifetime is so short. The principal decay mode is

$$\pi^0 \rightarrow 2\gamma. \quad (2.27)$$

The π^0 is observed only indirectly by observation of the two photons into which it decays. The photons are observed by letting them convert (in matter) into electron-positron pairs.

2.8 Strange particles

In the same year that Powell discovered the pion, G. D. Rochester and C. C. Butler observed so-called “strange” particles resulting from the collisions of cosmic rays with matter. Strange particles are hadrons which carry a quantum number called strangeness S . After the invention of the quark model, strange particles were understood to contain one or more strange quarks. The strange quark belongs to the second family of elementary particles in the standard model.

Originally, strange particles were thought to be peculiar because they were observed to be produced strongly by cosmic rays, but to decay only weakly. Abraham Pais proposed a mechanism of “associated production” of strange particles as a reason for this behavior. Associated production means that a single strange particle cannot be produced in a strong interaction, but that two strange particles must be produced at the same time. Then in 1952 Murray Gell-Mann and independently Nakano and Nishijima proposed the existence of a strangeness quantum number which is conserved in strong and electromagnetic, but not weak, interactions. (Actually, Nakano and Nishijima proposed a different, but equally useful, quantum number known as hypercharge.) If two strange particles are produced in one interaction, one having $S = 1$ and the second having $S = -1$, then strangeness can be conserved, and the interaction can proceed strongly. However, when a strange particle decays into ordinary particles, strangeness is violated, and the decay must be weak.

One of the earliest associated production reactions seen is

$$\pi^- + p \rightarrow \Lambda^0 + K^0, \quad (2.28)$$

where Λ^0 is a so-called hyperon (a baryon carrying strangeness) and has the quark content uds , while the K^0 is called a neutral kaon (or neutral K meson) and has the quark content $\bar{s}d$. We recall from Chapter 1 that the s quark has $S = -1$ and the \bar{s} quark has $S = 1$. It follows that the Λ^0 has $S = -1$ and the K^0 has $S = 1$.

The mass and lifetime of the Λ^0 are

$$M_\Lambda = 1115.6 \text{ MeV}, \quad \tau_\Lambda = 2.63 \times 10^{-10} \text{ s}. \quad (2.29)$$

Its principal decay mode is

$$\Lambda^0 \rightarrow N + \pi, \quad (2.30)$$

where N stands for a nucleon (p or n). The branching fractions are: 0.64 into $p + \pi^-$ and 0.36 into $n + \pi^0$. The Λ^0 has isospin $I = 0$, which means it comes in only one charge state (in this case, the charge is zero). Because of this fact, we often omit the superscript 0 on the symbol.

On the other hand, the K like the N , has isospin $I = 1/2$, which means it comes in two charge states. We have already introduced the K^0 ; the other is the K^+ .

Other hyperons were later observed, including an isospin triplet ($I = 1$) Σ^+ , Σ^0 , and Σ^- , with $S = -1$; an isospin doublet ($I = 1/2$) Ξ^0 and Ξ^- , with $S = -2$; and an isospin singlet ($I = 0$) Ω^- , with $S = -3$. The quark content of these particles is

$$\Sigma : \quad \Sigma^+ = uus, \quad \Sigma^0 = uds, \quad \Sigma^- = dds, \quad (2.31)$$

$$\Xi : \quad \Xi^0 = ssu, \quad \Xi^- = ssd, \quad (2.32)$$

$$\Omega : \quad \Omega^- = sss. \quad (2.33)$$

Still other strange particles are known, but we defer discussion of them to later chapters.

The K meson, like the nucleon N , has isospin $I = 1/2$, which means it comes in two charge states. We have already introduced the K^0 ; the other is the K^+ . The K mesons are fascinating to study. Because the K^0 meson has $S = 1$, it is distinct from its antiparticle, the \bar{K}^0 , which has $S = -1$. Therefore, unlike the pions, which belong to an isospin triplet (π^+ , π^0 , π^-), the kaons belong to two isospin doublets (K^+ , K^0) and (K^- , \bar{K}^0).

Theoretical and experimental study of the K^0 - \bar{K}^0 system has proved very fruitful. Gell-Mann and Abraham Pais pointed out that the K^0 and \bar{K}^0 should be able to convert into one another because strangeness is not conserved in weak interactions. The K^0 and \bar{K}^0 (the eigenstates of strangeness which are created in strong interactions), are not eigenstates of the full Lagrangian. As a result, the K^0 and \bar{K}^0 do not have definite masses and lifetimes, but are linear combinations of two other states, called K_S and K_L , each of which has a definite mass and lifetime. An initially produced K^0 (or \bar{K}^0) will after a time which is long compared to the K_S lifetime of 0.9×10^{-10} s, convert to a K_L , because the K_S amplitude in the wave function will have died away. The lifetime of the K_L is about 5×10^{-8} s, about 500 times as long as the lifetime of the K_S . It was further observed by James Cronin, Val Fitch, and their colleagues that CP (the product of charge conjugation and parity) was not conserved in the decay of the K_L (called K_2 at the time).

2.9 Hadron resonances

In the mid 1950's, experiments performed at an accelerator called the Cosmotron at Brookhaven National Laboratory revealed the presence of a pion-nucleon resonance. First, the proton beam at the accelerator is allowed to strike a nuclear target, producing secondary beams of positive and negative pions. (Neutral pions have too short a lifetime to form a useful secondary beam.) The charged pions, in turn, are made to scatter from a hydrogen target. The total cross section as a function of energy has a large peak, whose maximum is at a pion lab energy of

approximately 200 MeV. This energy corresponds to a total energy of the system (including rest energy) of around 1230 MeV. The resonance shows up in both π^+p and π^-p scattering.

There followed considerable controversy about how to interpret this resonance. On the one hand, the resonance can be considered a composite state composed of two elementary particles: the pion and the nucleon. On the other hand, the resonance can be considered as another hadron, no more and no less elementary than the pion and nucleon. Today, we regard the pion-nucleon resonance as a hadron (a baryon). It has the name Δ , or, more descriptively, $\Delta(1230)$. The Δ comes in four charge states, and so has isospin $I = 3/2$.

Unlike the proton, which is apparently stable, and the neutron and charged pions, which decay weakly, the Δ decays strongly with a width $\Gamma = 120$ MeV. (The width means the full width at half maximum.) The width and lifetime are related by $\Gamma = 1/\tau$. If we use Eq. (2.3), we find the width of the Δ corresponds to a mean life of about $\tau = 0.6 \times 10^{-23}$ s. A lifetime this short cannot be measured directly because the particle decays too quickly after being produced to travel a measurable distance. Therefore, the lifetime must be calculated from the measured width.

The Δ was only the first of many other baryon and meson resonances discovered in scattering experiments. We do not have the time to discuss the individual particles here. However, it is interesting to point out that many of the particles discovered in the 1950's and 1960's were observed in bubble chambers exposed to high-energy beams from accelerators.

The *bubble chamber* was invented by Donald Glaser, for which he received the Nobel prize. It is a device containing a superheated liquid. The liquid must contain very few impurities, or otherwise bubbles will form spontaneously in it. When a charged particle passes through the liquid, it ionizes atoms of the liquid along its path, and bubbles form along the track. These bubbles can be photographed, showing the path of the particle. After a photograph is taken, the chamber is compressed to remove the bubbles, and then is expanded again so as to be ready for the next charged particles.

After the invention of the bubble chamber, cloud chambers were only rarely used. A bubble chamber has several advantages over a cloud chamber. First, a liquid is denser than a gas, and so there are more interactions in a bubble chamber than in a cloud chamber of the same size. Second, a bubble chamber can be made from various liquids. If the chamber is filled with liquid hydrogen, when a beam particle undergoes a collision in the chamber, the target nucleus must be a proton. This fact makes for a straightforward analysis.

Many of the hadrons observed in bubble chambers were first seen in a large chamber containing liquid hydrogen. Luis Alvarez was the leader of the team of physicists making the discoveries, and he received the Nobel prize for this work.

At the present time, bubble chambers have been largely superseded by elaborate detectors containing wire spark chambers, Cerenkov counters, and other detecting elements. These complex detectors are necessary to process a large

number of interactions at high energy. Pavel Cerenkov received the Nobel prize for the discovery of the effect that bears his name, and Charpak for his invention of the spark chamber. The Cerenkov effect is the emission of light by charged particles in a medium traveling faster than the speed of light in the medium. A wire spark chamber is a fast counter which contains many elements and so allows one to obtain the path of a charged particle through space. Unfortunately, the modern detectors are too complicated for us to discuss in detail here.

The 1992 edition of the Particle Properties Data Booklet lists more than 100 mesons and more than 100 baryons, although not all are well established experimentally. There are far too many hadrons for us to be comfortable with the notion that they are all elementary particles. An early idea proposed by Geoffrey Chew is that the hadrons are all composites of one another. We do not discuss this idea because it has been superseded by the proposal that hadrons are composites of quarks.

2.10 Quarks

For about 50 years, the proton was believed to be an elementary particle. Then, around 1960 Robert Hofstadter and collaborators at Stanford and R. Wilson and collaborators at Cornell carried out high-energy (over 1000 MeV) electron scattering experiments by protons. These experiments showed that protons have a size greater than zero, and can be considered as opening up the possibility that protons are not elementary particles. Hofstadter received the Nobel prize for his work.

In the Stanford experiments, a *linear accelerator* was used to accelerate the electrons, while at Cornell, a circular *synchrotron* was used. In a linear accelerator, a high-frequency electromagnetic wave is timed so that as a bunch of electrons pass through an evacuated cavity, the electrons experience an electric field always in the same direction. In a synchrotron, a suitably chosen magnetic field causes a bunch of electrons to travel in an approximately circular path in an evacuated tube. Each time around the circle, the electrons are accelerated by a high-frequency electric field whose frequency is a multiple of the orbital frequency of the electrons. As the electrons gain energy, the magnetic field is made to increase to keep the electrons moving in the same orbit, as otherwise the electrons would spiral outward to larger orbits.

We digress here to say that the term “high energy” has meaning only in context. At the time the 1960 electron scattering experiments were carried out, the beam energy was considered to be high, because it was much higher than the energy of a few MeV available in particles from radioactive decay. However, the electron beam energy of about 1 GeV was considerably lower than the beam energies of up to 50 GeV in today’s electron accelerators and up to 1000 GeV in today’s proton accelerators. All the high-energy proton accelerators are synchrotrons.

The reason we need higher and higher energies to probe particle structure at smaller and smaller distances stems from basic quantum mechanics. All particles

have wavelike properties. The wavelength λ of a particle traveling with momentum of magnitude p is given by

$$\lambda = h/p. \quad (2.34)$$

But quantum mechanics says that we cannot resolve phenomena at smaller distances than the wavelength of the beam we use to observe the phenomena. We see from Eq. (2.34) that to obtain small wavelength, we need high momentum, which in turn means high energy. At very high energy, the mass of a particle is negligible compared to its momentum, and then, in units with $c = 1$, the momentum and energy of the particle are equal.

Another important reason to carry out particle collisions at high energy is to be able to create new particles. A certain portion E of the kinetic energy of the incident particle is “available” to be converted into the rest energy (or mass) m of a created particle in accordance with the Einstein equation

$$E = mc^2. \quad (2.35)$$

(In our units, this equation is simply $E = m$.) The higher the available energy, the larger the mass of the particle that can be created. With every increase in accelerator energy, we have the potential for discovering new particles that we have previously been unable to produce.

The energy available to create particles is the energy in the c.m. (center-of-mass, or, more precisely, center-of-momentum) frame of the system. The highest-energy accelerators are so-called *colliders*, in which two beams are accelerated in opposite directions and made to intersect. In a collider in which the particles in each beam have the same energy, the lab system is the same as the c.m. system. The available energy in a collision is twice the energy of one of the particles. On the other hand, if a high-energy beam is incident on a stationary target, the available energy goes as the square root of the energy of the incident particle. Some colliders have beams of unequal energies, and then relativistic kinematics may be used to find the energy available to create particles.

Experiments done early in the century needed only modest energy to explore atomic distance scales (10^{-10} m or 10^5 fm), and could be carried out on a table top by a single physicist. (We can use Eq. (2.5) to obtain the energy needed to explore a given distance.) Present-day experiments on the other hand, explore distances smaller than 10^{-18} m (10^{-3} fm), and are performed by teams of hundreds of physicists at accelerators which are several km in size and which cost hundreds of millions of dollars. At these large accelerators, particles of masses up to about 90 GeV have been created. An accelerator now under construction, the so-called Superconducting Super Collider (SSC), has an estimated cost of about 10 billion (1993) dollars. If it is completed on schedule, it will begin operating early in the 21st century.

We return to the Stanford and Cornell experiments. In each case, the beam energy was a little over 1 GeV, an energy sufficiently high to show that the proton

has a size greater than zero and that both its charge and magnetic moment are distributed over a region with a root-mean-square radius of about 0.8 fm. Despite uneasiness among some physicists at having an “elementary” particle with extended structure, the proton was still regarded as elementary by most workers in the area.

Then in 1964 Gell-Mann and George Zweig independently invented the quark model. According to their original picture, there are three kinds of quarks, now called “up,” “down,” and strange, out of which hadrons are made, with a baryon composed of three quarks and a meson of a quark and antiquark. Gell-Mann received the Nobel prize for his contributions to elementary particle theory.

Even after the invention of the quark model, many physicists did not believe in the reality of quarks and still regarded the proton as elementary. Even Gell-Mann on occasion discussed the possibility that quarks are mathematical objects with no physical reality. Among the physicists who took the quark model seriously from the beginning were Giacomo Morpurgo and Richard Dalitz.

In the 1960's, *deep-inelastic* electron scattering experiments on protons were carried out by teams of physicists headed by Friedman, Kendall, and Taylor, who shared the Nobel prize for their work. Deep-inelastic scattering is scattering in which the beam particle loses a considerable fraction of its energy in the scattering process, the lost energy going into excitation of the target. A considerable number of the electrons were observed to be scattered at large angles, a result which could be interpreted as electron scattering from pointlike constituents of the proton—the quarks. The reasoning behind the interpretation was similar to that used by Rutherford in deducing the existence of the atomic nucleus a half century earlier.

Elastic scattering experiments are insufficient to give evidence of the existence of quarks inside a proton. The reason is that, unlike an atom, which has a massive nucleus at its center, a proton contains three quarks of approximately the same mass, which are moving rapidly with respect to the c.m. of the system. Therefore, in elastic electron-proton scattering, only the average charge distribution of the proton can be measured. On the other hand, deep-inelastic scattering occurs when the electron has a close collision with one of the charged pointlike constituents of the proton, and scatters from it. The recoil of the constituent particle causes the proton to become “excited” and to emit hadrons such as pions. The excitation energy comes at the expense of the energy of the scattered electron.

The deep-inelastic scattering experiments suggested to many physicists that the proton is indeed composite. However, it was not until about 1975 that nearly all physicists were convinced that the proton is made of quarks.

The quark composition of the proton accounts for many of its properties, including size, mass, spin, and anomalous magnetic moment. We shall discuss in later chapters how these properties are obtained within the framework of the quark model.

There were only three kinds (flavors) of quarks in the original quark model, because no more were needed to form the hadrons known at that time. Not very

long afterward, however, an additional flavor (charm) was proposed, and a little later two more flavors (bottom and top) were proposed. Some years later, hadrons containing charmed and bottom quarks were experimentally observed (see Sections 2.11 and 2.13).

After the quark model was proposed, a paradox arose. In the model, the proton is made of two u quarks and one d quark. Furthermore, Quarks have spin $1/2$ in order to account for the fact that baryons have half-integral spin. But particles with half-integral spin are fermions, i.e., their wave function is antisymmetric under the interchange of the coordinates of any identical pair. Thus, the wave function of the two u quarks in a proton ought to be antisymmetric under their interchange. But a good body of experimental evidence (from measurements of the proton form factor and magnetic moment) indicates that the space and spin wave functions of the u quarks are in fact symmetric.

The resolution of this paradox came with the assigning of a new degree of freedom to the quarks, color, which can take on three values (say, red, green, and blue). The color wave function of the quarks in a proton (or in any baryon made of three quarks) is taken to be antisymmetric under the interchange of any two. Then, if any two quarks have identical flavors, the remaining part of the wave function must be symmetric to make the total wave function antisymmetric.

2.11 Charmed particles

In 1964, shortly after Gell-Mann and Zweig proposed that hadrons are made of three kinds of quarks (u , d , s), a number of physicists suggested that a fourth kind of quark, now called a charmed quark c , should also exist. The reasoning was not much more profound than the following: four leptons were known to exist (electron, muon, and two neutrinos), so why not also four quarks? This principle is called “quark-lepton symmetry.”

A few years later, in 1970, Glashow, Illiopolis, and Maiani gave what is perhaps a more compelling reason why the c quark should exist. Their argument is a technical one, based on the fact that certain kinds of weak interactions were not observed experimentally. The unobserved interactions are allowed in a theory with three quarks, but forbidden in a theory with four quarks; therefore, according to the argument, there must exist a fourth quark. Furthermore, it was discovered soon thereafter that the electroweak theory has a serious flaw (a gauge anomaly) unless each family of fermions containing two flavors of leptons also contains two flavors of quarks. This result gave a firm theoretical foundation to the idea of quark-lepton symmetry proposed some years earlier. Thus, in the early 1970's, a number of physicists expected a fourth quark to be observed in the near future.

It was in 1974 that two teams of experimental physicists, one working on the east coast and the other on the west, discovered evidence for a charmed quark. The first team, led by Samuel Ting, performed an experiment at the Alternating Gradient Synchrotron (AGS), a 30 GeV accelerator at Brookhaven National Laboratory on Long Island. The physicists observed $\mu^+ - \mu^-$ pairs emerging from

the nuclear target being bombarded by protons from the accelerator. A plot of the number of muon pairs produced as a function of the *invariant mass* of the pair yielded a clear peak at 3.097 GeV. The invariant mass is defined as the total energy of the pair in its own c.m. system. (Recall that energy and mass have the same units when $c = 1$.) The peak was interpreted as the result of the production of a new meson, which Ting called the J , of mass 3.097 GeV, which then decayed into two muons. This state is now known to be composed of a $c\bar{c}$, and so has hidden charm.

The second team was led by Burt Richter at the Stanford Linear Accelerator Center (SLAC). The two-mile-long linear accelerator at SLAC can accelerate both electrons and positrons, which are then introduced into an electron-positron circular collider. The group observed muon pairs produced at the collider as a function of the total energy (sum of the of the electron and positron energies). They found that at an energy of 3.097 GeV, there was a peak of muon pairs produced as a result of annihilation of the e^+e^- pairs. This team also interpreted their data as the result of the formation of a new heavy meson, which they called the ψ , followed by its decay into a muon pair. Both the east-coast and west-coast teams announced their results at the same time, but the east-coast team actually observed the meson several months earlier. The story is that word of the discovery informally reached the west-coast team, which then knew where to look. In any case, Ting and Richter shared the Nobel prize for the discovery. Unfortunately, the meson now has the complicated symbol J/ψ .

The J/ψ has a mass and width given by

$$m = 3,096.9 \pm 0.1 \text{ MeV}, \quad \Gamma = 86 \pm 6 \text{ keV}. \quad (2.41)$$

Its spin J , parity P , and charge conjugation parity C are given by $J^{PC} = 1^{--}$. Because of its large mass, the J/ψ has many decay modes into lighter particles. None of these modes is dominant. About 88% of its decays are into hadrons, about 6% into e^+e^- pairs, and about 6% into $\mu^+\mu^-$ pairs (the mode of decay which led to its discovery). It is interesting that the strong and electromagnetic decays of the J/ψ compete with each other. Evidentially, the strong decays are inhibited. We shall discuss a reason for this inhibition in a later chapter.

Soon after the discovery of the J/ψ , excited states with similar properties were observed at SLAC. Also observed was a state with somewhat lower energy, the so-called η_c . The spectrum of these higher-energy mesons was qualitatively the same as one would expect from calculating the energy levels of a heavy quark-antiquark pair in a potential.

The discovery of the J/ψ and its excited states convinced many physicists not only that quarks exist but that a new charmed quark had been discovered.

But the J/ψ and its excited states are each composed of a $c\bar{c}$ pair, and consequently contain only hidden charm (the net charm quantum number is zero). But it was not long afterward that mesons containing manifest charm were discovered at Stanford. The first charmed mesons to be seen were the D^+ and D^0 , with

quark content $c\bar{d}$ and $c\bar{u}$ respectively. Also observed was the charmed-strange meson D_s^+ , with quark content $c\bar{s}$. The antiparticles of these mesons were also seen. The masses of these mesons are

$$m_{D^+} = 1869 \pm 1 \text{ MeV}, \quad m_{D^0} = 1864 \pm 1 \text{ MeV}, \quad (2.42)$$

$$m_{D_s} = 1969 \pm 1 \text{ MeV}. \quad (2.43)$$

Each of these mesons decays only weakly. Their lifetimes are

$$\tau_{D^+} = (10.7 \pm 0.3) \times 10^{-13} \text{ s}, \quad \tau_{D^0} = (4.0 \pm 0.1) \times 10^{-13} \text{ s}, \quad (2.44)$$

$$\tau_{D_s} = (4.05 \pm 0.3) \times 10^{-13} \text{ s}. \quad (2.45)$$

Later, excited states were observed with the same quark content as the D and D_s .

Charmed baryons have also been discovered, the first being the Λ_c , which has isospin 0 and is composed of udc . The notation for a charmed baryon is to use the symbol for the corresponding strange baryon, except for a subscript c . The isospin 0 strange baryon is the Λ , composed of uds ; hence the name Λ_c for the baryon in which the strange quark has been replaced by a charmed quark. The Λ_c was observed in both proton synchrotrons and electron colliders. Its mass and lifetime are

$$m_{\Lambda_c} = 2284.9 \pm 0.6 \text{ MeV}, \quad \tau_{\Lambda_c} = (1.9 \pm 0.2) \times 10^{-13} \text{ s}. \quad (2.46)$$

2.12 Leptons

Soon after charmed particles were discovered at SLAC, Martin Perl, one of the members of the experimental team, found some peculiarities in the data. He observed decays of a particle with mass similar to that of the D meson, but which seemed to have different decay modes. Further study revealed that Perl's particle was not a meson, but a third charged lepton, the τ , with mass and lifetime given by

$$m_\tau = 1784 \pm 3 \text{ MeV}, \quad \tau_\tau = (3.05 \pm 0.06) \times 10^{-11} \text{ s}. \quad (2.47)$$

(Sorry that the symbol τ is used both for the lepton and for mean life.)

The τ can decay either into a μ or an electron and two neutrinos. These decay modes are

$$\tau^- \rightarrow \mu^- + \nu_\tau + \bar{\nu}_\mu, \quad \tau^- \rightarrow e^- + \nu_\tau + \bar{\nu}_e. \quad (2.48)$$

Each of these decays occurs about 18% of the time. Thus far, ν_τ neutrinos have not been produced copiously enough in the laboratory to enable physicists to verify that a beam of them colliding with matter will create τ leptons but not muons or electrons. However, there is indirect evidence that the ν_τ is distinct from both ν_e and ν_μ . The best evidence comes from the decay of the weak boson Z^0 , which we discuss in Section 2.15.

Because of its large mass, the τ can decay into hadrons as well as into other leptons. In fact, hadrons are created in τ decays more than 60% of the time. In every τ decay, whether or not hadrons are among the decay products, we expect a ν_τ to be created. A relatively simple hadronic decay mode of the τ is

$$\tau^- \rightarrow \pi^- + \nu_\tau, \quad (2.49)$$

and this decay occurs about 12% of the time. There are many other known decay modes, and these are given in the *Particle Properties Data Booklet*.

With the observation of the τ lepton and the indirect evidence for the existence of a distinct ν_τ , we have a evidence for third family of leptons. Are there any more families? If there are, we have no hint of their existence. Furthermore, if any more lepton families exist and have interaction strengths similar to those of the observed leptons, the additional neutrinos must be heavy or evidence of their existence would have been seen in Z^0 decays (Section 2.15).

2.13 Bottom particles

Once a third family of leptons was discovered, some physicists expected that it would just be a matter of time before a third family of quarks was also observed. In fact, a third quark family had been predicted in 1973 (before the discovery of charm or the τ lepton) by M. Kobayashi and T. Maskawa. These physicists considered the problem of how to account for the fact that CP is not a strictly conserved quantity. They pointed out that, with only four quark flavors, there is no way to break CP symmetry within the standard model, but with six flavors, the standard model can be made to include CP violation.

A few years later, in 1977, in an experiment at the Fermilab synchrotron, a team of physicists led by Lederman discovered a particle which they called the Υ . The group detected $\mu^+ - \mu^-$ pairs emerging from a platinum target that had been bombarded with 400 GeV protons. The physicists observed a peak at 9.46 GeV in the invariant mass of $\mu^+ - \mu^-$ pairs, in much the same way as Ting's group found a lower-energy peak, the J/ψ , three years earlier. The muon pairs were interpreted as coming from rapid decays of the new particle.

The Υ has properties which make it a natural candidate for a meson composed of $\bar{b}b$. Its mass and width are

$$m_\Upsilon = 9,460.3 \pm 0.2 \text{ MeV}, \quad \Gamma_\Upsilon = 52 \pm 2 \text{ keV}. \quad (2.50)$$

The Υ has only hidden bottomness, because the total bottom quantum number is zero.

The same team observed at least one, and possibly two, excited states of the Υ . Later, in electron-positron colliders at Cornell, Stanford, and Hamburg, additional excited states were seen. Some of these excited states have sufficient energy to decay into states with manifest bottomness. In particular, the $\Upsilon(10,580)$

(the number in parentheses is the mass in MeV) sometimes decays into a $B\bar{B}$ pair, where the B is a meson containing a light u or d quark and a \bar{b} quark (B^+ and B^0 respectively). The masses of these mesons are

$$m_{B^+} = 5279 \pm 2 \text{ MeV}, \quad m_{B^0} = 5279 \pm 2 \text{ MeV}. \quad (2.51)$$

The B_s meson, consisting of $s\bar{b}$, has also been observed, but its mass is not yet well known.

2.14 Gluons

Quantum chromodynamics (QCD) is a theory of quarks and gluons and their interactions. Presumably, gluons are confined to the interior of hadrons, and so will never be directly observed. Furthermore, unlike quarks, gluons do not have flavor quantum numbers, and so we cannot obtain evidence for their existence from the flavor quantum numbers of hadrons. However, the success of physicists in using QCD to make predictions which agree with experiment provides indirect evidence for the existence of gluons.

Some of the evidence for gluons comes from electron-positron annihilation. If e^+ and e^- annihilate into a quark-antiquark pair, the quark and antiquark go off in opposite directions. They cannot get very far without converting into hadrons because of the strong attractive color force acting between them. However, what we see in the lab are two “jets” of hadrons going off in opposite directions. Each jet is a fairly-well columnated spray of hadrons peaked in the direction of the original quark or antiquark which gave rise to it. The observed angular distributions and cross sections for these jets are similar to those calculated in QCD for the quark-antiquark pair.

Just as an accelerating charged particle radiates photons, accelerating colored quarks radiate gluons. In the pair creation of a quark and antiquark, sometimes, one of the produced particles will radiate an energetic gluon. This gluon, because it is colored, cannot escape, but converts into another jet of hadrons. A so-called “gluon jet” has somewhat different properties from a quark jet. Typically, the emitted hadrons are not so well columnated in a gluon jet. The observed frequency of emission and angular distribution of gluon jets agree with QCD calculations. Thus, physicists are confident that they have indirectly observed gluons.

2.15 Weak bosons

A number of the main ingredients of the electroweak sector of the standard model were already in place in the late 1960’s, although there have been numerous improvements since then. In particular, in the late 1960’s one could calculate the masses of the weak W and Z bosons in terms of properties of the weak interactions measured at low energy.

It was not until 1983 that the W and Z were actually observed at a high-energy proton-antiproton collider at CERN. Carlo Rubbia, the leader of the team

that discovered these particles, and Simon van der Meer, a physicist who achieved high luminosity in the colliding beams, shared the Nobel prize for this achievement.

The W and Z were observed to have masses approximately equal to those predicted by the theory. The present best values of these masses are:

$$M_W = 82.2 \pm 0.3 \text{ GeV}, \quad M_Z = 91.17 \pm 0.02 \text{ GeV}, \quad (2.53)$$

while their widths are

$$\Gamma_W = 2.1 \pm 0.1 \text{ GeV}, \quad \Gamma_Z = 2.49 \pm 0.01 \text{ GeV}. \quad (2.54)$$

Although the Z was discovered at a proton-antiproton collider, many of its properties were obtained by measurements at the CERN electron-positron collider called LEP. The decay of the Z into a neutrino-antineutrino pair can be calculated in the standard model. The width of the Z depends on how many standard species of neutrino exist in nature. The observed width agrees with the calculation if there are three different kinds of neutrinos. Thus, although the standard model does not tell us how many families there are, the measured width of the Z tells us that there are three. Of course, the experiment does not rule out the existence of a fourth family with a neutrino so heavy that the Z could not decay into it. Likewise, the measurement does not rule out the existence of neutrinos with nonstandard interactions.

3. A LITTLE ABOUT GROUP THEORY

3.1 Preliminaries

It is an apparent fact that nature exhibits many symmetries, both exact and approximate. A symmetry is an invariance property of a system under a set of transformations. For example, our faces have approximate reflection symmetry, because we look approximately the same in a photograph as in a mirror. As another example, a sphere has rotational symmetry because it looks the same no matter how it is rotated.

Symmetry transformations of physical systems have properties analogous to those of a mathematical *group*. These properties are: If we successively perform two symmetry transformations we obtain a unique symmetry transformation; the transformations are associative; and inverse and identity transformations exist.

We have already mentioned in chapter 1 a theorem, called *Noether's theorem*, which relates symmetry principles to conservation laws. Noether's theorem says that if a physical system can be described by a classical Lagrangian which is invariant under a continuous group of transformations, then the system has a conserved current. If a classical field is quantized, the resulting quantum field theory usually has the same symmetry. However, the quantized theory may have an *anomaly*, which breaks the classical symmetry. We briefly discuss anomalies in Section 4.8.

As two examples of symmetry, we note that the fundamental interactions of nature are apparently invariant under the group of translations and the group of rotations in three dimensions. Noether's theorem relates symmetry under translations to the law of conservation of momentum, and the symmetry under rotations to the law of conservation of angular momentum. The translation and rotation groups are examples of *Lie* groups, which we define in Section 3.3.

The Lagrangian of the standard model is invariant under the group of gauge transformations $SU(3) \times SU(2) \times U(1)$. (We define gauge transformations in chapter 4.) The standard model is also invariant under the proper Poincaré group, which includes translations in space and time, rotations, and proper Lorentz transformations. (Improper Lorentz transformations include space and time reflections.) The Poincaré group is a Lie group. The groups $SU(3)$, $SU(2)$, and $U(1)$ are *special unitary* groups, which are also Lie groups.

A unitary group is a group of unitary matrices, and a special unitary group is a group of unitary matrices with determinants equal to unity. In order to understand the standard model, we have to have some familiarity with the Lie groups and their Lie algebras, especially unitary groups. (We discuss unitary groups further in Section 3.5.)

In addition to the space–time symmetries of the proper Poincaré group, some theories have additional symmetries under finite transformation groups. Of these, we single out space reflection or parity P , time inversion T , and charge conjugation C . We do not give much discussion of these groups, but they play an important role in the standard model. As we have already mentioned, none of the groups P , C , and T is an exact symmetry of the standard model, but the combined symmetry CPT , taken in any order, is exact.

In this chapter we briefly discuss groups in general, then Lie groups and their algebras, and the unitary groups and their algebras. We concentrate on group representations (which we define in the Section 3.2), especially irreducible unitary representations. We can have unitary representations of many different groups, not only of unitary groups. The treatment in this chapter may seem to be a little condensed for those who only know a little about group theory. More details can be found in many places, for example, in a book on unitary symmetry (Lichtenberg, 1978). We do not discuss the Poincaré group in any detail.

A group G is a set of elements which satisfy four postulates:

- 1) A law of combination, often called a product, is defined so that if a and b belong to G , the product ab is a unique element of G .
- 2) Multiplication is *associative*, i.e., $a(bc) = (ab)c$.
- 3) An *identity* e exists such that $ea = ae = a$.
- 4) An *inverse* a^{-1} exists to any element a such that $a^{-1}a = aa^{-1} = e$.

The number of elements of a group may be finite, in which case the group is called a finite group, or infinite. If all the elements of a group commute with one another, the group is said to be *abelian*. Otherwise the group is *nonabelian*. A subgroup of a group is a subset of elements which is itself a group under the same multiplication law. Every group has at least two subgroups: itself and the group consisting only of the identity. These are called *improper* subgroups; any others are called *proper* subgroups. A group H is *homomorphic* to a group G if there is a mapping of the elements of G onto the elements of H . The groups are *isomorphic* if the mapping is one-to-one.

An element a belonging to G is said to be *conjugate* to an element b

in G if there exists an element u in G such that $a = ubu^{-1}$. Let H be a subgroup of G , and let h be in H and g be in G . Form the product elements $h' = ghg^{-1}$ for all h . Then the h' form a group H' which is isomorphic to H . If, for all g in G , the elements of H and H' are identical, then H is called an *invariant* or self-conjugate subgroup of G .

The group G is said to be the *direct product* of two groups H and H' if every h in H commutes with every h' in H' and if every g in G can be written uniquely as a product of an element in H and an element in H' . The direct product is written in the form $G = H \times H'$.

3.2 Group representations

A *representation* of a group is a homomorphism between the group and a group of linear operators which operate on a vector space. We can think of the vectors in this space as being the states (wave functions) of a quantum mechanical system. A finite-dimensional *matrix representation* of a group is a homomorphism between the group and a group of matrices. We often simply use the word “representation” to mean a matrix representation. If a representation is isomorphic to the group, it is said to be *faithful*. We shall consider only representations by square matrices. If G is a group with elements g , then we often denote the corresponding element of the representation by $D(g)$.

The matrices of a representation are a special case of linear operators which act on a vector space. If the matrices are n -by- n , the vectors (wave functions) are column matrices with n entries, and their hermitian conjugates are row matrices. The vectors are members of an n -dimensional vector space, and therefore the matrices are said to be n -dimensional.

A similarity transformation is a transformation by means of a matrix S which leaves unaltered the algebra of the transformed system. A similarity transformation acts differently on a representation D and on a vector V , namely

$$D' = SDS^{-1}, \quad V' = SV. \quad (3.1)$$

If a representation can be brought into the following form by a similarity transformation:

$$D(g) = \begin{pmatrix} D_1(g) & X(g) \\ 0 & D_2(g) \end{pmatrix} \quad (3.2)$$

for all g , then the representation is called *reducible*. If not, it is *irreducible*. If $X(g) = 0$, the representation is fully reducible. We shall restrict our considerations to cases in which reducible representations are fully reducible,

and we shall omit the word “fully.” A theorem (Schur’s lemma) states: A matrix which commutes with all matrices of an irreducible representation is a multiple of the unit matrix.

We next discuss the importance of irreducible unitary representations of groups within the Hamiltonian formalism, as using this formalism is somewhat easier than using the Lagrangian formalism. Let us consider an n -dimensional irreducible unitary representation of a group G . The unitary matrices act on a set of n linearly-independent vectors, which can be chosen to be orthonormal. The members of this orthonormal set (or basis) constitute a *multiplet*.

Let a unitary representation of a symmetry group (that is, a group of transformations which leaves the physical system invariant) be denoted by U_a , where a stands for all the parameters which specify individual group elements. If we have any transition matrix (ϕ, ψ) , where ϕ and ψ are state vectors (or wave functions) describing physical states, then the transformed states $\phi' = U_a\phi$ and $\psi' = U_a\psi$ satisfy the condition

$$(\phi', \psi') = (U_a\phi, U_a\psi) = (U_a^{-1}U_a\phi, \psi) = (\phi, \psi). \quad (3.3)$$

Thus, unitary transformations are important in quantum mechanics because they leave transition matrixes invariant.

If the Hamiltonian H of a physical system is invariant under a symmetry group G , then all members of a multiplet belonging to an irreducible unitary representation have the same energy, as we now show. Now consider the eigenvalue equation

$$H\psi_n = E_n\psi_n. \quad (3.4)$$

If we operate on this equation with U_a , we get

$$U_aH\psi_n = U_aHU_a^{-1}U_a\psi_n = E_nU_a\psi_n. \quad (3.5)$$

Now if H' and ψ'_n are defined as

$$H' = U_aHU_a^{-1}, \quad \psi'_n = U_a\psi_n, \quad (3.6)$$

our equation becomes

$$H'\psi'_n = E_n\psi'_n. \quad (3.7)$$

But because U_a is a symmetry group, by definition it leaves the Hamiltonian H invariant, so that $H' = H$. This implies that U_a commutes with the Hamiltonian:

$$HU_a = U_aH, \text{ or } [H, U_a] = 0, \quad (3.8)$$

where $[H, U_a] = HU_a - U_aH$ is called the *commutator* of H and U_a . Then Eq. (3.7) becomes simply

$$H\psi'_n = E_n\psi'_n, \quad (3.9)$$

so that the transformed wave functions ψ'_n are also eigenfunctions of the Hamiltonian with the same energy eigenvalue. But the transformed wave functions are in general linear combinations of all members of the original multiplet. Therefore, in order for Eq. (3.9) to be true, all members of the multiplet must have the same energy eigenvalue. We mention that if the representation is reducible, the new wave functions are not in general linear combinations of *all* the wave functions belonging to the representation, so that all the wave functions do not need to have the same energy.

It should be clear from the above arguments that if any operator A commutes with the U_a , then all members of a multiplet have the same eigenvalue of the operator A . Thus, for example, let us consider the rotation group $R(3)$. Not only is the Hamiltonian invariant under rotations, so that all members of a multiplet have the same energy, but U_a also commutes with the operator J^2 , so that all members of a multiplet have the same eigenvalue of J^2 , namely, $J(J+1)$.

3.3 Lie groups

We have noted that a group may have a finite or infinite number of elements. A *Lie* group has a continuously infinite number of elements characterized by a finite number of parameters which can vary continuously. Furthermore, if an element of a Lie group is characterized by a set of r parameters collectively denoted by a ($a = a_1, a_2 \dots a_r$), and another element is characterized by a set of parameters b , then the product element is characterized by a set of parameters c which are *analytic functions* of a and b .

As an example of a Lie group, consider the rotations in two dimensions. These are characterized by a parameter θ . The transformation is

$$\begin{aligned} x'_1 &= x_1 \cos \theta - x_2 \sin \theta, \\ x'_2 &= x_1 \sin \theta + x_2 \cos \theta. \end{aligned} \quad (3.10)$$

The transformation can be written in matrix form as

$$x' = R(\theta)x, \quad (3.11)$$

where

$$x = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}, \quad R = \begin{pmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{pmatrix}. \quad (3.12)$$

The rotation matrix $R(\theta)$ is the group element characterized by the single parameter θ . Rotations in two dimensions constitute an abelian group, but in more dimensions the rotation group is nonabelian. Note that the groups $R(n)$ are faithful representations of themselves in n dimensions.

The group multiplication law for rotations in 2 dimensions can be stated as follows: If

$$R(\theta) = R(\theta_2)R(\theta_1), \quad (3.13)$$

then

$$\theta = \theta_2 + \theta_1. \quad (3.14)$$

The rotation groups are *compact*. This means that the parameters vary over a *finite, closed* region. For example, the parameter θ of the two-dimensional rotation group varies over the interval $0 \leq \theta \leq 2\pi$.

On the other hand, the translation groups are not compact because the parameters are unbounded. For example, a translation in 1 dimension,

$$x' = x + a,$$

is characterized by a parameter a which can vary from $-\infty$ to ∞ . Likewise, the group of Lorentz transformations is not compact because the group is characterized by a parameter v (the velocity) which varies in the interval $0 \leq v < c$, which is open at one end. Rotations and Lorentz transformations are both subgroups of the Lorentz group.

The concepts of simple and semisimple Lie groups are important but somewhat complicated. An oversimplified definition, which is adequate for our purposes, is that a Lie group is *simple* if it is nonabelian and has no proper invariant Lie subgroups. It is *semisimple* if it is nonabelian and has no abelian invariant Lie subgroups. Clearly, a simple group is also semisimple. If a group is the direct product of two or more groups H, H', \dots , then the subgroups H, H', \dots are invariant. The direct product of simple and/or semisimple Lie groups is semisimple.

Recall that the local gauge group of the standard model is $SU(3) \times SU(2) \times U(1)$. This group is not semisimple because it has an abelian invariant subgroup $U(1)$. However, the group $SU(3) \times SU(2)$ is semisimple. The groups $SU(3)$ and $SU(2)$ are simple.

3.4 Lie algebras

Let us consider a Lie group of transformations. We obtain the *Lie algebra* of the group by considering group elements which differ only infinitesimally

from the identity. From these elements we can construct operators called *generators* which allow us to obtain a unitary representation of the group. More precisely, we obtain all the elements of the group which can be generated by continuous transformations from the identity. There is one generator for each parameter of the group. Methods for obtaining the generators of a Lie group have been discussed in many places (see, e.g., Lichtenberg, 1978).

Let the generators of a Lie group be X_i , $i = 1, 2, \dots, r$, where the group is characterized by r real parameters a_i . If the generators are Hermitian, a unitary representation of an arbitrary group element U_a is given by

$$U_a = e^{-i \sum a_i X_i}.$$

It can be shown that the X_i form a Lie algebra, which means that they satisfy the algebraic equations

$$[X_i, X_j] = i \sum_{k=1}^r c_{ij}^k X_k. \quad (3.15)$$

Here the *commutator* $[A, B] = AB - BA$ and the c_{ij}^k are constants called the *structure constants* of the group. (Some people call other constants $b_{ij}^k = ic_{ij}^k$, the group structure constants.) There is no significance to the fact that we write c_{ij}^k with both lower and upper indices in Eq. (3.15). We do this because in the future we shall use the *summation convention* of omitting the summation sign and summing over a repeated upper and lower index (in any order). The structure constants of a Lie algebra can differ with different choices of generators.

As we see from Eq. (3.15), a Lie algebra has the property that the commutator of any two members of the algebra (generators of the Lie group) is a linear combination of the members of the Lie algebra. We also see that the algebra is in general neither commutative nor associative. A *representation* of a Lie algebra is a set of matrices which obey the commutation relations of the algebra.

If a Lie group is abelian, all the commutators of its Lie algebra vanish, i.e. all its structure constants are zero. The maximum number of commuting generators of a Lie group is called the *rank* of the group. Since any generator commutes with itself, every Lie group is at least rank one. The k commuting generators of a rank k Lie group can be simultaneously diagonalized in a matrix representation.

If a Lie group of rank k is semisimple and compact, then one can construct from the members of its Lie algebra k nonlinear *invariant* operators, called *Casimir* operators, which commute with every member of the algebra.

As an example, let us consider the Lie algebra and Casimir operators of the familiar rotation group in three dimensions $R(3)$. This group is characterized by 3 parameters (for example, the Euler angles). Therefore, it has three generators, which can be taken to be the familiar angular momentum operators J_x , J_y , and J_z . They satisfy the Lie algebra

$$[J_x, J_y] = iJ_z, \quad \hbar = 1, \quad (3.16)$$

and cyclic permutations. This group is rank one because none of the J_i commutes with any other. It is also semisimple (actually, simple), so that it has one Casimir operator J^2 given by

$$J^2 = J_x^2 + J_y^2 + J_z^2. \quad (3.17)$$

A representation of the Casimir operator in n dimensions commutes with all the members of an irreducible representation in n dimensions. Therefore, by Schur's lemma, a representation of the Casimir operator is a multiple of the unit matrix.

It is convenient to denote the generators of $R(3)$ by J_1 , J_2 , and J_3 , and write

$$[J_i, J_j] = i\epsilon_{ij}^k J_k, \quad i, j, k = 1, 2, 3, \quad (3.18)$$

where $\epsilon_{ij}^k = \epsilon_{ijk}$ is completely antisymmetric in its indices and $\epsilon_{123} = 1$. Then we see that the structure constants of $R(3)$ are given by ϵ_{ijk} or $i\epsilon_{ijk}$, depending on whether the structure constants are defined with or without the i . It is easy to show from the definition of rotations that the number of parameters of $R(n)$ is $(n^2 - n)/2$.

3.5 Unitary groups and algebras

The unitary group in n dimensions $U(n)$ is the group of $n \times n$ matrices U_a satisfying

$$U_a^\dagger = U_a^{-1}, \quad (3.19)$$

where a stands for the parameters of the group, the dagger denotes the Hermitian conjugate matrix, and the superscript -1 denotes the inverse. By definition, for any matrix A , we have $(A^\dagger)_{ij} = A_{ji}^*$, with the asterisk denoting the complex conjugate.

A complex matrix in n dimensions is specified by $2n^2$ real numbers. If the matrix is unitary, there are n^2 relations among these numbers, so that $U(n)$ is characterized by n^2 parameters. The group $U(1)$ is one-dimensional

and is characterized by only one parameter. Each element of $U(1)$ is a phase $e^{i\theta}$.

The special unitary groups $SU(n)$ have matrices with determinants equal to unity. This provides another relation so that $SU(n)$ is characterized by $n^2 - 1$ parameters. The rank of $SU(n)$ is $n - 1$. The $SU(n)$ groups are semisimple and compact, so that $SU(n)$ has $n - 1$ Casimir operators.

Like $R(3)$, $SU(2)$ has 3 parameters and is of rank 1. In fact, the generators of $SU(2)$ satisfy the same Lie algebra as the generators of $R(3)$. This implies that the two groups are locally isomorphic (i.e., the mapping of a neighborhood of one onto a neighborhood of the other is one-to-one) and globally homomorphic. In fact, the homomorphism is two-to-one from $SU(2)$ onto $R(3)$.

The groups $U(n)$ and $SU(n)$ are matrix groups, and so are faithful representations of themselves. In discussing representations of the unitary groups, we usually confine ourselves to $SU(n)$. The reason is that the algebra of $U(n)$ is the same as the algebra of $SU(n) \times U(1)$, and all the representations of $U(1)$ are one dimensional. The group $SU(n)$ has $n - 1$ so-called fundamental representations. Of these, two are n -dimensional if $n > 2$. There is only one fundamental (two-dimensional) representation if $n = 2$.

The group $SU(n)$ also has a representation of $n^2 - 1$ dimensions, the same number as the number of generators of the group. This representation is called the *adjoint* representation.

We can construct n -dimensional representations of the algebra of $SU(n)$. For $n = 2$, we can choose these matrices to be the familiar Pauli spin matrices σ_1 , σ_2 , and σ_3 , given by

$$\sigma_1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad \sigma_2 = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \quad \sigma_3 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}. \quad (3.20)$$

The Pauli matrices satisfy the commutation relations

$$[\sigma_i, \sigma_j] = 2i\epsilon_{ij}^k \sigma_k. \quad (3.21)$$

Note the factor 2 difference between the structure constants when expressed in terms of the σ 's rather than in terms of the J 's, given in Eq. (3.10). This follows because $\sigma_i = 2J_i$, and is an example of the fact that the structure constants depend on the representation of the Lie algebra. Still another representation of the Lie algebra of $SU(2)$ or $R(3)$ is by the two-dimensional matrices σ_+ , σ_- , and σ_3 , where

$$\sigma_+ = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}, \quad \sigma_- = \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}. \quad (3.22)$$

The matrix σ_+ is called a raising operator and σ_- is called a lowering operator because of their action on the eigenvectors of the operator σ_3 (see Section 3.6). The matrices σ_+ and σ_- can be written in terms of σ_1 and σ_2 as follows:

$$\sigma_+ = (\sigma_1 + i\sigma_2)/2; \quad \sigma_- = (\sigma_1 - i\sigma_2)/2.$$

The Casimir operator of $SU(2)$ is the same as that of $R(3)$. We can write the two-dimensional Casimir operator in terms of the Pauli matrices:

$$J^2 = \frac{1}{4}\sigma^2,$$

where

$$\sigma^2 = \sum_{i=1}^3 \sigma_i^2 = 2(\sigma_+\sigma_- + \sigma_-\sigma_+) + \sigma_3^2. \quad (3.23)$$

We now turn to $SU(3)$. The generalization of the Pauli matrices are the so-called Gell-Mann matrices λ_i ($i = 1, 2, \dots, 8$), which are given by

$$\begin{aligned} \lambda_1 &= \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, & \lambda_2 &= \begin{pmatrix} 0 & -i & 0 \\ i & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, & \lambda_3 &= \begin{pmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \\ \lambda_4 &= \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix}, & \lambda_5 &= \begin{pmatrix} 0 & 0 & -i \\ 0 & 0 & 0 \\ i & 0 & 0 \end{pmatrix}, & & (3.24) \\ \lambda_6 &= \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}, & \lambda_7 &= \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & -i \\ 0 & i & 0 \end{pmatrix}, & \lambda_8 &= \frac{1}{\sqrt{3}} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -2 \end{pmatrix}. \end{aligned}$$

The λ_i satisfy the commutation relations

$$[\lambda_i, \lambda_j] = 2if_{ij}^k \lambda_k, \quad (3.25)$$

with summation implied. The f_{ijk} are themselves often called the structure constants of the group. They are given in Table 3.1.

There are two Casimir operators of $SU(3)$, one quadratic and the other cubic in the generators. We shall have occasion to use only the quadratic Casimir operator F^2 , which is given by

$$F^2 = \frac{1}{4}\lambda^2 = \frac{1}{4} \sum_{i=1}^8 \lambda_i^2. \quad (3.26)$$

Table 3.1. Nonvanishing structure constants of $SU(3)$. The f_{ijk} are anti-symmetric under permutation of any two indices.

ijk	f_{ijk}	ijk	f_{ijk}
123	1	345	1/2
147	1/2	367	-1/2
156	-1/2	458	$\sqrt{3}/2$
246	1/2	678	$\sqrt{3}/2$
257	1/2		

We now introduce a different representation for the generators, which is suitable to generalization to any $SU(n)$. We introduce the notation H_a , ($a = 1, 2, \dots, n-1$) for the mutually commuting generators (which can be simultaneously diagonalized) and the notation E_{ab} for the $n^2 - n$ nondiagonal generators. These are $n \times n$ matrices with matrix elements given by

$$(H_a)_{jk} = \delta_{jk} \left[\sum_{l=1}^a \delta_{jl} - a\delta_{j,a+1} \right], \quad (3.27)$$

$$(E_{ab})_{jk} = \delta_{aj}\delta_{bk}, \quad a \neq b, \quad (3.28)$$

where δ_{ab} are elements of the unit matrix. If $a > b$, E_{ab} is a lowering operator; if $a < b$, it is a raising operator. Also, $E_{ab} = E_{ba}^\dagger$. In $SU(2)$,

$$H_a = \sigma_3, \quad E_{12} = \sigma_+, \quad E_{21} = \sigma_-. \quad (3.29)$$

We see from Eqs. (3.27) and (3.28) that the $SU(3)$ generators are

$$\begin{aligned}
 H_1 &= \begin{pmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 0 \end{pmatrix}, & H_2 &= \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -2 \end{pmatrix}, & E_{12} &= \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \\
 E_{21} &= \begin{pmatrix} 0 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, & E_{13} &= \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, & E_{31} &= \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix}, \\
 E_{23} &= \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix}, & E_{32} &= \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}.
 \end{aligned} \quad (3.30)$$

It should be clear from the examples we have given for $SU(2)$ and $SU(3)$ that it is straightforward to write down the matrix generators of any $SU(n)$ in both the λ_i and (H_a, E_{ab}) representations (up to normalization constants).

We can use the lowering operators $E_{a+1,a}$ and their hermitian conjugates to obtain the Clebsch-Gordan coefficients of any $SU(n)$. We show in the next section how this is done.

The trace of a matrix is the sum of its diagonal elements. The $n^2 - 1$ generators of $SU(n)$ are traceless matrices in n dimensions. Any real $n \times n$ traceless matrix can be written as a linear combination of them, and any real $n \times n$ matrix can be written as a linear combination of them and the unit matrix in n dimensions.

3.6 Multiplets of unitary groups

The n -dimensional generators of $SU(n)$ operate on n -dimensional column vectors. Clearly, there are n linearly independent vectors, which we may denote by u_a , ($a = 1, 2 \dots n$). A convenient representation for these vectors is that the j -th row of u_a is equal to δ_{aj} . In $SU(2)$ the u_a are

$$u_1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \quad u_2 = \begin{pmatrix} 0 \\ 1 \end{pmatrix}. \quad (3.31)$$

In $SU(3)$ they are

$$u_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \quad u_2 = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, \quad u_3 = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}. \quad (3.32)$$

We can order the vectors from *highest* to *lowest*, such that u_a is *higher* than u_b if $a < b$. The vectors u_a of $SU(n)$ are said to belong to the first *fundamental* representation. Altogether, $SU(n)$ has $n - 1$ inequivalent so-called fundamental representations, two of which have n dimensions, except for $SU(2)$, which has only one fundamental representation. All the multiplets of $SU(n)$ can be built up from the vectors of the first fundamental representation using only the raising and lowering matrices of the n -dimensional representation of the Lie algebra. Sometimes, however, it is convenient to use more than one fundamental representation in building the multiplets.

The eigenvalues of the H_a operating on a state vector is called the *weight* \mathbf{m} of the vector. For example, in $SU(3)$, we see from Eqs. (3.30) and (3.32) that $(H_1, H_2)u_2 = (-1, 1)u_2$. therefore, the weight \mathbf{m} of u_2 is $\mathbf{m} = (m_1, m_2) = (-1, 1)$.

We see from the definitions of E_{ab} and the u_a , that

$$E_{ab}u_c = u_a\delta_{bc}. \quad (3.33)$$

With our ordering of the vectors such that u_a is higher than u_b if $a < b$, we can see that the operators E_{ab} are indeed raising or lowering operators for $a < b$ or $a > b$ respectively. We shall restrict ourselves to the the lowering operators $E_{a+1,a}$ and their hermitian conjugates, which are raising operators.

Let us use the u_a in $SU(2)$ and $SU(3)$ to build up some other multiplets of these groups. We start with $SU(2)$, which is a case which should be familiar. To make things more concrete, u_1 and u_2 can be the spin-up and spin-down wave functions (state vectors) of a particle of spin $1/2$. As another possibility, they can be the isospin wave functions of a u and d quark respectively. If we wish to consider N -particle wave functions, we start with the wave function $u_1(1)u_1(2)\dots u_1(N)$, where the numbers in parentheses stand for particle 1, particle 2, etc. A simpler notation is to omit the numbers in parentheses, and by convention write the wave function of particle 1 first, etc. We shall adopt this simpler notation in the following. We also introduce the notation that the lowering operator E_{ab} operating on an N -particle state is given by

$$E_{ab} = \sum_{i=1}^N E_{ab}(i). \quad (3.34)$$

It is best to begin with only two particles. Let $\chi_1 = u_1u_1$ and operate on χ_1 with the lowering operator

$$E_{21} = E_{21}(1) + E_{21}(2) = \sigma_-(1) + \sigma_-(2). \quad (3.35)$$

We get

$$E_{21}\chi_1 = u_1u_2 + u_2u_1 = \sqrt{2}\chi_2, \quad (3.36)$$

where we have defined χ_2 to be a normalized state. Repeating the operation, we get

$$E_{21}\chi_2 = \sqrt{2}u_2u_2 = \sqrt{2}\chi_3. \quad (3.37)$$

If we operate on χ_3 we get 0. Thus, starting from two doublets of $SU(2)$, we have obtained a triplet state χ_i , corresponding to spin or isospin 1. We next construct a state ϕ_1 orthogonal to χ_2 . We see that ϕ_1 must be given by

$$\phi_1 = (u_1u_2 - u_2u_1)/\sqrt{2}. \quad (3.38)$$

If we operate on ϕ_1 we get 0, as we expect, so that this state is a singlet, corresponding to spin or isospin 0. We can obtain the eigenvalues of the diagonal operators J_3 and J^2 by directly operating on the χ_i and ϕ_1 .

The coefficients multiplying the product wave functions $u_i u_j$ in the expressions for χ_i and ϕ_i are known as Clebsch-Gordan coefficients. In the case we have considered, these Clebsch-Gordan coefficients are unique, but in the case of the product of three or more wave functions, the Clebsch-Gordan coefficients can depend on somewhat arbitrary definitions of wave functions. We can see this as follows: If we start with the product $u_1 u_1 u_1$, we can use the lowering operator E_{21} to construct all the symmetric wave functions belonging to the same multiplet as $u_1 u_1 u_1$. The problem arises when we want to construct other multiplets. For example, consider the (unnormalized) wave function ξ given by

$$\xi = u_1 u_1 u_2 + u_1 u_2 u_1 + u_2 u_1 u_1. \quad (3.39)$$

There are *two independent* wave functions orthogonal to ξ , and the Clebsch-Gordan coefficients depend on what linear combination of these wave functions we choose. The choice in some instances is given by convention, but real questions of physics can influence what choice is convenient.

Let us now generalize to $SU(3)$. The χ_i and ϕ_1 from $SU(2)$ are also wave functions of $SU(3)$, but they are not complete multiplets. We can operate on χ_2 with E_{32} to get

$$\chi_4 = (u_1 u_3 + u_3 u_1)/\sqrt{2}. \quad (3.40)$$

Operating on χ_4 with E_{21} , we get

$$\chi_5 = (u_2 u_3 + u_3 u_2)/\sqrt{2}, \quad (3.41)$$

and operating on χ_5 with E_{32} , we get after normalizing

$$\chi_6 = u_3 u_3. \quad (3.42)$$

Thus, the χ_i are a sextet of $SU(3)$. Likewise, from ϕ_1 , we can obtain

$$\phi_2 = u_1 u_3 - u_3 u_1, \quad \phi_3 = u_2 u_3 - u_3 u_2, \quad (3.43)$$

so that the ϕ_i are a triplet of $SU(3)$.

We now define a new triplet of vectors \bar{u}_a given by

$$\bar{u}_1 = \phi_3, \quad \bar{u}_2 = -\phi_2, \quad \bar{u}_3 = \phi_1. \quad (3.44)$$

By explicit construction we find

$$E_{ab}\bar{u}_c = -\bar{u}_b\delta_{ac}. \quad (3.45)$$

This differs from the action of E_{ab} on u_c . The \bar{u}_a are the vectors belonging to the second fundamental representation of $SU(3)$.

3.7 Young tableaux

A *Young tableau* or *Young diagram* describes the symmetry of a collection of an integer ν identical particles, each of which can be in one of several available states. We shall confine our considerations to the description of Young diagrams in the case that the symmetry group is $U(n)$ or $SU(n)$ and the particles belong to the first fundamental representation of the group. Then the number of possible states of a particle is also n . An example is a collection of five electrons, each of which can be in one of two spin states (spin up or spin down). In this case $\nu = 5$, $n = 2$.

A Young tableau is a collection of boxes, one for each particle, arranged in rows and columns to represent the symmetry of the state: symmetric in rows, antisymmetric in columns. A *proper* tableau is one in which each row is no longer than the one above it and each column is no longer than the one to the left of it. When we refer to a Young tableau, we shall mean a proper tableau unless otherwise specified. An example of a proper tableau is the following:



$$\begin{array}{|c|c|c|c|} \hline & & & \\ \hline & & & \\ \hline & & & \\ \hline \end{array} \quad (3.46)$$

We begin by using Young tableaux in connection with $SU(2)$. We denote the basis vectors of the fundamental doublet of $SU(2)$ by u_1 and u_2 . These vectors may, for example, denote the two states of a particle with spin $1/2$. Another notation for these one-particle states is by means of a Young tableau with a single box. We make the identification

$$u_1 = \boxed{1}, \quad u_2 = \boxed{2} \quad (3.47)$$

The single box without a number stands for both members of the doublet. The same considerations hold for $U(2)$.

Now suppose we have a two-particle state. If it is a symmetric ψ_s , we denote it by a row, and if it is an antisymmetric state ψ_a , by a column:

$$\psi_s = \boxed{\quad\quad}, \quad \psi_a = \begin{array}{|c|} \hline \quad \\ \hline \quad \\ \hline \end{array} \quad (3.48)$$

These tableaux represent multiplets which are different than the two-dimensional fundamental multiplet. Consider first the symmetric state. If both particles are in the state u_1 or both are in the state u_2 , the corresponding tableaux are

$$\begin{array}{|c|c|} \hline 1 & 1 \\ \hline \end{array}, \quad \begin{array}{|c|c|} \hline 2 & 2 \\ \hline \end{array}.$$

There is one symmetric state with one particle having the state vector u_1 and the other having the vector u_2 , namely $(u_1u_2 + u_2u_1)/\sqrt{2}$. (We adopt the convention of writing the state vector of the first particle at the left, the vector of the second particle next, etc. This convention saves us the trouble of writing $[u_1(1)u_2(2)+u_2(1)u_1(2)]/\sqrt{2}$.) This symmetric state is represented by the tableau

$$\begin{array}{|c|c|} \hline 1 & 2 \\ \hline \end{array}.$$

The arrangement

$$\begin{array}{|c|c|} \hline 2 & 1 \\ \hline \end{array}$$

is obviously the same as the previous arrangement, and must not be counted. It is called a *nonstandard* arrangement. Thus, the symmetric state is a triplet. There is only one antisymmetric two-particle state $(u_1u_2 - u_2u_1)/\sqrt{2}$, corresponding to the arrangement

$$\begin{array}{|c|} \hline 1 \\ \hline 2 \\ \hline \end{array}.$$

The other arrangement

$$\begin{array}{|c|} \hline 2 \\ \hline 1 \\ \hline \end{array}$$

is nonstandard and must not be counted.

The above considerations for $U(2)$ or $SU(2)$ can be generalized to any $U(n)$ or $SU(n)$ and lead us to the following definition:

A *standard arrangement* of a tableau of $U(n)$ or $SU(n)$ is a proper tableau containing a positive integer i in each box ($1 \leq i \leq n$) such that the integers increase in going from top to bottom in a column and do not decrease in going left to right in a row. Hereafter, unless we explicitly state otherwise, an arrangement will mean a standard arrangement.

An important theorem which we do not prove is that the number N of standard arrangements of a Young tableau with positive integers no greater

than n is equal to the dimension of an irreducible representation of $U(n)$ or $SU(n)$.

We see that a Young tableau for $U(n)$ or $SU(n)$ consists of ν boxes in no more than n rows. The tableaux are limited to n rows because one cannot antisymmetrize a configuration of more than n particles when each particle has only n available states.

As an example, consider a collection of five electrons, each of which can be in one of two spin states (spin “up” u_1 or spin “down” u_2). In this case $\nu = 5$, $n = 2$, and the symmetry group of the spins is $SU(2)$. If we include the lepton number in our description (an additive quantum number), then the symmetry group is $U(2)$ or $SU(2) \times U(1)$. (We do not distinguish between these last two groups, as we use only the Lie algebra, which is the same for both.) The dimensionality is the same whether the lepton number is included in the description. Electrons must obey Fermi statistics, that is, their state vectors must be antisymmetric under the interchange of *all* the coordinates of any two electrons. However, the symmetry under the interchange of only the spins is given by any standard Young tableau.

Consider a three-electron state belonging to the Young tableau



This state has the same multiplicity as the one-electron state belonging to the tableau



but the lepton number is 3 in the first case and 1 in the second. If we are interested only in the $SU(2)$ of the spins and not in the lepton number, the two diagrams are equivalent in that they correspond to the same quantum numbers.

To avoid the complication of the quantum number associated with the $U(1)$ subgroup of $U(n)$, we often restrict ourselves to $SU(n)$. Then all columns with n boxes may be removed from a given tableau, as there is only one way to antisymmetrize a state of n particles, each of which has n degrees of freedom. The number of states N_n is equal to the result obtained by counting the number of ways one can put positive integers $\leq n$ in the remaining boxes, consistent with the rules that numbers must not decrease going from left to right in any row and must increase going from top to bottom in any column.

We see that in the case of $SU(n)$, a tableau with ν boxes can also denote states of the corresponding symmetry containing a different number of particles. We can therefore divorce the concept of the symmetry of a state corresponding to a Young tableau from the concept of particles belonging to the first fundamental representation.

A tableau with no more than $n - 1$ rows can be specified by a set of $n - 1$ integers p_i , which is the number of boxes in row i minus the number of boxes in row $i + 1$. The multiplicity N_n of any diagram is a function of the p_i . It is a complicated combinatorial problem to find the number of standard arrangements of a tableau with a given \mathbf{p} . The formula for $N_n(p_1, p_2 \dots p_{n-1}) = N_n(\mathbf{p})$ is known for any n but we write it down just for $n = 2$ and $n = 3$. For $SU(2)$ we have

$$N_2(p) = p + 1. \quad (3.49)$$

The number of states of a given angular momentum j is $2j + 1$. Then, using Eq. (3.40), we can make the identification

$$p = 2j. \quad (3.50)$$

For $SU(3)$ the formula is

$$N_3(\mathbf{p}) = (p_1 + 1)(p_2 + 1)(p_1 + p_2 + 2)/2. \quad (3.51)$$

These formulas give the number of states in a multiplet belonging to an irreducible representation of the group with the symmetry specified by the Young tableau \mathbf{p} .

We see that for $SU(2)$ the number N_2 can be any positive integer. However, for $SU(n)$ with $n > 2$, the numbers N_n include only a proper subset of the positive integers. For example, in $SU(3)$, the numbers N_3 have the values 1, 3, 6, 8, 10, 15, etc, as determined either by counting the standard arrangements of Young tableaux or from Eq. (3.51), substituting non-negative integers for p_1 and p_2 .

All the formulas $N_n(\mathbf{p})$ are symmetric under the interchange

$$p_i \leftrightarrow p_{n-i},$$

that is,

$$N_n(p_1, p_2 \dots p_{n-1}) = N_n(p_{n-1} \dots p_1). \quad (3.52)$$

Two Young tableaux which transform into each other under this transformation are called *conjugate tableaux*, and the irreducible representations which act on them are called *conjugate representations*. The first fundamental representation of $SU(3)$ is characterized by $(p_1, p_2) = (1, 0)$; the second fundamental representation is the conjugate representation, and is characterized by $(p_1, p_2) = (0, 1)$. However, it is common to characterize these (and other representations) by a single number which gives their multiplicity: in the case of the first and second representations, we use $\mathbf{3}$ and $\bar{\mathbf{3}}$ respectively. Conjugate representations have the same dimensionality, but conjugate tableaux do not necessarily have the same number of boxes. If a tableau is unchanged under the transformation of interchanging p_i and p_{n-i} , it is self-conjugate, and likewise for the corresponding representation. Since all representations of $SU(2)$ consist of a single row, all representations of $SU(2)$ are self-conjugate.

We now show how to build all irreducible representations of $SU(n)$ starting from the first fundamental one. First we note that a product of two or more representations is in general reducible. We can see this by considering the basis vectors on which they act. It is simplest to begin with $SU(2)$. The product states are written using Young tableaux as

$$\square \times \square$$

These stand for the four product states

$$u_1 u_1, \quad u_1 u_2, \quad u_2 u_1, \quad u_2 u_2.$$

But we know that to obtain the basis vectors of irreducible representations we must take the linear combinations which correspond to the symmetric and antisymmetric Young tableaux. This result is true in any $SU(n)$. We write

$$\square \times \square = \begin{array}{|c|c|} \hline \square & \square \\ \hline \end{array} + \begin{array}{|c|} \hline \square \\ \hline \square \\ \hline \end{array} \quad (3.53)$$

In $SU(2)$, the multiplicities are

$$SU(2) : \quad 2 \times 2 = 3 + 1. \quad (3.54)$$

In $SU(3)$, we have

$$SU(3) : \quad 3 \times 3 = 6 + \bar{3}, \quad (3.55)$$

and in any $SU(n)$ we have

$$SU(n) : \quad n \times n = n(n+1)/2 + n(n-1)/2. \quad (3.56)$$

In the above examples we have found the irreducible representations contained in the product of two irreducible representations. This decomposition is called the *Clebsch–Gordan series*. If the decomposition contains no representation more than once, the product is called *simply reducible*. If $n > 2$, the decomposition of $n \times \bar{n}$ is different from the decomposition of $n \times n$. We have

$$SU(n) : \quad n \times \bar{n} = (n^2 - 1) + 1, \quad (3.57)$$

which is different from the decomposition given in Eq. (3.57). In particular, in $SU(3)$ we have

$$3 \times \bar{3} = 8 + 1. \quad (3.58)$$

We now tell how to find the Clebsch–Gordan series for the product of any two representations of $SU(n)$. We do not give the proof, as it is quite complicated (even the recipe is complicated).

Recipe. We draw the Young tableaux of the two representations, marking each box of the second with the number of the row to which it belongs. We then attach the boxes of the second tableau in all possible ways to the first tableau, subject to the following rules for the combined tableaux:

- 1) Each tableau should be proper.
- 2) No tableau should have a column with more than n boxes, and we can remove all columns with n boxes.
- 3) The numbers must not decrease from left to right in a row.
- 4) The numbers must increase from top to bottom in a column.
- 5) We can make a path by moving along each row from the right, starting at the top. At each point of the path, the number of boxes encountered with the number i must be less or equal to the number of boxes with $i - 1$.

As an example, if we follow the rules, we find the irreducible representations contained in 8×8 of $SU(3)$ to be

$$8 \times 8 = 27 + 10 + \bar{10} + 8 + 8 + 1. \quad (3.59)$$

We see that the Clebsch–Gordan series contains two equivalent representations, namely, the two 8's. This means that the product of 8×8 is not simply reducible.

The product of any two representations of $SU(2)$ is simply reducible, but this result does not hold for any $n > 2$. Even for $SU(2)$, the product of three representations is not necessarily simply reducible. For example, in $SU(2)$, we have

$$2 \times 2 \times 2 = 4 + 2 + 2. \quad (3.60)$$

If two or more equivalent representations appear in the reduction of a product, the use of group theory alone is not sufficient to enable us to label the states. We must know something of the physics of the problem in order to obtain the most useful labeling. The Clebsch–Gordan coefficients are also not determined uniquely without additional input.

For example, the reduction given in Eq. (3.60) can come about from obtaining the irreducible multiplets from three electron spins. One of the two doublets arises by combining the spins of the first two electrons to give a triplet and then combining the third spin to give a doublet. The other doublet arises from combining the first two spins to form a singlet and then combining the third spin to give a doublet. Group theory says nothing about how the first two spins in fact combine, as that depends on the forces involved. As another example, in $SU(3)$ we have

$$3 \times 3 \times 3 = 10 + 8 + 8 + 1. \quad (3.61)$$

The way the two 8's are chosen in either the decomposition (3.59) or (3.61) depends on the physics.

3.8 Evaluating quadratic Casimir operators

The group $SU(n)$ has $n - 1$ Casimir operators, one of which is quadratic in the generators. The quadratic Casimir of $SU(n)$ may be evaluated by making use of the raising and lowering operators. For $SU(2)$ this operator is J^2 and is given by

$$J^2 = \frac{1}{4}(2\sigma_+\sigma_- + 2\sigma_-\sigma_+ + \sigma_z^2). \quad (3.62)$$

We take advantage of the fact that J^2 is an invariant operator to operate on the particular state for which σ_+ vanishes. We eliminate the term $\sigma_+\sigma_-$ by making use of the commutation relation

$$[\sigma_+, \sigma_-] = \sigma_z. \quad (3.63)$$

This enables us to write J^2 (when operating on our particular state) only in terms of the diagonal operator σ_z . Since σ_z is an additive quantum number, it is straightforward to evaluate J^2 for any representation. Details are given in books on quantum mechanics. The answer is $J^2 = j(j + 1)$, or, since $p = 2j$,

$$J^2 = \frac{1}{2}p\left(\frac{1}{2}p + 1\right). \quad (3.64)$$

The same method works for any $SU(n)$ except that the algebra is more complicated. In particular, for $SU(3)$ we get

$$F^2 = p_1 + p_2 + \frac{1}{3}(p_1^2 + p_2^2 + p_1 p_2). \quad (3.65).$$

4. A LITTLE ABOUT GAUGE FIELD THEORY

In this chapter we introduce and discuss field theory, emphasizing gauge field theory. By a gauge field theory, we mean a field theory which is invariant under a group of local gauge transformations, including the case in which the gauge group is nonabelian. We define both local and global gauge transformations later in this chapter. We restrict ourselves to a field theory describable by a Lagrangian, so that the Lagrangian of a gauge theory is invariant under gauge transformations.

We require, as is customary, that our field theory be *local*, i.e., that the basic interaction among fields should occur at a single space-time point, so as not to violate causality. However, we should remain open to the possibility that a theory with a small amount of non-locality may in fact provide a better description of nature. We also require our field theory to be invariant under the proper Poincaré group, so as not to violate invariance under translations, rotations, and Lorentz transformations. We also require the theory to have a positive definite Hamiltonian, so that a stable vacuum state should exist.

A field theory having the properties of locality, proper Poincaré invariance, and a stable vacuum we will call for short a *relativistic* field theory. Such a theory can be proved to be invariant under the combined symmetry operation *CPT* taken in any order, where *C* is charge conjugation, *P* is parity, and *T* is time reversal. This theorem is called the *CPT* theorem. The *CPT* theorem has as a consequence that every kind of particle has an antiparticle, which has the same mass and lifetime as the particle. If the particle has any kind of charge, or any internal additive quantum numbers different from zero, such as baryon number, the antiparticle has anticharge and opposite values of the additive quantum numbers, and so is distinct from the particle. Furthermore, it can also be proved of a relativistic field theory that, given the two possibilities of fermi and bose statistics, the quanta of fields with integral spin must be bosons, and the quanta of fields with half-integral spin must be fermions. This theorem is called the *spin-statistics* theorem. We do not prove either the *CPT* theorem or the spin-statistics theorem here. There is a classic book on this topic by R. F. Streater and A. S. Wightman called *PCT, Spin and Statistics, and All That*. Another source is chapters 7 and 8 of *Introduction to Quantum Field Theory* by P. Roman.

Experimentally, we know of no violation either of the *CPT* theorem or of the spin-statistics theorem. This gives us confidence that local field theory

is a good approximation to the real world, or at least is a good approximation to that part of the world which is presently accessible to experiment.

We require our field theory (i.e., the Lagrangian of the theory) to be invariant under a group of local gauge transformations in order to make the theory analogous to the successful theory of quantum electrodynamics (QED). The gauge group of QED is $U(1)$, which is an abelian group. In order to describe elementary particles, we must generalize the idea of gauge invariance to invariance under nonabelian gauge groups. In particular, in considering the strong and electroweak interactions, we generalize to the nonabelian gauge groups $SU(3)$ and $SU(2)$. Good treatments of gauge field theory are by Huang (1992) and by Aitchison and Hey (1989).

4.1 From coordinates to fields

The Lagrangian of a field is a generalization of a Lagrangian of a collection of particles to a continuously infinite number of degrees of freedom. We first review the Lagrangian

$$L = L(q_1, \dots, q_n, \dot{q}_1, \dots, \dot{q}_n) \quad (4.1)$$

for a collection of particles with generalized coordinates q_i and their time derivatives \dot{q}_i . We define an action A by

$$A = \int_{t_1}^{t_2} L dt. \quad (4.2)$$

The principle of stationary action says $\delta A = 0$ for an arbitrary variation which vanishes at the end points. This variation leads to Lagrange's equations of motion. Let us go through this exercise:

$$\delta A = \int_{t_1}^{t_2} \delta L dt = \sum_i \int_{t_1}^{t_2} \left(\frac{\partial L}{\partial q_i} \delta q_i + \frac{\partial L}{\partial \dot{q}_i} \delta \dot{q}_i \right) dt = 0.$$

We write $\delta \dot{q}_i = \frac{\partial}{\partial t} \delta q_i$ and integrate by parts, obtaining

$$\delta A = \sum_i \int_{t_1}^{t_2} \left(\frac{\partial L}{\partial q_i} - \frac{\partial}{\partial t} \frac{\partial L}{\partial \dot{q}_i} \right) \delta q_i dt = 0,$$

since the δq_i vanish at the end points. Because the δq_i are arbitrary in the interval, we get the Lagrangian equations of motion:

$$\frac{\partial L}{\partial q_i} - \frac{\partial}{\partial t} \frac{\partial L}{\partial \dot{q}_i} = 0. \quad (4.3)$$

The Hamiltonian H is

$$H = \sum_{i=1}^n p_i \dot{q}_i - L, \quad (4.4)$$

where p_i is defined by

$$p_i = \frac{\partial L}{\partial \dot{q}_i}. \quad (4.5)$$

Suppose the Lagrangian is a sum of terms \mathcal{L}_i , each of which depends only on one coordinate and its time derivative. Then we write

$$L = \sum_i \mathcal{L}_i(q_i, \dot{q}_i). \quad (4.6)$$

If all \mathcal{L}_i are the same functions of their arguments, we can drop the subscript on \mathcal{L}_i and write

$$L = \sum_i \mathcal{L}(q_i, \dot{q}_i) \quad (4.7)$$

and

$$H = \sum_i \mathcal{H}(p_i, q_i), \quad (4.8)$$

where

$$\mathcal{H}(p_i, q_i) = p_i \dot{q}_i - \mathcal{L}. \quad (4.9)$$

Let us rewrite Eq. (4.7), showing explicitly that q_i and \dot{q}_i depend explicitly on t :

$$L = \sum_i \mathcal{L}[q_i(t), \dot{q}_i(t)]. \quad (4.10)$$

Suppose now that the number of degrees of freedom increases to a continuously infinite number, so that the index i becomes a parameter x . Then Eq. (4.10) becomes

$$L = \int dx \mathcal{L}[q(x, t), \dot{q}(x, t)], \quad (4.11)$$

where here \mathcal{L} is a Lagrangian *density* (although usually we call it simply the Lagrangian). Because \mathcal{L} now depends on the continuous parameter x , it may contain a partial derivative with respect to x . Therefore, we generalize \mathcal{L} and write

$$\mathcal{L} = \mathcal{L}\left[q(x, t), \frac{\partial q(x, t)}{\partial t}, \frac{\partial q(x, t)}{\partial x}\right]. \quad (4.12)$$

When q is a function of a continuous variable, we call it a *field*, in this case a scalar field, and usually denote it by the symbol ϕ .

We now generalize to the case where x is not a single parameter but is replaced by a three-vector \mathbf{x} , which stands for three parameters. We let

$$x^\mu = (t, x, y, z) = (x^0, x^1, x^2, x^3) = (t, \mathbf{x}), \quad (4.13)$$

Also we let $g_{\mu\nu} = g^{\mu\nu}$ be the metric tensor with diagonal elements $(1, -1, -1, -1)$ and zeros elsewhere. We define $x_\mu = g_{\mu\nu}x^\nu$, where we are using the convention that a repeated upper and lower index (in either order) is summed over. It follows from this definition that $x_\mu = (t, -x, -y, -z)$. The quantity x^μ is called a contravariant four-vector, and x_μ is called a covariant four-vector. If A^μ and B^μ are two four-vectors, then the product $A^\mu B_\mu = A_\mu B^\mu$ is a Lorentz scalar. We sometimes write a four-vector A^μ as A for short, and denote scalar products as A^2 or $A \cdot B$.

The momentum four-vector is

$$p^\mu = (E, p_x, p_y, p_z) = (p^0, p^1, p^2, p^3). \quad (4.14)$$

In quantum mechanics we make the replacements

$$E \rightarrow i\partial/\partial t, \quad \mathbf{p} \rightarrow -i\nabla. \quad (4.15)$$

Then

$$p_\mu = (i\frac{\partial}{\partial t}, i\nabla) = (i\frac{\partial}{\partial x^0}, i\frac{\partial}{\partial x^1}, i\frac{\partial}{\partial x^2}, i\frac{\partial}{\partial x^3}),$$

or

$$p_\mu = i\frac{\partial}{\partial x^\mu} = i\partial_\mu. \quad (4.16)$$

We now return to the Lagrangian density given in Eq. (4.12) and consider the case in which x gets replaced by the three-vector \mathbf{x} . We also use the notation ϕ instead of q . Then the Lagrangian becomes

$$\mathcal{L} = \mathcal{L}(\phi, \frac{\partial\phi}{\partial t}, \frac{\partial\phi}{\partial x}, \frac{\partial\phi}{\partial y}, \frac{\partial\phi}{\partial z}) = \mathcal{L}(\phi, \partial_\mu\phi). \quad (4.17)$$

In Eq. (4.17), $\phi = \phi(x^\mu)$, but we have omitted the argument x^μ .

The action A is given by

$$A = \int dtL = \int d^4x\mathcal{L}. \quad (4.18)$$

As before, the equation of motion is obtained by letting the action be stationary. We get

$$\frac{\partial \mathcal{L}}{\partial \phi} - \partial_\mu \frac{\partial \mathcal{L}}{\partial(\partial_\mu \phi)} = 0. \quad (4.19)$$

This is shorthand for

$$\frac{\partial \mathcal{L}}{\partial \phi} = \frac{\partial}{\partial t} \frac{\partial \mathcal{L}}{\partial \dot{\phi}} + \frac{\partial}{\partial x} \frac{\partial \mathcal{L}}{\partial(\partial \phi / \partial x)} + \frac{\partial}{\partial y} \frac{\partial \mathcal{L}}{\partial(\partial \phi / \partial y)} + \frac{\partial}{\partial z} \frac{\partial \mathcal{L}}{\partial(\partial \phi / \partial z)}.$$

We define a conjugate field π as

$$\pi = \frac{\partial \mathcal{L}}{\partial \dot{\phi}}, \quad (\dot{\phi} = \partial_0 \phi). \quad (4.20)$$

Then the Hamiltonian (density) is

$$\mathcal{H} = \pi \dot{\phi} - \mathcal{L}. \quad (4.21)$$

A Lagrangian density must be a Lorentz scalar. But the field itself may have more than one component and transform differently from a scalar. The fields of relevance to us here are scalar fields (e.g. the Higgs field), whose quanta are particles of spin 0; spinor fields (e.g. the electron field), whose quanta are particles of spin 1/2; and vector fields (e.g. the electromagnetic field or the gluon fields), whose quanta are particles of spin 1. We shall not discuss the gravitational field, which has spin 2.

4.2 Scalar fields

We consider a complex scalar field ϕ with mass m and a Lagrangian (density) given by

$$\mathcal{L} = \partial_\mu \phi^* \partial^\mu \phi - m^2 \phi^* \phi - \lambda (\phi^* \phi)^2, \quad (4.22)$$

where the constant λ is the self-interaction coupling strength. In obtaining the equations of motion we treat ϕ and ϕ^* as independent fields. Using Eq. (4.19) we get

$$\partial_\mu \partial^\mu \phi + m^2 \phi + 2\lambda \phi^* \phi^2 = 0, \quad (4.23)$$

$$\partial_\mu \partial^\mu \phi^* + m^2 \phi^* + 2\lambda \phi^{*2} \phi = 0. \quad (4.24)$$

If $\lambda = 0$, ϕ and ϕ^* satisfy a Klein–Gordon equation for a free field.

In order to obtain the Hamiltonian (density), we rewrite \mathcal{L} as follows:

$$\mathcal{L} = \dot{\phi}^* \dot{\phi} - \nabla \phi^* \cdot \nabla \phi - m^2 \phi^* \phi - \lambda(\phi^* \phi)^2.$$

The conjugate (momentum) fields are

$$\pi = \frac{\partial \mathcal{L}}{\partial \dot{\phi}} = \dot{\phi}^*, \quad \pi^* = \frac{\partial \mathcal{L}}{\partial \dot{\phi}^*} = \dot{\phi}. \quad (4.25)$$

Then \mathcal{H} is

$$\mathcal{H} = |\pi|^2 + |\nabla \phi|^2 + m^2 |\phi|^2 + \lambda |\phi|^4. \quad (4.26)$$

It is instructive to write \mathcal{L} in terms of real fields η and ζ , defined by

$$\phi = (\eta + i\zeta)/\sqrt{2}. \quad (4.27)$$

After substitution into Eq. (4.22) and some algebra we get

$$\mathcal{L} = \frac{1}{2} \partial_\mu \eta \partial^\mu \eta + \frac{1}{2} \partial_\mu \zeta \partial^\mu \zeta - \frac{1}{2} m^2 (\eta^2 + \zeta^2) - \frac{1}{4} \lambda (\eta^2 + \zeta^2)^2. \quad (4.28)$$

This Lagrangian describes the sum of two real free fields, each of mass m , plus a term proportional to λ , which describes the interactions of the fields with themselves and with each other.

To obtain the Hamiltonian, we first substitute Eq. (4.27) into Eq. (4.25), obtaining

$$\pi = (\dot{\eta} - i\dot{\zeta})/\sqrt{2}, \quad \pi^* = (\dot{\eta} + i\dot{\zeta})/\sqrt{2},$$

or

$$\dot{\eta} = (\pi + \pi^*)/\sqrt{2}, \quad \dot{\zeta} = (\pi^* - \pi)/\sqrt{2}. \quad (4.29)$$

If we substitute Eqs. (4.29) and (4.27) into (4.26) and introduce the notation

$$\pi_\eta = \dot{\eta}, \quad \pi_\zeta = \dot{\zeta}, \quad (4.30)$$

we get

$$\mathcal{H} = \frac{1}{2} \dot{\eta}^2 + \frac{1}{2} \dot{\zeta}^2 + \frac{1}{2} (\nabla \eta)^2 + \frac{1}{2} (\nabla \zeta)^2 + \frac{1}{2} m^2 (\eta^2 + \zeta^2) + \frac{1}{4} (\eta^2 + \zeta^2)^2. \quad (4.31)$$

We obtain the equations of motion by substituting Eq. (4.27) into either (4.23) or (4.24). We get

$$\partial_\mu \partial^\mu \eta + m^2 \eta + \lambda (\eta^2 + \zeta^2) \eta = 0, \quad (4.32)$$

$$\partial_\mu \partial^\mu \zeta + m^2 \zeta + \lambda(\eta^2 + \zeta^2)\zeta = 0. \quad (4.33)$$

4.3 Spinor fields

We consider the Lagrangian for a free Dirac spinor field ψ with mass m . The Lagrangian (density) is given by

$$\mathcal{L} = \bar{\psi}(i\gamma^\mu \partial_\mu - m)\psi, \quad (4.34)$$

where the γ^μ are 4×4 Dirac matrices which we define shortly. We can get the equation of motion by varying the action with respect to $\bar{\psi}$, treating it as independent of ψ . We obtain the free Dirac equation in covariant form:

$$(i\gamma^\mu \partial_\mu - m)\psi = 0. \quad (4.35)$$

This equation may be more familiar in a form which is not obviously Lorentz invariant, namely

$$(\alpha \cdot \mathbf{p} + \beta m)\psi = E\psi, \quad (4.36)$$

where $\mathbf{p} = -i\nabla$, $E = i\partial/\partial t$, and

$$\alpha = \begin{pmatrix} 0 & \sigma \\ \sigma & 0 \end{pmatrix}, \quad \beta = \begin{pmatrix} I & 0 \\ 0 & -I \end{pmatrix}, \quad (4.37)$$

with the elements of α and β being 2×2 matrices. Here I is the 2×2 identity matrix and $\sigma = (\sigma_1, \sigma_2, \sigma_3)$, where the σ_i are the usual Pauli matrices. We multiply Eq. (4.36) on the left by β , obtaining

$$(\beta\alpha \cdot \mathbf{p} + m)\psi = \beta E\psi, \quad (4.38)$$

since $\beta^2 = 1$. If we compare (4.38) with (4.35), remembering that $i\partial_\mu = p_\mu = (E, -\mathbf{p})$, we see that

$$\gamma^i = \beta\alpha_i \quad (i = 1, 2, 3), \quad \gamma^0 = \beta. \quad (4.39)$$

The field $\bar{\psi}$ is related to ψ by

$$\bar{\psi} = \psi^\dagger \gamma^0, \quad (4.40)$$

where \dagger denotes Hermitian conjugate.

We can verify explicitly that the γ^μ satisfy the following anticommutation relations

$$\{\gamma^\mu, \gamma^\nu\} = 2g^{\mu\nu}, \quad (4.41)$$

where $g^{\mu\nu} = g_{\mu\nu}$ is a diagonal matrix with diagonal elements $(1, -1, -1, -1)$ and the anticommutator $\{A, B\} = AB + BA$. We define

$$\gamma^5 = \gamma_5 = i\gamma^0\gamma^1\gamma^2\gamma^3. \quad (4.42)$$

Then

$$\{\gamma^5, \gamma^\mu\} = 0. \quad (4.43)$$

We have the following explicit representation for the γ matrices:

$$\gamma^0 = \begin{pmatrix} I & 0 \\ 0 & -I \end{pmatrix}, \quad \gamma^i = \begin{pmatrix} 0 & \sigma_i \\ -\sigma_i & 0 \end{pmatrix}, \quad \gamma^5 = \begin{pmatrix} 0 & I \\ I & 0 \end{pmatrix}. \quad (4.44)$$

Other representations of the γ matrices exist in the literature. The above considerations also apply to the massless Dirac spinor field ($m = 0$). If a spinor is massless, we can define fields which are eigenstates of γ_5 , which is called the chirality operator. The eigenstates, called states of definite chirality, have eigenvalues ± 1 . If the eigenvalue is 1, the state is said to be right handed; if the eigenvalue is -1 , the state is called left handed. A state having a definite mass greater than zero is not an eigenstate of chirality, but is a linear combination of left-handed and right-handed states with equal weight.

We have confined our attention to Dirac spinors. Even if a Dirac spinor has no electric charge, it is distinct from its antiparticle. There is also the possibility of an uncharged Majorana spinor, which is the same as its antiparticle.

4.4 Vector fields

We first consider the free electromagnetic field, which we denote by A^μ . Photons, which are the quanta of this field, have spin 1 and negative parity. We define $F_{\mu\nu}$ by

$$F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu. \quad (4.45)$$

The A^μ and $F^{\mu\nu}$ are given in terms of the ‘‘scalar’’ (actually, the time component of a 4-vector) potential ϕ , the three-vector potential \mathbf{A} , and the electric and magnetic fields \mathbf{E} and \mathbf{B} by

$$A^\mu = (\phi, \mathbf{A}), \quad F^{\mu\nu} = \begin{pmatrix} 0 & -E_x & -E_y & -E_z \\ E_x & 0 & -B_z & B_y \\ E_y & B_z & 0 & -B_x \\ E_z & -B_y & B_x & 0 \end{pmatrix}. \quad (4.46)$$

According to the definition (4.45), we have

$$\mathbf{E} = -\nabla\phi - \dot{\mathbf{A}}, \quad \mathbf{B} = \nabla \times \mathbf{A}. \quad (4.47)$$

Also, the definition (4.45) implies that

$$\nabla \cdot \mathbf{B} = 0, \quad \nabla \times \mathbf{E} = -\dot{\mathbf{B}}. \quad (4.48)$$

These are two of Maxwell's equations in empty space.

We get the other two of Maxwell's equations from the Lagrangian equations of motion. The Lagrangian is

$$\mathcal{L} = -\frac{1}{4}F_{\mu\nu}F^{\mu\nu} \quad (4.49)$$

or

$$\mathcal{L} = -\frac{1}{2}(\partial_\mu A_\nu \partial^\mu A^\nu - \partial_\mu A_\nu \partial^\nu A^\mu), \quad (4.50)$$

where we have used (4.45) and interchanged dummy indices in two terms. This Lagrangian yields the equations of motion

$$\partial_\mu F^{\mu\nu} = 0, \quad (4.51)$$

which correspond to

$$\nabla \times \mathbf{B} = \dot{\mathbf{E}}, \quad \nabla \cdot \mathbf{E} = 0. \quad (4.52)$$

These are the third and fourth of Maxwell's equations in empty space. In the presence of a charge density ρ and a current density \mathbf{j} , these equations become

$$\nabla \times \mathbf{B} = \dot{\mathbf{E}} + \mathbf{j}, \quad \nabla \cdot \mathbf{E} = \rho. \quad (4.53)$$

If we define a four-vector current (density)

$$j^\mu = (\rho, \mathbf{j}), \quad (4.54)$$

then in regions where there exists a current we have

$$\partial_\mu F^{\mu\nu} = j^\nu. \quad (4.55)$$

The current is conserved if

$$\partial_\nu j^\nu = 0. \quad (4.56)$$

We can prove that the current is conserved by taking the partial derivative of Eq. (4.55) and interchanging dummy indices in one of the two terms, obtaining Eq. (4.56). Let us write Eq. (4.56) in three-vector notation, using Eq. (4.54). We get

$$\frac{\partial \rho}{\partial t} + \nabla \cdot \mathbf{j} = 0. \quad (4.57)$$

Let us integrate Eq. (4.57) over a volume of space so large that it contains all the charge and current and use Gauss' theorem, which says

$$\int_{\text{vol}} \nabla \cdot \mathbf{j} d^3 \mathbf{x} = \int_{\text{surf}} \mathbf{j} \cdot d^2 \mathbf{x}.$$

Then, because \mathbf{j} vanishes on the surface enclosing the volume, we get

$$\dot{e} = 0, \quad (4.58)$$

where the charge e is given by

$$e = \int \rho d^3 \mathbf{x}. \quad (4.59)$$

Our result (4.58) says that if the four-vector current density is conserved, so is the charge.

Next consider a complex vector field with mass. The Lagrangian density is

$$\mathcal{L} = -\frac{1}{2} F_{\mu\nu}^* F^{\mu\nu} + m^2 A_\mu^* A^\mu. \quad (4.60)$$

This can be rewritten as

$$\mathcal{L} = -\partial_\mu A_\nu^* F^{\mu\nu} + m^2 A_\mu^* A^\mu. \quad (4.61)$$

We obtain the Lagrangian equations of motion by treating A_μ^* and A_μ as independent. Because

$$\frac{\partial \mathcal{L}}{\partial A_\mu^*} = m^2 A^\mu, \quad \partial_\mu \frac{\partial \mathcal{L}}{\partial (\partial_\mu A_\nu^*)} = \partial_\mu F^{\mu\nu},$$

we get the equations of motion:

$$\partial_\mu F^{\mu\nu} + m^2 A^\nu = 0. \quad (4.62)$$

If we set $m = 0$ and assume A_μ is real, we recover Maxwell's equations. This is simpler than starting with a Lagrangian containing real fields.

Let us write the complex field A^μ in terms of real fields as

$$A^\mu = (A_1^\mu + iA_2^\mu)/\sqrt{2}. \quad (4.63)$$

Then the mass term in Eq. (4.51) becomes

$$m^2 A_\mu^* A^\mu = \frac{1}{2} m^2 (A_{1\mu} A_1^\mu + A_{2\mu} A_2^\mu). \quad (4.64)$$

We see from this equation that for a real vector field, m^2 is twice the coefficient of the term in the Lagrangian which is quadratic in the field, whereas, as we see from (4.61), for a complex vector field, m^2 is equal to the coefficient of the quadratic term in the Lagrangian.

4.5 Global and local gauge symmetries

Let us consider the Lagrangian (density) for a free Dirac field, say the field of an electron. As we have seen in Section 4.3, the Lagrangian is given by

$$\mathcal{L} = \bar{\psi}(i\gamma^\mu \partial_\mu - m)\psi, \quad (4.65)$$

where ψ is a function of the space-time four-vector x .

Now let us make a unitary transformation U_θ on ψ and $\bar{\psi}$:

$$\psi' = U_\theta \psi, \quad \bar{\psi}' = U_\theta^* \bar{\psi}, \quad U_\theta = e^{-i\theta}, \quad (4.66)$$

where θ is any constant independent of x . We find by direct substitution that \mathcal{L} is invariant under the transformation, namely

$$\mathcal{L}' = \bar{\psi}'(i\gamma^\mu \partial_\mu - m)\psi' = \mathcal{L}. \quad (4.67)$$

The U_θ belong to the unitary group in one dimension $U(1)$, which is therefore a symmetry group of the free Dirac equation.

The Lagrangian for a scalar field with self interaction

$$\mathcal{L} = \partial_\mu \phi^* \partial^\mu \phi - m_\phi^2 \phi^* \phi - \lambda(\phi^* \phi)^2 \quad (4.68)$$

is also invariant under this transformation, as can be easily checked.

The transformation $U_\theta = e^{-i\theta}$ was originally introduced without the i , and corresponded to a change of scale or gauge. For this reason it was called

a *gauge* transformation, and the name stuck after the introduction of the i . More properly, it should be called a *phase* transformation. Because U_θ is independent of the space–time point x (now a four-vector), the transformation is called a *global* gauge transformation.

The symmetry of a Lagrangian under a global gauge transformation is, according to Noether’s theorem, connected to a conservation law. In this case, there is an additive number conservation law. Examples from the standard model are lepton family number and baryon number. These conservation laws are strictly true in the classical version of the standard model, but are violated by anomalies in the quantized version (Section 4.8).

Now consider a gauge transformation

$$U_\theta = e^{-i\theta(x)}. \quad (4.69)$$

Because this transformation varies from point to point in space–time, it is called a *local* gauge transformation. It is reasonable to ask that the Lagrangian of a field be invariant under such a transformation because the phase is not observable in quantum mechanics. Therefore, the phase of the field at one space–time point should have nothing to do with the phase at another point. We do not, however, let $\theta(x)$ be quite so arbitrary because we require that its derivative exist. The group is still $U(1)$ independently of whether θ depends on x .

The Lagrangian for the free Dirac field and the Lagrangian for the self-interacting scalar field are not invariant under local gauge transformations because these Lagrangians contain derivatives with respect to the fields. In each case the transformed Lagrangian picks up an extra term arising from the derivative of $\theta(x)$. We need to find a way to compensate for this term by replacing the derivative ∂_μ by a so-called *covariant* derivative D_μ . The covariant derivative involves a new field, a vector field $A_\mu(x)$, called a gauge field. We define D_μ by

$$D_\mu = \partial_\mu + ieA_\mu(x), \quad (4.70)$$

where e is a constant.

In what follows, for brevity we omit writing down explicitly the x dependence of θ and A_μ . If we replace ∂_μ by D_μ in the Lagrangian for the free Dirac field we obtain the new Lagrangian

$$\mathcal{L} = \bar{\psi}(i\gamma^\mu D_\mu - m)\psi. \quad (4.71)$$

Then this new Lagrangian will be invariant under local gauge transformations if we let A_μ transform in such a way as to cancel the extra term arising from

the derivative with respect to θ . It can be readily verified that A_μ must transform like

$$A'_\mu = A_\mu + \frac{1}{e}\partial_\mu\theta. \quad (4.72)$$

This transformation of A_μ should work for any Lagrangian which does not contain any derivatives higher than the first. In particular, the Lagrangian for a scalar field will also become gauge invariant if ∂_μ is replaced by D_μ when operating on ϕ and by D_μ^* when operating on ϕ^* .

The field A_μ is a legitimate vector field, and so its kinetic part $-\frac{1}{4}F_{\mu\nu}F^{\mu\nu}$ should be added to the Lagrangian, where $F_{\mu\nu}$ is given by Eq. (4.45). However, we cannot include a mass term of the form $\frac{1}{2}m_A^2 A_\mu A^\mu$ in the Lagrangian because such a term is not gauge invariant. Then the gauge-invariant Lagrangian for a Dirac field and massless vector field is

$$\mathcal{L}_\psi = \bar{\psi}(i\gamma^\mu D_\mu - m)\psi - \frac{1}{4}F_{\mu\nu}F^{\mu\nu}. \quad (4.73)$$

Similarly, the gauge-invariant Lagrangian for a scalar field and a massless vector field is

$$\mathcal{L}_\phi = D_\mu^*\phi^*D^\mu\phi - m_\phi^2\phi^*\phi - \lambda(\phi^*\phi) - \frac{1}{4}F_{\mu\nu}F^{\mu\nu}. \quad (4.74)$$

It is instructive to rewrite \mathcal{L}_ψ as

$$\mathcal{L}_\psi = \bar{\psi}(i\gamma^\mu\partial_\mu - m)\psi - \frac{1}{4}F_{\mu\nu}F^{\mu\nu} - e\bar{\psi}\gamma^\mu A_\mu\psi. \quad (4.75)$$

The last term is an interaction term which can be written

$$\mathcal{L}_{int} = -ej^\mu A_\mu, \quad j^\mu = \bar{\psi}\gamma^\mu\psi, \quad (4.76)$$

where j^μ is called a *current*. This current is a *neutral vector* current, and it is coupled to a neutral vector field A_μ . The coupling strength e of the interaction is called a *charge*. If e is the electric charge, A_μ is the electromagnetic field.

The current given in (4.76) can be shown to transform like a four-vector under Lorentz transformations. With the aid of the γ matrices and spinor fields, currents with other transformation properties can be constructed: $\bar{\psi}\psi$ transforms as a scalar, $\bar{\psi}\gamma_5\psi$ like a pseudoscalar, $\bar{\psi}\gamma^\mu\gamma_5\psi$ like an axial vector, and $\bar{\psi}\sigma^{\mu\nu}\psi$ like a second-rank tensor, where

$$\sigma^{\mu\nu} = \frac{1}{2}i[\gamma^\mu, \gamma^\nu]. \quad (4.77)$$

It is seen that gauge invariance is a powerful principle, as use of it leads from a free field theory to a field theory with an interaction whose form is determined just from the requirement of gauge invariance. In the case in which there is a single four-vector field introduced, the gauge group of the Lagrangian is $U(1)$. This is the invariance group of electromagnetism. As far as we know, it is an exact, manifest symmetry. The gauge principle we have discussed in this section leads to a vector current coupled to a vector gauge field.

In accordance with Noether's theorem, local gauge invariance is connected to a conservation law: namely, the conservation of the four-vector current. The conservation of the current has as a consequence the conservation of charge. Of course, conservation of charge number holds, as this conservation law follows from invariance under either a global or local gauge transformation. But if the transformation is local, we have in addition the conservation of the strength of the charge, as measured by the strength of the coupling of the current to the gauge field. In the previous section, we demonstrated the conservation of the four-vector current and the charge for Maxwell's equations, which are gauge invariant.

With a slight modification, the local gauge principle can also accommodate an axial vector current. The Weinberg–Salam model contains parity-violating currents with both vector and axial vector parts. An anomaly prevents the axial vector current from being conserved in the quantized version of the theory. There is an additional complication in that the gauge symmetry of the model is spontaneously broken. We discuss spontaneous symmetry breaking in Section 4.7.

4.6 Nonabelian gauge fields

We can consider a group of gauge transformations other than $U(1)$. For example, suppose we have n complex scalar fields ϕ_i , each of which satisfies a Lagrangian including self-interactions. We can collect the ϕ_i in a column vector ϕ and consider the transformation

$$\phi' = U_\theta \phi, \quad U_\theta = e^{-i\mathbf{T}\cdot\theta}, \quad (4.78)$$

where \mathbf{T} is an r -component vector, each component of which is an $n \times n$ matrix. Here $\mathbf{T}\cdot\theta$ is short for $T_a\theta^a$ and the θ^a are constants. (We are using the summation convention.) In the case of the free Dirac Lagrangian, the four-component spinor ψ field becomes a $4n$ -component spinor, and the transformation operator is U_θ just as in the scalar case.

The transformations U_θ form a group. An interesting case is one in which the group is a nonabelian Lie group G . This will be true if the matrices T_a do not commute but satisfy a Lie algebra

$$[T_a, T_b] = if_{ab}^c T_c. \quad (4.79)$$

The U_θ form an n -dimensional unitary representation of G , and the T_a are generators of the group.

If the constants θ_a are independent of x , the Lagrangians of the ϕ and ψ fields are invariant under the group of transformations. However, if the θ_a are functions of x , then the Lagrangians are not invariant. However, just as in the case where the group was $U(1)$, we can find a covariant derivative so that the Lagrangians are invariant under the transformations. We shall use a somewhat different procedure and restrict ourselves to infinitesimal gauge transformations

$$U_\theta = e^{-i\mathbf{T}\cdot\theta} \simeq 1 - i\mathbf{T}\cdot\theta. \quad (4.80)$$

Instead of a single gauge field A^μ we now need r vector gauge fields A_a^μ or \mathbf{A}^μ , one for each generator (or one for each parameter) of the Lie group. As before, we introduce a covariant derivative, in the present case given by

$$D_\mu = \partial_\mu + ig\mathbf{T}\cdot\mathbf{A}_\mu, \quad (4.81)$$

where g is a constant analogous to the charge e . The Dirac Lagrangian formally looks the same as in Eq. (4.71), except that in the present case ψ has $4n$ components and D_μ is given by (4.81).

In order to find out how \mathbf{A}_μ must transform under a gauge transformation, we substitute (4.81) into (4.71), make the gauge transformation (4.80), and keep terms only up to first order in θ . After some algebra we find that the Lagrangian is gauge invariant if

$$(A'_\mu)^c = A_\mu^c + \frac{1}{g}\partial_\mu\theta^c + \theta^a A_\mu^b f_{ab}^c. \quad (4.82)$$

Note that if the structure constants all vanish, then we get the usual transformation properties of A'_μ . For example, in the case in which the symmetry is just the direct product of different $U(1)$ groups, the theory is essentially as easy to treat as for a single $U(1)$.

Let us return to the nonabelian case. We also wish to add the kinetic energy of the gauge fields to the Lagrangian. It turns out, however, that the term $\mathbf{F}_{\mu\nu}\cdot\mathbf{F}^{\mu\nu}$ is not invariant under a gauge transformation unless

$$F_{\mu\nu}^a = \partial_\mu A_\nu^a - \partial_\nu A_\mu^a - gf_{bc}^a A_\mu^b A_\nu^c. \quad (4.83)$$

(Again, if the group is abelian, the structure constants vanish, and the expression is the usual one.)

The Dirac Lagrangian in the nonabelian case is

$$\mathcal{L} = \bar{\psi}(i\gamma^\mu D_\mu - m)\psi - \frac{1}{4}\mathbf{F}_{\mu\nu} \cdot \mathbf{F}^{\mu\nu}. \quad (4.84)$$

Again the gauge fields must be massless, because the mass terms are not gauge invariant.

In view of the fact that $\mathbf{F}_{\mu\nu}$ is given by Eq.(4.83), the term in the Lagrangian proportional to $\mathbf{F}_{\mu\nu} \cdot \mathbf{F}^{\mu\nu}$ contains self-interactions of the gauge fields as well as their kinetic energy.

If we are to make a field theory invariant under local $SU(n)$ gauge transformations, we require $n^2 - 1$ gauge fields, one each for the parameters of the group. These fields transform under the $n^2 - 1$ dimensional representation of $SU(n)$. This representation is called the adjoint representation. Associated with the symmetry are $n^2 - 1$ conserved currents.

Nonabelian local gauge transformations were first treated by Yang and Mills (1954) for the group $SU(2)$. The gauge fields A_μ^a are sometimes called Yang–Mills fields.

4.7 Spontaneous symmetry breaking

Among the symmetries we recognize in nature, some are only approximate. As we have already remarked, we commonly classify approximate symmetry in three categories:

- 1) The Lagrangian is a sum of two terms, one of which respects the symmetry (i.e., is invariant under the transformation), and the second of which does not, but is small.
- 2) The Lagrangian respects the symmetry, but the ground state (lowest energy or vacuum state) of the system does not.
- 3) Both the Lagrangian and the classical ground state respect the symmetry, but the quantized version of the theory contains an anomaly which breaks the symmetry.

An example of a case in which the symmetry is violated in the Lagrangian is parity violation. Both the strong and electromagnetic interactions conserve parity, but the weak interaction does not. As a consequence, the effect of parity violation at low energy is small in most reactions except in special cases in which neither the strong nor the electromagnetic interaction contributes.

The second case, in which the Lagrangian respects the symmetry but the vacuum does not, is called *spontaneous symmetry breaking*. The symmetry is called a *hidden* symmetry, in contrast to a *manifest* symmetry. We here treat the case of hidden symmetry, and in the next section we briefly discuss what we mean by a symmetry which is broken by an anomaly.

If the vacuum state does not respect the symmetry of the Lagrangian, then we shall see that the vacuum must be degenerate, i.e., there must exist more than one solution to the Lagrangian equations of motion with the same energy. Also, there can exist higher-energy states of the system which do not respect the symmetry. However, at sufficiently high temperature we expect the symmetry to be “restored”, i.e., to be manifest.

Let us consider a complex scalar field with a Lagrangian given by

$$\mathcal{L} = \partial_\mu \phi^* \partial^\mu \phi - m^2 \phi^* \phi - \lambda(\phi^* \phi)^2. \quad (4.85)$$

This Lagrangian makes good sense in the limit $\lambda = 0$, as then the Lagrangian describes a free field with mass m . If $\lambda \ll 1$, then the term $\lambda(\phi^* \phi)^2$ can be considered as a perturbation and so ϕ can continue to be interpreted as field with mass m , only now weakly interacting.

Suppose now that we replace m^2 by $-\mu^2$, so that the Lagrangian is

$$\mathcal{L} = \partial_\nu \phi^* \partial^\nu \phi + \mu^2 \phi^* \phi - \lambda(\phi^* \phi)^2. \quad (4.86)$$

The Hamiltonian corresponding to this Lagrangian is

$$\mathcal{H} = |\dot{\phi}|^2 + |\nabla \phi|^2 - \mu^2 |\phi|^2 + \lambda |\phi|^4. \quad (4.87).$$

If $\lambda = 0$, \mathcal{H} is not positive definite, and, what is worse, becomes arbitrarily large and negative as $|\phi|$ increases. Therefore, \mathcal{L} is not an acceptable Lagrangian for $\lambda = 0$, and so one cannot do perturbation theory in λ .

We now look for an acceptable interpretation for the Lagrangian given in Eq. (4.86). Let us write the Hamiltonian of Eq. (4.87) as

$$\mathcal{H} = H_0 + V, \quad (4.88)$$

where

$$H_0 = |\dot{\phi}|^2 + |\nabla \phi|^2, \quad V = -\mu^2 |\phi|^2 + \lambda |\phi|^4. \quad (4.89)$$

Let $|\phi| = \chi$. Then

$$V = -\mu^2 \chi^2 + \lambda \chi^4. \quad (4.90)$$

It can be seen that V has a minimum at

$$\chi = \chi_0, \quad \chi_0 = \mu/\sqrt{2\lambda}. \quad (4.91)$$

Therefore, the vacuum or lowest-energy state ϕ_0 of the classical field occurs when the field has the value $|\phi_0| = \chi_0$.

Now the Lagrangian (4.86) is invariant under global gauge transformations of the form $U_\theta = e^{-i\theta}$. However, the vacuum state ϕ_0 is not invariant under the transformation, because the transformed state is $\phi'_0 = e^{-i\theta}\phi_0$. The energy of ϕ'_0 is the same as the energy of ϕ_0 ; hence, the degeneracy. Because θ can take on an infinite number of values, the vacuum is infinitely degenerate.

In order to give a physical interpretation of the Lagrangian (4.74) we make the change of variables

$$\phi(x) = [v + \eta(x) + i\zeta(x)]/\sqrt{2}, \quad (4.92)$$

where v is a constant given by

$$v = \sqrt{2}\chi_0 = \mu/\sqrt{\lambda}. \quad (4.93)$$

Then, substituting (4.92) into (4.86) and using (4.93) to eliminate μ , we obtain

$$\mathcal{L} = \frac{1}{2}\partial_\nu\eta\partial^\nu\eta - \frac{1}{2}(2\lambda v^2)\eta^2 + \frac{1}{2}\partial_\nu\zeta\partial^\nu\zeta - \frac{1}{4}\lambda(\eta^2 + \zeta^2)^2 - \lambda v\eta(\eta^2 + \zeta^2), \quad (4.94)$$

where we have dropped the constant $\lambda v^4/4$.

We see that the \mathcal{L} of Eq. (4.94) is the Lagrangian for two real scalar fields η and ζ including interactions. Furthermore, the field η has mass $m_\eta^2 = 2\lambda v^2$, while the field ζ is massless. Thus, although the complex field ϕ has no obvious physical interpretation (it appears to have imaginary mass), the real fields η and ζ can be interpreted physically: the first has mass and the second is massless.

Because $\mathcal{L}(\phi)$ is invariant under a global gauge transformation, so is $\mathcal{L}(\eta, \zeta)$. But the vacuum state ϕ_0 is not invariant, and so the symmetry of \mathcal{L} is spontaneously broken. In our case the spontaneous symmetry breaking gives rise to a massless field ζ , whose quanta are massless scalar particles. It is a rather general phenomenon that the spontaneous breaking of a *global* symmetry gives rise to a massless scalar particle. This particle is called a *Goldstone boson*.

We next consider spontaneous symmetry breaking for a Lagrangian which is invariant under a *local* gauge transformation. Again we consider the case of a complex scalar field ϕ with an apparently imaginary mass. The Lagrangian is (do not confuse the constant μ with the index μ)

$$\mathcal{L} = -\frac{1}{4}F_{\mu\nu}F^{\mu\nu} + D_\nu^*\phi^*D^\nu\phi + \mu^2\phi^*\phi - \lambda(\phi^*\phi)^2. \quad (4.95)$$

This Lagrangian is invariant under the local gauge transformation

$$\phi' = e^{-i\theta}\phi, \quad A'_\nu = A_\nu + \frac{1}{e}\partial_\nu\theta, \quad (4.96)$$

where θ is a function of x . To get a physical interpretation, we make the change of variables

$$\phi = (v + \eta)e^{i\zeta/v}/\sqrt{2}. \quad (4.97)$$

If η and ζ are small, (4.97) reduces to (4.92) in lowest order. At the same time we make the following change of variables for A_ν :

$$A_\nu = B_\nu - \frac{1}{ev}\partial_\nu\zeta. \quad (4.98)$$

We next make the gauge transformation

$$\phi' = e^{-i\zeta/v}\phi, \quad A'_\nu = A_\nu + \frac{1}{ev}\partial_\nu\zeta. \quad (4.99)$$

Then we get

$$\phi' = (v + \eta)/\sqrt{2}, \quad A'_\nu = B_\nu. \quad (4.100)$$

Substituting (4.99) into (4.95), we get

$$\begin{aligned} \mathcal{L} = & -\frac{1}{4}F_{\mu\nu}F^{\mu\nu} + \frac{1}{2}(e^2v^2)B_\nu B^\nu + \frac{1}{2}\partial_\nu\eta\partial^\nu\eta - \frac{1}{2}(2\lambda v^2)\eta^2 \\ & + \frac{1}{2}e^2B_\nu B^\nu(2v\eta + \eta^2) - \lambda v\eta^3 - \frac{1}{4}\lambda\eta^4, \end{aligned} \quad (4.101)$$

where $\mu^2 = v^2\lambda$ and again we have dropped a constant term. In (4.101),

$$F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu = \partial_\mu B_\nu - \partial_\nu B_\mu. \quad (102)$$

The first two terms are the Lagrangian for a vector field B_ν with mass $m_B^2 = e^2v^2$. The next two terms are the Lagrangian for a scalar field η with mass

$m_\eta^2 = 2\lambda v^2$. The remaining terms are interaction terms. The ζ field has disappeared.

We started with a Lagrangian with four independent degrees of freedom, two for the massless vector field A_ν and two for the complex scalar field ϕ . We end up with a massive vector field B_ν , which has three degrees of freedom, and a massive real scalar field η which has one degree of freedom. When we considered a global gauge transformation, we found a massless Goldstone boson in addition to the massive scalar field. Here the Goldstone boson has disappeared, but its degree of freedom has given the vector field B_ν a mass. This is called the Higgs mechanism. The remaining massive scalar is called a Higgs boson.

4.8 A few words about quantization

Thus far, the gauge fields we have discussed can be considered as classical fields. However, modern successful theories of elementary particles are quantum theories. It is therefore necessary to quantize the fields. Two main methods of quantization have been treated in the literature: *canonical quantization* and *path-integral quantization*. Path-integral quantization appears to be the more useful of the two for quantizing nonabelian gauge field theories.

Here, we do not undertake the rather lengthy process of quantizing gauge fields, but note that Aitchison and Hey (1989) discuss canonical quantization and Huang (1992) discusses path-integral quantization. A more detailed and advanced treatment of quantization methods is given in the book by Itzykson and Zuber (1980). The object of quantization is to obtain the so-called *Feynman rules* for evaluating the theory in a perturbation expansion. Each term in the perturbation series corresponds to a *Feynman diagram* or *Feynman graph*. In discussing the consequences of a field theory we customarily simply write down the Feynman graph corresponding to a particular term in the perturbation expansion of the theory rather than writing down the mathematical formula for the term. We discuss Feynman graphs in more detail in Chapter 6.

The quanta of free fields may be either massless or massive, except that the quanta of a gauge field connected with a manifest gauge symmetry are massless. The quanta of scalar fields (scalar particles) have spin 0 and positive parity. If the parity is negative, the particles are pseudoscalars. The particles of vector fields (vector particles) have spin 1 and negative parity. If the parity is positive, the particles are axial vectors. The quanta of the spinor fields we discuss have spin 1/2. If they are different from their antiparticles, they are Dirac spinors; if they are the same as their antiparticles, they

are Majorana spinors. Because of the conservation of angular momentum, spinors are produced in pairs, or, alternatively, one spinor may be destroyed and another created.

The perturbation expansion of a relativistic gauge field theory contains infinite terms. A procedure to eliminate these infinities is called *renormalization*. At present, when the principal way we have to evaluate gauge field theory is by perturbation theory, such a theory must be renormalizable to be useful. Here, we do not discuss how to renormalize a theory. See, for example, Itzykson and Zuber (1980).

It may happen that the classical version of a field theory has a certain symmetry, but the quantized version does not. If this occurs, we say that the theory has an *anomaly*. In practice, this means that certain Feynman graphs, called anomalous graphs, do not respect the symmetry. This shows up as a nonvanishing divergence of a current which is expected to be conserved. In some cases one may obtain an estimate of how much the symmetry is violated in a particular process by approximately evaluating the relevant anomalous graph. In other cases, however, one can obtain such an estimate only in an approximate nonperturbative treatment of the quantized theory.

If a gauge theory has any anomalies which violate gauge invariance, then the renormalizability of the theory may also be destroyed. We reject such a theory at present for the practical reason that we don't know how to solve it. (Future developments in mathematics may change this situation.) On the other hand, the quantized version of a theory may have an anomaly which preserves renormalizability and violates only a global symmetry of the classical theory. If the anomaly leads to an effect which is large enough to be observable, one has a test of the quantized version of the theory.

5. THE STANDARD MODEL

The standard model is a nonabelian local gauge field theory which is invariant under the gauge group $SU(3) \times SU(2) \times U(1)$. The model consists of two parts: the Weinberg–Salam model, which is based on the spontaneously broken gauge group $SU(2) \times U(1)$ and quantum chromodynamics, which is based on the manifestly invariant gauge group $SU(3)$.

5.1 The Weinberg–Salam model

We first consider the Weinberg–Salam model, which is sometimes called the Glashow–Weinberg–Salam model. (Other people have also made substantial contributions to it.) This model is a nonabelian gauge theory with spontaneous symmetry breaking. The model is sometimes called a unified theory of weak and electromagnetic interactions. However, these interactions are not completely unified because the model contains two different coupling constants, one associated with $SU(2)$ and the other with $U(1)$. Nevertheless, the model certainly relates the weak and electromagnetic interactions, and so it is rightly called a theory of electroweak interactions. Although it would seem as if we could treat the groups $SU(2)$ and $U(1)$ separately because they enter only as a direct product, in actual fact, we cannot, owing to the fact that the groups are mixed in the vacuum state because of spontaneous symmetry breaking. The manifest symmetry group $U(1)_{\text{em}}$ of electromagnetism is not the $U(1)$ of $SU(2) \times U(1)$ but contains parts of both $SU(2)$ and $U(1)$.

The $SU(2)$ of the direct product-group is sometimes denoted by $SU(2)_L$ because it concerns left-handed currents (to be defined shortly) or $SU(2)_i$ because it is the group of weak isospin. The $U(1)$ of the direct-product group is sometimes called $U(1)_y$ because it is the group of weak hypercharge (I and Y are also used).

Since the time of Fermi’s theory of weak interactions, it was known from experiment that a charged weak current exists. By a charged current we here mean a current which carries electric charge. An example of a charged weak current is the current carried by an electron and its antineutrino as they move along. It was later proposed that the weak interaction should be a gauge theory, which would have as a consequence that the charged current be coupled to charged vector (gauge) bosons (now called W^\pm). One difficulty with this picture is that it was believed that gauge bosons must be massless,

like the photon. However, the range of weak interactions is very short, and so the force carrier must be very heavy.

Another difficulty with the gauge idea is that the two generators of the charged currents do not form a Lie algebra. As a consequence, there is no group of gauge transformations which leaves the Lagrangian invariant. To remedy this situation, it was proposed that the W^\pm and the photon are a triplet of bosons connected with the broken gauge group $SU(2)$. (It was believed the group had to be broken to allow the weak bosons to have mass.) However, Glashow realized that this solution was not plausible because the electromagnetic interaction conserves parity but the weak interaction does not. Therefore, Glashow proposed that a neutral weak boson (now called the Z^0), should also exist. A gauge theory with three weak bosons and the photon, requires a four-parameter gauge group. In 1961 Glashow proposed that the four-parameter group $SU(2) \times U(1)$ is the gauge group of the electroweak interaction. Because the Z is neutral, it must be coupled to a weak *neutral* current. So Glashow predicted the existence of the weak neutral current in 1961, 12 years before it was discovered.

As we have discussed, the quanta of gauge fields are supposed to be massless. However, Glashow realized that if the three weak bosons he proposed actually exist, they must have large masses or they would have been observed well before 1961. (The weak bosons were, in fact, first observed in the early 1980s.) Glashow was not able to give a reason why weak bosons should have masses greater than zero. Subsequently, Higgs (1964) and others found a mechanism for spontaneous symmetry breaking, which gives masses to gauge bosons. This mechanism involves introducing a complex scalar Higgs field, as we have discussed in Chapter 4. The application of the Higgs mechanism to break the electroweak gauge symmetry spontaneously and to give masses to the weak bosons is due to Weinberg (1967) and Salam (1968) and also to Ward (Salam and Ward, 1964). In the Weinberg–Salam model, the Higgs field is also used to generate fermion masses.

We discuss the Lagrangian of the Weinberg–Salam model primarily for the electron e and electron neutrino ν (we omit the subscript e on the symbol for the electron neutrino). The Lagrangian contains similar terms for each family of leptons. The Lagrangian for each family of quarks is also similar but a little more complicated because, unlike neutrinos in the model, all quarks have mass and charge. So, for the moment, we are going to deal with only a single generation.

Before writing down the Lagrangian, we introduce some notation. We let the symbol for a particle stand for its field and write a lepton current l_μ

for the electron and its neutrino:

$$l_\mu = \bar{e}\gamma_\mu(1 - \gamma_5)\nu \quad (5.1)$$

This (weak) lepton current differs from the electromagnetic current $j_\mu = \bar{e}\gamma_\mu e$ in two important ways: First, unlike the neutral electromagnetic current, l_μ is a charged current, as can be seen from the fact that it contains the charged field \bar{e} , but instead of the charged field e to neutralize it, it contains rather the neutral field ν . Second, because of the factor $\gamma_\mu(1 - \gamma_5)$, l_μ contains both a vector and an axial vector part, and thus does not conserve parity. The current l_μ is phenomenological and is based on what we knew about the weak interactions of electrons and neutrinos after the discovery in 1956 that these interactions did not conserve parity and after the realization that the structure of the weak current was vector minus axial vector.

We now introduce projection operators P_R and P_L which operate on four-component Dirac spinors. These operators, when operating on spinors with zero mass, project them onto positive (right-handed) and negative (left-handed) chirality states respectively. The projection operators are given by

$$P_R = \frac{1}{2}(1 + \gamma_5), \quad P_L = \frac{1}{2}(1 - \gamma_5). \quad (5.2)$$

If a spinor ψ has mass, then $\psi_L = P_L\psi$ is not entirely a negative-helicity state, and $\psi_R = P_R\psi$ is not entirely a positive-helicity state. Nevertheless, it is customary to call ψ_L and ψ_R left-handed and right-handed respectively.

There are some interesting properties of the chiral spinors ψ_L and ψ_R . Consider what happens to ψ_L under conjugation.

$$\overline{\psi_L} = (P_L\psi)^\dagger\gamma^0 = \psi^\dagger P_L^\dagger\gamma^0. \quad (5.3)$$

However, γ_5 is hermitian and anticommutes with γ^0 , so $P_L^\dagger\gamma^0 = \gamma^0 P_R$ and

$$\overline{\psi_L} = \overline{\psi}P_R. \quad (5.4)$$

From this, we can immediately see that the fermion mass term in the Lagrangian couples left- and right-handed spinors, *i.e.*,

$$\bar{\psi}\psi = \overline{\psi_L}\psi_R + \overline{\psi_R}\psi_L \quad (5.5)$$

and

$$\overline{\psi_L}\psi_L = \overline{\psi_R}\psi_R = 0. \quad (5.6)$$

However, if we place a single gamma matrix between the two spinors then the handedness of both spinors must be the same. For example,

$$\bar{\psi}_L \gamma^\mu \psi_R = \bar{\psi} P_R \gamma^\mu P_R \psi = \bar{\psi} \gamma^\mu P_L P_R \psi = 0. \quad (5.7)$$

The kinetic term of the fermion Lagrangian can be written as

$$\bar{\psi} \gamma^\mu \partial_\mu \psi = \bar{\psi}_L \gamma^\mu \partial_\mu \psi_L + \bar{\psi}_R \gamma^\mu \partial_\mu \psi_R. \quad (5.8)$$

We now introduce the left- and right-handed spinors

$$e_L = P_L e, \quad e_R = P_R e, \quad \nu_L = P_L \nu, \quad \nu_R = P_R \nu. \quad (5.9)$$

In the standard model, ν is massless and $\nu_R = 0$. The assumption that the neutrino is massless is made only for simplicity. If the neutrino turns out to have a mass greater than 0, then a ν_R can easily be introduced. (If $m_\nu > 0$, then the statement $\nu_R = 0$ is not invariant under Lorentz transformations.)

Using Eqs. (5.2) and (5.9) and the fact that for any projection operator P , we have $P^2 = P$, we can write l_μ as

$$l_\mu = 2\bar{e}_L \gamma_\mu \nu_L, \quad (5.10)$$

where we have also made use of the fact that γ^5 anticommutes with all the other γ matrices.

In the last chapter we worked hard to understand how to construct a non-Abelian gauge theory by starting with a globally invariant theory. To promote the global invariance to a local invariance, we had to replace ordinary derivatives for the fermion and scalar fields by the covariant derivative. We also had to add terms to the Lagrangian for the gauge fields themselves. This means that we have a simple program for constructing a model. First, decide on a gauge group. Next, decide on the representations for the fermions. Finally, decide on the representations for the scalar particles and arrange to break the symmetry, if necessary. If you make all the right choices, you might win a trip to Stockholm.

Let's try to understand the details of this for the Standard Model. Recall that the gauge group is $SU(3) \times SU(2) \times U(1)$. To select the fermions, we just have to specify a representation for each component of the gauge group. It should be clear to you by now that the quarks all belong to the **3** representation of color and the leptons, which don't carry color, are color singlets. For the weak isospin group, we chose to put left-handed fermions

in doublets and right-handed fermions in singlets. This accounts for the observed parity violation of the weak interaction. The $U(1)$ group is known as weak hypercharge, and its representations are one dimensional. All we have to specify is the magnitude of the hypercharge. To do this we will try to impose the weak analog of the Gell-Mann Nisijima formula. (See Eqs. (1.4) and (1.5).)

$$Q = i_z + y_W/2, \quad (5.11)$$

where Q is the charge in units of e , i_z is the weak isospin eigenvalue and y_W is the weak hypercharge. Let's work out the hypercharges of the quark and lepton multiplets.

For the neutrino, electron doublet, we have $Q_\nu = 0 = i_z + y_W/2 = 1/2 + y_W/2 \Rightarrow y_W = -1$. For the quark doublet that contains u_L and d_L , $Q_u = 2/3 = 1/2 + y_W/2 \Rightarrow y_W = 1/3$. For the e_R , $Q_e = -1 = i_z + y_W/2 = 0 + y_W/2 \Rightarrow y_W = -2$. Similarly, we can calculate the necessary hypercharges for the u_R and d_R . To summarize, we list all the representations, first in terms of the field, then we label the representations by (color representation, weak isospin representation, weak hypercharge).

$$\begin{array}{ll} \begin{pmatrix} \nu_L \\ e_L \end{pmatrix} & (1, 2, -1) \\ e_R & (1, 1, -2) \\ \begin{pmatrix} u_L \\ d_L \end{pmatrix} & (3, 2, 1/3) \\ u_R & (3, 1, 4/3) \\ d_R & (3, 1, -2/3) \end{array} \quad (5.12)$$

It should be clear from the fact the the weak hypercharge of u_R and u_L differ that weak and strong hypercharge are very different quantum numbers. After we have arranged for the symmetry breaking and identify the photon we must go back and show that it comes into the covariant derivative multiplied by $e(i_z + y_W/2)$. For the next several pages, we are going to be ignoring the terms in the weak Lagrangian that depend upon the quarks. The basic points that we are trying to make can be most easily seen in the lepton sector. At the end of the section, we will add the quarks and see what complications arise.

Let's introduce some simplifying notation.

$$L \equiv \begin{pmatrix} \nu_L \\ e_L \end{pmatrix}, \quad \bar{L} \equiv (\bar{\nu}_L, \bar{e}_L). \quad (5.13)$$

and

$$R \equiv \epsilon_R. \quad (5.14)$$

We can introduce globally invariant terms in the Lagrangian density that are kinetic terms for the free leptons.

$$\mathcal{L}_L = \bar{L}i\gamma^\mu\partial_\mu L. \quad (5.15)$$

and

$$\mathcal{L}_R = \bar{R}i\gamma^\mu\partial_\mu R. \quad (5.16)$$

These two terms are invariant under a global $SU(2)$ transformation for the lepton doublet L of the form

$$U_\theta = e^{-i\tau\cdot\theta/2}, \quad \tau\cdot\theta = \tau_a\theta^a. \quad (5.17)$$

(Of course, the singlet R does not change under an $SU(2)$ transformation.)

This theory is also invariant under global $U(1)$ transformations of the form

$$U_\phi = e^{-iy\phi}. \quad (5.18)$$

We make the theory locally invariant by introducing the $SU(2)$ gauge field \mathbf{A}_μ which is a 3-component vector in the internal space of $SU(2)$ as well as being a Lorentz 4-vector and the $U(1)$ gauge field B_μ . Later, after we spontaneously break the symmetry, we shall see the relation between the fields \mathbf{A}_μ , B_μ and the fields for W^\pm , Z^0 and γ .

To make the theory locally gauge invariant, we replace partial derivatives by covariant derivatives. Note the form of the covariant derivative depends on the representation of the field upon which it acts. We are thus led to write the Lagrangian for the left-handed and right-handed fields as

$$\mathcal{L}_L = \bar{L}i\gamma^\mu(\partial_\mu + \frac{1}{2}ig\tau\cdot\mathbf{A}_\mu + \frac{1}{2}ig'yB_\mu)L, \quad (5.19)$$

$$\mathcal{L}_R = \bar{R}i\gamma^\mu(\partial_\mu + \frac{1}{2}ig'yB_\mu)R, \quad (5.20)$$

where g and g' are the coupling constants for $SU(2)$ and $U(1)$, respectively. We note that the factor of $1/2$ multiplying g' is purely conventional and that the value of y is different for L and R . Also, with this convention for the coupling, under a $U(1)$ local gauge transformation,

$$B_\mu \rightarrow B'_\mu = B_\mu + \frac{2}{g'}\partial_\mu\phi(x). \quad (5.21)$$

The transformation rule for the $SU(2)$ gauge fields can be found in Chapter 4.

We can write $\mathcal{L}_L + \mathcal{L}_R$ in a more succinct fashion by introducing a covariant derivative

$$D_\mu = \partial_\mu + ig\mathbf{T} \cdot \mathbf{A}_\mu + \frac{1}{2}ig'yB_\mu, \quad (5.22)$$

where \mathbf{T} is the weak isospin operator. This operator is equal to $\frac{1}{2}\tau$ when operating on an $SU(2)$ doublet and has the value 0 when operating on an $SU(2)$ singlet. Then the gauge Lagrangian is

$$\mathcal{L}_L + \mathcal{L}_R = \bar{L}i\gamma^\mu D_\mu L + \bar{R}i\gamma^\mu D_\mu R, \quad (5.23)$$

where here D_μ is given by Eq. (5.22). Note that there are no fermion mass terms in Eq. (5.23). A mass term is not gauge invariant because such a term contains equal amounts of left- and right-handed fields, which transform differently under gauge transformations. (It is necessary for the left and right fields to transform differently in order that parity not be conserved.)

To the two terms $\mathcal{L}_L + \mathcal{L}_R$ we add the Lagrangian for the gauge fields

$$\mathcal{L}_G = -\frac{1}{4}\mathbf{A}_{\mu\nu} \cdot \mathbf{A}^{\mu\nu} - \frac{1}{4}B_{\mu\nu}B^{\mu\nu}, \quad (5.24)$$

where

$$B_{\mu\nu} = \partial_\mu B_\nu - \partial_\nu B_\mu, \quad (5.25)$$

$$A_{\mu\nu}^a = \partial_\mu A_\nu^a - \partial_\nu A_\mu^a - g\epsilon_{bc}^a A_\mu^b A_\nu^c. \quad (5.26)$$

This second formula is sometimes written schematically as

$$\mathbf{A}_{\mu\nu} = \partial_\mu \mathbf{A}_\nu - \partial_\nu \mathbf{A}_\mu - g\mathbf{A}_\mu \times \mathbf{A}_\nu. \quad (5.27)$$

Because A_μ is nonabelian, \mathcal{L}_G contains self-interaction terms of this field as well as kinetic energy terms.

If the electroweak Lagrangian consisted only of the sum of \mathcal{L}_L , \mathcal{L}_R , and \mathcal{L}_G , it would not be spontaneously broken, and the gauge fields and fermions would remain massless. To break the symmetry spontaneously, we need to proceed to the next step of our model building program, *i.e.*, add a Higgs field Φ . We have to decide what representation to choose for Φ . Since we don't want to break the color group, Φ should be a color singlet. The smallest

nontrivial representation of $SU(2)$ is the doublet, so we consider two complex scalar fields

$$\Phi = \begin{pmatrix} \varphi^+ \\ \varphi^0 \end{pmatrix}. \quad (5.28)$$

The superscripts refer to the electric charges of the Higgs fields. The reason we need two complex Higgs fields rather than one is that we need three degrees of freedom to generate spontaneously the masses of three gauge fields (the W^\pm and the Z^0). The remaining degree of freedom will show up as a neutral massive Higgs field. (It also turns out that this doublet can give mass to the fermions, so it economically serves two purposes, as we will soon see.) You might wonder if we could more economically use a real $SU(2)$ triplet, *i.e.*, a vector to break the symmetry. Unfortunately, since this field can carry no $U(1)$ hypercharge (it is real so can't undergo a phase transformation), it can only break the $SU(2)$ and not the $U(1)$. There would remain two massless gauge bosons. This Higgs choice would also not give mass to the fermions, so we still would have to add a doublet, a rather ugly state of affairs.

The Lagrangian of the Higgs field, including its interaction with the gauge fields, is

$$\mathcal{L}_H = (D_\mu \Phi)^\dagger D^\mu \Phi + \mu^2 \Phi^\dagger \Phi - \lambda (\Phi^\dagger \Phi)^2 \quad (5.29)$$

We shall see that the Lagrangian \mathcal{L}_H will generate masses of the weak gauge field, but before we do that there are some additional invariant terms that we can add to our Lagrangian. These are the terms that can give rise to fermion masses.

The Higgs–fermion coupling is often called a Yukawa coupling, as it was Yukawa who first postulated a coupling between spin 0 particles (mesons) and spin 1/2 particles (nucleons). In general, it takes the form $\bar{\psi}_1 \psi_2 \phi$, where we can now take any combinations of the fermion fields and the Higgs field that has a globally invariant product. Recall that before we had the Higgs field a mass term was ruled out by global $SU(2)$ invariance since all the left-handed spinors are in doublets and the right-handed parts are in singlets. A mass term which multiplies left- and right-handed parts together is necessarily an $SU(2)$ doublet, but now when we multiply by a Higgs doublet, we can form an $SU(2)$ singlet. A term of the form $\bar{L}\Phi R$ is globally invariant if the Higgs field has hypercharge +1. (This notation is fairly condensed. The $SU(2)$ indices on \bar{L} and Φ are contracted to form a singlet. The Dirac indices on \bar{L} and R are contracted to form a Lorentz scalar.)

To summarize, the Yukawa coupling between the Higgs and the electron has the form

$$\mathcal{L}_Y = -\rho(\bar{L}\Phi R + \bar{R}\Phi^\dagger L), \quad (5.30)$$

where ρ is the coupling strength, and we have added the complex conjugate of the term we formed above. The subscript Y on \mathcal{L}_Y stands for Yukawa. As we shall see, this term leads to spontaneous generation of the electron mass. We introduce similar Yukawa couplings of the Higgs to other massive fermions as we shall see at the end of this section.

The electroweak Lagrangian \mathcal{L}_{ew} is given by

$$\mathcal{L}_{\text{ew}} = \mathcal{L}_L + \mathcal{L}_R + \mathcal{L}_G + \mathcal{L}_H + \mathcal{L}_Y, \quad (5.31)$$

where the various terms are given in Eqs. (5.23), (5.24), (5.29), and (5.30). In \mathcal{L}_{ew} , the term containing μ^2 is of the wrong sign for μ to be interpreted as a mass. Therefore, this Lagrangian is not written in a form which is suitable for doing perturbation theory. To obtain a physical interpretation, we make a change of variables for the field Φ :

$$\Phi = e^{i\zeta \cdot \tau / 2v} \begin{pmatrix} 0 \\ (v + \eta) / \sqrt{2} \end{pmatrix}, \quad (5.32)$$

where v is a positive constant and $\zeta_1, \zeta_2, \zeta_3$, and η are four real fields replacing the two complex fields of Eq. (5.28).

Now we make an $SU(2)$ gauge transformation

$$\Phi' = e^{-i\zeta \cdot \tau / 2v} \Phi. \quad (5.33)$$

The fields L and \mathbf{A}_μ also get transformed. However, we do not need to substitute the explicit transformations into the Lagrangian because we did not specify the form of the old fields. We can therefore interpret the L and \mathbf{A}_μ as the new fields after the transformation. The net effect of all this is merely to substitute

$$\Phi = \begin{pmatrix} 0 \\ (v + \eta) / \sqrt{2} \end{pmatrix} \quad (5.34)$$

into the Lagrangian of Eq. (5.31). The ζ field has apparently disappeared, but its degrees of freedom will appear in another guise: namely, to give masses to the three weak gauge bosons.

To see this we make yet another change in variables, writing

$$W_\mu^\pm = (A_\mu^1 \mp iA_\mu^2) / \sqrt{2}. \quad (5.35)$$

Then, after using (5.34) and (5.35) and the fact that the Higgs field has hypercharge $y = 1$, we obtain for the Higgs term in the Lagrangian

$$\begin{aligned}\mathcal{L}_H &= \frac{1}{2}\partial^\mu\eta\partial_\mu\eta + \frac{1}{4}g^2W^{-\mu}W_\mu^+(v+\eta)^2 \\ &\quad + \frac{1}{8}(gA^{3\mu} - g'B^\mu)(gA_\mu^3 - g'B_\mu)(v+\eta)^2 \\ &\quad + \frac{1}{2}v^2\lambda(v+\eta)^2 - \frac{1}{4}\lambda(v+\eta)^4.\end{aligned}\tag{5.36}$$

The quantities v and η enter the Lagrangian symmetrically, except in the derivative term, because they were put in symmetrically. However, they have very different interpretations: v is a constant and η is a real neutral scalar field. Using Eq. (5.34) and the identity $e = e_L + e_R$, we obtain the Higgs Yukawa term:

$$\mathcal{L}_Y = -\rho(v+\eta)\bar{e}e/\sqrt{2}.\tag{5.37}$$

We can identify various terms in Eqs. (5.36) and (5.37) as spontaneously generated mass terms. Comparing these equations with previous Lagrangians for massive fields (see Chapter 4), we obtain the result that the weak boson, the electron, and the Higgs have masses given by

$$m_W = gv/2, \quad m_e = \rho v/\sqrt{2}, \quad m_\eta = v\sqrt{2\lambda}.\tag{5.38}$$

In the Lagrangian (5.36) the combination $gA_\mu^3 - g'B_\mu$ appears quadratically, multiplied by a constant. Therefore, this linear combination of fields has mass, and we identify it, after normalization, with the Z^0 field

$$Z_\mu = \frac{gA_\mu^3 - g'B_\mu}{(g^2 + g'^2)^{1/2}}.\tag{5.39}$$

This real field has mass

$$m_Z = v(g^2 + g'^2)^{1/2}/2.\tag{5.40}$$

Let us introduce the *Weinberg angle* θ_W by the definition

$$\tan\theta_W = g'/g.\tag{5.41}$$

We drop the subscript on θ_W for the remainder of this section. We get from Eqs. (5.38) and (5.39):

$$m_W = m_Z \cos\theta.\tag{5.42}$$

Also Eq. (5.39) becomes

$$Z_\mu = A_\mu^3 \cos \theta - B_\mu \sin \theta. \quad (5.43)$$

We introduce a linear combination of fields which is orthogonal to Z_μ :

$$\begin{aligned} A_\mu &= A_\mu^3 \sin \theta + B_\mu \cos \theta \\ &= \frac{g' A_\mu^3 + g B_\mu}{(g^2 + g'^2)^{1/2}} \end{aligned} \quad (5.44)$$

We note from Eq. (5.36) that no mass term appears for this field and that it has no interaction with the Higgs field η . We identify A_μ with the ordinary electromagnetic field.

Now that we have identified the electromagnetic field, we must go back and verify that its coupling to particles is proportional to $T_3 + Y/2$ which was previously identified with the charge. To see this, we return to Eq. (5.22) where we saw that the covariant derivative contains the terms

$$igT^3 A_\mu^3 + \frac{1}{2}ig'Y B_\mu. \quad (5.45)$$

Using Eqs (5.39) and (5.44) to express this in terms of the physical Z_μ and A_μ fields, we find the couplings

$$i \frac{gg'}{(g^2 + g'^2)^{1/2}} \left(T^3 + \frac{Y}{2} \right) A_\mu + i \frac{1}{(g^2 + g'^2)^{1/2}} \left(g^2 T^3 - \frac{g'^2 Y}{2} \right) Z_\mu. \quad (5.46)$$

So we see that the photon coupling is, indeed, proportional to the combination of weak isospin and hypercharge that we have been calling the charge.

If we express all of $\mathcal{L}_L + \mathcal{L}_R$ in terms of the new fields and eliminate g' by using the Weinberg angle, we get

$$\begin{aligned} \mathcal{L}_L + \mathcal{L}_R &= \bar{\nu}_L i \gamma^\mu \partial_\mu \nu_L + \bar{e} i \gamma^\mu \partial_\mu e - g(\bar{\nu}_L \gamma^\mu e_L W_\mu^+ + \bar{e}_L \gamma^\mu \nu_L W_\mu^-) / \sqrt{2} \\ &\quad + \frac{1}{2} g \frac{\cos 2\theta}{\cos \theta} \bar{e}_L \gamma^\mu e_L Z_\mu - g \frac{\sin^2 \theta}{\cos \theta} \bar{e}_R \gamma^\mu e_R Z_\mu \\ &\quad - \frac{1}{2} \frac{g}{\cos \theta} \bar{\nu}_L \gamma^\mu \nu_L Z_\mu + g \sin \theta \bar{e} \gamma^\mu e A_\mu. \end{aligned} \quad (5.47)$$

Of course, the electromagnetic field A_μ interacts with the electron field, which is charged, but not with the neutrino field, which is neutral. The magnitude e of the electron charge is

$$e = g \sin \theta, \quad \alpha = e^2 / 4\pi = 1/137. \quad (5.48)$$

Note that the symbol e here stands for the magnitude of the electron charge, but in other places stands for the electron field. It should be clear from the context which is meant.

Returning to \mathcal{L}_Y , we see from Eq. (5.37) that it contains the term $-\rho\eta\bar{e}e\sqrt{2}$, which describes the Yukawa interaction of the Higgs field with the electron field. Because of Eq. (5.38), the coupling strength can be written

$$\rho\sqrt{2} = 2m_e/v. \quad (5.49)$$

Thus, the interaction of the Higgs field with an electron (or, more generally, with any fermion) is proportional to the mass of the electron (fermion).

The original electroweak Lagrangian for the electron, neutrino, gauge fields, and Higgs field contained the parameters g , g' , μ , λ , and ρ . Instead of these parameters we can use combinations of them. (In fact, above we have always used v rather than μ as one of the two parameters in \mathcal{L}_H .) An equally good set is α , θ , m_e , m_W , and m_η . The parameters α and m_e are of course well known. The masses of the W and Z bosons have been measured, and their ratio gives $\cos\theta$. Alternatively, m_W and θ can be deduced from experimental information on low-energy weak interactions: m_W can be related to the measured coupling constant G_F appearing in Fermi's old theory of weak interactions, and θ can be related to the measured ratio of the neutral to the charged weak current. The relationship between m_W and G_F is

$$m_W^2 = \pi\alpha/(\sin^2\theta G_F\sqrt{2}). \quad (5.50a)$$

An equivalent and useful relationship between G and g and M_W is

$$\frac{G_F}{\sqrt{2}} = \frac{g^2}{8M_W^2}. \quad (5.50b)$$

In fact, the low energy measurements enabled physicists to predict the masses of the W and Z before these bosons were discovered. We discuss these topics again in Chapter 8.

The quantity λ (or alternatively, m_η) is still unknown. There is no term in the Lagrangian which relates either of these two quantities to low-energy experiments, and, despite many experiments to search for the Higgs boson, it has not been observed.

We have considerably oversimplified our discussion of the measurement of the parameters in the electroweak Lagrangian. First of all, the measured or physical charges and masses are not the ones that enter into the Lagrangian.

The parameters of the Lagrangian are so-called *bare* charges and masses, which have to be *renormalized* to obtain the physical charges and masses. Only in lowest-order in perturbation theory are the bare masses and the physical masses the same. Secondly, in comparing measurements to theory, we have to make so-called *radiative corrections* to take into account the fact that low-energy (infrared) photons are not observed in the experiments. We discuss these topics in a little more detail in Chapter 8.

Thus far, we have written the Lagrangian for only one pair of fermions, the electron and its neutrino. Similar terms occur in the model for each family of leptons, and an additional parameter is required for each charged lepton. Quarks also participate in weak interactions, and a similar term in the Lagrangian occurs for each left-handed quark doublet and each right-handed quark singlet.

There are some differences between the electroweak Lagrangian for leptons and for quarks. First, all quarks have mass, and so have right-handed as well as left-handed couplings. Therefore, in the standard model with three families there are six masses or Yukawa couplings for the quarks, but only three for the leptons. Second, all quarks have charge, and so are coupled to the electromagnetic field. Also, because the quark charges are fractional, the values of their weak hypercharge are different from the values for leptons as we see in Eq. (5.12). Third, the negatively-charged quarks which are mass eigenstates, denoted by d , s , and b , are not the states which participate as left-handed doublets in the electroweak Lagrangian. It is not difficult to see how this can come about. Recall that the Yukawa terms involve one fermion spinor, one antifermion spinor and the Higgs field. Once there are several generations, there is no reason that the fermion and antifermion spinors have to belong to the same generation. So, the Yukawa coupling becomes a matrix and once the spontaneous symmetry breaking occurs, we have to find the eigenvalues of the matrix to identify the quarks that are mass eigenstates (*i.e.*, make a change of basis). Since the charge $+2/3$ and charge $-1/3$ quarks have completely independent Yukawa couplings (as we shall see below), there is no relationship between the two changes of basis, and the u quark is found to interact with some combination of d , s and t , the mass eigenstates.

The left-handed doublets for three quark families are

$$L_1 = \begin{pmatrix} u_L \\ d'_L \end{pmatrix}, \quad L_2 = \begin{pmatrix} c_L \\ s'_L \end{pmatrix}, \quad L_3 = \begin{pmatrix} t_L \\ b'_L \end{pmatrix}, \quad (5.51)$$

where d' , s' , and b' are linear combinations of d , s , and b . A unitary matrix V connects the primed and unprimed quantities. The matrix equation is

conventionally written

$$\begin{pmatrix} d' \\ s' \\ b' \end{pmatrix} = \begin{pmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{pmatrix} \begin{pmatrix} d \\ s \\ b \end{pmatrix}. \quad (5.52)$$

A unitary matrix in n dimensions is characterized by n^2 real parameters. However, owing to the fact that certain phases of fields are unobservable, the number of parameters is reduced to $(n-1)^2$ parameters, of which $\frac{1}{2}n(n-1)$ are angles and $\frac{1}{2}(n-1)(n-2)$ are CP -violating phases. In two dimensions, there is only one parameter, the Cabibbo (1963) angle. The formalism in the two-dimensional case is due to Glashow *et al.* (1970). In three dimensions, the mixing matrix V is the CKM or Cabibbo–Kobayashi–Maskawa (Kobayashi and Maskawa, 1973) matrix, and is characterized by three angles θ_{ij} ($i < j$) and one phase δ , which violates the combined discrete symmetry CP . The CKM matrix may be parametrized in several ways. One common way is given by the Particle Data Group (Hikasa *et al.*, Phys. Rev. D45, #11 Part II, 1992):

$$V = \begin{pmatrix} c_{12}c_{13} & s_{12}c_{13} & s_{13}p_- \\ -s_{12}c_{23} - c_{12}s_{23}s_{13}p_+ & c_{12}c_{23} - s_{12}s_{23}s_{13}p_+ & s_{23}c_{13} \\ s_{12}s_{23} - c_{12}c_{23}s_{13}p_+ & -c_{12}s_{23} - s_{12}c_{23}s_{13}p_+ & c_{23}c_{13} \end{pmatrix}, \quad (5.53)$$

where $c_{ij} = \cos \theta_{ij}$, $s_{ij} = \sin \theta_{ij}$, and $p_{\pm} = e^{\pm i\delta}$.

In discussing weak decays of particles of only the first two families, we can approximate the CKM matrix by a 2×2 matrix containing only one parameter, the Cabibbo angle θ_{12} :

$$V \simeq \begin{pmatrix} c_{12} & s_{12} \\ -s_{12} & c_{12} \end{pmatrix} \quad (5.54)$$

because empirically the angles θ_{13} and θ_{23} are small.

In the quark sector, the left and right Lagrangians are similar to those for the lepton sector, given by Eqs. (5.19) and (5.20), except that in \mathcal{L}_L the L gets replaced by the L_i of Eq. (5.51) for each family and the \mathcal{L}_R contains two terms for each family. The gauge and Higgs Lagrangians \mathcal{L}_G and \mathcal{L}_H are unaltered, but the Yukawa Lagrangian of Eq. (5.30) gets additional terms. It is instructive to write \mathcal{L}_Y explicitly for the first quark family. We get

$$\mathcal{L}_Y = -\rho_d(\bar{L}_{ud}\Phi R_d + \bar{R}_d\Phi^\dagger L_{ud}) - \rho_u(\bar{L}_{ud}\Phi^C R_u + \bar{R}_u\Phi^{C\dagger} L_{ud}), \quad (5.55)$$

where ρ_u and ρ_d are the coupling strengths (parameters) and Φ^C is given by

$$\Phi^C = \begin{pmatrix} \varphi^{0*} \\ -\varphi^{+*} \end{pmatrix}. \quad (5.56)$$

If neutrinos have mass, we can add right-handed fields for them and include three Yukawa terms to generate their masses. We can also include a matrix similar to the CKM matrix to describe their left-handed couplings. This will entail the introduction of another 7 parameters. We emphasize, however, that this is not the only way to introduce neutrino masses into the standard model.

5.2 Quantum chromodynamics

Quantum chromodynamics is a nonabelian local gauge theory based on the gauge group $SU(3)$. The theory of QCD results from a straightforward application to $SU(3)$ of the ideas which Yang and Mills (1954) first applied to $SU(2)$.

The group $SU(3)$ was chosen because of evidence that quarks come in three colors. The earliest evidence came from baryons. The proton and other low-energy baryons apparently have wave functions which are symmetric under the interchange of a pair of quarks of the same flavor. This is most easily seen for the Δ^{++} baryon, which has spin 3/2 and is composed of three u quarks. A spin-3/2 wave function is totally symmetric under the interchange of any two spin coordinates. Also, there is evidence that the three quarks are in states with zero orbital angular momentum, so that the spatial wave function is also symmetric. But if quarks have half-integral spin, their wave function should be antisymmetric under the interchange of any pair (the quarks should behave like fermions, not bosons). To rectify this situation, it was postulated that the quarks contain an additional degree of freedom, called color by Gell-Mann (although he was not the first to propose the new degree of freedom). Because a baryon contains three quarks, it is plausible that a quark can have any of three colors, and the wave function is antisymmetric in the color degree of freedom. Other evidence for the existence of three colors is given in later chapters.

Since $SU(3)$ is characterized by eight parameters, QCD has eight gauge fields. The quanta of these fields are called gluons. These gluons play a role analogous to the role of photons in QED. The fermions of the theory are the quarks, which play a role analogous to the role of charged leptons in QED.

Color is the QCD analogy of electric charge, and is really strong charge, that is, the charge of the strong interaction. However, while there is only

one kind of electric charge (plus its anticharge), there are three kinds of strong charges and their anticharges, or, as we usually say, three colors and three anticolors. A consequence of the nonabelian nature of the $SU(3)$ gauge theory is that the gluons are also required to carry color (or rather, both color and anticolor). This is in contrast to the QED case, in which the photon does not carry electric charge.

The symmetry of QCD is a manifest symmetry, and this implies that the gauge fields are massless. Naively, then, we might expect that the strong forces have a long range and that we can observe free massless gluons. In fact, strong forces are short range, and no free gluons have been observed. Free quarks also have not been observed although at least some of them apparently have small masses. A way out of this dilemma is to assume that the forces are so strong that they confine any colored particle to the interior of hadrons. Although there have been many attempts to prove that QCD confines color, this has not been accomplished. However, the result of all the effort has been to make it plausible that QCD does in fact confine colored particles. We discuss the problem of confinement in Chapter 8.

To construct QCD, we start with a Lagrangian describing the behavior of a free massless quark field ψ , which has 12 components: 4 components for each of the 3 colors. The free Lagrangian is

$$\mathcal{L} = \bar{\psi} i \gamma^\mu \partial_\mu \psi. \quad (5.57)$$

The mass of the quark is assumed to be generated by spontaneous symmetry breaking of the electroweak interaction, whether by means of coupling to a Higgs field or by some other mechanism. If one is considering QCD by itself, without the electroweak interaction, it is acceptable to add a mass term as input to the Lagrangian.

We next require that the Lagrangian be invariant under local nonabelian gauge transformations of $SU(3)$, namely

$$\psi' = U\psi, \quad U = e^{-i\mathbf{F}\cdot\theta} = e^{-iF_a\theta^a}, \quad (5.58)$$

where the F_a are the eight generators of $SU(3)$. In order to make the Lagrangian invariant under such gauge transformations, we must replace the derivative by the covariant derivative

$$D_\mu = \partial_\mu + i g_s \mathbf{F} \cdot \mathbf{G}_\mu, \quad (5.59)$$

where g_s is the strong coupling constant and the G_μ^a are eight gauge (gluon) fields. We also add the kinetic energy to the gauge fields and add a mass

term which presumably arises from the electroweak sector of the standard model. Then we obtain the Lagrangian

$$\mathcal{L}_{\mathcal{QCD}} = \bar{\psi}(i\gamma^\mu D_\mu - m)\psi - \frac{1}{4}\mathbf{G}_{\mu\nu} \cdot \mathbf{G}^{\mu\nu}, \quad (5.60)$$

where

$$G_{\mu\nu}^a = \partial_\mu G_\nu^a - \partial_\nu G_\mu^a - g_s f_{bc}^a G_\mu^b G_\nu^c \quad (5.61)$$

and the f_{bc}^a are the structure constants of $SU(3)$. Of course, if we add the Lagrangian $\mathcal{L}_{\mathcal{QCD}}$ to the electroweak Lagrangian, we omit the term containing the mass m , as this mass already appears in the electroweak Lagrangian and we don't need the $\bar{\psi}i\gamma^\mu\partial_\mu\psi$ terms since they are also in the electroweak covariant derivative. Another way of saying this is that the complete covariant derivative is of the form

$$D_\mu = \partial_\mu + ig_s \mathbf{F} \cdot \mathbf{G}_\mu + ig \mathbf{T} \cdot \mathbf{A}_\mu + \frac{1}{2}ig' y B_\mu, \quad (5.62)$$

where the multiplicities of the representations corresponding to the matrices \mathbf{F} and \mathbf{T} and the value of y are given for each representation in Eq. (5.12). The field ψ in the Lagrangian $\mathcal{L}_{\mathcal{QCD}}$ has 12 components for one quark flavor. The Lagrangian for six quark flavors contains a sum of six terms, each of which is like the Lagrangian $\mathcal{L}_{\mathcal{QCD}}$ with its own field ψ_q of mass m_q . Alternatively, one may regard ψ in Eq. (5.60) as a 72-dimensional field and m as a diagonal mass matrix with 6 distinct values.

Because of the last term in (5.61), the Lagrangian (5.60) contains, in addition to an interaction of quarks with the gauge field G_μ , also interactions of the gauge fields with themselves. Let us make an analogy between charge of QED and color of QCD. The QED gauge field (electromagnetic field) couples to all charged fields, but not to itself because it is neutral. In the QCD case, however, the gauge fields couple to all colored fields including themselves. Note that in order for this self coupling to exist (*i.e.*, in order for the gluon fields to carry color), the group must be nonabelian. We see this directly from Eq. (5.61), where the self-coupling term containing g_s vanishes if the structure constants vanish.

Because the burden of accounting for quark masses is placed on the Weinberg–Salam model, QCD contains only one input parameter, the coupling constant g_s , or the more commonly used $\alpha_s = g_s^2/(4\pi)$. This makes QCD a very pretty theory indeed. However, despite its beauty, QCD has proved very difficult to evaluate, especially in the nonperturbative regime.

As we shall see in Chapter 7, the effective coupling constant α_s depends on the energy and/or four-momentum transfer at which it is measured. At an energy of 30 GeV, $\alpha_s \simeq 0.15$. Because of a property of QCD known as *asymptotic freedom* (Politzer 1973, Gross and Wilczek 1973), α_s decreases as the four-momentum transfer increases. In Chapter 7 we describe asymptotic freedom and discuss why it holds.

An additional coupling constant, which gives rise to a strong CP -violating term, is sometimes included in QCD. There is no empirical evidence for such a term, and we set it equal to zero. A large number of physicists have tried to understand why the coupling constant that violates CP is either zero or too small to have been measured so far. However, discussion of this subject is beyond our scope here, and we refer the reader to papers by Peccei and Quinn (1977), Weinberg (1978), and Wilczek (1978) as well as a treatment in the book by Huang (1992).

5.3 Symmetries of the standard model

The Lagrangian of the standard model is invariant under the local gauge group $SU(3) \times SU(2) \times U(1)$. However, as we have seen, the vacuum state does not have this full symmetry, but rather the local gauge symmetry $SU(3) \times U(1)_{\text{em}}$. This is the manifest local gauge symmetry of the physical world (excluding gravity) at energies at or below the mass of the Z boson. At energies sufficiently higher than that, we expect the full gauge symmetry to be restored.

The standard model has a number of other symmetries. We have not really discussed Poincaré invariance, but it can be directly shown that the Lagrangian of the standard model is invariant under the proper Poincaré group. The proper Poincaré group consists of translations, rotations, and Lorentz transformations, but excludes the discrete symmetry operations P , C , and T . As with all local field theories invariant under the proper Poincaré group and having a stable vacuum, the standard model is invariant under CPT . However, the weak sector of the model violates C and P separately. Furthermore, as we have noted, with three families, the weak sector violates CP invariance unless the phase δ of the CKM matrix is zero. If CP is violated, then, because of CPT invariance, T is also violated. It is interesting that with only two families, the CKM matrix has only one parameter, so that CP is conserved (although C and P are separately violated independently of the number of families because of the structure of the theory). With a fourth family, the CKM matrix contains 9 parameters, but no new symmetries are broken.

The strong and electromagnetic sectors of the standard model are invariant under C , P , and T separately, although in principle the strong sector can contain a term violating CP and therefore T . It is not known why the coupling strength of the strong CP -violating term is either zero or too small to have been measured so far.

The Lagrangian of the standard model is also invariant under global gauge transformations associated with the conservation of baryon number and the three lepton family numbers. As we have noted in Section 2.4, because of an anomaly the quantized theory exactly conserves only $A - L$. Theoretical estimates of the breaking of A and L separately lead to the conclusion that violations are too small to be observable under present conditions, in agreement with the fact that no violation has yet been observed. However, we do not know of any fundamental principle which requires these invariances and conservation laws to be exact.

In addition, the strong and electromagnetic (but not the weak) sectors, are invariant under global gauge transformations associated with quark and lepton flavor number conservation. Isospin conservation in the strong sector appears to be an approximate symmetry associated with the fact that the mass difference between the u and d quarks is much smaller than the constituent masses of these quarks. The u and d constituent masses, in turn, are probably related to the scale parameter Λ of QCD, which we discuss in the next chapter. Lastly, we mention that QCD has a broken chiral symmetry. We do not discuss chiral symmetry in these notes, but refer to a treatment in the book by Huang (1992).

5.4 Parameters of the standard model

The Weinberg–Salam model with three families contains as parameters 2 coupling constants, the masses of the W and Higgs bosons, the masses of 3 charged leptons, the masses of 6 quarks, and the 4 parameters of the CKM matrix, for a total of 17 parameters. If the neutrinos have masses different from zero and can be included in the standard model in the same fashion as quarks, then up to 7 additional parameters will be necessary, 3 for the neutrino masses and 4 for a neutrino mixing matrix analogous to the CKM matrix. But, as we have already remarked, there are other ways to incorporate neutrino masses and mixing into a modified standard model. If a fourth family exists, there will be at least an additional 9 parameters: 4 fermion masses (the fourth neutrino must be heavy or it would have been observed) and 5 more parameters in the CKM mixing matrix (even without neutrino mixing). Most of these parameters are connected to the Higgs

sector, which is considered the least satisfying part of the model.

In addition, there are one or two parameters associated with QCD. The first is the strong-interaction coupling constant. The second is a parameter associated with strong CP violation, but in the present version of the standard model this parameter is set equal to zero. The reason that QCD contains so few parameters is that the burden of giving the quarks masses is placed in the Higgs sector of the Weinberg–Salam model.

Many physicists are not happy with a theory containing at least 18 parameters (19, if we count the strong CP phase which is extremely close to zero). There have been many attempts to invent a better theory, so far without marked success.

Chapter 1

The Basics

Def: A *group* is a pair (G, \circ) , where G is a set, and \circ is a binary operation (“multiplication”) defined on G such that:

1. G is closed under \circ :

$$a \circ b \in G \quad \forall a, b \in G.$$

2. \circ is associative:

$$(a \circ b) \circ c = a \circ (b \circ c) \quad \forall a, b, c \in G.$$

3. \exists a right identity element, $e \in G$, such that:

$$a \circ e = a \quad \forall a \in G.$$

4. For some right identity e , \exists for each $a \in G$ at least one right inverse, $a^{-1} \in G$, such that

$$a \circ a^{-1} = e.$$

We say “ (G, \circ) is a group”, or simply, “ G is a group.” The operation, \circ , is referred to as the group multiplication, or, simply, multiplication.

Some examples to illustrate this definition:

- The set of integers under addition is a group (denoted Z).
- The set of people is not a group (need an operation).
- The set of non-singular 2×2 matrices under matrix multiplication is a group (denoted $GL(2)$, for “general linear group in 2 dimensions”).

There are several immediate consequences of the group axioms:

Physics 195
Supplementary Notes
Groups, Lie algebras, and Lie groups
020922 F. Porter

This note defines some mathematical structures which are useful in the discussion of angular momentum in quantum mechanics (among other things).

Def: A pair (G, \circ) , where G is a non-empty set, and \circ is a binary operation defined on G , is called a **group** if:

1. Closure: If $a, b \in G$, then $a \circ b \in G$.
2. Associativity: If $a, b, c \in G$, then $a \circ (b \circ c) = (a \circ b) \circ c$.
3. Existence of right identity: There exists an element $e \in G$ such that $a \circ e = a$ for all $a \in G$.
4. Existence of right inverse: For some right identity e , and for any $a \in G$, there exists an element $a^{-1} \in G$ such that $a \circ a^{-1} = e$.

The \circ operation is typically referred to as “multiplication”.

The above may be termed a “minimal” definition of a group. It is amusing (and useful) to prove that:

1. The right identity element is unique.
2. The right inverse element of any element is unique.
3. The right identity is also a left identity.
4. The right inverse is also a left inverse.
5. The solution for $x \in G$ to the equation $a \circ x = b$ exists and is unique, for any $a, b \in G$.

We will usually drop the explicit \circ symbol, and merely use juxtaposition to denote group multiplication. Note that both G (the set) and \circ (the “multiplication table”) must be specified in order to specify a group. Where the operation is clear, we will usually just refer to “ G ” as a group.

Def: An **abelian** (or **commutative**) group is one for which the multiplication is commutative:

$$ab = ba \quad \forall a, b \in G. \tag{1}$$

Def: The **order** of a group is the number of elements in the set G . If this number is infinite, we say it is an “infinite group”.

In the discussion of infinite groups of relevance to physics (in particular, Lie groups), it is useful to work in the context of a richer structure called an algebra. For background, we start by giving some mathematical definitions of the underlying structures:

Def: A **ring** is a triplet $\langle R, +, \circ \rangle$ consisting of a non-empty set of elements (R) with two binary operations ($+$ and \circ) such that:

1. $\langle R, + \rangle$ is an abelian group.
2. (\circ) is associative.
3. Distributivity holds: for any $a, b, c \in R$

$$a \circ (b + c) = a \circ b + a \circ c \quad (2)$$

and

$$(b + c) \circ a = b \circ a + c \circ a \quad (3)$$

Conventions:

We use 0 (“zero”) to denote the identity of $\langle R, + \rangle$. We speak of $(+)$ as addition and of (\circ) as multiplication, typically omitting the (\circ) symbol entirely (*i.e.*, $ab \equiv a \circ b$).

Def: A ring is called a **field** if the non-zero elements of R form an abelian group under (\circ) .

Def: An abelian group $\langle V, \oplus \rangle$ is called a **vector space** over a field $\langle F, +, \circ \rangle$ by a scalar multiplication $(*)$ if for all $a, b \in F$ and $v, w \in V$:

1. $a * (v \oplus w) = (a * v) \oplus (a * w)$ distributivity
2. $(a + b) * v = (a * v) \oplus (b * v)$ distributivity
3. $(a \circ b) * v = a * (b * v)$ associativity
4. $1 * v = v$ unit element ($1 \in F$)

Conventions:

We typically refer to elements of V as “vectors” and elements of F as “scalars.” We typically use the symbol $+$ for addition both of vectors and scalars. We also generally omit the $*$ and \circ multiplication symbols. Note that this definition is an abstraction of the definition of vector space given in the note on Hilbert spaces, page 1.

Def: An **algebra** is a vector space V over a field F on which a multiplication (\circ) between vectors has been defined (yielding a vector in V) such that for all $u, v, w \in V$ and $a \in F$:

1. $(au) \circ v = a(u \circ v) = u \circ (av)$
2. $(u+v) \circ w = (u \circ w) + (v \circ w)$ and $w \circ (u+v) = (w \circ u) + (w \circ v)$

(Once again, we often omit the multiplication sign, and hope that it is clear from context which quantities are scalars and which are vectors.)

We are interested in the following types of algebras:

Def: An algebra is called **associative** if the multiplication of vectors is associative.

We note that an associative algebra is, in fact, a ring. Note also that the multiplication of vectors is not necessarily commutative. An important non-associative algebra is:

Def: A **Lie algebra** is an algebra in which the multiplication of vectors obeys the further properties (letting u, v, w be any vectors in V):

1. Anticommutativity: $u \circ v = -v \circ u$.
2. Jacobi Identity: $u \circ (v \circ w) + w \circ (u \circ v) + v \circ (w \circ u) = 0$.

We may construct the idea of a “group algebra”: Let G be a group, and V be a vector space over a field F , of dimension equal to the order of G (possibly ∞). Denote a basis for V by the group elements. We can now define the multiplication of two vectors in V by using the group multiplication table as “structure constants”: Thus, if the elements of G are denoted by g_i , a vector $u \in V$ may be written:

$$u = \sum a_i g_i$$

We require that, at most, a finite number of coefficients a_i are non-zero. The multiplication of two vectors is then given by:

$$\left(\sum a_i g_i\right) \left(\sum b_j g_j\right) = \sum \left(\sum_{g_i g_j = g_k} a_i b_j\right) g_k$$

[Since only a finite number of the $a_i b_j$ can be non-zero, the sum $\sum_{g_i g_j = g_k} a_i b_j$ presents no problem, and furthermore, we will have closure under multiplication.]

Since group multiplication is associative, our group algebra, as we have constructed it, is an associative algebra.

Theorem: If $x \in G$ and $x \circ x = x$, and if e is a right identity such that property (4) holds, then $x = e$.

Proof:

$$\begin{aligned}
 x &= x \circ e && e = \text{right identity} \\
 &= x \circ (x \circ x^{-1}) && \text{property (4)} \\
 &= (x \circ x) \circ x^{-1} && \text{associativity} \\
 &= x \circ x^{-1} && \text{by assumption} \\
 &= e && \text{property (4)}
 \end{aligned}$$

QED

Theorem: The right inverse is also a left inverse: If (G, \circ) is a group with identity e , and $a \circ a^{-1} = e$, then $a^{-1} \circ a = e$.

Proof: Let $f = a^{-1} \circ a$. Then:

$$\begin{aligned}
 f \circ f &= (a^{-1} \circ a) \circ (a^{-1} \circ a) \\
 &= a^{-1} \circ (a \circ (a^{-1} \circ a)) && \text{associativity} \\
 &= a^{-1} \circ ((a \circ a^{-1}) \circ a) && \text{associativity} \\
 &= a^{-1} \circ (e \circ a) && \text{right inverse} \\
 &= (a^{-1} \circ e) \circ a && \text{associativity} \\
 &= a^{-1} \circ a && \text{right identity} \\
 &= f && \text{assumption} \\
 &= e && \text{previous theorem}
 \end{aligned}$$

QED

Hence, we may drop the “right” and simply say “inverse”.

Several other properties can also be quickly proven:

Theorem: The right identity is unique.

Theorem: The right identity is also a left identity.

Theorem: The inverse is unique.

Theorem: The solutions to $a \circ x = b$ and to $x \circ a = b$, where $a, b \in G$ exist ($x \in G$), and are unique.

Theorem: The inverse of a product $a \circ b$ is:

$$(a \circ b)^{-1} = b^{-1} \circ a^{-1}.$$

This may be readily extended to higher order products.

We usually drop the explicit notation for the group multiplication, and use concatenation to denote multiplication, unless doing so would be unclear.

Some important groups have an additional property:

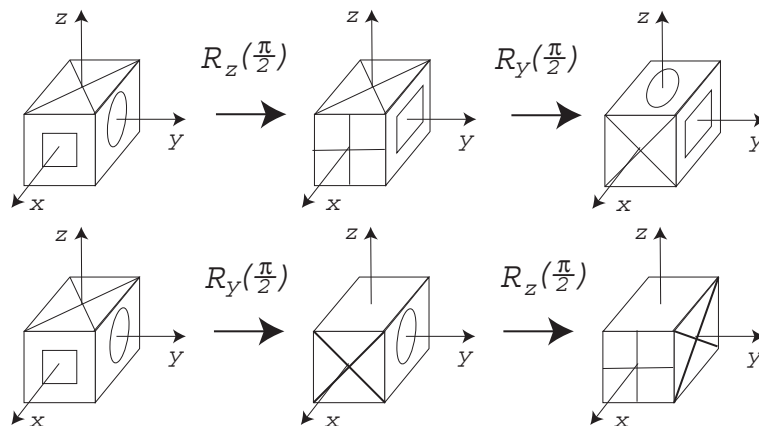


Figure 1.1: Illustration that the result of successive rotations in three dimensions depends on the order of the rotations.

Def: If G is a group such that

$$ab = ba \quad \forall a, b \in G,$$

then G is called an *abelian*, or *commutative*, group.

For example, Z is an abelian group. $GL(2)$ is a non-abelian group:

$$\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \neq \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}.$$

For a more “physically relevant” example, the group of rotations in two dimensions is abelian, but the group of rotations in three dimensions is non-abelian:

$$R_z(\pi/2)R_y(\pi/2) \neq R_y(\pi/2)R_z(\pi/2).$$

See Fig. 1.

Some groups are almost trivial:

Def: If G is a group such that the powers of one element generate the group, then the group is called *cyclic*:

$$a, a^2, a^3, \dots, a^n = e.$$

For example, the group $Z_n = \{0, \dots, n-1\}$ under modulo n addition is a cyclic group. The powers of 1 generate the group, with $1^n = 0 = e$. Obviously, a cyclic group is also abelian.

The number of elements in a group is an important basic parameter:

Def: If there are a finite number, n , of elements in a group, then it is said to be a *finite group*, of order n . Otherwise, it is an *infinite group*.

For an infinite group, the infinity may be denumerable (for example, Z), or non-denumerable (for example, $GL(2)$) For a finite group, we may explicitly give a multiplication table, or *Cayley table*, as a table with n columns and n rows. For example consider a group of order five, with elements a, b, c, d , and e , where e is the identity. A possible multiplication table for such a group is shown in Table 1.1.

Table 1.1: An example of a multiplication table for a group of order five. The row labels indicate the left multiplicand and the column labels the right multiplicand. Thus, for example, the product $db = a$ may be found in the last row of the table.

L \ R	e	a	b	c	d
e	e	a	b	c	d
a	a	b	c	d	e
b	b	c	d	e	a
c	c	d	e	a	b
d	d	e	a	b	c

We can remark several things concerning this table:

- By the existence and uniqueness of the solution to $ax = b$, the multiplication table must be a Latin square – every element occurs exactly once in each row or column. This is a statement of the so-called “rearrangement lemma”: If $pb = pc$, then $b = c$.
- Since our example is symmetric about the diagonal, it specifies an abelian group.
- Noting that $b = a^2$, $c = ab = a^3$, $d = ac = a^4$, and $e = ad = a^5$, we see that this is a cyclic group.
- Finally, we may remark that there exists no group of order five which is not cyclic. In fact, we have given the only group multiplication table for order five, up to renaming of the elements.

1.1 Permutation Group

We introduce here a very important class of groups, known as the *permutation* or *symmetry* groups. We denote by S_n the group of all possible permutations or rearrangements of n objects. As there are $n!$ ways of rearranging n objects (taken to be distinct), S_n is a group of order $n!$.

Let us develop some of the notational tools used in discussing S_n . We imagine that we have a set of n “slots”, arranged in a line, into which we are going to place our n objects, one per slot. For example, we use the array:

$$(1, 2, 3, \dots, n) \quad (1.1)$$

to denote that object “1” is in the first slot (the first position in the array), object “2” is in the second slot, etc. A permutation of these objects may be written as

$$p = \begin{pmatrix} 1 & 2 & 3 & \dots & n \\ p_1 & p_2 & p_3 & \dots & p_n \end{pmatrix}. \quad (1.2)$$

In this case, object “1” in slot one has been replaced by object “ p_1 ”, object “2” in slot two has been replaced by object “ p_2 ”, etc.

The identity element is just to “do nothing”:

$$e = \begin{pmatrix} 1 & 2 & 3 & \dots & n \\ 1 & 2 & 3 & \dots & n \end{pmatrix}. \quad (1.3)$$

The inverse of element p above is:

$$p^{-1} = \begin{pmatrix} p_1 & p_2 & p_3 & \dots & p_n \\ 1 & 2 & 3 & \dots & n \end{pmatrix}. \quad (1.4)$$

That is, we have the product $e = p^{-1}p$:

$$p^{-1}p = \begin{pmatrix} p_1 & p_2 & p_3 & \dots & p_n \\ 1 & 2 & 3 & \dots & n \end{pmatrix} \begin{pmatrix} 1 & 2 & 3 & \dots & n \\ p_1 & p_2 & p_3 & \dots & p_n \end{pmatrix}. \quad (1.5)$$

This notation is a bit more cumbersome than we need, since we don’t really need to keep track of the slots, only what objects are replacing which other objects. For example, consider the permutation in S_5 :

$$p = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 4 & 1 & 5 & 2 & 3 \end{pmatrix}. \quad (1.6)$$

In this case, object one is being replaced by object 4, which is being replaced by object two, and object two is being replaced by object one. Also, objects 3 and 5 are being switched. We could write this as $(1 \rightarrow 4 \rightarrow 2 \rightarrow 1)$ and $(3 \rightarrow 5 \rightarrow 3)$. We call these sub-rearrangements “cycles”, and shorten the notation to $p = (142)(35)$. Permutation p consists of a “3-cycle” and a “2-cycle”. Note that $(142) = (214)$, but (142) is not the same as (124) . The inverse permutation is:

$$p^{-1} = \begin{pmatrix} 4 & 1 & 5 & 2 & 3 \\ 1 & 2 & 3 & 4 & 5 \end{pmatrix} = (412)(35) = (124)(35). \quad (1.7)$$

We may check in cycle notation:

$$pp^{-1} = [(142)(35)][(124)(35)] = e \quad (1.8)$$

For example, in the first operation, object 1 is replaced by object 2. In the second operation, object 2 is replaced by object 1, putting object 1 back into its original position. On final simplification in notation – we may drop the “one-cycles” as understood, for example:

$$(123)(4)(5)(6) = (123). \quad (1.9)$$

As an example, the reader is encouraged to construct the multiplication table for S_3 , shown in Table 1.2. Notice that this is a non-abelian group.

Table 1.2: The multiplication table for permutation group S_3 .

L \ R	e	(12)	(13)	(23)	(123)	(132)
e	e	(12)	(13)	(23)	(123)	(132)
(12)	(12)	e	(132)	(123)	(23)	(13)
(13)	(13)	(123)	e	(132)	(12)	(23)
(23)	(23)	(132)	(123)	e	(13)	(12)
(123)	(123)	(13)	(23)	(12)	(132)	e
(132)	(132)	(23)	(12)	(13)	e	(123)

The reader is cautioned that different conventions exist for the notation for the elements of the symmetry groups. Here, I adopt the convention of Wu-Ki Tung and of Hamermesh; and not that of Mathews & Walker.

This concludes our introduction to the most basic elements of group theory. We now proceed to slightly more sophisticated notions.

1.2 Classes

We first introduce the idea of equivalence of two elements of a group:

Def: Given a group G , two elements $a, b \in G$ are said to be *equivalent* if there exists an element $u \in G$ such that

$$u^{-1}au = b. \quad (1.10)$$

The equivalence of two elements is denoted $a \sim b$.

We remark that this defines a true equivalence relation, since the following properties of an equivalence are satisfied:

1. Reflexivity: $a \sim a$, $\forall a \in G$. To see this, simply take $u = e$.
2. Symmetry: If $a \sim b$, then $b \sim a$. Since, if $u^{-1}au = b$, then $a = v^{-1}bv$, where $v = u^{-1}$.

3. Transitivity: If $a \sim b$, and $b \sim c$, then $a \sim c$. The reader should verify this.

If we consider groups of operators, the equivalence of two group elements may be described as: If you first “do u ”, then “do a ” and finally “undo u ”, and the result of all this is the operation b , then a and b are equivalent. More concretely, consider the group of all rotations in three dimensions. A rotation by 45° about the x -axis is equivalent to a rotation by 45° about the y axis. To see this, take u to be a rotation about the z -axis by 90° :

$$R_x(45^\circ) = R_z(-90^\circ)R_y(45^\circ)R_z(90^\circ). \quad (1.11)$$

More generally, any two rotations by the same angle are equivalent. This gives a nice intuitive feel for what equivalence means: Since we can find a rotation relating any two given axes of rotations, rotations by the same angle about these two axes are equivalent. We remark that in the future we will consider smaller groups of rotations which may not contain the necessary rotation from one axis to another. In this case the rotations by the same angle will no longer be equivalent.

The notion of equivalence will permit a great simplification in the study of group structure, through the use of equivalence classes:

Def: The subsets of G consisting of elements of G which are equivalent to each other are called the *classes* of G .

Some remarks are in order:

1. The simplification we will achieve will be attained through the treatment of a class as a single object, where the distinctions among its members is (often) unimportant.
2. Different classes of a group are, by transitivity, disjoint sets. Every element of the group belongs to some class, that is, the union of all classes is the entire group.
3. The identity element is always in a class by itself, since

$$u^{-1}eu = e, \quad \forall u \in G. \quad (1.12)$$

4. In an abelian group, every element is in a class by itself, since in this case

$$u^{-1}au = a, \quad \forall u \in G. \quad (1.13)$$

1.3 Subgroups

Another important concept in the study of group structure is the possibility that a group may contain other groups as subsets:

Def: If (G, \circ) is a group, and $S \subset G$ is a non-empty subset of G , such that (S, \circ) is a group, then S is called a *subgroup* of G .

For example, $\{e\}$ is a subgroup of G , and G itself is a subgroup of G . A subgroup which is neither of these “trivial” cases is referred to as a *proper subgroup* of G . For a more interesting example, consider once again the group of all rotations in three dimensions. Pick any axis. The set of all rotations about the chosen axis is a proper subgroup of the entire rotation group.

For finite groups there is an important theorem concerning the order of subgroups:

Theorem: (Lagrange) The order of a subgroup of a finite group is a divisor of the order of the group.

Thus, any group of prime order has only two subgroups, $\{e\}$ and G , and no proper subgroups. For example, consider the group of rotations about a given axis by angles $2\pi(m/n)$, where $n > 1$ is a prime number, and $m = 0, 1, \dots, n-1$. This group is of prime order, hence has no proper subgroups according to the theorem. The reader should quickly verify that this is indeed the case.

The proof of Lagrange’s theorem is instructive, and introduces some additional concepts:

Proof: Consider group G of order $g < \infty$, and subgroup H of order h . If $H = G$, then the theorem is trivially satisfied, with $h = g$.

Suppose $H \neq G$. Let a be an element in G that is not in H . Denote the elements of H by

$$e = H_1, H_2, H_3, \dots, H_h. \quad (1.14)$$

Form the products

$$\{ae = a, aH_2, aH_3, \dots, aH_h\} = “aH” = \{aH_i | i = 1, 2, \dots, h\}. \quad (1.15)$$

Each product must be distinct, since if $aH_i = aH_j$ then $a^{-1}(aH_i = aH_j)$ yields $H_i = H_j$. Furthermore, no product aH_i is contained in H , since if $aH_i \in H$ for some i , then $(aH_i)H_i^{-1} = a \in H$ (since $H_i^{-1} \in H$). But $a \notin H$ by assumption.

Thus, we have two disjoint sets of h distinct elements, H and aH , which are contained in G . If $\{z | z \in H \text{ or } z \in aH\} = G$, then $g = 2h$, and the theorem holds. Otherwise, there must be an element $b \in G$ such that $b \notin H$ and $b \notin aH$. In this case, we proceed as before, forming the set

$$bH = \{bH_i | i = 1, 2, \dots, h\}, \quad (1.16)$$

again finding that bH and H are disjoint sets.

Furthermore, aH and bH are disjoint sets, since if $aH_i = bH_j$ for some i and j , then $aH_iH_j^{-1} = bH_jH_j^{-1} = b \in aH$. But $b \notin aH$ by assumption. If the sets H , aH , and bH comprise all of the elements of G then $g = 3h$ and the theorem holds. Otherwise, we repeat the process of finding disjoint

subsets with h elements each, until we have exhausted the elements of G . Thus, G is the sum of a finite number of distinct sets containing h elements each:

$$G = H + aH + bH + \dots + kH, \quad (1.17)$$

and hence $g = mh$ where m is an integer (called the **index** of the subgroup H under the group G). This completes the proof.

The sets of elements of the form aH , where $a \in G$ and H is a subgroup of G are called the **left cosets** of H in G . We could just as easily have proven the theorem using **right cosets**, that is sets of the form Ha .

We may note that for a finite group G , any element a will have some lowest power p , called its **order**, such that $a^p = e$. This is true because the sequence a, a^2, a^3, \dots cannot continue to generate new elements for a finite group; it must eventually repeat. The sequence $a, a^2, a^3, \dots, a^p = e$ is called the **period** of a . Notice now that the period of a is the smallest subgroup of G which contains a . Since it is a subgroup, the order of a must be a divisor of the order of G , for any finite group G . Thus, we find in particular that any finite group of prime order *must* be a cyclic group (and hence also abelian).

It is useful to keep in mind these facts as we examine the structure of groups.

Def: If a subgroup $S \subset G$ is such that

$$g^{-1}Sg = S, \quad \forall g \in G, \quad (1.18)$$

then S is called an **invariant subgroup** of G .

The notation Sg , where S is a set of elements and g is an element, means the set of elements obtained by multiplying every element of S by g . An invariant subgroup consists of classes – if it contains part of a class, it must contain the entire class. For example, any subgroup of an abelian group is an invariant subgroup. For an invariant subgroup we also have that:

$$gS = Sg, \quad \forall g \in G. \quad (1.19)$$

That is, the left and right cosets of S in G are identical.

To get a better intuition into the notion of an invariant subgroup, the reader should ponder the following examples of subgroups of the rotation group:

1. Consider the group of all proper and improper rotations (that, is, we include the spatial inversion, or parity operator, \mathcal{P}). The group of all proper rotations is an invariant subgroup of this group.
2. Consider the group of all (proper) rotations. The subgroup of all rotations about a specified axis is not an invariant subgroup.

Finally, the concepts of “simple” and “semi-simple” groups will be useful in the classification of groups in terms of basic subgroup structure:

Def: A group is called *simple* if it does not contain any proper invariant subgroups. A group is called *semi-simple* if it does not contain any abelian invariant subgroups.

1.4 Some Groups

We'll conclude this chapter with a table of some groups that we encounter frequently:

Symbol	Elements	Operation
Z	integers	addition
Z_n	$0, 1, \dots, n - 1$	addition, modulo n
Q	rationals	addition
Q^*	rationals, except 0	multiplication
R	reals	addition
R^*	reals, except 0	multiplication
C	complex	addition
C^*	complex, except 0	multiplication
S	complex on unit circle	multiplication
S_n	n th roots of unity	multiplication
S_n	permutations of n objects	successive permutations
$GL(n)$	non-singular complex $n \times n$ matrices	matrix multiplication
$GL(nR)$	non-singular real $n \times n$ matrices	matrix multiplication
$SL(n)$	$GL(n)$ with determinant one	matrix multiplication
$U(n)$	$n \times n$ unitary matrices	matrix multiplication
$SU(n)$	$U(n)$ with determinant one	matrix multiplication
$O(n)$	$n \times n$ real unitary matrices	matrix multiplication
$SO(n) \equiv O^+(n)$	$O(n)$ with determinant one	matrix multiplication

The “ G ” in the symbols stands for “general”, the “ L ” is for “linear”, the “ U ” is for “unitary”, and the “ O ” is for “orthogonal”. For the matrix groups, the “ S ” is for “special”, and means determinant one.

1.5 Exercises

- Which of the following define groups? If not a group, give at least one axiom which is violated.
 - The set of all real numbers, excluding zero, under division. That is, if a and b are non-zero real numbers, the proposed binary group operation is $c = a \circ b \equiv a/b$.
 - The set of all Hermitian $n \times n$ matrices, under matrix multiplication.
 - The set of all Hermitian $n \times n$ matrices, under matrix addition.
 - The set of all operations (rotations and reflections) which leave a tetrahedron invariant. For convenience, you may wish to imagine a coordinate system in which the origin is at the “center” of the tetrahedron (this is a “fixed” point under the symmetry operations).
- Prove the five theorems stated at the bottom of page 2.

3. Write down the multiplication table, using cycle notation, for symmetry group S_4 . You don't need to do the whole table if you find it too tedious, but at least do all columns, and enough rows to show an example of each cycle structure.
4. Decompose symmetry group S_4 into classes.
5. List all of the proper subgroups of symmetry group S_4 .
6. Find all of the invariant subgroups of symmetry group S_4 .

Chapter 2

Crystallographic Point Groups

Crystals are characterized by periodic structure, or symmetry. The notion of a *point group* has to do with the symmetry operations on a geometric object, with a fixed point, that is, those operations such as rotations and reflections which leave the object invariant. Those point symmetries which are also consistent with periodic crystal structure (translational symmetry) yield the *Crystallographic Point Groups*. Thus, for example, the symmetry of the cube is consistent with crystal structure, while that of the icosahedron is not. To be explicit, we have the definition:

Def: A three-dimensional lattice of points is said to have *translational symmetry* if there exists some set of “primitive” translations, $\{\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3\}$ such that every point in the lattice may be reached from another point by a translation of the form:

$$\mathbf{T}(t_1, t_2, t_3) = t_1\mathbf{a}_1 + t_2\mathbf{a}_2 + t_3\mathbf{a}_3, \quad (2.1)$$

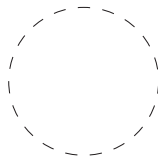
where t_1, t_2, t_3 are integers.

There are a total of 32 Crystallographic Point Groups (we’ll just say “point groups” here for brevity). These are listed in this chapter.

2.1 Notation

We use the notation $R_x(\theta)$ to indicate a rotation by angle θ about the x axis. We use the notation M_x to indicate a mirror plane in which the normal to the mirror is along the x -axis.

There exists more than one system of notation for the groups we discuss here. The simple **Schönflies** notation will suffice here. In this notation, a C_n

Figure 2.1: The uniaxial group C_1 .

is used to designate an n -fold rotation axis. For example, C_4 implies a four-fold rotation axis (with associated rotations of $\pm\pi/2$, π , $e = 2\pi$). The notation may also be used to indicate a particular rotation, e.g., C_3 indicates a rotation by $2\pi/3$ about the designated axis, while C_3^{-1} indicates a rotation by $-2\pi/3$.

We may add the inversion operator (parity) \mathcal{I} to obtain the improper rotations. In Schönflies notation, $S_2 = \mathcal{I}$. Closely associated with this is the mirror reflection, indicated by σ (for a reflection in a plane). If the reflection is through a plane perpendicular to a specified rotation axis, then an h subscript is added, giving σ_h . It is readily seen that a reflection followed by a rotation by π about the axis perpendicular to the mirror plane is the same as an inversion:

$$S_2 = \mathcal{I} = C_2\sigma_h. \quad (2.2)$$

We generalize the S_2 notation to include n -fold axes. Hence,

$$S_n = C_n\sigma_h = \sigma_h C_n. \quad (2.3)$$

It is helpful to have a pictorial representation. We will adopt something known as “stereograms”. This will be developed in the course of listing the 32 point groups.

2.2 The 32 Crystallographic Point Groups

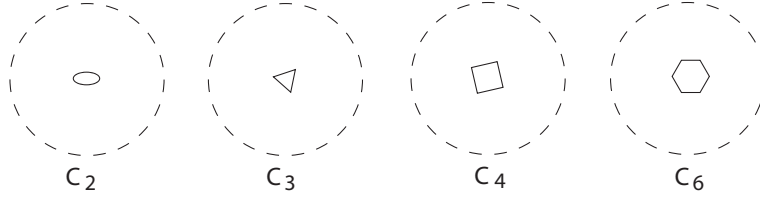
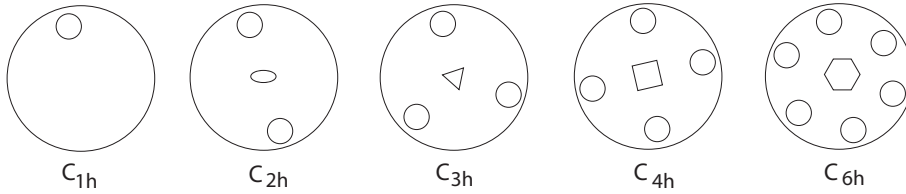
We start with a theorem, which limits the possible symmetry axes that need to be considered:

Theorem: The only possible proper rotations consistent with translational symmetry are C_n , $n = 1, 2, 3, 4, 6$.

Proof of this is left to the reader.

We are ready to make the list. We’ll start with the groups with the least symmetry. We start with what are known as the “uniaxial” groups, those with a single n -fold axis. The first is C_1 , see Fig. 2.1.

In this figure, we see a dashed circle and a \times . For now, the reader may wish to imagine that there is a disk, represented by the circle, with a peg sticking up above the plane of the disk at the \times . Thus, there is no symmetry, either under rotations or mirrors (we suppose that there is no symmetry under splitting the

Figure 2.2: The uniaxial groups C_2 , C_3 , C_4 , and C_6 .Figure 2.3: The uniaxial groups C_{1h} , C_{2h} , C_{3h} , C_{4h} , and C_{6h} .

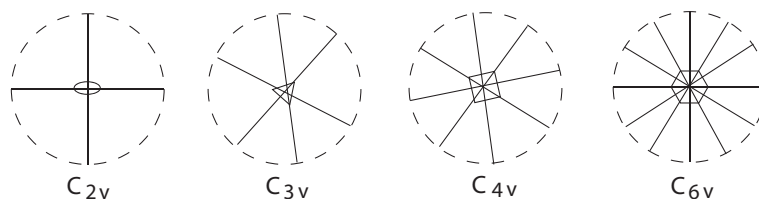
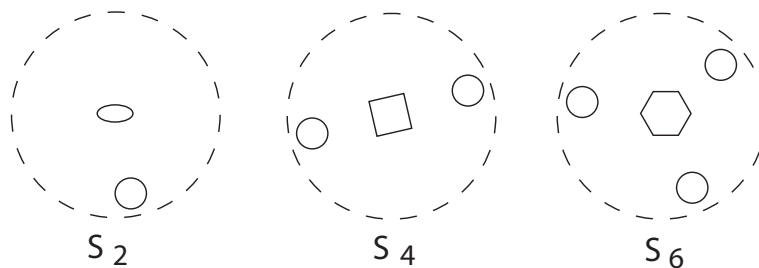
peg lengthwise in half). An anisotropic array of lattice sites may have this symmetry (or lack of symmetry).

Next, Fig. 2.2 shows the remaining uniaxial groups containing no mirror planes. Again, you may imagine that the \times marks are pegs sticking up from the plane of the disk. Alternatively, you may imagine that the disk has only one peg, and the additional \times marks show what happens to the peg under the actions of the group. Thus, for example, C_3 has a three-fold symmetry: rotations by $\pm 2\pi/3$ leave it invariant.

Next, we add a mirror plane, first in the plane of the “disk”, see Fig. 2.3. We have added two new features to our graphical notation: The outline of the disk has become solid, rather than dashed. This indicates that there is a mirror plane in the plane of the disk. Also, we have added small circle symbols (each here overlapping a \times , but this isn’t always the case). You may think of the circle as indicating a peg sticking out below the plane of the disk. Thus, the illustration for C_{1h} indicates symmetry with respect to a mirror reflection through the plane of the disk. The “ h ” in the Schönflies notation indicates a horizontal mirror plane, where “horizontal” is the plane of the disk.

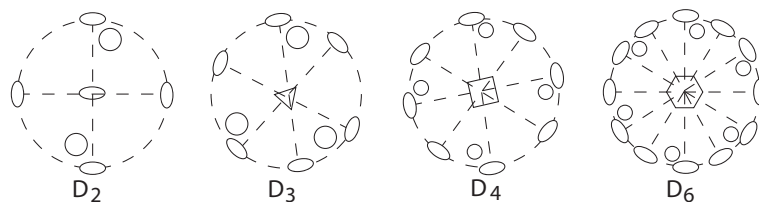
We get four more groups by taking away the horizontal mirror plane and adding a vertical mirror plane, see Fig. 2.4. Note that C_{1v} is the same as C_{1h} , since there is no uniquely defined principal axis. Also, note that adding one vertical mirror plane implies others in general. For example, in the case of C_{2v} , adding a mirror plane (M_y) in the $x-z$ plane gives one also in the $y-z$ plane (assuming the principal two-fold rotation axis is the z -axis), since:

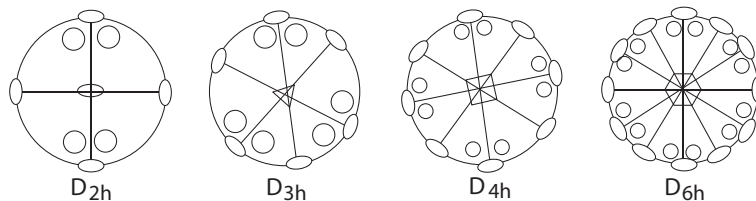
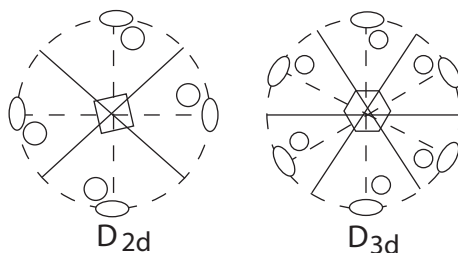
$$R_z(\pi)M_y = M_x. \quad (2.4)$$

Figure 2.4: The uniaxial groups C_{2v} , C_{3v} , C_{4v} , and C_{6v} .Figure 2.5: The uniaxial groups S_2 , S_4 , and S_6 .

We may also consider the improper rotations, starting with the simple inversion, S_2 . The inversion followed by a rotation by angle π is the same as a mirror, e.g., C_{1h} , hence does not generate a new group. The inversion followed by a rotation by $\pi/2$ generates S_4 . Ultimately, we end up with three new groups by considering the improper rotations with a single principal axis, see Fig. 2.5. Note that S_2 and S_6 contain the inversion symmetry, while S_4 does not. Also, be aware that the notation for these groups has nothing to do with the overlapping notation for the permutation groups.

This completes the “uniaxial” groups. We next consider adding a two-fold rotation axis perpendicular to the principal axis of C_n , obtaining the “dihedral” groups, D_n . These are shown in Fig. 2.6. Note that adding such a two-fold axis to C_1 just gives C_2 , as C_1 doesn’t have any well-defined axis in the first place.

Figure 2.6: The dihedral groups D_2 , D_3 , D_4 , and D_6 .

Figure 2.7: The dihedral groups D_{2h} , D_{3h} , D_{4h} , and D_{6h} .Figure 2.8: The dihedral groups D_{2d} and D_{3d} .

We may likewise add a two-fold rotation axis perpendicular to the principal axis of C_{nh} , obtaining the additional dihedral groups, D_{nh} , see Fig. 2.7. Adding such an axis to C_{1h} just gives C_{2v} , rather than a new group.

We may also add a two-fold axis perpendicular to the principal axis for the improper rotation groups S_n . In the case of S_2 , we obtain C_{2h} rather than a new group. Thus, we have two new dihedral groups, called D_{2d} and D_{3d} . They are graphed in Fig. 2.8.

This brings the total to 27 groups so far. There are an additional five groups, known as the “cubic groups” that do not have a principal axis with all other axes perpendicular to it. All of these remaining five have a three-fold axis equidistant from three mutually perpendicular two- or four-fold axes. The first group is the group of proper rotations of the tetrahedron (that is, those rotations which take a tetrahedron with some orientation into a tetrahedron with the same orientation, with indistinguishable faces). This group is labeled T .

The second of these groups is the full tetrahedral symmetry group, T_d , including mirror planes. The third group, T_h , is obtained by adding the inversion operation to T . Note that the tetrahedron is not invariant under the operations of this group. The fourth group, O , is the group of proper rotations of the octahedron. Equivalently, it is the group of proper rotations of the cube, noting that the faces of the octahedron may be identified with the vertices of a cube, and vice versa. The final group is the full symmetry group O_h of the octahedron, obtained by adding the inversion to O or to T_d .

2.3 Exercises

1. Give the multiplication table for D_{3h} .
2. List the classes of C_{6v} .
3. Consider the symmetry group, C_{4v} , of the square, consisting of the rotations about the axis perpendicular to the square, and reflections about the vertical, horizontal, and diagonal axes in the plane of the square (but no mirror plane in the plane of the square). List all of the group elements, and classify them into classes of equivalent elements. Find all subgroups and identify the invariant subgroups.
4. List the elements of the tetrahedral symmetry group T_d , and categorize by class.
5. We have looked at the permutation group S_4 and the tetrahedral symmetry group T_d . Show that these two groups are isomorphic, giving an explicit mapping between the elements of the two groups.

Chapter 3

Representation Theory

Groups may be very abstract objects and operations in general, and it would be convenient if we could always put them in some standard, equivalent form, and in particular a form that lends itself to easy manipulation. This is the motivation for the following discussion.

Def: Let (G, \circ) and $(H, *)$ be two groups. These groups are called *isomorphic* ($G \cong H$) if:

1. There exists a one-to-one mapping $\phi : G \rightarrow H$ from G onto H such that
2. $\phi(a \circ b) = \phi(a) * \phi(b), \forall a, b \in G$.

Note that (1) implies that G and H have the same order, and (2) implies that the multiplication tables of G and H are “identical”, up to relabeling of elements (as specified by the mapping ϕ). Statements about the structure of group G are equivalent to statements about the structure of group H , and we may choose to study either case for convenience.

A somewhat less rigid correspondence is given by:

Def: Let (G, \circ) and $(H, *)$ be two groups. A mapping ϕ from G into H is called a *homomorphism* if: $\phi(a \circ b) = \phi(a) * \phi(b), \forall a, b \in G$.

In this case, the mapping may be many-to-one. An extreme case occurs with H being a group of order one (the identity element), and all elements of G are mapped into this element of H .

Def: An isomorphism of a group into itself is called an *automorphism*. A homomorphism of a group into itself is called an *endomorphism*.

In physical applications, the most prevalent isomorphisms and homomorphisms of abstract groups are into matrix groups:

Def: A (matrix) *representation* of a group G is a group of matrices (with group multiplication given by matrix multiplication) obtained by a homomorphism of G into the set of $n \times n$ matrices. A matrix representation which is an isomorphism is called a *faithful representation* of the group.

It is perhaps worth demonstrating that if G is represented by non-singular matrices D , then $D(e) = I$ is the identity matrix: We start by noting that $D(e)^2 = D(ee) = D(e)$. Now,

$$\begin{aligned} D(g^{-1}) &= D(g^{-1})I \\ &= D(g^{-1})D(g)D(g)^{-1} \\ &= D(g^{-1}g)D(g)^{-1} \\ &= D(e)D(g)^{-1}. \end{aligned}$$

Multiplying by $D(e)$, we find, $D(e)D(g^{-1}) = D(e)D(g)^{-1}$. Multiply both sides now by $D(e)^{-1}$, to find that $D(g^{-1}) = D(g)^{-1}$. Thus, we see that, in a non-singular representation, the matrix representing the inverse of a group element is just the inverse of the matrix representing the original group element. In particular,

$$\begin{aligned} D(e) &= D(g)D(g^{-1}) \\ &= D(g)D(g)^{-1} \\ &= I. \end{aligned}$$

3.1 Poincaré Group

An important example of a group representation is the representation of the Poincaré group with a set of 5×5 matrices. The Poincaré group is also known as the inhomogeneous Lorentz group, and is denoted by \bar{L} . It is the group consisting of all homogeneous Lorentz transformations (velocity boosts, rotations, and reflections, including time-reversal), plus all translations in spacetime. The abstract group \bar{L} may be represented by the set of all 5×5 matrices of the form:

$$\Lambda(M, z) = \begin{pmatrix} & & & & z_1 \\ & & & & z_2 \\ & M & & & z_3 \\ & & & & z_4 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}, \quad (3.1)$$

where M is a 4×4 matrix which “preserves the invariant interval”¹ when multiplying 4-vectors, and $z = (z_1, z_2, z_3, z_4)$ is any element of R^4 .

¹The invariant interval (squared) between vectors a and b is $(a - b)^2 = (a_4 - b_4)^2 - (a_1 - b_1)^2 - (a_2 - b_2)^2 - (a_3 - b_3)^2$.

Thus, an inhomogeneous Lorentz transformation is a transformation of the form:

$$\{x'\} = \Lambda(M, z)\{x\} = \{Mx + z\}, \quad (3.2)$$

where we use the artifice for any four-vector x :

$$\{x\} \equiv \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ 1 \end{pmatrix}. \quad (3.3)$$

This permits us to express an inhomogeneous transformation as a linear transformation.

Our representation is actually isomorphic to \bar{L} . Note, however, that the $n \times n$ identity matrix (pick any n) is also a representation for \bar{L} , although no longer an isomorphism. In fact, the $n \times n$ unit matrix is a representation for any group, although “trivial”.

There are some important subgroups of the Poincaré group, such as:

1. The group Tr , of all pure translations in spacetime is a proper subgroup of \bar{L} . A representation for Tr is the set of matrices of the form:

$$\Lambda(I, z) = \begin{pmatrix} 1 & 0 & 0 & 0 & z_1 \\ 0 & 1 & 0 & 0 & z_2 \\ 0 & 0 & 1 & 0 & z_3 \\ 0 & 0 & 0 & 1 & z_4 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}, \quad (3.4)$$

2. The group L , of homogeneous Lorentz transformations is another subgroup of \bar{L} , with a representation:

$$\Lambda(M, 0) = \begin{pmatrix} & & & & 0 \\ & & & & 0 \\ & M & & & 0 \\ & & & & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}. \quad (3.5)$$

Note that both $\Lambda(M, 0)$ and M itself provide representations for L .

Intuitively, we might suppose that Tr is in fact an invariant subgroup of \bar{L} . For example, if we first perform a rotation, then do a translation, then “undo” the rotation, we think the overall result should be a translation. See Fig. 3.1. However, our intuition may become strained when we include boosts and reflections, so let us see whether we can make a convincing demonstration.

If Tr is an invariant subgroup of \bar{L} , we must show that:

$$\Lambda^{-1}(M, z')\Lambda(I, z)\Lambda(M, z') \in Tr, \quad \forall \Lambda(M, z') \in \bar{L}. \quad (3.6)$$

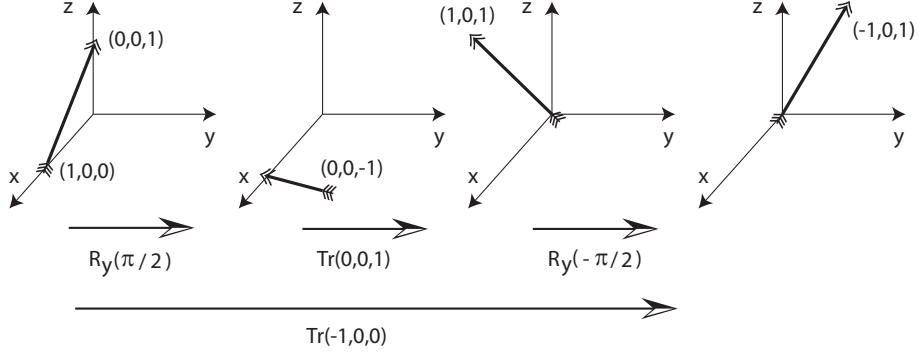


Figure 3.1: Illustration suggesting that $R_y(-\pi/2)Tr(0,0,1)R_y(\pi/2) = Tr(-1,0,0)$.

So far, we have avoided knowing the “multiplication table” for \bar{L} . But now life would be much easier if we knew it. So, what is

$$\Lambda(M', z')\Lambda(M'', z'')? \quad (3.7)$$

We could find the answer by considering the faithful representation:

$$\Lambda(M, z) = \begin{pmatrix} & & z_1 \\ & M & z_2 \\ & & z_3 \\ & & z_4 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}, \quad (3.8)$$

and seeing what ordinary matrix multiplication gives us. The reader is encouraged to try this.

However, it is perhaps more instructive to remember that the elements of \bar{L} are transformations in spacetime, and approach the question by looking at the action of $\Lambda \in \bar{L}$ on an arbitrary 4-vector. Thus, recalling that $\Lambda(M, z)\{x\} = \{Mx + z\}$, we have:

$$\begin{aligned} \Lambda(M', z')\Lambda(M'', z'')\{x\} &= \Lambda(M', z')\{M''x + z''\} \\ &= \{M'(M''x + z'') + z'\} \\ &= \Lambda(M'M'', M'z'' + z')\{x\}. \end{aligned} \quad (3.9)$$

This relation holds for any 4-vector x , hence we have the multiplication table:

$$\Lambda(M', z')\Lambda(M'', z'') = \Lambda(M'M'', M'z'' + z'). \quad (3.10)$$

To see whether Tr is an invariant subgroup of \bar{L} , we also need to know the inverse of $\Lambda \in \bar{L}$: What is $\Lambda^{-1}(M, z)$? Since the identity element is obviously

the transformation where we “do nothing”, $\Lambda(I, 0)$, we must find $\Lambda^{-1} \in \bar{L}$ such that

$$\Lambda^{-1}(M, z)\Lambda(M, z) = \Lambda(I, 0). \quad (3.11)$$

Since $\Lambda^{-1} \in \bar{L}$, we must be able to parameterize it with a matrix M' and a translation z' : $\Lambda^{-1}(M, z) = \Lambda(M', z')$, so that

$$\Lambda(M', z')\Lambda(M, z) = \Lambda(I, 0). \quad (3.12)$$

Use the multiplication table to obtain:

$$\begin{aligned} \Lambda(I, 0) &= \Lambda(M', z')\Lambda(M, z) \\ &= \Lambda(M'M, M'z + z'). \end{aligned} \quad (3.13)$$

Thus,

$$\Lambda^{-1}(M, z) = \Lambda(M^{-1}, -M^{-1}z). \quad (3.14)$$

We are finally ready to prove that Tr is an invariant subgroup of \bar{L} :

$$\begin{aligned} \Lambda^{-1}(M, z')\Lambda(I, z)\Lambda(M, z') &= \Lambda(M^{-1}, -M^{-1}z')\Lambda(M, z + z') \\ &= \Lambda(M^{-1}M, M^{-1}(z + z') - M^{-1}z') \\ &= \Lambda(I, M^{-1}z) \in Tr \quad \forall \Lambda(M, z') \in \bar{L}. \end{aligned} \quad (3.15)$$

Hence, Tr is an invariant subgroup of \bar{L} .

Given any element $\Lambda(M, 0) \in L$, we have an isomorphism of Tr into Tr :

$$\Lambda(I, z) \rightarrow \Lambda(I, M^{-1}z). \quad (3.16)$$

That is, we have an automorphism on Tr .

3.2 Regular Representation

For any finite group $G = \{g_1, g_2, \dots, g_n\}$, with multiplication table $g_i g_j = g_k$, we may construct an isomorphic representation by a set of $n \times n$ matrices. Consider the following expression:

$$g_i g_j = g_m \Delta_{ij}^m. \quad (3.17)$$

A formal sum over $m = 1, 2, \dots, n$ is implied here. In fact, only one term in the sum is non-zero, with

$$\Delta_{ij}^m = \delta_k^m, \quad (3.18)$$

where δ_k^m is the Kronecker delta, as follows from the group multiplication table (index k is a function of i and j). We have the theorem:

Theorem: The *regular representation*, formed by the matrices $(\Delta_i)_j^k = \Delta_{ij}^k$, $i = 1, 2, \dots, n$, forms an isomorphic representation of G .

Proof: To understand a bit better what is going on, note that the matrices consist entirely of zeros and ones, where the matrix representing group element $a \in G$ has ones in locations such that a times an element of G specified by the column index gives an element of G given by the row index. Consider now:

$$ag_k = g_m \Delta_{a_k}^m, \quad (3.19)$$

where we use the group element a also as its index in the set of elements. Notice that

$$(\Delta_a)_k^m = \delta_{a_k}^m, \quad (3.20)$$

where $g_{a_k} \equiv ag_k$ determines index a_k . Suppose $ab = c$, for $a, b, c \in G$. Let us check that this multiplication table is preserved by our representation:

$$\begin{aligned} (\Delta_a)_m^k (\Delta_b)_j^m &= \delta_{a_m}^k \delta_{b_j}^m \\ &= \delta_{a_{b_j}}^k \\ &= \delta_{c_j}^k \\ &= (\Delta_c)_j^k. \end{aligned} \quad (3.21)$$

This follows since

$$\begin{aligned} g_{a_{b_j}} &\equiv ag_{b_j} \\ &= abg_j \\ &= cg_j \\ &= g_{c_j}, \end{aligned} \quad (3.22)$$

that is, $a_{b_j} = c_j$. The remaining aspects of the proof are left to the reader.

We deal in the following with finite-dimensional representations.

3.3 Equivalence of Representations

Def: Two $n \times n$ matrix representations, D and D' , where $D(g)$ and $D'(g)$ are the matrices representing group element g in the two representations, are said to be *equivalent* if there exists a non-singular matrix S such that

$$D'(g) = S^{-1}D(g)S, \quad \forall g \in G. \quad (3.23)$$

It is readily checked that this defines a true equivalence relation; the reader is encouraged to do so. Note that the matrix S need not be a member of either representation. If D and D' are equivalent representations, we write $D \sim D'$.

A transformation of this form may be regarded as simply a change in basis for the vector space upon which our matrices operate. Hence, equivalent representations are identical as far as the intrinsic internal group structure is concerned. Presuming we are really interested in this intrinsic structure, we

would like to be able to concentrate on those statements which are independent of “coordinate” system. That is, we are interested in studying quantities which are invariant with respect to similarity transformations.

In principle, there are n such invariant quantities, corresponding to the n eigenvalues. However, we typically don’t need to study all n . In fact, just one invariant, the trace (sum of the eigenvalues), contains sufficient information for many purposes. Recall

$$\mathrm{Tr} [D(g)] = \sum_{i=1}^n D_{ii}(g). \quad (3.24)$$

This is invariant under similarity transformations:

$$\begin{aligned} \mathrm{Tr} [D'(g)] &= \mathrm{Tr} [S^{-1}D(g)S] \\ &= \mathrm{Tr} [SS^{-1}D(g)] \\ &= \mathrm{Tr} [D(g)]. \end{aligned} \quad (3.25)$$

3.4 Characters

The trace of a representation matrix plays a very important role, so it gets a special name:

Def: The trace of $D(g)$ is called the *character* of g in the representation D .

The character is usually denoted with the Greek letter chi:

$$\chi(g) = \mathrm{Tr} [D(g)]. \quad (3.26)$$

We have seen that equivalent representations have the same set of characters. We have the further fact:

Theorem: Given a representation D , any two group elements belonging to the same class have the same character.

The proof of this is straightforward: Suppose that g_1 and g_2 belong to the same class in G . Then there exists an element $h \in G$ such that

$$h^{-1}g_1h = g_2. \quad (3.27)$$

Thus, in representation D , we must have:

$$D(h^{-1})D(g_1)D(h) = D(g_2). \quad (3.28)$$

Note also that $D(h)D(h^{-1}) = D(e)$ (we are not assuming that $D(h)^{-1}$ exists here, as our representation could be singular). Thence,

$$\begin{aligned} \chi(g_2) &= \mathrm{Tr} [D(g_2)] \\ &= \mathrm{Tr} [D(h^{-1})D(g_1)D(h)] \end{aligned}$$

$$\begin{aligned}
&= \operatorname{Tr} [D(h)D(h^{-1})D(g_1)] \\
&= \operatorname{Tr} [D(e)D(g_1)] \\
&= \operatorname{Tr} [D(eg_1)] \\
&= \operatorname{Tr} [D(g_1)] \\
&= \chi(g_1).
\end{aligned} \tag{3.29}$$

Thus, given a representation D , we can completely specify the character structure by evaluating the character for one member of each class – the character is a “class function”.

We are often interested in more than one representation for a given group G . In this case, we can add labels to the representations to distinguish them, for example, $D^{(a)}, D^{(b)}, \dots$. We similarly label the characters, e.g., $\chi^{(a)}, \chi^{(b)}, \dots$. If we have a class, say C_i , in representation (a) , we may refer to the “character of the class” as $\chi^{(a)}(C_i)$.

3.5 Unitary Representations

Unitary matrices are especially nice. They preserve the lengths of vectors when operating on a complex vector space. The inverse is easy to compute: If U is a unitary matrix, then $U^{-1} = U^\dagger$, where the \dagger means to take the transpose of the complex conjugate matrix. Thus, it is quite nice to learn that:

Theorem: If G is a finite group, then every non-singular representation (that is, representation by non-singular matrices) is equivalent to a unitary representation.

Thus, at least for finite groups, it is sufficient to consider representations by unitary matrices. The proof of this theorem is instructive:

Proof: We suppose we are given a (non-singular, but possibly non-unitary) $n \times n$ representation, D , of G . We may regard an element of the representation as a linear operator on an n -dimensional vector space. Define a scalar product on the vector space by

$$(x, y) \equiv \sum_{i=1}^n x_i^* y_i, \tag{3.30}$$

where x and y are any pair of vectors.

Suppose that we have a matrix U with the property that:

$$(Ux, Uy) = (x, y), \quad \forall x, y. \tag{3.31}$$

That is, U “preserves the scalar product”. Let us see what this condition requires for U :

$$(Ux, Uy) = \sum_{i=1}^n (Ux)_i^* (Uy)_i$$

$$\begin{aligned}
&= \sum_{i=1}^n \left[\sum_{j=1}^n U_{ij}^* x_j^* \right] \left[\sum_{k=1}^n U_{ik} y_k \right] \\
&= \sum_{i=1}^n \sum_{j=1}^n \sum_{k=1}^n (U^\dagger)_{ji} U_{ik} x_j^* y_k \\
&= \sum_{j=1}^n \sum_{k=1}^n (U^\dagger U)_{jk} x_j^* y_k \\
&= \sum_{j=1}^n x_j^* y_j \quad (\text{by assumption}). \tag{3.32}
\end{aligned}$$

But x and y are arbitrary vectors, hence we must have

$$(U^\dagger U)_{jk} = \delta_{jk}, \tag{3.33}$$

or $U^\dagger = U^{-1}$, that is U must be a unitary matrix.

We wish to show that our given representation, D , is equivalent to some unitary representation, say D' . For this to be true there must exist a transformation T such that

$$D' = T^{-1}DT, \tag{3.34}$$

where we mean that this transformation is applied to every element of the representation. If we can find a transformation T such that

$$(D'(a)x, D'(a)y) = (x, y), \quad \forall x, y \text{ and } \forall a \in G, \tag{3.35}$$

then by the above discussion we will have found a unitary representation.

We construct a suitable transformation by the following technique, which introduces an approach that will be useful elsewhere as well. Let g be the order of G . Define an ‘‘average’’ over the elements of the group:

$$\{x, y\} \equiv \frac{1}{g} \sum_{a \in G} (D(a)x, D(a)y). \tag{3.36}$$

In a sense $\{x, y\}$ is the average scalar product over all group elements, with respect to representation D , acting on the vectors x and y . Notice that $\{x, y\}$ itself defines a scalar product, since

1. $\{x, x\} \geq 0$ and $\{x, x\} = 0$ if and only if $x = 0$;
2. $\{x, y\} = \{y, x\}^*$;
3. $\{x, cy\} = c\{x, y\}$;
4. $\{x_1 + x_2, y\} = \{x_1, y\} + \{x_2, y\}$.

We remark that it is the first property that requires D to be a non-singular representation.

Now let b be any element of G , and consider:

$$\begin{aligned}
 \{D(b)x, D(b)y\} &= \frac{1}{g} \sum_{a \in G} (D(a)D(b)x, D(a)D(b)y) \\
 &= \frac{1}{g} \sum_{a \in G} (D(ab)x, D(ab)y) \\
 &= \frac{1}{g} \sum_{a \in G} (D(a)x, D(a)y) \\
 &= \{x, y\}.
 \end{aligned} \tag{3.37}$$

The third step is valid because the multiplication table is a Latin square: Summing products ab over all $a \in G$ is the same as summing ab over all $ab \in G$; the only difference is the ordering in the sum. This “invariance” of the sum is a property that will often come in handy.

Thus, we have shown that $D(b)$ is a unitary operator (that is, it preserves the scalar product) with respect to the $\{, \}$ scalar product. It is not necessarily a unitary operator with respect to the $(,)$ scalar product, that is D is not necessarily a unitary matrix representation. Somehow, we would like to find a transformation which takes this desired unitary property under the $\{, \}$ scalar product back into the $(,)$ scalar product. In other words, we wish to transform from a basis suitable for $(,)$ to one suitable for $\{, \}$.

Consider a set of n orthonormal vectors with respect to $(,)$:

$$(u_i, u_j) = \delta_{ij}, \tag{3.38}$$

and a set orthonormal with respect to $\{, \}$:

$$\{v_i, v_j\} = \delta_{ij}. \tag{3.39}$$

Let T be the transformation operator which takes u 's to v 's:

$$v_i = Tu_i. \tag{3.40}$$

An arbitrary vector x may be expanded in the u basis as:

$$x = \sum_{i=1}^n x_i u_i, \tag{3.41}$$

where the x_i are the components in the u basis. Consider the transformed vector Tx :

$$Tx = T \sum_{i=1}^n x_i u_i = \sum_{i=1}^n x_i T u_i = \sum_{i=1}^n x_i v_i. \tag{3.42}$$

Thus, the components of the transformed vector in the new basis v are the same as the components of the un-transformed vector in the old basis u . We have

$$\begin{aligned} \{Tx, Ty\} &= \sum_{i=1}^n \sum_{j=1}^n x_i^* y_j \{v_i, v_j\} \\ &= \sum_{i=1}^n x_i^* y_i \\ &= (x, y). \end{aligned} \tag{3.43}$$

Now consider the representation

$$D' \equiv T^{-1}DT, \tag{3.44}$$

which is equivalent to D . Evaluate the scalar product:

$$\begin{aligned} (D'(a)x, D'(a)y) &= (T^{-1}D(a)Tx, T^{-1}D(a)Ty) \\ &= \{D(a)Tx, D(a)Ty\} \quad (\text{since } \{x, y\} = (T^{-1}x, T^{-1}y)) \\ &= \{Tx, Ty\} \quad (\text{since } D(a) \text{ is a unitary operator wrt } \{, \}) \\ &= (x, y). \end{aligned} \tag{3.45}$$

Hence, $D'(a)$, for any $a \in G$, is a unitary operator with respect to the $(,)$ scalar product. Therefore, $D'(a)$ is a unitary matrix. This completes the proof.

3.6 Reducible and Irreducible Representations

Def: Given any two representations, $D^{(1)}$ and $D^{(2)}$, of a group G , we may construct a new representation simply by forming the matrix direct sum:

$$D(g) = \begin{pmatrix} D^{(1)}(g) & 0 \\ 0 & D^{(2)}(g) \end{pmatrix} \equiv D^{(1)}(g) \oplus D^{(2)}(g). \tag{3.46}$$

A representation which is equivalent to a representation of this form is called *fully reducible*. A representation which is equivalent to a representation of the form:

$$\begin{pmatrix} D^{(1)}(g) & A(g) \\ 0 & D^{(2)}(g) \end{pmatrix} \tag{3.47}$$

is called *reducible*. A representation which is not reducible is called *irreducible*.

Note that the definition of reducibility is equivalent to the statement that there exists a proper invariant subspace, V_1 , in the Euclidean space operated on by the representation. The further restriction of full reducibility is equivalent to the statement that the orthogonal complement of V_1 is also an invariant subspace.

Theorem: If a reducible representation, $D(G)$, is equivalent to a unitary representation, then $D(G)$ is fully reducible.

Proof: Let $U(G)$ be a unitary representation which is equivalent to $D(G)$, and let V be the Euclidean space operated on by U . By assumption, there exists a proper invariant subspace $V_1 \subset V$ under the actions of $U(G)$. Define an orthonormal basis for V , consisting of the vectors $\{\mathbf{e}_i, i = 1 \dots n\}$, such that the first n_1 basis vectors span V_1 . Let V_2 be the orthogonal complement of V_1 , spanned by basis vectors $\{\mathbf{e}_i, i = n_1+1 \dots n\}$. We demonstrate that V_2 is also an invariant subspace under $U(G)$. Since $U(G)$ is unitary, we have, for any $g \in G$:

$$(U(g)\mathbf{e}_i, U(g)\mathbf{e}_j) = (\mathbf{e}_i, \mathbf{e}_j) \quad (3.48)$$

Suppose $\mathbf{e}_j \in V_1$ and $\mathbf{e}_i \in V_2$. Then $U(g)\mathbf{e}_j \in V_1$, since V_1 is invariant. Further, since $(\mathbf{e}_i, \mathbf{e}_j) = 0$, $U(g)\mathbf{e}_i$ is orthogonal to any vector in V_1 , since we could pick any $\mathbf{e}_j \in V_1$, and the set of all vectors $\{U(g)\mathbf{e}_j | \mathbf{e}_j \in V_1\}$ spans V_1 . Thus, $U(g)\mathbf{e}_i$ is in V_2 . Therefore, V_2 is also an invariant subspace under $U(G)$. QED

Since we will be dealing here with representations which are equivalent to unitary representations, we may assume that our representations are either fully reducible or irreducible. In our study of group structure, two equivalent irreducible representations are not counted as distinct.

The irreducible representations (or “irreps”, for short) are important because an arbitrary representation can be expressed as a *direct sum* of irreps. For illustration,

$$D(g) = \begin{pmatrix} D^{(1)}(g) & & & 0 \\ & D^{(2)}(g) & & \\ & & D^{(3)}(g) & \\ 0 & & & D^{(3)}(g) \end{pmatrix} = D^{(1)}(g) \oplus D^{(2)}(g) \oplus 2D^{(3)}(g). \quad (3.49)$$

Note that the reduction of a representation to irreps may include some irreps multiple times.

There are some important properties of irreps, under the name of “Schur’s lemmas”:

Theorem: If D and D' are irreps of G , and if matrix A satisfies

$$D(g)A = AD'(g), \quad \forall g \in G, \quad (3.50)$$

then either $D \sim D'$ or $A = 0$.

Proof: Note that A may not be a square matrix, as the dimensions of representations D and D' could be different. We may consider D and D' to be sets of operators on vector spaces V and V' , respectively. The range of A is

$$R_A = \{x \in V : x = Ax', \text{ where } x' \in V'\}. \quad (3.51)$$

R_A is an invariant subspace of V , since

$$\begin{aligned} D(g)x &= D(g)Ax', \quad \text{for any } x \in R_A \\ &= AD'(g)x', \quad (\text{by assumption}) \\ &\in R_A, \quad \text{since } D'(g)x' \in V'. \end{aligned} \quad (3.52)$$

But since D is an irrep, this means that either $R_A = V$ or $R_A = \{0\}$ (that is, $A = 0$).

Now consider

$$N' \equiv \{x' \in V' : Ax' = 0\}. \quad (3.53)$$

This is referred to as the *null space* of A in V' . It is an invariant subspace of D' in V' since, if $x' \in N'$, then

$$AD'(g)x' = D(g)Ax' = D(g)0 = 0. \quad (3.54)$$

But D' is irreducible, therefore either $N' = V'$ (hence $A = 0$) or $N' = \{0\}$. If $N' = \{0\}$, then the equation $Ax' = Ay'$ implies $x' = y'$, and the mapping A is one-to-one and onto.

We have so far shown that either A provides an isomorphism between V and V' or $A = 0$. If an isomorphism, then A is invertible, and

$$D(g) = AD'(g)A^{-1}, \quad \forall g \in G. \quad (3.55)$$

That is, D and D' are equivalent representations in this case. We remark that two irreps can be equivalent only if they have the same dimension.

Theorem: D is an irrep if and only if, given matrix A such that

$$AD(g) = D(g)A, \quad \forall g \in G, \quad (3.56)$$

then A is a constant times the identity matrix.

Proof: Consider the eigenvalue equation

$$Ax = \lambda x, \quad (3.57)$$

where $x \in V$. If x is an eigenvector with eigenvalue λ , then

$$AD(g)x = D(g)Ax = \lambda D(g)x \quad \forall g \in G. \quad (3.58)$$

That is, $D(g)x$ is also an eigenvector of A belonging to eigenvalue λ . The subspace of eigenvectors belonging to λ is invariant with respect to D . Hence, there are three possibilities: either D is reducible, or this subspace is V , or the subspace consists only of $x = 0$. If the subspace is V , then A has only one eigenvalue, and $A = \lambda I$. If the subspace is $x = 0$, then $A = 0$.

Note that the second theorem provides a test for irreducibility: Given a representation D , we look for a matrix A such that $AD(g) = D(g)A$ for all $g \in G$, and see whether it must be true that $A = \lambda I$. For example, consider an abelian group G . Certainly any one-dimensional representation is irreducible, since any number A is a constant times 1. Suppose we have a representation of dimension larger than one. Since G is abelian, we must have

$$D(a)D(b) = D(b)D(a), \quad \forall a, b \in G. \quad (3.59)$$

Thus, pick $A = D(a)$ for some element a of G . Then $AD(b) = D(b)A$ for all $b \in G$. But if $A = \lambda I$ for every $a \in G$, then D is reducible. Suppose there exists $a \in G$ such that $A = D(a) \neq \text{constant} \times I$. Then by the lemma, D must be reducible. We have just shown that all irreps of an abelian group must be one-dimensional.

To pursue this example further with a concrete case, consider the abelian group $Z_5 = \{0, 1, 2, 3, 4\}$ (with group multiplication given by addition modulo five). The inequivalent irreducible matrix representations are shown in Table 3.1; note that they are all one-dimensional.

Table 3.1: The irreducible representations of Z_5 .

$g \setminus \text{irrep}$	$D^{(1)}$	$D^{(2)}$	$D^{(3)}$	$D^{(4)}$	$D^{(5)}$
0	1	1	1	1	1
1	1	$e^{2\pi i/5}$	$e^{4\pi i/5}$	$e^{6\pi i/5}$	$e^{8\pi i/5}$
2	1	$e^{4\pi i/5}$	$e^{8\pi i/5}$	$e^{2\pi i/5}$	$e^{6\pi i/5}$
3	1	$e^{6\pi i/5}$	$e^{2\pi i/5}$	$e^{8\pi i/5}$	$e^{4\pi i/5}$
4	1	$e^{8\pi i/5}$	$e^{6\pi i/5}$	$e^{4\pi i/5}$	$e^{2\pi i/5}$

3.7 Orthogonality Theorems

The Schur's lemmas are also useful in proving the very important "orthogonality relations". These theorems are important tools in determining the essential structure of the irreps for a group.

The first theorem may appropriately be referred to as the "general orthogonality relation".

Theorem: If $D^{(i)}$ and $D^{(j)}$ are irreps, where $i \neq j$ label inequivalent irreps, then

$$\sum_{g \in G} D^{(i)}(g)_{\mu\nu} D^{(j)}(g^{-1})_{\alpha\beta} = \frac{h}{\ell_i} \delta_{ij} \delta_{\alpha\nu} \delta_{\beta\mu}, \quad (3.60)$$

where h is the order of the group, and ℓ_i is the dimension of representation $D^{(i)}$.

Proof: Let A be any $\ell_i \times \ell_j$ matrix (ℓ_i rows and ℓ_j columns). Define

$$M_A \equiv \sum_g D^{(i)}(g)AD^{(j)}(g^{-1}). \quad (3.61)$$

Note the use once again of the technique of summing over the group. Now consider

$$\begin{aligned} D^{(i)}(b)M_A &= \sum_g D^{(i)}(b)D^{(i)}(g)AD^{(j)}(g^{-1}) \\ &= \sum_g D^{(i)}(b)D^{(i)}(g)AD^{(j)}(g^{-1})D^{(j)}(b^{-1})D^{(j)}(b) \\ &= \left[\sum_g D^{(i)}(bg)AD^{(j)}((bg)^{-1}) \right] D^{(j)}(b) \\ &= M_A D^{(j)}(b), \quad (\text{by the rearrangement lemma}). \end{aligned} \quad (3.62)$$

By Schur's lemma, either $D^{(i)} \sim D^{(j)}$ (that is, $i = j$) or $M_A = 0$.

If $i \neq j$, then $M_A = 0$. Picking A such that $A_{\nu\alpha} = 1$ and all other elements are zero, we obtain:

$$\sum_g D^{(i)}(g)_{\mu\nu} D^{(j)}(g^{-1})_{\alpha\beta} = 0, \quad \forall \mu, \nu, \alpha, \beta. \quad (3.63)$$

If $i = j$, then we may simplify the notation, letting $D^{(i)} = D^{(j)} = D$. We have:

$$D(b)M_A = M_A D(b), \quad \forall b \in G. \quad (3.64)$$

By the other Schur's lemma, this means that M_A is a multiple of the identity:

$$\sum_g D(g)AD(g^{-1}) = \lambda_A I, \quad (3.65)$$

where the value of the multiple depends on A . Pick matrix A so that $A_{\nu\alpha} = 1$, with all other elements zero. Then

$$\sum_g D(g)_{\mu\nu} D(g^{-1})_{\alpha\beta} = \delta_{\mu\beta} \lambda_{\nu\alpha}, \quad (3.66)$$

where the Kronecker delta gives the components of the identity matrix, and $\lambda_{\nu\alpha}$ is the constant multiplying the identity.

To determine $\lambda_{\nu\alpha}$, set $\mu = \beta$ and sum over μ :

$$\begin{aligned} \sum_{\mu=1}^{\ell_i} \lambda_{\nu\alpha} &= \sum_g \sum_{\mu} D(g)_{\mu\nu} D(g^{-1})_{\alpha\mu} \\ \ell_i \lambda_{\nu\alpha} &= \sum_g [D(g^{-1})D(g)]_{\alpha\nu} \end{aligned}$$

$$\begin{aligned}
&= \sum_g D(g^{-1})_{\alpha\nu} \\
&= h\delta_{\alpha\nu},
\end{aligned} \tag{3.67}$$

where we have used the fact that any irrep that is equivalent to a unitary representation is non-singular, and hence $D(e) = I$. This completes the proof.

Our theorem holds whether the irreps are unitary or not. For a unitary representation we can restate the general orthogonality relation in a more convenient form. For a unitary irrep, we have

$$D^{(j)}(g^{-1}) = D^{(j)}(g)^{-1} = D^{(j)}(g)^\dagger. \tag{3.68}$$

For unitary irreps we can thus rewrite the general orthogonality relation as

$$\sum_{g \in G} D^{(i)}(g)_{\mu\nu} D^{(j)}(g)_{\beta\alpha}^* = \frac{h}{\ell_i} \delta_{ij} \delta_{\alpha\nu} \delta_{\beta\mu}, \tag{3.69}$$

In obtaining some consequences of this theorem, it is useful to regard the group G as generating an h -dimensional vector space, and to interpret $D^{(i)}(g)_{\mu\nu}$ as the “ g th” component of a vector in this space. The labels i, μ, ν identify a particular vector. The theorem tells us that all such distinct vectors in the space are orthogonal. Let us count how many distinct vectors there are: For a given representation, there are ℓ_i^2 pairs μ, ν , so the number of distinct vectors is

$$\sum_{i=1}^{n_r} \ell_i^2,$$

where n_r is the number of (inequivalent) irreducible representations. Since it is an h -dimensional space, we cannot have more than h linearly independent vectors, hence,

$$\sum_{i=1}^{n_r} \ell_i^2 \leq h. \tag{3.70}$$

In fact, we will soon see that equality holds.²

This equality is a very useful fact to know in approaching the problem of finding irreducible representations. For example, suppose we have a group of order 6. In this case the possible dimensions of the irreducible representations are: (i) $\{\ell_i\} = \{1, 1, 1, 1, 1, 1\}$, corresponding to an abelian group, isomorphic to Z_6 ; and (ii) $\{\ell_i\} = \{1, 1, 2\}$, which can be shown to correspond to the lowest-order non-abelian group. There are no other possibilities for a group of order 6.

The general orthogonality relation yields some subsidiary orthogonality relations for characters, which are very important in evaluating and using character tables.

²We will shortly prove equality by considering the identity element in the regular representation, and showing that its reduction into irreps is such that each irrep occurs in the regular representation a number of times that is equal to the dimension of the irrep.

Theorem: (*First Orthogonality Relation*) Given a group G of order h , with n_c classes and N_k elements in class k , then for unitary irreps $D^{(i)}$ and $D^{(j)}$:

$$\sum_{k=1}^{n_c} \chi^{(i)}(C_k)^* \chi^{(j)}(C_k) N_k = h \delta_{ij}. \quad (3.71)$$

Proof: Start with the general orthogonality relation for irreducible (unitary) representations $D^{(i)}$ and $D^{(j)}$:

$$\sum_{g \in G} D^{(i)}(g)_{\mu\nu}^* D^{(j)}(g)_{\alpha\beta} = \frac{h}{\ell_i} \delta_{ij} \delta_{\mu\alpha} \delta_{\nu\beta}.$$

We are interested in characters, so let $\mu = \nu, \alpha = \beta$, and sum over μ and α :

$$\begin{aligned} \sum_{g \in G} \chi^{(i)}(g)^* \chi^{(j)}(g) &= \frac{h}{\ell_i} \delta_{ij} \sum_{\mu=1}^{\ell_i} \sum_{\alpha=1}^{\ell_i} \delta_{\mu\alpha} \delta_{\mu\alpha} \\ &= h \delta_{ij}. \end{aligned}$$

We complete the proof by replacing the summation $\sum_{g \in G}$ with $\sum_{k=1}^{n_c} N_k$. QED

As with the general orthogonality relation, we may make a geometrical interpretation: Distinct vectors in a space of dimension equal to the number of classes (n_c) are orthogonal (with respect to “weight” N_k). But now the distinct vectors are labelled only by the index (i), and so there are only n_r (the number of irreducible representations) of them. Since the number of distinct vectors cannot exceed the dimension of the space, we have $n_r \leq n_c$.

Our first orthogonality relation tells us that, for irreducible representations $D^{(i)}$, the vectors,

$$\chi^{(i)} = \left(\chi^{(i)}(C_1), \chi^{(i)}(C_2), \dots, \chi^{(i)}(C_{n_c}) \right), \quad i = 1, \dots, n_r, \quad (3.72)$$

form a set of n_r orthogonal vectors (with respect to weight N_k) and hence span an n_r -dimensional subspace of an n_c -dimensional space. Note that the weight N_k poses no essential difficulty, since we could always absorb it into the definition of the vectors if we choose: $\chi^{(i)}(C_k) \rightarrow \chi^{(i)}(C_k) \sqrt{N_k}$.

An arbitrary vector in our subspace may be expanded according to:

$$\chi = \sum_{i=1}^{n_r} a_i \chi^{(i)}. \quad (3.73)$$

In fact, the character of an arbitrary representation may be so expanded, since

$$D = \oplus_{i=1}^{n_r} a_i D^{(i)}, \quad (3.74)$$

which, upon taking the trace of both sides, yields Eq. 3.73. By components, this is:

$$\chi(C_k) = \sum_{i=1}^{n_r} a_i \chi^{(i)}(C_k). \quad (3.75)$$

We can define the inner product between any two vectors by:

$$\lambda \cdot \chi = \sum_{k=1}^{n_c} \lambda(C_k) \chi(C_k)^* N_k. \quad (3.76)$$

To find the expansion coefficients, a_i , take:

$$\begin{aligned} \chi \cdot \chi^{(j)} &= \sum_{i=1}^{n_r} a_i \chi^{(i)} \cdot \chi^{(j)} \\ &= \sum_{i=1}^{n_r} a_i \sum_{k=1}^{n_c} \chi^{(i)}(C_k) \cdot \chi^{(j)}(C_k)^* N_k \\ &= \sum_{i=1}^{n_r} a_i h \delta_{ij} \quad \text{first orthogonality relation} \\ &= a_j h. \end{aligned} \quad (3.77)$$

Thus, we have:

$$a_i = \frac{1}{h} \sum_{k=1}^{n_c} N_k \chi(C_k) \chi^{(i)}(C_k)^*. \quad (3.78)$$

This permits us to prove the following:

Theorem: In the regular representation D of a group of order h , each irreducible representation appears exactly ℓ_i times. Furthermore,

$$\sum_{i=1}^{n_r} \ell_i^2 = h. \quad (3.79)$$

Proof: Recall that the regular representation consists of the matrices (with k, j labelling components):

$$\{\Delta_{ij}^k, i = 1, \dots, h\},$$

where, if $g_i g_j = g_k$ then

$$\Delta_{ij}^m = \begin{cases} 1 & m = k \\ 0 & \text{otherwise.} \end{cases}$$

Thus,

$$\chi(g) = \begin{cases} h & g = e \\ 0 & \text{otherwise,} \end{cases} \quad (3.80)$$

since the regular representation of the identity is the $h \times h$ identity matrix, and if $g \neq e$, then all diagonal elements of the regular representation are zero, by the fact that if $gf = f$, then $g = e$.

Now consider the expansion of the regular representation in terms of irreducible representations:

$$\chi = \sum_{i=1}^{n_r} a_i \chi^{(i)}.$$

Using

$$a_i = \frac{1}{h} \sum_{k=1}^{n_c} \chi(C_k) \chi^{(i)}(C_k)^* N_k,$$

we find that $a_i = \ell_i$, because the irreducible representation of the identity is the $\ell_i \times \ell_i$ unit matrix. Hence, each irreducible representation occurs exactly ℓ_i times in the regular representation.

Finally, since $h = \chi(e) = \sum_{i=1}^{n_r} \ell_i \chi^{(i)}(e)$, we find $\sum_{i=1}^{n_r} \ell_i^2 = h$, which completes the proof. QED

We are now ready to obtain the “second orthogonality relation”:

Theorem: (*Second Orthogonality Relation*) Given a group G of order h , with n_r irreducible unitary representations, we have:

$$\sum_{i=1}^{n_r} \chi^{(i)}(C_k)^* \chi^{(i)}(C_m) = \frac{h}{N_k} \delta_{km}, \quad (3.81)$$

and $n_r = n_c$.

Proof: From the general orthogonality relation, we have $\sum_{i=1}^{n_r} \ell_i^2 = h$ orthonormal (up to a factor of h/ℓ_i) vectors labelled by i, μ, ν , with h components labelled by g . Since there are h vectors, and the space is h -dimensional, this is a complete orthonormal set, which we can express by:

$$\sum_{i=1}^{n_r} \sum_{\mu=1}^{\ell_i} \sum_{\nu=1}^{\ell_i} \frac{\ell_i}{h} D^{(i)}(g)_{\mu\nu}^* D^{(i)}(g')_{\mu\nu} = \delta_{gg'}. \quad (3.82)$$

That is, the sum of the projection operators onto each of the orthogonal directions is the identity matrix. Sum this expression over $g \in C_k$, and $g' \in C_m$ to obtain:

$$\sum_{i=1}^{n_r} \sum_{\mu=1}^{\ell_i} \sum_{\nu=1}^{\ell_i} \frac{\ell_i}{h} \sum_{g \in C_k} D^{(i)}(g)_{\mu\nu}^* \sum_{g' \in C_m} D^{(i)}(g')_{\mu\nu} = \delta_{km} N_k. \quad (3.83)$$

To sum the left-hand side we use:

$$\sum_{g \in C_k} D^{(i)}(g) = \frac{N_k}{\ell_i} \chi^{(i)}(C_k) I. \quad (3.84)$$

Let us quickly demonstrate this. Let

$$S \equiv \sum_{g \in C_k} D^{(i)}(g) \quad (3.85)$$

Consider, for any $a \in G$:

$$\begin{aligned}
D^{(i)}(a^{-1})SD^{(i)}(a) &= \sum_{g \in C_k} D^{(i)}(a^{-1})D^{(i)}(g)D^{(i)}(a) \\
&= \sum_{g \in C_k} D^{(i)}(a^{-1}ga) \\
&= S,
\end{aligned} \tag{3.86}$$

where the final step follows because $a^{-1}ga \in C_k$ and the sum is the same for any $a \in g$. Thus, $SD^{(i)}(a) = D^{(i)}(a)S$ for all $a \in G$, and by Schur's lemma S must therefore be a multiple of the identity. Deriving the constant is left to the reader. We thus obtain:

$$\begin{aligned}
\sum_{i=1}^{n_r} \sum_{\mu=1}^{\ell_i} \sum_{\nu=1}^{\ell_i} \frac{\ell_i}{h} \sum_{g \in C_k} D^{(i)}(g)_{\mu\nu}^* \sum_{g' \in C_m} D^{(i)}(g')_{\mu\nu} &= \sum_{i=1}^{n_r} \sum_{\mu=1}^{\ell_i} \sum_{\nu=1}^{\ell_i} \frac{\ell_i}{h} \frac{N_k N_m}{\ell_i^2} \chi^{(i)}(C_k)^* \chi^{(i)}(C_m) \delta_{\mu\nu} \\
&= \sum_{i=1}^{n_r} \sum_{\mu=1}^{\ell_i} \frac{N_k N_m}{h \ell_i} \chi^{(i)}(C_k)^* \chi^{(i)}(C_m) \\
&= \sum_{i=1}^{n_r} \frac{N_k N_m}{h} \chi^{(i)}(C_k)^* \chi^{(i)}(C_m).
\end{aligned}$$

Substituting into Eq. 3.83, this gives the desired orthogonality relation, Eq. 3.81.

Now, we can interpret Eq. 3.81 as stating that vectors in a n_c -dimensional subspace of an n_r -dimensional space are orthogonal. Hence, $n_r \geq n_c$. But we already have $n_c \geq n_r$, from our discussion following the first orthogonality relation, hence $n_r = n_c$. QED

These theorems are of great help in reducing the effort required to construct and check character tables, which we discuss next.

3.8 Character Tables

If a group G has classes C_1, C_2, \dots, C_{n_c} , then it must have n_c irreducible representations $D^{(1)}, D^{(2)}, \dots, D^{(n_c)}$, with characters $\chi^{(1)}(C_k), \chi^{(2)}(C_k), \dots, \chi^{(n_c)}(C_k)$, $k = 1, \dots, n_c$. We can summarize this in a *character table*, see Table 3.2.

There are several useful things to note concerning a character table:

1. It is a square table, with $n_c = n_r$ rows and $n_c = n_r$ columns.
2. The rows must be orthogonal (remembering to take the complex conjugate of one of them), from the second orthogonality relation.
3. The columns must be orthogonal, with N_k as weighting factors, according to the first orthogonality relation (again remembering complex conjugates).

Table 3.2: Skeleton of a character table.

N_k	$\ell_i \rightarrow$ class \downarrow ; irrep \rightarrow	$\ell_i = 1$ $\chi^{(1)}$	ℓ_2 $\chi^{(2)}$	\cdots \cdots	ℓ_{n_c} $\chi^{(n_c)}$
$N_1 = 1$	$C_1 = \{e\}$	$\ell_1 = 1$	ℓ_2	\cdots	ℓ_{n_c}
N_2	C_2	1	\cdots	\cdots	\cdots
\vdots	\vdots	\vdots	\vdots	\vdots	\vdots
N_{n_c}	C_{n_c}	1	\cdots	\cdots	\cdots

4. By convention, we let C_1 be the class consisting of the identity element. In every irrep the matrix for the identity is the identity matrix. Therefore:

$$\chi^{(k)}(C_1) = \ell_k, \quad (3.87)$$

where ℓ_k is the dimension of irrep k .

5. As demonstrated earlier, we must have

$$\sum_{i=1}^{n_c} \ell_i^2 = h. \quad (3.88)$$

6. The simplest representation of any group is to represent every element by the number one (the “unit” or “identity” representation). This is an irrep, which we by convention here denote $D^{(1)}$. Then the first column of the character table is a string of ones.

Various other facts may be derived and used, but this set is already quite powerful in reducing the amount of work required to construct the character table for a group.

3.9 Decomposition of Reducible Representations

Suppose that we have a representation of a group, which may be reducible. If we have found the character table we may decompose our representation into a direct sum of irreps:

We start by writing the decomposition as:

$$D = a_1 D^{(1)} \oplus a_2 D^{(2)} \oplus \cdots \oplus a_{n_r} D^{(n_r)}, \quad (3.89)$$

where n_r is the number of irreps, and the a_i are non-negative integers to be determined. Noting that characters are just traces, we have that the character for class C_k must be:

$$\begin{aligned} \chi(C_k) &= \text{Tr}[D(g \in C_k)] \\ &= \sum_{i=1}^{n_r} a_i \chi^{(i)}(C_k). \end{aligned} \quad (3.90)$$

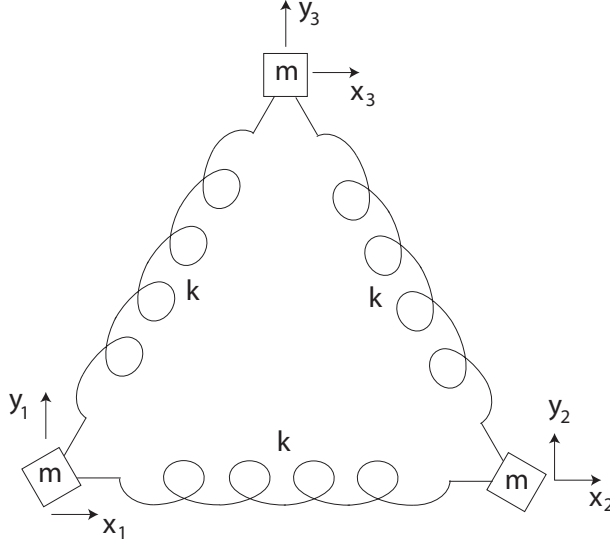


Figure 3.2: The three springs example, showing the coordinate system. Each coordinate pair has its origin at the center of its respective mass in the equilibrium position.

Finally, we use the first orthogonality relation to isolate a particular coefficient, obtaining,

$$a_j = \frac{1}{h} \sum_{k=1}^{n_c} \chi^{(j)*}(C_k) \chi(C_k) N_k. \quad (3.91)$$

3.10 Example Application

For our first example of a physical application, we consider an arrangement of springs and masses which have a particular symmetry in the equilibrium position. We'll consider here the case of an equilateral triangle, expanding on the example in Mathews & Walker chapter 16.

Suppose that we have a system of three equal masses, m , located (in equilibrium) at the vertices of an equilateral triangle. The three masses are connected by three identical springs of strength k . See Fig. 3.2. The question we wish to answer is: If the system is constrained to move in a plane, what are the normal modes? We'll use group theory to analyze what happens when a normal mode is excited, potentially breaking the equilateral triangular symmetry to some lower symmetry.

Let the coordinates of each mass, relative to the equilibrium position, be $x_i, y_i, i = 1, 2, 3$. The state of the system is given by the 6-dimensional vector:

$\eta = (x_1, y_1, x_2, y_2, x_3, y_3)$, as a function of time. The kinetic energy is:

$$T = \frac{m}{2} \sum_{i=1}^6 \dot{\eta}_i^2. \quad (3.92)$$

Likewise, the potential energy, for small perturbations about equilibrium, is given by:

$$V = \frac{k}{2} \left\{ (x_2 - x_1)^2 + \left[-\frac{1}{2}(x_3 - x_2) + \frac{\sqrt{3}}{2}(y_3 - y_2) \right]^2 + \left[\frac{1}{2}(x_1 - x_3) + \frac{\sqrt{3}}{2}(y_1 - y_3) \right]^2 \right\}. \quad (3.93)$$

Or, we may write:

$$V = \frac{k}{2} \sum_{i,j=1}^6 U_{ij} \eta_i \eta_j, \quad (3.94)$$

where

$$U = \frac{1}{4} \begin{pmatrix} 5 & \sqrt{3} & -4 & 0 & -1 & -\sqrt{3} \\ \sqrt{3} & 3 & 0 & 0 & -\sqrt{3} & -3 \\ -4 & 0 & 5 & -\sqrt{3} & -1 & \sqrt{3} \\ 0 & 0 & -\sqrt{3} & 3 & \sqrt{3} & -3 \\ -1 & -\sqrt{3} & -1 & \sqrt{3} & 2 & 0 \\ -\sqrt{3} & -3 & \sqrt{3} & -3 & 0 & 6 \end{pmatrix}. \quad (3.95)$$

The equations of motion ($F = ma$) are:

$$m\ddot{\eta}_i = -\frac{\partial V}{\partial \eta_i} = -k \sum_{j=1}^6 U_{ij} \eta_j. \quad (3.96)$$

In a normal mode,

$$\eta = \mathbf{A} e^{i\omega t}, \quad (3.97)$$

where \mathbf{A} is a constant 6-vector, and hence,

$$-m\omega^2 \eta_i = -k \sum_{j=1}^6 U_{ij} \eta_j, \quad (3.98)$$

or,

$$\sum_{j=1}^6 U_{ij} \eta_j = \lambda \eta_i, \quad \text{where } \lambda = \frac{m\omega^2}{k}. \quad (3.99)$$

That is, the normal modes are the eigenvectors of U , with frequencies given in terms of the eigenvalues. In principle, we need to solve the secular equation $|U - \lambda I| = 0$, a sixth-order polynomial equation, in order to get the eigenvalues. Let's see how group theory can help make this tractable, by incorporating the symmetry of the system.

Each eigenvector “generates” an irreducible representation when acted upon by elements of the symmetry group. Consider a coordinate system in which U is diagonal (such a coordinate system must exist, since U is Hermitian):

$$U = \begin{pmatrix} \lambda_a & & & & & \\ & \ddots & & & & \\ & & \lambda_a & & & \\ & & & \lambda_b & & \\ & & & & \ddots & \\ & & & & & \lambda_b \\ & & & & & & \ddots \end{pmatrix}, \quad (3.100)$$

where the first n_a coordinate vectors in this basis belong to eigenvalue λ_a , and transform among themselves according to irreducible representation $D^{(a)}$, and so forth.

What is the appropriate symmetry group? Well, it must be the group, C_{3v} , of operations which leaves an equilateral triangle invariant. This group is generated by taking products of a rotation by $2\pi/3$, which we will call R , and a reflection about the y -axis, which we will call P . The entire group is then given by the 6 elements $\{e, R, R^2, P, PR, PR^2\}$. Note that this group is isomorphic with the group of permutations of three objects, S_3 . The classes are:

$$\{e\}, \{R, R^2\}, \{P, PR, PR^2\}. \quad (3.101)$$

As there are three classes, there must be three irreducible representations, and hence their dimensions must be 1, 1, and 2. Thus, we can easily construct the character table in Table 3.3.

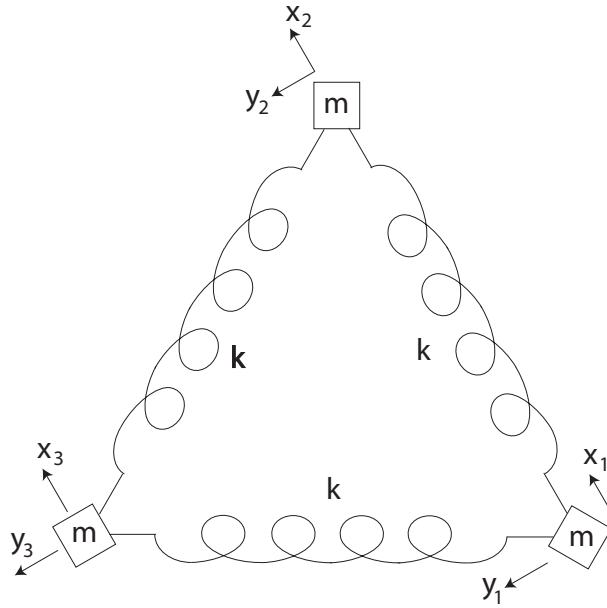
The first row is given by the dimensions of the irreps, since these are the traces of the identity matrices in those irreps. The first column is all ones, since this is the trivial irrep where every element of C_{3v} is represented by the number 1. The second and third row of the second column may be obtained by orthogonality with the first row (remembering the N_k weights), noticing that in a one-dimensional representation the traces are the same as the representation. In particular, the representation of R must be a cube root of one, and the representation of P must be a square root of one. Finally, the second and third rows of the final column are readily determined using the orthogonality relations. Note that in this example, we don't actually need to construct the non-trivial representations to determine the character table. In general, it may be necessary to construct a few of the matrices explicitly.

There is a 6-dimensional representation of C_{3v} which acts on our 6-dimensional coordinate space. We wish to decompose this representation into irreducible representations (why? because that will provide a breakdown of the normal modes by their symmetry under C_{3v}). It is sufficient to know the characters, which we obtain by explicitly considering the action of one element from each class.

Clearly, $\eta = D(e)\eta$, hence $D(e)$ is the 6×6 identity matrix. Its character is $\chi(C_1) = 6$.

Table 3.3: Character table for C_{3v} .

N_k	$l_i \rightarrow$ class \downarrow ; irrep \rightarrow	$l_i = 1$ $\chi^{(1)}$	$l_2 = 1$ $\chi^{(2)}$	$l_3 = 2$ $\chi^{(3)}$
1	$\{e\}$	1	1	2
2	$\{R, R^2\}$	1	1	-1
3	$\{P, PR, PR^2\}$	1	-1	0

Figure 3.3: The three springs example, showing result of a rotation by $2\pi/3$.

Now consider a rotation by $2\pi/3$, see Fig. 3.3. The 6×6 matrix representing this rotation is:

$$D(R) = \begin{pmatrix} 0 & 0 & r \\ r & 0 & 0 \\ 0 & r & 0 \end{pmatrix}, \quad (3.102)$$

where r is the 2×2 rotation matrix:

$$r = \begin{pmatrix} \cos \frac{2\pi}{3} & -\sin \frac{2\pi}{3} \\ \sin \frac{2\pi}{3} & \cos \frac{2\pi}{3} \end{pmatrix} = \frac{1}{2} \begin{pmatrix} -1 & -\sqrt{3} \\ \sqrt{3} & -1 \end{pmatrix}. \quad (3.103)$$

We see that the trace is zero, that is $\chi(C_2) = 0$.

The action of P is to interchange masses 1 and 2, and reflect the x coordi-

nates:

$$D(P) = \begin{pmatrix} 0 & p & 0 \\ p & 0 & 0 \\ 0 & 0 & p \end{pmatrix}, \quad (3.104)$$

where p is the 2×2 reflection matrix:

$$p = \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix}. \quad (3.105)$$

We see that the trace is again zero, that is $\chi(C_3) = 0$.

With these characters, we are now ready to decompose D into the irreps of C_{3v} . We wish to find the coefficients a_1, a_2, a_3 in:

$$D = a_1 D^{(1)} \oplus a_2 D^{(2)} \oplus a_3 D^{(3)}. \quad (3.106)$$

They are given by:

$$a_j = \frac{1}{h} \sum_{k=1}^{n_c} N_k \chi^{(j)*}(C_k) \chi(C_k). \quad (3.107)$$

The result is:

$$\begin{aligned} a_1 &= \frac{1}{6}(1 \cdot 1 \cdot 6 + 2 \cdot 1 \cdot 0 + 3 \cdot 1 \cdot 0) = 1 \\ a_2 &= \frac{1}{6}(1 \cdot 1 \cdot 6 + 2 \cdot 1 \cdot 0 + 3 \cdot -1 \cdot 0) = 1 \\ a_3 &= \frac{1}{6}(1 \cdot 2 \cdot 6 + 2 \cdot -1 \cdot 0 + 3 \cdot 0 \cdot 0) = 2. \end{aligned} \quad (3.108)$$

That is,

$$D = D^{(1)} \oplus D^{(2)} \oplus 2D^{(3)}. \quad (3.109)$$

In the basis corresponding to the eigenvalues we thus have:

$$U = \begin{pmatrix} \lambda_1 & & & & & \\ & \lambda_2 & & & & \\ & & \lambda_{31} & & 0 & \\ & & & \lambda_{31} & & \\ & & 0 & & \lambda_{32} & \\ & & & & & \lambda_{32} \end{pmatrix}, \quad (3.110)$$

where λ_1 corresponds to $D^{(1)}$, λ_2 to $D^{(2)}$, and $\lambda_{31}, \lambda_{32}$ to two instances of $D^{(3)}$. Thus, we already know that there are no more than four distinct eigenvalues, that is, some of the six modes have the same frequency.

Let's see that we can find the actual frequencies without too much further work. Consider $D(g)U$ in this diagonal coordinate system. In this basis we must have:

$$D(g) = \begin{pmatrix} D^{(1)}(g) & & & & & \\ & D^{(2)}(g) & & & 0 & \\ & & 0 & & D^{(3)}(g) & \\ & & & & & D^{(3)}(g) \end{pmatrix}, \quad (3.111)$$

and hence,

$$D(g)U = \begin{pmatrix} \lambda_1 D^{(1)}(g) & & & & & \\ & \lambda_2 D^{(2)}(g) & & & & \\ & & 0 & & & \\ & & & \lambda_{31} D^{(3)}(g) & & \\ & & & & & \\ & & & & & \lambda_{32} D^{(3)}(g) \end{pmatrix}. \quad (3.112)$$

We don't know what this coordinate system is, but we may consider quantities which are independent of coordinate system, such as the trace:

$$\text{Tr}[D(g)U] = \lambda_1 \chi^{(1)}(g) + \lambda_2 \chi^{(2)}(g) + (\lambda_{31} + \lambda_{32}) \chi^{(3)}(g). \quad (3.113)$$

Referring to Eqn. 3.95 we find, for $g = e$:

$$\text{Tr}[D(e)U] = \text{Tr} U = \frac{1}{4}(5 + 3 + 5 + 3 + 2 + 6) = 6. \quad (3.114)$$

For $g = R$:

$$\begin{aligned} \text{Tr}[D(R)U] &= \text{Tr} \frac{1}{2} \begin{pmatrix} 0 & & & -1 & -\sqrt{3} \\ & 0 & & \sqrt{3} & -1 \\ & -1 & -\sqrt{3} & & \\ \sqrt{3} & -1 & & 0 & \\ & & & & 0 \\ 0 & & -1 & -\sqrt{3} & \\ & & \sqrt{3} & -1 & 0 \end{pmatrix} \times \\ &\quad \frac{1}{4} \begin{pmatrix} 5 & \sqrt{3} & -4 & 0 & -1 & -\sqrt{3} \\ \sqrt{3} & 3 & 0 & 0 & -\sqrt{3} & -3 \\ -4 & 0 & 5 & -\sqrt{3} & -1 & \sqrt{3} \\ 0 & 0 & -\sqrt{3} & 3 & \sqrt{3} & -3 \\ -1 & -\sqrt{3} & -1 & \sqrt{3} & 2 & 0 \\ -\sqrt{3} & -3 & \sqrt{3} & -3 & 0 & 6 \end{pmatrix} \\ &= \frac{1}{8}(1 + 3 - 3 + 3 + 4 + 0 + 1 - 3 + 3 + 3) = \frac{3}{2} \end{aligned} \quad (3.115)$$

For $g = P$:

$$\begin{aligned} \text{Tr}[D(P)U] &= \text{Tr} \begin{pmatrix} 0 & 0 & -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ -1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix} \frac{1}{4} \begin{pmatrix} 5 & \sqrt{3} & -4 & 0 & -1 & -\sqrt{3} \\ \sqrt{3} & 3 & 0 & 0 & -\sqrt{3} & -3 \\ -4 & 0 & 5 & -\sqrt{3} & -1 & \sqrt{3} \\ 0 & 0 & -\sqrt{3} & 3 & \sqrt{3} & -3 \\ -1 & -\sqrt{3} & -1 & \sqrt{3} & 2 & 0 \\ -\sqrt{3} & -3 & \sqrt{3} & -3 & 0 & 6 \end{pmatrix} \\ &= \frac{1}{4}(4 + 0 + 4 + 0 - 2 + 6) = 3. \end{aligned} \quad (3.116)$$

This gives us the three equations:

$$\begin{aligned} 6 &= \lambda_1 + \lambda_2 + 2(\lambda_{31} + \lambda_{32}) \\ \frac{3}{2} &= \lambda_1 + \lambda_2 - (\lambda_{31} + \lambda_{32}) \\ 3 &= \lambda_1 - \lambda_2. \end{aligned} \quad (3.117)$$

Hence,

$$\lambda_1 = 3 \quad (3.118)$$

$$\lambda_2 = 0 \quad (3.119)$$

$$\lambda_{31} + \lambda_{32} = \frac{3}{2}. \quad (3.120)$$

To determine λ_{31} and λ_{32} , we could consider another invariant, such as

$$\text{Tr}U^2 = \lambda_1^2 + \lambda_2^2 + 2(\lambda_{31}^2 + \lambda_{32}^2). \quad (3.121)$$

Alternatively, we may use some physical insight: There must be three degrees of freedom with eigenvalue 0, corresponding to an overall rotation of the system and overall translation of the system in two directions. Thus, choose $\lambda_{31} = 0$ and then $\lambda_{32} = 3/2$.

The frequencies are $\omega = \sqrt{\lambda k/m}$. The highest frequency is $\omega = \sqrt{3k/m}$, corresponding to the “breathing mode” in which the springs all expand or contract in unison. Note that this is the mode corresponding to the identity representation; the symmetry of the triangle is not broken in this mode.

3.11 Another example

Let us consider another simple example (again an expanded discussion of an example in Mathews & Walker, chapter 16), to try to get a more intuitive picture of the connection between eigenfunctions and irreducible representations:

Consider a square “drumhead”, and the connection of its vibrational modes with representations of the symmetry group of the square. We note that two eigenfunctions which are related by a symmetry of the square must have the same eigenvalue – otherwise this would not be a symmetry. The symmetry group of the square (see Fig. 3.4) is generated by a 4-fold axis, plus mirror planes joining the sides and vertices.

This group has the elements:

$$\{e, M_a, M_b, M_\alpha, M_\beta, R_{\pm\pi/2}, R_\pi\}. \quad (3.122)$$

Thus, the order is $h = 8$. The classes are readily seen to be:

$$\begin{aligned} C_1 &= \{e\} \\ C_2 &= \{M_a, M_b\} \\ C_3 &= \{M_\alpha, M_\beta\} \\ C_4 &= \{R_\pi\} \\ C_5 &= \{R_{\pi/2}, R_{-\pi/2}\} \end{aligned} \quad (3.123)$$

We must have $\sum_{i=1}^{n_r} \ell_i^2 = 8$, but $n_r = 5$, and therefore $\ell_1 = \ell_2 = \ell_3 = \ell_4 = 1$, and $\ell_5 = 2$ are the dimensions of the irreducible representations.

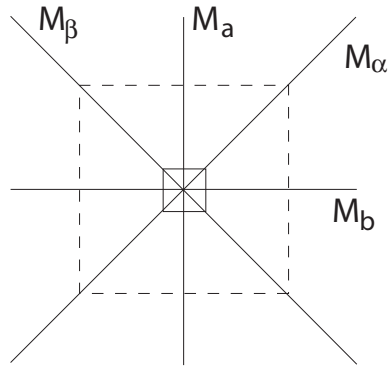


Figure 3.4: The symmetry group of the square.

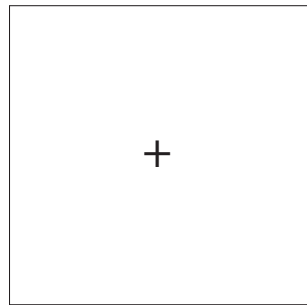


Figure 3.5: The lowest excitation of the square drumhead. The “plus” in the center is supposed to indicate that the whole drumhead is oscillating back and forth through the plane of the square.

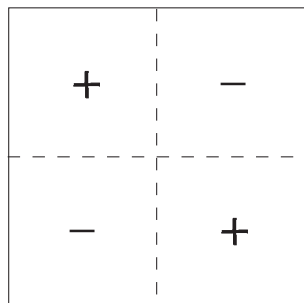


Figure 3.6: A higher excitation of the square drumhead.

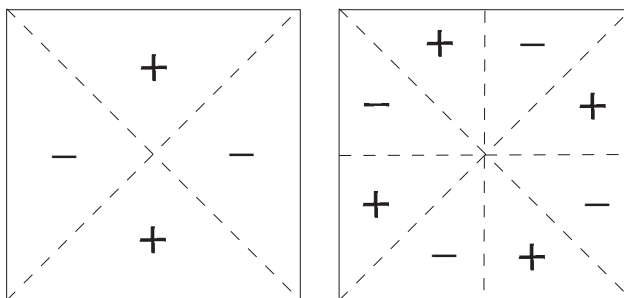


Figure 3.7: Two additional excitations of the square drumhead, corresponding to one-dimensional representations.

Let us consider some vibrational modes and see what representations they generate: The lowest mode is just when the whole drumhead vibrates back and forth, Fig. 3.5.

The action of any group element on this mode is to transform it into itself, hence, this mode generates the trivial representation where all elements are represented by the number 1.

Another mode is shown in Fig. 3.6.

This mode is also non-degenerate, hence it must generate also a one-dimensional representation, but it is no longer the trivial representation, since it is not invariant under the action of all of the elements of the group. For example, $R_{\pi/2}$ yields a minus sign on this mode.

Likewise, the modes shown in Fig. 3.7 are non-degenerate and generate new one-dimensional irreducible representations. It may be seen that these one-dimensional irreps are all inequivalent, as the actions of the group elements differ in the different irreps.

Finally, we have the degenerate modes illustrated in Fig. 3.8. These two modes transform among themselves under the group operations, hence generate a two-dimensional irreducible representation.

We might wonder about the modes illustrated in Fig. 3.9. These also gen-

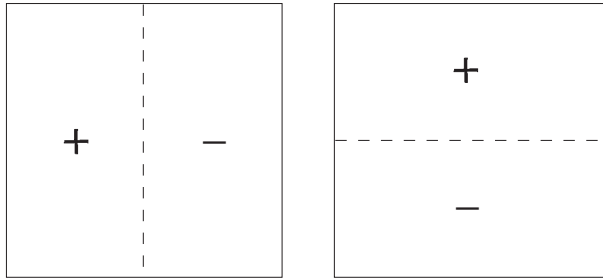


Figure 3.8: Two degenerate modes, generating a two-dimensional representation.

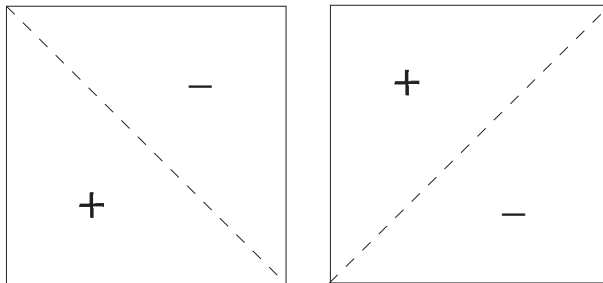


Figure 3.9: Another pair of degenerate modes, generating a two-dimensional representation.

erate a two-dimensional irrep. However, the reader is encouraged to show that this irrep is equivalent to the one above.

We might also wonder about the modes illustrated in Fig. 3.10. These also generate a two-dimensional representation. In this case, however, the representation is reducible.

There are also modes which generate the same irreps already considered, but corresponding to higher excitations. For example, see Fig. 3.11

3.12 Direct Product Theory

The direct product of two matrices A and B is the set of product elements obtained by multiplying every element of A by every element of B . It is convenient to think of these products as arranged in a “direct product matrix” form. For example, if A is $n \times n$, and B is $m \times m$, the direct product matrix is $nm \times nm$. The multiplication of direct product matrices is defined so that they can describe successive transformations in a “product” space. A product space is formed out of two spaces so that a transformation in the product space is a combination of transformations done separately in each of the ordinary spaces such that the rule of successive transformations is obeyed in each of the ordinary

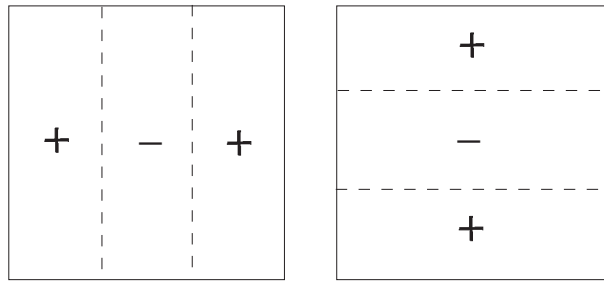


Figure 3.10: Another pair of degenerate modes, generating a two-dimensional representation, but this time a reducible representation.

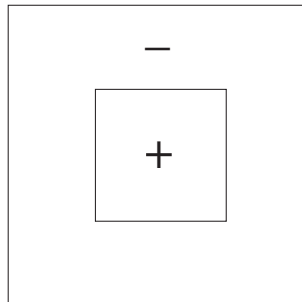


Figure 3.11: A higher excitation, generating the identity representation.

spaces separately.

Thus, if A, A' are operators in space a , and B, B' are operators in space b , then $A'' = AA'$ is the successive operation of A' , then A in space a , and $B'' = BB'$ is the successive operation of B' , then B in space b . Suppose we define “direct product” operators $C = A \otimes B$, and $C' = A' \otimes B'$. Then we require that

$$\begin{aligned} C'' &= CC' \\ &= (A \otimes B)(A' \otimes B') \\ &= AA' \otimes BB' \\ &= A'' \otimes B'' \end{aligned} \quad (3.124)$$

By components, this is:

$$\begin{aligned} C''_{ik,jm} &= \sum_{p,q} C_{ik,pq} C'_{pq,jm} \\ &= \sum_{p,q} A_{ip} B_{kq} A'_{pj} B'_{qm}, \quad \text{since...} \\ &= \sum_{p,q} A_{ip} A'_{pj} B_{kq} B'_{qm} \\ &= A''_{ij} B''_{km}. \end{aligned} \quad (3.125)$$

One possible way to write out the direct product matrix is:

$$C = A \otimes B = \begin{pmatrix} a_{11}B & a_{12}B & \cdots \\ a_{21}B & a_{22}B & \cdots \\ \vdots & \vdots & \ddots \end{pmatrix}. \quad (3.126)$$

If we are given two groups, $G_a = (\{a_i\}, \circ)$ of order h_a and $G_b = (\{b_i\}, *)$ of order h_b , the *direct product group*, $G_a \otimes G_b$, is formed by the elements consisting of all ordered pairs (a_i, b_j) with multiplication defined by:

$$(a_i, b_j)(a_k, b_\ell) \equiv (a_i \circ a_k, b_j * b_\ell). \quad (3.127)$$

As usual, we typically drop the explicit operation symbols in the hopes that the appropriate operation is understood from context. The reader is urged to demonstrate that we have in fact defined a group here.

We list some facts concerning direct product groups:

1. In the groups $G_a = \{a_i\}$ and $G_b = \{b_i\}$, the indices i and j run over some index sets, not necessarily finite or even countable. The order of $G_a \otimes G_b$ is the product of the orders of the two groups, i.e., $h_a h_b$. This may be infinite.
2. If e_a is the identity element of G_a and e_b is the identity element of G_b , then the set of elements in $G_a \otimes G_b$ of the form (e_a, b_i) yields a subgroup isomorphic with G_b , and those of the form (a_i, e_b) yield a subgroup isomorphic with G_a .

Table 3.4: Character table for C_{3v} .

N_k	$\ell_i \rightarrow$ class \downarrow ; irrep \rightarrow	$\ell_i = 1$ $\chi^{(1)}$	$\ell_2 = 1$ $\chi^{(2)}$	$\ell_3 = 2$ $\chi^{(3)}$
1	$\{e\}$	1	1	2
2	$\{R, R^2\}$	1	1	-1
3	$\{P, PR, PR^2\}$	1	-1	0

Table 3.5: Character table for the inversion group, \mathcal{I} .

N_k	$\ell_i \rightarrow$ class \downarrow ; irrep \rightarrow	$\ell_i = 1$ $\chi^{(1)}$	$\ell_2 = 1$ $\chi^{(2)}$
1	$\{e\}$	1	1
1	$\{i\}$	1	-1

- The classes of the direct product group are given by the direct products of the classes of the original groups.
- The direct products of the matrices representing G_a and G_b provide representations of $G_a \otimes G_b$ under the matrix multiplication rule for direct product matrices.
- If $D_a^{(i)}(a_r)$ and $D_b^{(j)}(b_s)$ are irreps of G_a and G_b , respectively, then

$$D_c^{(ij)}(c_{rs}) \equiv D_a^{(i)}(a_r) \otimes D_b^{(j)}(b_s) \quad (3.128)$$

is an irrep of $G_c = G_a \otimes G_b$. Furthermore, there are no additional irreps besides those constructed in this way. Note that this, plus the previous item on representations, implies that the character table of the product group is:

$$\chi_c^{(ij)}(c_{rs}) = \chi_a^{(i)}(a_r) \chi_b^{(j)}(b_s). \quad (3.129)$$

Let's look at an example of the construction of a character table for a direct product group. Suppose we have the symmetry group C_{3v} of the equilateral triangle. We have already obtained the character table for this group in our example on springs in Section 3.10. This table is repeated in Table 3.4. Recall that R stands for a rotation by $2\pi/3$, and P is one of the mirrors containing the rotation axis and a vertex.

Now suppose that we wish to add to this group the operation of inversion. The resulting group is called D_{3d} . The inversion group, \mathcal{I} , is a two-element group, consisting of the identity e and the inversion operator i . The only possible character table for a group of order two is shown in Table 3.5.

We wish to obtain the character table for the product group $D_{3d} = C_{3v} \otimes \mathcal{I}$. Recall that the character of a representation is the trace of a matrix, so we must

Table 3.6: Character table for D_{3d} .

N_k	$\ell_i \rightarrow$ class \downarrow ; irrep \rightarrow	$\ell_i = 1$ $\chi^{(1)}$	$\ell_2 = 1$ $\chi^{(2)}$	$\ell_3 = 2$ $\chi^{(3)}$	$\ell_4 = 1$ $\chi^{(4)}$	$\ell_5 = 1$ $\chi^{(5)}$	$\ell_6 = 2$ $\chi^{(6)}$
1	$C_1 = \{e\}$	1	1	2	1	1	2
2	$C_2 = \{R, R^2\}$	1	1	-1	1	1	-1
3	$C_3 = \{P, PR, PR^2\}$	1	-1	0	1	-1	0
1	$C_4 = \{ie\}$	1	1	2	-1	-1	-2
2	$C_5 = \{iR, iR^2\}$	1	1	-1	-1	-1	1
3	$C_6 = \{iP, iPR, iPR^2\}$	1	-1	0	-1	1	0

determine the trace of a direct product matrix. If $c = a \otimes b$ is the matrix direct product of matrices a and b , then

$$\begin{aligned}
 \chi(c) &= \chi(a \otimes b) \\
 &= \sum_{k\ell} (a \otimes b)_{k\ell, k\ell} \\
 &= \sum_k a_{kk} \sum_\ell b_{\ell\ell} \\
 &= \chi(a)\chi(b).
 \end{aligned} \tag{3.130}$$

There will be $2 \times 3 = 6$ irreps for our product group (we have doubled the number of classes of D_3 . The character table must be as shown in Table 3.6.

The order of D_{3d} is $h = 12$, which agrees with the sum of the squares of the dimensions of the irreps $\ell_k = 1, 1, 1, 1, 2, 2$. We remark also that the character table looks like the direct product of the input character tables:

$$\begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} \otimes \begin{pmatrix} 1 & 1 & 2 \\ 1 & 1 & -1 \\ 1 & -1 & 0 \end{pmatrix}. \tag{3.131}$$

Expressing a group as a direct product of smaller groups provides a useful method for studying the irreps of the larger group. Note that $\mathcal{I} = \{e, i\}$ is an abelian invariant subgroup of D_{3d} ($gag^{-1} \in \mathcal{I}$, $\forall a \in \mathcal{I}$ and $\forall g \in D_{3d}$). Therefore, D_{3d} is not a simple group (since it contains a proper invariant subgroup), nor is it semi-simple (since the invariant subgroup is abelian).

We may write the suggestive notation $C_{3v} = D_{3d}/\mathcal{I}$ and refer to C_{3v} as the “factor group” or “quotient group”. Since C_{3v} is the group that leaves the triangle invariant, we refer to it as the “little group” of D_{3d} (or the “little group of the triangle”).

3.13 Generating Additional Representations

Given one or more representations D of a group G , there are various ways of generating additional representations. We have already seen the direct sum method. Let us now see some others. Relax, for now, the assumption of unitary representations. First, there are three simple operations on representation D which will give us (possibly) new representations of the same dimension:

1. *Adjoint Representation:* Given a group G with a (invertible) representation $D(a), a \in G$, consider the set of matrices obtained by taking the inverse transpose of $D(a)$: $[D(a)^{-1}]^T$. This is also a representation of G , since,

$$\begin{aligned} [D(ab)^{-1}]^T &= [D(b)^{-1}D(a)^{-1}]^T \\ &= [D(a)^{-1}]^T[D(b)^{-1}]^T, \end{aligned}$$

hence, the multiplication table is preserved. This is called the “adjoint representation”, \bar{D} .

2. *Complex Conjugate Representation:* Given a representation D , consider the matrices formed by taking the complex conjugate of the elements of $D(a)$: $[D(a)]^*$. We have,

$$\begin{aligned} D(ab)^* &= [D(a)D(b)]^* \\ &= D(a)^*D(b)^*, \end{aligned}$$

so this also defines a representation. It is called the “complex conjugate representation”, D^* .

3. Finally, we also obtain a representation by taking the complex conjugate of the adjoint representation:

$$[D(a)^{-1}]^\dagger = \bar{D}(a)^*. \quad (3.132)$$

We note that $D, \bar{D}, D^*, \bar{D}^*$ are all either reducible or irreducible representations, which may, or may not, be equivalent. Thus, this is one thing to try towards finding new (irreducible) representations for G . Note that if we have a unitary representation, which is always possible for a finite group,

$$\bar{D}(g) = [D(g)^{-1}]^T = [D(g)^\dagger]^T = D^*(g), \quad (3.133)$$

hence, the adjoint representation is identical with the complex conjugate representation.

If the representation is real, then $D(g) = D(g)^*$, and $\chi(g)$ is real. If, instead, we know that $\chi(g)$ is real, then $\chi(g) = \text{Tr}[D(g)] = \text{Tr}[D(g)^*]$, and therefore D and D^* are equivalent. If, on the other hand, $\chi(g)$ is complex, then D and D^* are not equivalent representations. We can state these observations in the form of a theorem:

Theorem: D and D^* are equivalent representations if and only if their characters are real.

Let us revisit briefly our general orthogonality relation. Since we stated it for finite groups, we have been justified in assuming we can always deal with unitary representations. However, we might happen to deal at some point with a non-unitary representation. In general, the orthogonality relation for irreducible representations $D^{(i)}$, and $D^{(j)}$ reads:

$$\sum_{g \in G} D^{(i)}(g)_{\mu\nu} D^{(j)}(g^{-1})_{\alpha\beta} = \frac{h}{\ell_i} \delta_{ij} \delta_{\mu\beta} \delta_{\nu\alpha}. \quad (3.134)$$

In terms of this general relation, we can repeat the earlier derivation of the “first orthogonality relation”, setting $\nu = \mu$, $\beta = \alpha$, and summing over μ and α , to obtain:

$$\sum_{g \in G} \chi^{(i)}(g) \chi^{(j)}(g^{-1}) = h \delta_{ij}. \quad (3.135)$$

But, in the adjoint representation, $\bar{D}(g) = [D(g)^{-1}]^T = D(g^{-1})^T$, and hence $\bar{\chi}(g) = \chi(g^{-1})$. Therefore,

$$\sum_{g \in G} \chi^{(i)}(g) \bar{\chi}^{(j)}(g) = h \delta_{ij}, \quad (3.136)$$

or, in terms of classes:

$$\sum_{k=1}^{n_c} N_k \chi^{(i)}(C_k) \bar{\chi}^{(j)}(C_k) = h \delta_{ij}. \quad (3.137)$$

Likewise, our second orthogonality relation in general is:

$$\sum_{i=1}^{n_r} \chi^{(i)}(C_k) \chi^{(i)}(C_\ell^{-1}) = \frac{h}{N_k} \delta_{k\ell}, \quad (3.138)$$

where C_ℓ^{-1} means take the inverses of the elements in class C_ℓ . Or, using again $\bar{\chi}(g) = \chi(g^{-1})$

$$\sum_{i=1}^{n_r} \chi^{(i)}(C_k) \bar{\chi}^{(i)}(C_\ell) = \frac{h}{N_k} \delta_{k\ell}, \quad (3.139)$$

With these forms, the expansion coefficients for the decomposition of an arbitrary representation, D , into irreps become:

$$a_m = \frac{1}{h} \sum_{k=1}^{n_c} N_k \chi(C_k) \bar{\chi}^{(m)}(C_k). \quad (3.140)$$

3.14 Kronecker Products and Clebsch-Gordan Coefficients

Given two representations $D^{(i)}$ and $D^{(j)}$ of a group G , we may construct a new representation, $D^{(i \times j)}$, by taking the direct product matrices:

$$D^{(i \times j)}(g) = D^{(i)}(g) \otimes D^{(j)}(g). \quad (3.141)$$

In components:

$$D^{(i \times j)}(g)_{\alpha\beta, \mu\nu} = D^{(i)}(g)_{\alpha\mu} D^{(j)}(g)_{\beta\nu}. \quad (3.142)$$

It is left to the reader to verify that $D^{(i \times j)}$ is in fact a representation for G . It is called a *product representation* or a *Kronecker product*. As with direct product groups, we find:

$$\chi^{(i \times j)}(g) = \chi^{(i)}(g) \chi^{(j)}(g). \quad (3.143)$$

Let us now assume that $D^{(i)}$ and $D^{(j)}$ are irreducible representations. The product representation $D^{(i \times j)}$ may, however, be reducible. We would like to find the decomposition into irreps:

$$D^{(i \times j)} = a_1 D^{(1)} \oplus \dots \oplus a_{n_r} D^{(n_r)}. \quad (3.144)$$

This reduction is called the *Clebsch-Gordan series*. For the coefficients, we have:

$$\begin{aligned} a_m &= \frac{1}{h} \sum_{k=1}^{n_c} N_k \chi^{(i \times j)}(C_k) \bar{\chi}^{(m)}(C_k) \\ &= \frac{1}{h} \sum_{g \in G} \chi^{(i \times j)}(g) \bar{\chi}^{(m)}(g) \\ &= \frac{1}{h} \sum_{g \in G} \chi^{(i)}(g) \chi^{(j)}(g) \bar{\chi}^{(m)}(g). \end{aligned} \quad (3.145)$$

For example, in the note on rotations in quantum mechanics (section 10), the Clebsch-Gordan series for the group $SU(2)$ (an isomorphic representation of the rotation group in quantum mechanics) is obtained:

$$D^{(i \times j)} = \bigoplus_{m=|i-j|}^{i+j} D^{(m)}. \quad (3.146)$$

We'll proceed now to define the notion of "Clebsch-Gordan coefficients" (note that the a_m coefficients in the reduction above are sometimes referred to as Clebsch-Gordan coefficients; this will not be our usage). We start by expressing the Clebsch-Gordan series in a different notation:

$$D^{(i \times j)} = \bigoplus_{m=1}^{n_r} a_m D^{(m)} = \bigoplus_{m=1}^{n_r} (ijm) D^{(m)}. \quad (3.147)$$

3.14. KRONECKER PRODUCTS AND CLEBSCH-GORDAN COEFFICIENTS 57

That is, $(ijm) \equiv a_m$, and our coefficient names now contain explicitly the information of which product representation we are looking at. Notice that there is a symmetry, $(ijm) = (jim)$, since $D^{(m)}$ will appear the same number of times in $D^{(i \times j)}$ as in $D^{(j \times i)}$.

For physical applications (e.g., quantum mechanical angular momentum), we are especially interested in determining the basis functions for the representations in the Kronecker product. For the irrep $D^{(i)}$ (that is, for the vector space acted upon by this representation) we have the basis functions:

$$\{\psi_\alpha^{(i)}; \alpha = 1, 2, \dots, \ell_i\}, \quad (3.148)$$

where ℓ_i is the dimension of irrep i . Likewise, for irrep $D^{(j)}$ we have basis functions:

$$\{\phi_\beta^{(j)}; \beta = 1, 2, \dots, \ell_j\}. \quad (3.149)$$

Since we are considering the product representation $D^{(i)} \otimes D^{(j)}$ we may ask for the ℓ_m functions

$$\{\omega_\gamma^{(m)}; \gamma = 1, \dots, \ell_m\} \quad (3.150)$$

that are linear combinations of the products $\psi_\alpha^{(i)}$ and $\phi_\beta^{(j)}$, and which form a basis for the irrep $D^{(m)}$. Such a set of functions $\{\omega_\gamma^{(m)}\}$ exists only if $D^{(m)}$ is contained in $D^{(i)} \otimes D^{(j)}$, that is, only if $(ijm) > 0$.

Now, (ijm) may be one, but $(ijm) > 1$ is also possible, in which case there will be more than one such sets of functions. In general, there will be precisely (ijm) independent sets of functions $\{\omega_\gamma^{(m)}\}$ formed from the products $\psi_\alpha^{(i)} \phi_\beta^{(j)}$. We'll label them:

$$\{\omega_\gamma^{(m\tau_m)}; \tau_m = 1, \dots, (ijm)\} \quad (3.151)$$

More explicitly, these are functions of the form:

$$\omega_\gamma^{(m\tau_m)} = \sum_{\alpha=1}^{\ell_i} \sum_{\beta=1}^{\ell_j} \psi_\alpha^{(i)} \phi_\beta^{(j)} (i\alpha, j\beta | m\tau_m \gamma) \quad (3.152)$$

The quantities $(i\alpha, j\beta | m\tau_m \gamma)$ are called *Clebsch-Gordan coefficients*.

It is important to understand that all we are really doing here is describing a transformation of basis between alternative bases in the space operated on by the product representation. We remark also that in the case of the quantum mechanical rotation group, the numbers τ_m are never greater than one - the rotation group is said to be *simply reducible*.

The total number of functions $\omega_\gamma^{(m\tau_m)}$ must be the same as the total number of product functions $\psi_\alpha^{(i)} \phi_\beta^{(j)}$:

$$\sum_{m=1}^{n_r} (ijm) \ell_m = \ell_i \ell_j. \quad (3.153)$$

Hence, the Clebsch-Gordan coefficients make a $\ell_i \ell_j \times \ell_i \ell_j$ matrix. As our expansion for $\omega_\gamma^{(m\tau_m)}$ is just a basis transformation, we can write the inverse transformation:

$$\psi_\alpha^{(i)} \phi_\beta^{(j)} = \sum_{\gamma, m, \tau_m} (m\tau_m \gamma | i\alpha, j\beta) \omega_\gamma^{(m\tau_m)}. \quad (3.154)$$

Substituting back into the original equation (3.152), we find:

$$\omega_\gamma^{(m\tau_m)} = \sum_{\alpha, \beta} \sum_{\gamma', m', \tau'_m} (m'\tau'_m \gamma' | i\alpha, j\beta) (i\alpha, j\beta | m\tau_m \gamma) \omega_{\gamma'}^{(m'\tau'_m)}, \quad (3.155)$$

or,

$$\sum_{\alpha, \beta} (m'\tau'_m \gamma' | i\alpha, j\beta) (i\alpha, j\beta | m\tau_m \gamma) = \delta_{mm'} \delta_{\tau_m \tau'_m} \delta_{\gamma \gamma'}. \quad (3.156)$$

Alternatively, substituting Eqn. 3.152 into Eqn. 3.154, we obtain:

$$\sum_{m, \tau_m, \gamma} (i\alpha', j\beta' | m\tau_m \gamma) (m\tau_m \gamma | i\alpha, j\beta) = \delta_{\alpha\alpha'} \delta_{\beta\beta'}. \quad (3.157)$$

At least for unitary representations, it may be shown that the matrix of Clebsch-Gordan coefficients is a matrix which puts $D^{(i)} \times D^{(j)}$ into reduced form.

3.15 Angular Momentum in Quantum Mechanics

The theory of angular momentum in quantum mechanics is developed in detail in the note on this subject linked to the Ph 129 page. Here, we'll summarize a few of the key elements relative to our discussion of group theory. As the rotation group is an infinite group, we'll also remark on the extension of our discussion to infinite groups.

As an explicit function, the spherical harmonic $Y_{\ell m}$ is given by:

$$Y_{\ell m}(\theta, \phi) = \frac{(-1)^m}{2^\ell \ell!} \sqrt{\frac{2\ell+1}{4\pi} \frac{(\ell+m)!}{(\ell-m)!}} e^{im\phi} \frac{1}{(\sin \theta)^m} \left(\frac{d}{d \cos \theta} \right)^{\ell+m} (1 - \cos^2 \theta)^\ell, \quad (3.158)$$

where $0 \leq \theta \leq \pi$, and $0 \leq \phi < 2\pi$. However, it is perhaps more profound to define the $Y_{\ell m}$ in terms of the matrices $D^\ell(R)$ which give the irreducible representations of the rotation group.

Consider the rotation R expressed in terms of the Euler angles α, β, γ :

$$R = R(\alpha, \beta, \gamma) = R_z(\gamma) R_y(\beta) R_z(\alpha). \quad (3.159)$$

Choosing $\alpha = 0$, a vector along the z -axis may be rotated to $\theta = \beta$ and $\phi = \gamma$. We consider the rotation matrices with components:

$$\begin{aligned} D^\ell(\gamma, \beta, \alpha)_{mm'} &= e^{-i(m\gamma+m'\alpha)} d_{mm'}^\ell(\beta) \\ &= \langle \ell m | D^\ell(R) | \ell, m' \rangle, \end{aligned} \quad (3.160)$$

with the interpretation that these are the rotation matrices acting on a vector space of functions corresponding to angular momentum ℓ . We then define the spherical harmonics:

$$Y_{\ell m}(\theta, \phi) \equiv \sqrt{\frac{2\ell+1}{4\pi}} D_{m0}^{*\ell}(\phi, \theta, 0). \quad (3.161)$$

Note that since $m' = 0$, the spherical harmonics describe states with integer angular momentum only.

As mentioned before, the Clebsch-Gordan coefficients describe a change of basis. Consider a system of two “particles”, with spins j_1 and j_2 . We may describe the (angular momentum) state of these particles according to:

$$|j_1 m_1 j_2 m_2\rangle, \quad (3.162)$$

where m_1 and m_2 are the z -components of the spins. However, we might also describe the state by specifying the total angular momentum, j , and the total component along the z -axis, m ($= m_1 + m_2$):

$$|j_1 j_2 j m\rangle. \quad (3.163)$$

This situation corresponds to the reduction of a product representation $D^{(j_1)} \otimes D^{(j_2)}$ into irreducible representations $D^{(j)}$. The Clebsch-Gordan coefficients (also known in this context as vector addition or Wigner coefficients) merely tell us how to transform from one basis to the other. For example,

$$|j_1 j_2 j m\rangle = \sum_{m_1, m_2} (j_1 m_1, j_2 m_2 | j m) |j_1 m_1 j_2 m_2\rangle, \quad (3.164)$$

where we have omitted the $\tau_j = 1$.

As mentioned already, we may show that for the rotation group ($SU(2)$) the Clebsch-Gordan series is

$$D^{(j_1 \times j_2)} = \bigoplus_{j=|j_1-j_2|}^{j_1+j_2} D^{(j)}. \quad (3.165)$$

The proof of this relies on

$$\chi^{(j_1 \times j_2)} = \chi^{(j_1)} \chi^{(j_2)}, \quad (3.166)$$

and on the orthogonality relations:

$$\frac{1}{h} \sum_{g \in G} \chi^{(i)}(g) \bar{\chi}^{(j)}(g) = \delta_{ij}. \quad (3.167)$$

But this is an infinite group, $h = \infty$. We are faced with the issue of defining “ $\frac{1}{h} \sum_{g \in G}$ ”.

To generalize this to the case of an infinite group, notice that $\frac{1}{h} \sum_{g \in G}$ is a kind of average over the elements of the group. Since the rotation group depends

on three continuously varying parameters, we expect our sum to become some kind of integral. The question in constructing the appropriate integral is how to weight the various regions in parameter space. That is, we need to define a notion of the size of a set of rotations; we need a measure function, $\mu(\{R\})$, where $\{R\}$ is some set of rotations.

The measure function must satisfy the property that no rotation gets any bigger weight than any other. Consider a set $\{R\}$ of rotations. We can obtain another set of rotations, $\{R_0R\}$ by applying a specified rotation, R_0 , to each element of this set. The rotated set should be the same size as the original set. We require the following invariance of the measure (considering rotation group $O^+(3)$):

$$\mu(\{R\}) = \mu(\{R_0R\}), \quad \forall R_0 \in O^+(3). \quad (3.168)$$

For $O^+(3)$, the invariant measure, normalized such that the integral over the set of all rotations is one, is:

$$\mu(dR) = \frac{1}{8\pi^2} \sin \theta d\theta d\psi d\phi, \quad (3.169)$$

for rotations parameterized by the Euler angles:

$$\begin{aligned} 0 &\leq \psi < 2\pi \\ 0 &\leq \theta \leq \pi \end{aligned} \quad (3.170)$$

$$0 \leq \phi < 2\pi. \quad (3.171)$$

For $SU(2)$, the range of ϕ becomes $0 \leq \phi < 4\pi$, and the normalized invariant measure is:

$$\mu(dR) = \frac{1}{16\pi^2} \sin \theta d\theta d\psi d\phi. \quad (3.172)$$

See the note on angular momentum in quantum mechanics for a more detailed discussion.

In general, we may define such a measure on a group. If the group is “compact” (e.g., the rotation group is compact because it may be parameterized by parameters on a compact set), then it is possible to define a measure such that the measure $\mu(G)$ over the entire group is finite. In this case, many of our proofs may be readily modified to apply to the infinite groups. For example, every representation of a compact group is equivalent to a unitary representation. The rearrangement lemma, so handy in several of our proofs, reads,

$$\int_G f(u)\mu(du) = \int_G f(uv)\mu(du), \quad \forall v \in G. \quad (3.173)$$

In the case of $SU(2)$, this is:

$$\begin{aligned} \frac{1}{16\pi^2} \int_0^{2\pi} d\psi \int_{-1}^1 d \cos \theta \int_0^{4\pi} d\phi f [R(\psi, \theta, \phi)] = \\ \frac{1}{16\pi^2} \int_0^{2\pi} d\psi \int_{-1}^1 d \cos \theta \int_0^{4\pi} d\phi f [R(\psi_0, \theta_0, \phi_0)R(\psi, \theta, \phi)]. \end{aligned} \quad (3.174)$$

The invariant measure on a group is referred to as the *Haar measure*.

We'll expand on these notions in our note on Lie groups.

3.16 Exercises

1. For the Poincare group \bar{L} , show that any element $\Lambda(M, z)$ can be written as a product of a pure homogeneous transformation followed by a pure translation. Also show that it can be written as a pure translation followed by a pure homogeneous transformation.
2. Show that the object $\{x, y\}$ defined in Eqn. 3.36 is a scalar product.
3. Carry out the steps to demonstrate the decomposition of a representation into irreps,

$$D = a_1 D^{(1)} \oplus a_2 D^{(2)} \oplus \dots \oplus a_{n_r} D^{(n_r)}, \quad (3.175)$$

with coefficients:

$$a_j = \frac{1}{h} \sum_{k=1}^{n_c} \chi^{(j)*}(C_k) \chi(C_k) N_k. \quad (3.176)$$

4. Derive the constant in Eqn. 3.84, that is, determine λ in:

$$\sum_{g \in C_k} D^{(i)}(g) = \lambda I. \quad (3.177)$$

5. Show that the two irreps generated according to Figs. 3.8 and 3.9 are equivalent.
6. Consider the group of all rotations in two dimensions: $G = \{R(\theta) : 0 \leq \theta < 2\pi\}$. As a linear operator on vectors in a two-dimensional Euclidean space, the elements of G may be represented (faithfully, or isomorphically) by the set of 2×2 matrices of the form:

$$D(\theta) = \begin{pmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{pmatrix}. \quad (3.178)$$

Show that this group can be decomposed into two one-dimensional representations, i.e., that you can find a transformation such that every element of G can be represented in the form:

$$D(\theta) = \begin{pmatrix} f(\theta) & 0 \\ 0 & g(\theta) \end{pmatrix}, \quad (3.179)$$

where the new representation is still faithful. You should find explicit expressions for f and g .

Hint: You want to find a similarity transformation, which just corresponds to a change in basis. You might consider the basis transformation so

often encountered in quantum mechanics and in optics, corresponding to describing states in terms of “circular polarization” instead of “linear polarization”.

7. Consider the dihedral group D_3 , which is isomorphic with the group of permutations of three objects, S_3 . Let V_2 be a two-dimensional Euclidean space spanned by orthonormal vectors $\mathbf{e}_x, \mathbf{e}_y$. Give the representation, D , of the elements of D_3 with respect to this basis. That is, express the transformed vectors $D(g)\mathbf{e}_i$ in terms of the original basis, and hence obtain representation D .
8. Consider the symmetry group, C_{4v} , of the square, consisting of rotations about the axis perpendicular to the square, and reflections about the vertical, horizontal, and diagonal axes in the plane of the square (but no mirror plane in the plane of the square).
 - (a) Construct a suitable set of irreducible representations of C_{4v} . That is, up to equivalence, construct all of the irreducible representations of this group.
 - (b) Give the character table for C_{4v} .
9. In problem 6 you consider the reducibility of a two-dimensional representation of the group of rotations in two dimensions. We may remark that this is an abelian group. Let us generalize that result: Consider a group, G , with a unitary representation D , consisting of unitary matrices $D(g)$, $g \in G$. If G is an abelian group, show that any such representation is, by a similarity transformation, equivalent to a representation by diagonal matrices (i.e., by a direct sum of 1×1 matrices). Note that we have already used group theory (Schur’s lemma) to argue the truth of this. In this problem, I want you to use what you know about matrix theory to demonstrate the result.
10. Construct the character table for the tetrahedral symmetry group T_d . You may wish to keep a copy of your result for problem 12.
11. Let’s take a peek at the relation of irreducible representations and the invariant subspaces of a vector space: Let V be the 6-dimensional function space consisting of polynomials of degree 2 in the two real variables x and y :

$$f(x, y) = ax^2 + bxy + cy^2 + dx + ey + h, \quad (3.180)$$

where a, b, c, d, e, h are complex numbers. If x, y transforms under the dihedral group D_3 (problem 7) as the coordinates of a 2-vector, then we obtain a 6-dimensional representation of D_3 on V . Identify the invariant subspaces of V under D_3 , and the corresponding irreducible representations contained in this six dimensional representation (don’t be afraid to use your intuition to make sure that what you find is sensible).

12. At last we are ready for a real physics application of group theory. We looked at the example of masses joined by springs in the shape of an equilateral triangle in this note. Now, let us consider the problem of four masses joined by springs. The four masses are at the corners of a tetrahedron, and the springs form the edges of the tetrahedron. Thus, there are six springs connecting the four masses. All four masses are equal, and all six springs are identical.

We wish to determine the frequencies of the normal modes for this system. Notice that to solve the secular equation, $|V - \lambda I| = 0$, presents a formidable image. A little physical intuition can reduce it somewhat, but it would take real cleverness to solve it completely. This cleverness comes in the form of group theory! Group theory permits one to incorporate in a systematic and deliberate way everything we know about the symmetry of the problem, hence reducing it to a simpler problem.

The problem is still not trivial – you should spend time thinking about convenient approaches in setting things up, and about ways to avoid doing unnecessary work. Above all, be careful, and check your results as you proceed. You already obtained the character table for the tetrahedral symmetry group in problem 10. This problem takes you the rest of the way through solving for the frequencies of the normal modes.

- (a) First step: Set up a 12-dimensional vector (coordinate system) describing the system, and derive the equations of motion, arriving finally at a set of linear equations that could be solved, in principle, to yield the frequencies of the normal modes.
 - (b) Second step: Obtain the character table for the twelve-dimensional representation of the tetrahedral symmetry group that acts on your 12-dimensional vector describing the system. Decompose this representation into irreducible representations.
 - (c) Final step: Obtain a small number of trace equations which you can use to solve to obtain the frequencies of the normal modes. Give the frequencies of the normal modes, and their degeneracies. Do your answers make physical sense?
13. The “quaternion” group consists of eight elements,

$$Q = \{1, -1, i, -i, j, -j, k, -k\}, \quad (3.181)$$

with multiplication table defined by (q is any element of Q):

$$\begin{aligned} 1q &= q \\ (-1)^2 &= 1 \\ (-1)q = q(-1) &= -q \\ i^2 = j^2 = k^2 = ijk &= -1 \end{aligned} \quad (3.182)$$

Find the character table for this group. Compare this character table with the character table for dihedral group D_4 . Are these two groups isomorphic?

14. As a follow-on to the drumhead example in this note, consider the symmetry group of the regular pentagon, as given by a five-fold axis and several mirror planes. Do not include the mirror plan containing the plane of the pentagon itself (although you may amuse yourself by considering what happens if you add this operation).
- List the group elements. Denote rotations with R 's, and mirror operations with M 's. Draw a picture! List the classes.
 - Construct the character table for the irreducible representations of this group.
 - Consider the mode of oscillation of a pentagonal drumhead where a nodal line extends from a vertex to the midpoint of the opposite side. Define (with pictures) a basis for the space generated by this mode and its degenerate partners. Give an explicit matrix for one element of each class of the group for the representation of the pentagonal symmetry group that is generated by these degenerate modes.
 - Decompose the representation found in part (c) into irreducible representations.
15. We would like to consider the (qualitative) effects on the energy levels of an atom which is moved from freedom to an external potential (a crystal, say) with cubic symmetry. Let us consider a one-electron atom and ignore spin for simplicity. Recall that the wave function for the case of the free atom looks something like $R_{nl}(r)Y_{lm}(\theta, \phi)$, and that all states with the same n and l quantum numbers have the same energy, *i.e.*, are $(2l + 1)$ -fold degenerate. The Hamiltonian for a free atom must have the symmetry of the full rotation group, as there are no special directions. Thus, we recall some properties of this group for the present discussion. First, we remark that the set of functions $\{Y_{lm} : m = -l, -l + 1, \dots, l - 1, l\}$ for a given l forms the basis for a $(2l + 1)$ -dimensional subspace which is invariant under the operations of the full rotation group. [A set $\{\psi_i\}$ of vectors is said to span an *invariant subspace* V_s under a given set of operations $\{P_j\}$ if $P_j\psi_i \in V_s \forall i, j$.] Furthermore, this subspace is “irreducible,” that is, it cannot be split into smaller subspaces which are also invariant under the rotation group.

Let us denote the linear transformation operator corresponding to element R of the rotation group by the symbol \hat{P}_R , *i.e.*:

$$\hat{P}_R f(\vec{x}) = f(R^{-1}\vec{x})$$

The way to think about this equation is to regard the left side as giving a “rotated function,” which we evaluate at point \vec{x} . The right side tells us

that this is the same as the original function evaluated at the point $R^{-1}\vec{x}$, where R^{-1} is the inverse of the rotation matrix corresponding to rotation R . Since $\{Y_{lm}\}$ forms an invariant subspace, we must have:

$$\hat{P}_R Y_{lm} = \sum_{m'=-l}^l Y_{lm'} D^l(R)_{m'm}$$

The expansion coefficients, $D^l(R)_{m'm}$, can be regarded as the elements of a matrix $D^l(R)$. As we have discussed in general, and as you may see more explicitly in the note on rotations in QM, D^l corresponds to an irreducible representation of the rotation group.

- (a) Prove, or at least make plausible, the fact that D^l is an irreducible representation of the rotation group. (Hint: You might show first that it is a representation and then show irreducibility by finding a contradiction with the supposition of reducibility).

Thus, for a free atom, we have that the degenerate eigenfunctions of a given energy must transform according to an irreducible representation of this group. If the eigenfunctions transform according to the l^{th} representation, the degeneracy of the energy level is $(2l + 1)$ (assuming no additional, “accidental” degeneracy).

- (b) We will need the character table of this group. Since all elements in the same class have the same character, we pick a convenient element in each class by considering rotations about the z -axis, $R = (\alpha, z)$ (means rotate by angle α about the z -axis). Thus:

$$\hat{P}_{(\alpha,z)} Y_{lm} = e^{-im\alpha} Y_{lm}$$

(which you should convince yourself of).

Find the character “table” of the rotation group, that is, find $\chi^\ell(\alpha)$, the character of representation D^ℓ for the class of rotations through angle α . If you find an expression for the character in the form of a sum, do the sum, expressing your answer in as simple a form as you can. Note that the answer is given in the text, just fill in the missing steps to your satisfaction.

- (c) Let us put our atom into a potential with cubic symmetry. Now the symmetry of the free Hamiltonian is broken, and we are left with the discrete symmetry of the cube. The symmetry group of proper rotations of the cube is a group of order 24 with 5 classes. Call this group “ O ”.

Construct the character table for O .

- (d) Consider in particular how the f -level ($l = 3$) of the free atom may split when it is placed in the “cubic potential”. The seven eigenfunctions which transform according to the irreducible representation D^3

of the full group will most likely not transform according to an irreducible representation of O . On the other hand, since the operations of O are certainly operations of D^3 , the eigenfunctions will generate some representation of O .

Determine the coefficients in the decomposition.

$$D^3 = a_1 O^1 \oplus a_2 O^2 \oplus a_3 O^3 \oplus a_4 O^4 \oplus a_5 O^5,$$

where O^i are the irreducible representations of O . Hence, show how the degeneracy of the 7-fold level may be reduced by the cubic potential. Give the degeneracies of the final levels.

Note that we cannot say anything here about the magnitude of any splittings (which could “accidentally” turn out to be zero!), or even about the ordering of the resulting levels – that depends on the details of the potential, not just its symmetry.

Chapter 3

Representation Theory

Solutions to Problems

3.1 Exercises

1. For the Poincare group \bar{L} , show that any element $\Lambda(M, z)$ can be written as a product of a pure homogeneous transformation followed by a pure translation. Also show that it can be written as a pure translation followed by a pure homogeneous transformation.
2. Show that the object $\{x, y\}$ defined in Eqn. ?? is a scalar product.
3. Carry out the steps to demonstrate the decomposition of a representation into irreps,

$$D = a_1 D^{(1)} \oplus a_2 D^{(2)} \oplus \dots \oplus a_{n_r} D^{(n_r)}, \quad (3.1)$$

with coefficients:

$$a_j = \frac{1}{h} \sum_{k=1}^{n_c} \chi^{(j)*}(C_k) \chi(C_k) N_k. \quad (3.2)$$

4. Derive the constant in Eqn. ??, that is, determine λ in:

$$\sum_{g \in C_k} D^{(i)}(g) = \lambda I. \quad (3.3)$$

Solution: We can take the trace of both sides of this equation:

$$\text{Tr} \left[\sum_{g \in C_k} D^{(i)}(g) \right] = \text{Tr} [\lambda I] \quad (3.4)$$

$$\sum_{g \in C_k} \chi^{(i)}(g) = \lambda \ell_i \quad (3.5)$$

$$N_k \chi^{(i)}(C_k) = \lambda \ell_i, \quad (3.6)$$

or

$$\lambda = \frac{N_k}{\ell_i} \chi^{(i)}(C_k). \quad (3.7)$$

This is the result asserted in Eqn. ??.

5. Show that the two irreps generated according to Figs. ?? and ?? are equivalent.
6. Consider the group of all rotations in two dimensions: $G = \{R(\theta) : 0 \leq \theta < 2\pi\}$. As a linear operator on vectors in a two-dimensional Euclidean space, the elements of G may be represented (faithfully, or isomorphically) by the set of 2×2 matrices of the form:

$$D(\theta) = \begin{pmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{pmatrix}. \quad (3.8)$$

Show that this group can be decomposed into two one-dimensional representations, i.e., that you can find a transformation such that every element of G can be represented in the form:

$$D(\theta) = \begin{pmatrix} f(\theta) & 0 \\ 0 & g(\theta) \end{pmatrix}, \quad (3.9)$$

where the new representation is still faithful. You should find explicit expressions for f and g .

Hint: You want to find a similarity transformation, which just corresponds to a change in basis. You might consider the basis transformation so often encountered in quantum mechanics and in optics, corresponding to describing states in terms of “circular polarization” instead of “linear polarization”.

7. Consider the dihedral group D_3 , which is isomorphic with the group of permutations of three objects, S_3 . Let V_2 be a two-dimensional Euclidean space spanned by orthonormal vectors $\mathbf{e}_x, \mathbf{e}_y$. Give the representation, D , of the elements of D_3 with respect to this basis. That is, express the transformed vectors $D(g)\mathbf{e}_i$ in terms of the original basis, and hence obtain representation D .
8. Consider the symmetry group, C_{4v} , of the square, consisting of rotations about the axis perpendicular to the square, and reflections about the vertical, horizontal, and diagonal axes in the plane of the square (but no mirror plane in the plane of the square).
 - (a) Construct a suitable set of irreducible representations of C_{4v} . That is, up to equivalence, construct all of the irreducible representations of this group.

Solution: The order of C_{4v} is $h = 8$, with $n_c = 5$ classes and elements:

- i. $C_1 = \{e\}$, the identity.
- ii. $C_2 = \{M_x, M_y\}$, mirror planes perpendicular to the square, containing the horizontal (x) and vertical (y) symmetry axes of the square, respectively.
- iii. $C_3 = \{M_u, M_v\}$, mirror planes perpendicular to the square, containing the lines $x = y$ and $x = -y$, respectively.
- iv. $C_4 = \{R_+, R_-\}$, rotations by ± 90 degrees about the principal axis of the square.
- v. $C_5 = \{R\}$, rotation by 180 degrees about the principal axis of the square.

There are $n_r = n_c = 5$ irreps. We must have the sum of the squares of the dimensions equal to 8. Given that we know that the identity rep is of dimension 1, the possibilities are either $\{\ell_i\} = \{1, 1, 1, 1, 1, 1, 1, 1\}$ or $\{1, 1, 1, 1, 4\}$. We may apply the reasoning we used in the case of the square drumhead in class to obtain:

- i. $D^{(1)}(g) = 1, \forall g \in C_{4v}$.
- ii. $D^{(2)}(\{e, R, R_{\pm}\}) = 1, D^{(2)}(\{M_x, M_y, M_u, M_v\}) = -1$.
- iii. $D^{(3)}(\{e, R, M_x, M_y\}) = 1, D^{(3)}(\{R_{\pm}, M_u, M_v\}) = -1$.
- iv. $D^{(4)}(\{e, R, M_u, M_v\}) = 1, D^{(4)}(\{R_{\pm}, M_x, M_y\}) = -1$.
- v. $D^{(5)}(e) = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, D^{(5)}(M_x) = \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix},$
 $D^{(5)}(M_y) = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}, D^{(5)}(M_u) = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix},$
 $D^{(5)}(M_v) = \begin{pmatrix} 0 & -1 \\ -1 & 0 \end{pmatrix}, D^{(5)}(R) = \begin{pmatrix} -1 & 0 \\ 0 & -1 \end{pmatrix},$
 $D^{(5)}(R_{\pm}) = \begin{pmatrix} 0 & \mp 1 \\ \pm 1 & 0 \end{pmatrix}.$

- (b) Give the character table for C_{4v} .

Solution: Just take the traces of the irreps in part (a) to get the character table. The first four columns are just the irreps themselves.

9. In problem 6 you consider the reducibility of a two-dimensional representation of the group of rotations in two dimensions. We may remark that this is an abelian group. Let us generalize that result: Consider a group, G , with a unitary representation D , consisting of unitary matrices $D(g)$, $g \in G$. If G is an abelian group, show that any such representation is, by a similarity transformation, equivalent to a representation by diagonal matrices (i.e., by a direct sum of 1×1 matrices). Note that we have already used group theory (Schur's lemma) to argue the truth of this. In this problem, I want you to use what you know about matrix theory to demonstrate the result.
10. Construct the character table for the tetrahedral symmetry group T_d . You may wish to keep a copy of your result for problem 12.

11. Let's take a peek at the relation of irreducible representations and the invariant subspaces of a vector space: Let V be the 6-dimensional function space consisting of polynomials of degree 2 in the two real variables x and y :

$$f(x, y) = ax^2 + bxy + cy^2 + dx + ey + h, \quad (3.10)$$

where a, b, c, d, e, h are complex numbers. If x, y transforms under the dihedral group D_3 (problem 7) as the coordinates of a 2-vector, then we obtain a 6-dimensional representation of D_3 on V . Identify the invariant subspaces of V under D_3 , and the corresponding irreducible representations contained in this six dimensional representation (don't be afraid to use your intuition to make sure that what you find is sensible).

12. At last we are ready for a real physics application of group theory. We looked at the example of masses joined by springs in the shape of an equilateral triangle in this note. Now, let us consider the problem of four masses joined by springs. The four masses are at the corners of a tetrahedron, and the springs form the edges of the tetrahedron. Thus, there are six springs connecting the four masses. All four masses are equal, and all six springs are identical.

We wish to determine the frequencies of the normal modes for this system. Notice that to solve the secular equation, $|V - \lambda I| = 0$, presents a formidable image. A little physical intuition can reduce it somewhat, but it would take real cleverness to solve it completely. This cleverness comes in the form of group theory! Group theory permits one to incorporate in a systematic and deliberate way everything we know about the symmetry of the problem, hence reducing it to a simpler problem.

The problem is still not trivial – you should spend time thinking about convenient approaches in setting things up, and about ways to avoid doing unnecessary work. Above all, be careful, and check your results as you proceed. You already obtained the character table for the tetrahedral symmetry group in problem 10. This problem takes you the rest of the way through solving for the frequencies of the normal modes.

- (a) First step: Set up a 12-dimensional vector (coordinate system) describing the system, and derive the equations of motion, arriving finally at a set of linear equations that could be solved, in principle, to yield the frequencies of the normal modes.
- (b) Second step: Obtain the character table for the twelve-dimensional representation of the tetrahedral symmetry group that acts on your 12-dimensional vector describing the system. Decompose this representation into irreducible representations.
- (c) Final step: Obtain a small number of trace equations which you can use to solve to obtain the frequencies of the normal modes. Give the frequencies of the normal modes, and their degeneracies. Do your answers make physical sense?

13. The “quaternion” group consists of eight elements,

$$Q = \{1, -1, i, -i, j, -j, k, -k\}, \quad (3.11)$$

with multiplication table defined by (q is any element of Q):

$$\begin{aligned} 1q &= q \\ (-1)^2 &= 1 \\ (-1)q = q(-1) &= -q \\ i^2 = j^2 = k^2 = ijk &= -1 \end{aligned} \quad (3.12)$$

Find the character table for this group. Compare this character table with the character table for dihedral group D_4 . Are these two groups isomorphic?

14. As a follow-on to the drumhead example in this note, consider the symmetry group of the regular pentagon, as given by a five-fold axis and several mirror planes. Do not include the mirror plan containing the plane of the pentagon itself (although you may amuse yourself by considering what happens if you add this operation).
- List the group elements. Denote rotations with R 's, and mirror operations with M 's. Draw a picture! List the classes.
 - Construct the character table for the irreducible representations of this group.
 - Consider the mode of oscillation of a pentagonal drumhead where a nodal line extends from a vertex to the midpoint of the opposite side. Define (with pictures) a basis for the space generated by this mode and its degenerate partners. Give an explicit matrix for one element of each class of the group for the representation of the pentagonal symmetry group that is generated by these degenerate modes.
 - Decompose the representation found in part (c) into irreducible representations.
15. We would like to consider the (qualitative) effects on the energy levels of an atom which is moved from freedom to an external potential (a crystal, say) with cubic symmetry. Let us consider a one-electron atom and ignore spin for simplicity. Recall that the wave function for the case of the free atom looks something like $R_{nl}(r)Y_{lm}(\theta, \phi)$, and that all states with the same n and l quantum numbers have the same energy, *i.e.*, are $(2l + 1)$ -fold degenerate. The Hamiltonian for a free atom must have the symmetry of the full rotation group, as there are no special directions. Thus, we recall some properties of this group for the present discussion. First, we remark that the set of functions $\{Y_{lm} : m = -l, -l + 1, \dots, l - 1, l\}$ for a given l forms the basis for a $(2l + 1)$ -dimensional subspace which is invariant under the operations of the full rotation group. [A set $\{\psi_i\}$ of vectors is

said to span an *invariant subspace* V_s under a given set of operations $\{P_j\}$ if $P_j\psi_i \in V_s \forall i, j$.] Furthermore, this subspace is “irreducible,” that is, it cannot be split into smaller subspaces which are also invariant under the rotation group.

Let us denote the linear transformation operator corresponding to element R of the rotation group by the symbol \hat{P}_R , *i.e.*:

$$\hat{P}_R f(\vec{x}) = f(R^{-1}\vec{x})$$

The way to think about this equation is to regard the left side as giving a “rotated function,” which we evaluate at point \vec{x} . The right side tells us that this is the same as the original function evaluated at the point $R^{-1}\vec{x}$, where R^{-1} is the inverse of the rotation matrix corresponding to rotation R . Since $\{Y_{lm}\}$ forms an invariant subspace, we must have:

$$\hat{P}_R Y_{lm} = \sum_{m'=-l}^l Y_{lm'} D^l(R)_{m'm}$$

The expansion coefficients, $D^l(R)_{m'm}$, can be regarded as the elements of a matrix $D^l(R)$. As we have discussed in general, and as you may see more explicitly in the note on rotations in QM, D^l corresponds to an irreducible representation of the rotation group.

- (a) Prove, or at least make plausible, the fact that D^l is an irreducible representation of the rotation group. (Hint: You might show first that it is a representation and then show irreducibility by finding a contradiction with the supposition of reducibility).

Thus, for a free atom, we have that the degenerate eigenfunctions of a given energy must transform according to an irreducible representation of this group. If the eigenfunctions transform according to the l^{th} representation, the degeneracy of the energy level is $(2l + 1)$ (assuming no additional, “accidental” degeneracy).

- (b) We will need the character table of this group. Since all elements in the same class have the same character, we pick a convenient element in each class by considering rotations about the z -axis, $R = (\alpha, z)$ (means rotate by angle α about the z -axis). Thus:

$$\hat{P}_{(\alpha, z)} Y_{\ell m} = e^{-im\alpha} Y_{\ell m}$$

(which you should convince yourself of).

Find the character “table” of the rotation group, that is, find $\chi^\ell(\alpha)$, the character of representation D^ℓ for the class of rotations through angle α . If you find an expression for the character in the form of a sum, do the sum, expressing your answer in as simple a form as you can. Note that the answer is given in the text, just fill in the missing steps to your satisfaction.

- (c) Let us put our atom into a potential with cubic symmetry. Now the symmetry of the free Hamiltonian is broken, and we are left with the discrete symmetry of the cube. The symmetry group of proper rotations of the cube is a group of order 24 with 5 classes. Call this group “ O ”.

Construct the character table for O .

- (d) Consider in particular how the f -level ($l = 3$) of the free atom may split when it is placed in the “cubic potential”. The seven eigenfunctions which transform according to the irreducible representation D^3 of the full group will most likely not transform according to an irreducible representation of O . On the other hand, since the operations of O are certainly operations of D^3 , the eigenfunctions will generate some representation of O .

Determine the coefficients in the decomposition.

$$D^3 = a_1 O^1 \oplus a_2 O^2 \oplus a_3 O^3 \oplus a_4 O^4 \oplus a_5 O^5,$$

where O^i are the irreducible representations of O . Hence, show how the degeneracy of the 7-fold level may be reduced by the cubic potential. Give the degeneracies of the final levels.

Note that we cannot say anything here about the magnitude of any splittings (which could “accidentally” turn out to be zero!), or even about the ordering of the resulting levels – that depends on the details of the potential, not just its symmetry.

Chapter 4

Lie Groups and Lie Algebras

In this note we'll investigate two additional notions:

1. The addition of a continuity structure on the group;
2. The addition of an algebraic structure on the group.

The former is the subject of Lie groups, and the latter is the subject of Lie algebras. These are quite different concepts. However, we put them together here because in physics we are heavily concerned with the conjunction of the two ideas.¹

4.1 Lie Groups

Formally, we have

Def: A *Lie group* is a group, G , whose elements form an analytic manifold such that the composition $ab = c$ ($a, b, c \in G$) is an analytic mapping of $G \times G$ into G and the inverse $a \rightarrow a^{-1}$ is an analytic mapping of G into G .

That is, a Lie group is a group with a continuity structure: derivatives may be taken. Typically, we describe Lie groups by elements that are determined differentially by some set of continuously varying real parameters. If there are r such parameters, we have an “ r -parameter Lie group”.

We won't here develop the theory of Lie groups from an abstract level. Instead, we'll directly mostly think in terms of representations by matrices, where the matrices are specified by some number of continuously varying real parameters (up to possibly discrete points of discontinuity in some situations).

¹The reader may wish to refer back to the note on Hilbert Spaces from Ph 129a for some concepts.

As with finite groups, it is convenient when we can deal with unitary representations. This is guaranteed to be possible in the following case:

Theorem: Every finite-dimension representation of a compact Lie group is equivalent to a unitary representation, and is either irreducible or fully reducible.

By “compact” here we mean that the parameters that specify an element of the Lie group vary over a compact set (i.e., over a closed set of finite extent). The proof of this parallels the proof given for finite groups that we gave in the note on representation theory, but now using the notion of an invariant integration over the group. Compactness ensures that this integral will be finite.

The notions of compactness and invariance of the group integral are topological concepts. There is a further topological property we will sometimes assume, that the group is “connected”. By this, we mean that we can get to any element of the group from the identity via a sequence of small steps.

For some examples:

- The group $O^+(3)$ (representing proper rotations in three dimensions) is a compact, connected, 3-parameter Lie group.
- The group $O(3)$ (proper and improper rotations in three dimensions) is a compact, but not a connected group. It contains two disjoint categories of elements, those with determinant $+1$, and those with determinant -1 , and it is not possible to continuously go from one to the other. This may be regarded as the direct product group:

$$O(3) = O^+(3) \otimes \mathcal{I}, \quad (4.1)$$

where \mathcal{I} is the inversion group.

- The Lorentz group (of proper homogeneous Lorentz transformations) is connected, but not compact. This is a little more subtle – the lack of compactness is due to the fact that there is a limit point of a sequence of group elements that is not an element (consider a sequence of velocity boosts in which $v \rightarrow 1$).
- The improper, homogeneous Lorentz group is neither connected nor compact.

We will sometimes also restrict discussion to simple compact Lie groups, recalling that a simple group is one that contains no proper invariant subgroup.

If we have a compact Lie group, then we can define the invariant integral over the group and also work with unitary representations without loss of generality. The general orthogonality relation of finite groups may be generalized to include compact Lie groups. For unitary irreducible representations $D^{(i)}$ and $D^{(j)}$ we have:

$$\int_G D^{(i)}(g)_{\mu\nu} D^{(j)*}(g)_{\alpha\beta} \mu(dg) = \frac{1}{\ell_i} \delta_{ij} \delta_{\mu\alpha} \delta_{\nu\beta}. \quad (4.2)$$

We have assumed that the invariant integral over the group is normalized to one:

$$\int_G \mu(dg) = 1. \quad (4.3)$$

Let's consider an example. In the note on representation theory, we defined the spherical harmonic functions in terms of irreducible representations of the rotation group:

$$Y_{\ell m}(\theta, \phi) \equiv \sqrt{\frac{2\ell+1}{4\pi}} D_{m0}^{\ell*}(\phi, \theta, 0). \quad (4.4)$$

Suppose we wish to know the orthogonality properties of the $Y_{\ell m}$'s. We compute:

$$\begin{aligned} \int_{(4\pi)} Y_{\ell m}(\theta, \phi) Y_{\ell' m'}^*(\theta, \phi) d \cos \theta d\phi &= \quad (4.5) \\ \frac{\sqrt{(2\ell+1)(2\ell'+1)}}{4\pi} \int_{(4\pi)} D_{m0}^{\ell*}(\phi, \theta, 0) D_{m'0}^{\ell'}(\phi, \theta, 0) d \cos \theta d\phi \\ \frac{\sqrt{(2\ell+1)(2\ell'+1)}}{4\pi} \int_{(4\pi)} D_{m0}^{\ell*}(\phi, \theta, \alpha) D_{m'0}^{\ell'}(\phi, \theta, \alpha) d \cos \theta d\phi \\ \frac{\sqrt{(2\ell+1)(2\ell'+1)}}{8\pi^2} \int_{(8\pi^2)} D_{m0}^{\ell*}(\phi, \theta, \alpha) D_{m'0}^{\ell'}(\phi, \theta, \alpha) d \cos \theta d\phi d\alpha \end{aligned} \quad (4.6)$$

We have used here the invariance of the integral when adding the rotation by angle α about the x -axis, and averaging over this rotation. The result is now in the form of the general orthogonality relation:

$$\frac{1}{8\pi^2} \int_{(8\pi^2)} D_{mn}^{\ell*}(\phi, \theta, \alpha) D_{m'n'}^{\ell'}(\phi, \theta, \alpha) d \cos \theta d\phi d\alpha = \frac{1}{2\ell+1} \delta_{\ell\ell'} \delta_{mm'} \delta_{nn'}. \quad (4.7)$$

Therefore,

$$\int_{(4\pi)} Y_{\ell m}(\theta, \phi) Y_{\ell' m'}^*(\theta, \phi) d \cos \theta d\phi = \delta_{\ell\ell'} \delta_{mm'}. \quad (4.8)$$

A perhaps less-familiar but very important example may be found in classical mechanics: Consider a system with generalized coordinates $q_i, i = 1, 2, \dots, n$ and corresponding generalized momenta $p_i = \partial_{q_i} L$, where L is the Lagrangian. Hamilton's equations are:

$$\dot{p}_i = -\partial_{q_i} H, \quad (4.9)$$

$$\dot{q}_i = \partial_{p_i} H, \quad (4.10)$$

where H is the Hamiltonian. We may rewrite this in terms of the $2n$ -dimensional vector:

$$x \equiv \begin{pmatrix} q_1 \\ \vdots \\ q_n \\ p_1 \\ \vdots \\ p_n \end{pmatrix}, \quad (4.11)$$

as:

$$\dot{x} = J \frac{\partial H}{\partial x}, \quad (4.12)$$

with

$$J = \begin{pmatrix} 0 & I \\ -I & 0 \end{pmatrix}. \quad (4.13)$$

That is, J is a $2n \times 2n$ matrix written in terms of $n \times n$ submatrices 0 and I .

A canonical transformation is a transformation from x to y where

$$y = \begin{pmatrix} Q_1 \\ \vdots \\ Q_n \\ P_1 \\ \vdots \\ P_n \end{pmatrix}, \quad (4.14)$$

such that

$$\dot{y} = J \frac{\partial H[x(y)]}{\partial y}. \quad (4.15)$$

That is, Hamilton's equations are preserved under a canonical transformation.

We have

$$\dot{y}_i = \sum_j \frac{\partial y_i}{\partial x_j} \dot{x}_j, \quad (4.16)$$

which may be written in matrix form:

$$\dot{y} = M \dot{x}, \quad (4.17)$$

where

$$M_{ij} \equiv \frac{\partial y_i}{\partial x_j}. \quad (4.18)$$

Hence,

$$\dot{y} = M J \frac{\partial H}{\partial x}. \quad (4.19)$$

Now

$$\frac{\partial H}{\partial x_i} = \sum_j \frac{\partial H}{\partial y_j} \frac{\partial y_j}{\partial x_i} = \sum_j \frac{\partial H}{\partial y_j} M_{ji}, \quad (4.20)$$

or,

$$\frac{\partial H}{\partial x} = M^T \frac{\partial H}{\partial y}. \quad (4.21)$$

We conclude that

$$\dot{y} = MJM^T \frac{\partial H}{\partial y}, \quad (4.22)$$

and that the transformation is canonical if

$$MJM^T = J. \quad (4.23)$$

A matrix M which satisfies the condition of Eqn. 4.23 is said to be *symplectic*. The reader is encouraged to verify that the set of $2n \times 2n$ symplectic matrices forms a group, called the *symplectic group*, denoted $Sp(2n)$.

We remark that the evolution of the system in time corresponds to a sequence of canonical transformations, and hence the time evolution corresponds to the application of successive symplectic matrices. This finds practical application in various situations, for example in accelerator physics.

We turn now to another feature of unitary representations. Let U be a unitary matrix. Write

$$U = e^{iA} \equiv \sum_{n=0}^{\infty} \frac{(iA)^n}{n!}, \quad (4.24)$$

where we leave it to the reader to investigate convergence. Now,

$$\begin{aligned} U^{-1} &= U^\dagger = (e^{iA})^\dagger \\ &= \left[\sum_{n=0}^{\infty} \frac{(iA)^n}{n!} \right]^\dagger \\ &= \sum_{n=0}^{\infty} \frac{[(-iA^*)^n]^T}{n!} \\ &= \sum_{n=0}^{\infty} \frac{[(-iA^\dagger)^n]}{n!} \\ &= e^{-iA^\dagger}. \end{aligned} \quad (4.25)$$

But we also know that,

$$U^{-1} = e^{-iA}, \quad (4.26)$$

since A commutes with itself, and hence exponentials of multiples of A may be treated like ordinary numbers in products. Therefore, we may take $A = A^\dagger$. That is, A is a hermitian matrix.

Note that if we also have $\det U = 1$, then A can be taken to be traceless: The matrix A is hermitian, hence diagonalizable by a unitary transformation. Let

$$\Delta = SAS^{-1} = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_n), \quad (4.27)$$

be a diagonal equivalent of A , where S is unitary. Then, $A = S^{-1}\Delta S$, or

$$\begin{aligned}
 1 &= \det \left(e^{iS^{-1}\Delta S} \right) \\
 &= \det \sum_{k=0}^{\infty} \frac{1}{k!} (iS^{-1}\Delta S)^k \\
 &= \det S^{-1} \left[\sum_{k=0}^{\infty} \frac{1}{k!} (i\Delta)^k \right] S \\
 &= \det(S^{-1})\det(S)\det e^{i\Delta} \\
 &= \exp \left(i \sum_{j=1}^n \lambda_j \right). \tag{4.28}
 \end{aligned}$$

Thus, the sum of the eigenvalues is equal to $2\pi m$, where m is an integer. Notice that if $m \neq 0$, we can define a new diagonal matrix $\Delta' = \Delta - 2\pi m\delta_{11}$, where δ_{11} is the matrix with the $i, j = 1, 1$ element equal to one, and all other elements zero. The trace of Δ' is zero. Hence $A' \equiv S^{-1}\Delta'S$ is also traceless. But $\exp(i\Delta') = \exp(i\Delta)$, and therefore

$$U = S^{-1}e^{i\Delta'}S = \exp(iS^{-1}\Delta'S) = e^{iA'}, \tag{4.29}$$

where A' is hermitian and traceless.

Suppose D is a unitary representation of a group G . Then the elements of the group representation may be written in the form:

$$D(g) = \exp [i\epsilon^\alpha(g)X_\alpha], \tag{4.30}$$

where the summation convention on repeated indices is used, $\{X_\alpha\}$ is a set of constant hermitian matrices, and $\{\epsilon_\alpha\}$ is a set of real parameters.

We are in particular concerned here with Lie groups (with unitary representations assumed here). In that case, if G is an r -parameter Lie group, we can find a set of r matrices $X_\alpha, \alpha = 1, 2, \dots, r$ such that Eqn. 4.30 holds. We refer to these matrices as the *infinitesimal generators* of the group. In this case, we have the “fundamental theorem of Lie”:

Theorem: The local structure of a Lie group is completely specified by the commutation relations among the generators X_α :

$$[X_\alpha, X_\beta] = C_{\alpha\beta}^\gamma X_\gamma, \quad \alpha, \beta = 1, 2, \dots, r, \tag{4.31}$$

where the coefficients $C_{\alpha\beta}^\gamma$ (called the *structure constants* of the group) are independent of the representation.

We investigate the proof of this, or rather of the Baker-Campbell-Hausdorff theorem, in exercise 6.

The reader is encouraged to check that the structure constants must satisfy:

$$C_{\alpha\beta}^\gamma = -C_{\beta\alpha}^\gamma, \tag{4.32}$$

and (with summation convention over repeated indices)

$$C_{\alpha\beta}^{\delta}C_{\delta\gamma}^{\epsilon} + C_{\gamma\alpha}^{\delta}C_{\delta\beta}^{\epsilon} + C_{\beta\gamma}^{\delta}C_{\delta\alpha}^{\epsilon} = 0. \quad (4.33)$$

The matrices X_{α} may be regarded as operators on a vector space. If we are doing quantum mechanics, and we have a hermitian set of operators, they correspond to observables.

The commutator may be regarded as defining a kind of product, and the matrices $\{X_{\alpha}\}$ as generating a vector space, which is closed under this product. This brings us to the subject of Lie algebras, in the next section.

4.2 Lie Algebras

In the discussion of infinite groups of relevance to physics (in particular, Lie groups), it is useful to work in the context of a richer structure called an algebra. For background, we start by giving some mathematical definitions of the underlying structures:

Def: A **ring** is a triplet $\langle R, +, \circ \rangle$ consisting of a non-empty set of elements (R) with two binary operations ($+$ and \circ) such that:

1. $\langle R, + \rangle$ is an abelian group.
2. R is closed under \circ .
3. (\circ) is associative.
4. Distributivity holds: for any $a, b, c \in R$

$$a \circ (b + c) = a \circ b + a \circ c \quad (4.34)$$

and

$$(b + c) \circ a = b \circ a + c \circ a \quad (4.35)$$

Conventions:

We use 0 (“zero”) to denote the identity of $\langle R, + \rangle$. We speak of $(+)$ as addition and of (\circ) as multiplication, typically omitting the (\circ) symbol entirely (*i.e.*, $ab \equiv a \circ b$).

Def: A ring is called a **field** if the non-zero elements of R form an abelian group under (\circ) .

Def: An abelian group $\langle V, \oplus \rangle$ is called a **vector space** over a field $\langle F, +, \circ \rangle$ by a scalar multiplication $(*)$ if for all $a, b \in F$ and $v, w \in V$:

1. $a * (v \oplus w) = (a * v) \oplus (a * w)$ distributivity
2. $(a + b) * v = (a * v) \oplus (b * v)$ distributivity
3. $(a \circ b) * v = a * (b * v)$ associativity
4. $1 * v = v$ unit element ($1 \in F$)

Conventions:

We typically refer to elements of V as “vectors” and elements of F as “scalars.” We typically use the symbol $+$ for addition both of vectors and scalars. We also generally omit the $*$ and \circ multiplication symbols. Note that this definition is an abstraction of the definition of vector space given in the note on Hilbert spaces, page 6.

Def: An **algebra** is a vector space V over a field F on which a multiplication (\times) between vectors has been defined (yielding a vector in V) such that for all $u, v, w \in V$ and $a \in F$:

1. $(au) \times v = a(u \times v) = u \times (av)$
2. $(u+v) \times w = (u \times w) + (v \times w)$ and $w \times (u+v) = (w \times u) + (w \times v)$

(Once again, we often omit the multiplication sign, and hope that it is clear from context which quantities are scalars and which are vectors.)

We are sometimes interested in the following types of algebras:

Def: An algebra is called **associative** if the multiplication of vectors is associative.

We may construct the idea of a “group algebra”: Let G be a group, and V be a vector space over a field F , of dimension equal to the order of G (possibly ∞). Denote a basis for V by the group elements. We can now define the multiplication of two vectors in V by using the group multiplication table as “structure constants”: Thus, if the elements of G are denoted by g_i , a vector $u \in V$ may be written:

$$u = \sum a_i g_i$$

We require that, at most, a finite number of coefficients a_i are non-zero. The multiplication of two vectors is then given by:

$$\left(\sum a_i g_i \right) \left(\sum b_j g_j \right) = \sum \left(\sum_{g_i g_j = g_k} a_i b_j \right) g_k$$

[Since only a finite number of the $a_i b_j$ can be non-zero, the sum $\sum_{g_i g_j = g_k} a_i b_j$ presents no problem, and furthermore, we will have closure under multiplication.]

Since group multiplication is associative, our group algebra, as we have constructed it, is an associative algebra.

We note that an associative algebra is, in fact, a ring. Note also that the multiplication of vectors is not necessarily commutative. An important non-associative algebra is:

Def: A **Lie algebra** is an algebra in which the multiplication of vectors obeys the further properties (letting u, v, w be any vectors in V):

1. Anticommutivity: $u \times v = -v \times u$.
2. Jacobi Identity: $u \times (v \times w) + w \times (u \times v) + v \times (w \times u) = 0$.

We concentrate on Lie algebras henceforth in this note, in particular on Lie algebras associated with a Lie group. The generators, $\{X_\alpha\}$, of a Lie group generate a Lie algebra, where multiplication of vectors is defined as the commutator. Just as for groups, we have the notion of a *regular representation* (or also “adjoint representation”) of the Lie algebra. We may rewrite the identity for the structure constants:

$$C_{\alpha\beta}^\delta C_{\delta\gamma}^\epsilon + C_{\gamma\alpha}^\delta C_{\delta\beta}^\epsilon + C_{\beta\gamma}^\delta C_{\delta\alpha}^\epsilon = 0. \quad (4.36)$$

in the suggestive form:

$$C_{\alpha\beta}^\delta (C_\delta)_\gamma^\epsilon + (-C_\beta)_\delta^\epsilon (-C_\alpha)_\gamma^\delta + (-C_\alpha)_\delta^\epsilon (C_\beta)_\gamma^\delta = 0. \quad (4.37)$$

Interpreting, e.g., C_α as a matrix with elements $(C_\alpha)_\delta^\epsilon$, where δ is the column index, we find:

$$[C_\alpha, C_\beta] = C_{\alpha\beta}^\delta C_\delta. \quad (4.38)$$

The matrices C_α formed from the structure constants have the same commutation relations as the generators X_α of the Lie group, and hence form a *representation of the Lie algebra*, called the regular or adjoint representation.

The problem of classifying Lie groups is essentially the problem of finding the numbers $\{C\}$ satisfying the requirements of Eqns. 4.32 and 4.33 above, and then finding the r constant matrices which satisfy the commutation relations. This problem was solved by Cartan in 1913. We list the simple Lie groups here:

The “classical Lie groups” are (except as noted, $\ell = 1, 2, \dots$):

1. The group of unitary unimodular (i.e., determinant equal to one) $(\ell + 1) \times (\ell + 1)$ matrices, denoted A_ℓ or $SU(\ell + 1)$. This is an $\ell(\ell + 2)$ -parameter Lie group, as the reader is encouraged to demonstrate.
2. The group of orthogonal unimodular $(2\ell + 1) \times (2\ell + 1)$ matrices, denoted B_ℓ or $SO(2\ell + 1)$ or $O^+(2\ell + 1)$. This is an $\ell(2\ell + 1)$ -parameter Lie group, as the reader is encouraged to demonstrate.
3. The group of orthogonal unimodular $(2\ell) \times (2\ell)$ matrices, for $\ell > 2$, denoted D_ℓ or $SO(2\ell)$ or $O^+(2\ell)$. This is an $\ell(2\ell - 1)$ -parameter Lie group, as the reader is encouraged to demonstrate. It may be noted that for $\ell \leq 2$ the group is not simple.
4. The group of symplectic $(2\ell) \times (2\ell)$ matrices, denoted C_ℓ or $Sp(2\ell)$. This is an $\ell(2\ell + 1)$ -parameter Lie group, as the reader is encouraged to demonstrate.

In addition, there are five “exceptional groups”: G_4 with 14 parameters, F_4 with 52 parameters, E_6 with 78 parameters, E_7 with 133 parameters, and E_8 with 248 parameters.

Consider briefly the example of the rotation group and associated Lie algebra in quantum mechanics.² In three dimensions, a rotation about the $\hat{\alpha}$ unit axis by angle ϕ can be expressed in the form:

$$R_{\hat{\alpha}}(\phi) = e^{-i\beta \cdot T}, \quad (4.39)$$

where $\beta \cdot T \equiv \beta_x T_x + \beta_y T_y + \beta_z T_z$, $\beta = \beta(\hat{\alpha}, \phi)$, and $T_{x,y,z}$ are the infinitesimal generators of rotations in three dimensions:

$$T_x \equiv \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & -i \\ 0 & i & 0 \end{pmatrix}, T_y \equiv \begin{pmatrix} 0 & 0 & i \\ 0 & 0 & 0 \\ -i & 0 & 0 \end{pmatrix}, T_z \equiv \begin{pmatrix} 0 & -i & 0 \\ i & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}. \quad (4.40)$$

We may consider the application of successive rotations (which must be a rotation):

$$\begin{aligned} e^{-i\alpha \cdot T} e^{-i\beta \cdot T} &= e^{-i\gamma \cdot T} \\ &= \sum_{m=0}^{\infty} \frac{(-i\alpha \cdot T)^m}{m!} \sum_{n=0}^{\infty} \frac{(-i\beta \cdot T)^n}{n!} \\ &= 1 - i(\alpha + \beta) \cdot T + \frac{(-i\alpha \cdot T)^2}{2!} + \frac{(-i\beta \cdot T)^2}{2!} + (-i\alpha \cdot T)(-i\beta \cdot T) + O[(\alpha, \beta)^3] \\ &= 1 - i(\alpha + \beta) \cdot T + \frac{[-i(\alpha + \beta) \cdot T]^2}{2!} - \frac{[\alpha \cdot T, \beta \cdot T]}{2!} + O[(\alpha, \beta)^3] \\ &= \exp \left\{ -i(\alpha + \beta) \cdot T - \frac{[\alpha \cdot T, \beta \cdot T]}{2!} + O[(\alpha, \beta)^3] \right\}. \end{aligned} \quad (4.41)$$

Thus, to this order in the expansion, we need to have the values of commutators such as $[T_x, T_y]$, but not of products $T_x T_y$. This statement is true to all orders, as stated in the celebrated Campbell-Baker-Hausdorff theorem. Hence, every order is linear in the T 's, and therefore γ exists. This is also why we can learn most of what we need to know about Lie groups by studying the commutation relations of the generators, as indicated in the general ‘‘fundamental theorem of Lie’’.

It may be remarked that for a general, abstract Lie algebra, we should not even think of the product $[A, B]$ as $AB - BA$, since the product AB may not be defined, while the ‘‘Lie product’’ denoted $[A, B]$ may be. Of course, if we have a matrix representation for the generators, then AB is defined. In physics we typically deal with matrix representations, so referring to the Lie product as a commutator is justified.

For our three-dimensional rotation generators, the Lie products are found by evaluating the commutation relations of the matrices, with the result:

$$[T_\alpha, T_\beta] = i\epsilon_{\alpha\beta\gamma} T_\gamma, \quad (4.42)$$

²Again, this example is considerably expanded upon in the note on the rotation group in quantum mechanics, linked to the Ph 129 web page.

where $\epsilon_{\alpha\beta\gamma}$ is the “antisymmetric tensor” (in three dimensions), or “Levi-Civita antisymmetric symbol”, defined by:

$$\epsilon_{\alpha\beta\gamma} \equiv \begin{cases} +1 & \alpha, \beta, \gamma \text{ an even permutation of } 1, 2, 3, \\ -1 & \alpha, \beta, \gamma \text{ an odd permutation of } 1, 2, 3, \\ 0 & \text{otherwise.} \end{cases} \quad (4.43)$$

With these commutation relations, we may define an abstract Lie algebra, with generators (basis vectors) t_1, t_2, t_3 satisfying the Lie products:

$$[t_\alpha, t_\beta] = i\epsilon_{\alpha\beta\gamma}t_\gamma, \quad (4.44)$$

We complete the Lie algebra by considering linear combinations of the t 's, requiring:

$$[a \cdot t + b \cdot t, c \cdot t] = [a \cdot t, c \cdot t] + [b \cdot t, c \cdot t] \quad (4.45)$$

and

$$[a \cdot t, b \cdot t] = -[b \cdot t, a \cdot t]. \quad (4.46)$$

Our Lie algebra satisfies the Jacobi identity:

$$[a \cdot t, [b \cdot t, c \cdot t]] + [b \cdot t, [c \cdot t, a \cdot t]] + [c \cdot t, [a \cdot t, b \cdot t]] = 0. \quad (4.47)$$

The matrices T_x, T_y, T_z generate a *representation* of this Lie algebra with *dimension* three, since the matrices are 3×3 and hence operators on a 3-dimensional vector space. We note that the vector space of the Lie algebra itself is also three-dimensional, but this is not required, and the two vector spaces should not be confused.

Recalling quantum mechanics, we know that it is useful to define

$$t_+ \equiv t_1 + it_2 \quad (4.48)$$

$$t_- \equiv t_1 - it_2. \quad (4.49)$$

We may obtain the commutation relations

$$[t_3, t_+] = t_+ \quad (4.50)$$

$$[t_3, t_-] = -t_- \quad (4.51)$$

$$[t_+, t_-] = 2t_3. \quad (4.52)$$

We suppose that the t 's are represented by linear transformations, J , acting on some vector space V , where V is of finite dimension, but not necessarily three dimensions. We make the correspondence $t_\pm \rightarrow J_\pm$, $t_3 \rightarrow J_3$. Since none of these generators commute, only one of J_\pm, J_3 can be diagonalized at a time. We have the definition:

Def: The number of generators of a Lie algebra that can simultaneously be “diagonalized” is called the *rank* of the Lie group.

Thus, the rotation group is of rank 1.

We pick J_3 to be in diagonal form with respect to some basis $\{v\}$. We label the basis vectors by the diagonal element (eigenvalue) k :

$$J_3 v_k = k v_k. \quad (4.53)$$

By repeated action of J_{\pm} on v_k it may be demonstrated that k is either integer or $\frac{1}{2}$ -integer, with some maximal value j , and the eigenvalues of J_3 are $-j, -j+1, \dots, j$. This demonstration is commonly performed in quantum mechanics courses. There are $2j+1$ distinct eigenvalues, so the dimension of our representation is $\geq 2j+1$. If we define our space to be the space spanned by $\{v_k, k = -j, -j+1, \dots, j\}$ then our space is said to be *irreducible* – there is no proper subspace of V which is mapped onto itself by the various J 's.

As remarked earlier, for a compact Lie group we may find a unitary representation, and hence we may represent the generators of the associated Lie algebra by hermitian matrices. Assuming we have done so, we find

$$\begin{aligned} [X_{\alpha}, X_{\beta}]^{\dagger} &= (X_{\alpha} X_{\beta} - X_{\beta} X_{\alpha})^{\dagger} \\ &= X_{\beta}^{\dagger} X_{\alpha}^{\dagger} - X_{\alpha}^{\dagger} X_{\beta}^{\dagger} \\ &= X_{\beta} X_{\alpha} - X_{\alpha} X_{\beta} \\ &= -[X_{\alpha}, X_{\beta}]. \end{aligned} \quad (4.54)$$

We thus have

$$\begin{aligned} C_{\alpha\beta}^{\delta*} X_{\delta}^{\dagger} &= C_{\alpha\beta}^{\delta*} X_{\delta} \\ &= [X_{\alpha}, X_{\beta}]^{\dagger} \\ &= -[X_{\alpha}, X_{\beta}] \\ &= -C_{\alpha\beta}^{\delta} X_{\delta}. \end{aligned} \quad (4.55)$$

That is, $C_{\alpha\beta}^{\delta*} = -C_{\alpha\beta}^{\delta}$, and the structure constants are thus pure imaginary for a unitary representation.

We may introduce the concept of an operator for “raising and lowering indices” or a “metric tensor”, by defining:

$$g_{\mu\nu} = g_{\nu\mu} \equiv C_{\mu\alpha}^{\beta} C_{\nu\beta}^{\alpha}. \quad (4.56)$$

It may be shown that for a semi-simple Lie group $\det g \neq 0$, where g is the matrix formed by the elements $g_{\mu\nu}$. Thus, in this case, g has an inverse, which we define by:

$$g^{\mu\nu} g_{\nu\rho} = \delta_{\rho}^{\mu}, \quad (4.57)$$

where we have written the Kronecker function with one index raised.

The metric tensor may be used for raising or lowering indices, for example:

$$g^{\alpha\beta} g^{\mu\nu} g_{\nu\beta} = g^{\alpha\beta} \delta_{\beta}^{\mu} = g^{\alpha\mu}. \quad (4.58)$$

We have here “raised” the indices on $g_{\nu\beta}$. In general, given a quantity with lower indices, we may define a corresponding quantity with upper indices according to:

$$A^\alpha \equiv g^{\alpha\beta} A_\beta. \quad (4.59)$$

Or, given a quantity with raised indices, we may define a corresponding quantity with lower indices:

$$A_\alpha \equiv g_{\alpha\beta} A^\beta. \quad (4.60)$$

In particular, we may define structure constants with all lower indices:

$$C_{\alpha\beta\gamma} = C_{\alpha\beta}^\delta g_{\delta\gamma}. \quad (4.61)$$

The $C_{\alpha\beta\gamma}$ so defined is antisymmetric under interchange of any pair of indices. Note that, if $C_{\alpha\beta}^\delta$ is pure imaginary, then g is real, and $C_{\alpha\beta\gamma}$ is pure imaginary.

Now consider the quantity

$$F \equiv g_{\alpha\beta} X^\alpha X^\beta = X^\alpha X_\alpha = X_\alpha X^\alpha, \quad (4.62)$$

where the X_α are the infinitesimal generators of the Lie algebra. Consider the commutator of F with any generator:

$$\begin{aligned} [F, X_\gamma] &= g^{\alpha\beta} [X_\alpha X_\beta, X_\gamma] \\ &= g^{\alpha\beta} \{X_\alpha [X_\beta, X_\gamma] + [X_\alpha, X_\gamma] X_\beta\} \\ &= g^{\alpha\beta} (C_{\beta\gamma}^\delta X_\alpha X_\delta + C_{\alpha\gamma}^\delta X_\delta X_\beta) \\ &= g^{\alpha\beta} C_{\beta\gamma}^\delta X_\alpha X_\delta + g^{\beta\alpha} C_{\beta\gamma}^\delta X_\delta X_\alpha \\ &= g^{\alpha\beta} C_{\beta\gamma}^\delta (X_\alpha X_\delta + X_\delta X_\alpha) \\ &= g^{\alpha\beta} g^{\delta\epsilon} C_{\beta\gamma\epsilon} (X_\alpha X_\delta + X_\delta X_\alpha) \\ &= C_{\beta\gamma\epsilon} (X^\beta X^\epsilon + X^\epsilon X^\beta) \\ &= C_{\epsilon\gamma\beta} (X^\epsilon X^\beta + X^\beta X^\epsilon) \\ &= -C_{\beta\gamma\epsilon} (X^\epsilon X^\beta + X^\beta X^\epsilon) \\ &= 0, \end{aligned} \quad (4.63)$$

since it is equal to its negative. Thus, F commutes with every generator, hence commutes with every element of the algebra. By Schur’s lemma, F must be a multiple of the identity, since if F commutes with every generator, then it must commute with every element of the group in some irreducible representation. An operator which commutes with every generator is known as a *Casimir operator*.

For example, consider again the rotation group in quantum mechanics. The structure constants are

$$C_{\alpha\beta}^\gamma = i\epsilon_{\alpha\beta\gamma}. \quad (4.64)$$

The metric tensor is thus

$$\begin{aligned} g_{\mu\nu} &= C_{\mu\alpha}^\beta C_{\nu\beta}^\alpha \\ &= -\epsilon_{\mu\alpha\beta} \epsilon_{\nu\beta\alpha} \\ &= 2\delta_{\mu\nu}. \end{aligned} \quad (4.65)$$

Hence, $J^\alpha = 2J_\alpha$, and $J^2 = J_1^2 + J_2^2 + J_3^2$ is a Casimir operator, a multiple of the identity. To determine the multiple, we consider the action of J^2 on a basis vector. This may be accomplished by writing it in the form $J^2 = J_z^2 + \frac{1}{2}(J_+J_- + J_-J_+)$, where $J_\pm \equiv J_x \pm iJ_y$. This exercise yields the familiar result

$$J^2 v_k = j(j+1)v_k, \quad (4.66)$$

where $2j+1$ is the dimension of the representation. Thus,

$$J^2 = j(j+1)I. \quad (4.67)$$

4.3 Example: $SU(3)$

The group $SU(3)$ consists of the set of unitary unimodular 3×3 matrices. In the exercises, you show that it is an eight parameter group. Thus, we know that the associated Lie algebra must have eight linearly independent generators. That is, we wish to find a set of eight linearly independent traceless hermitian 3×3 matrices. It is readily demonstrated that the vector space of such matrices is in fact eight dimensional, that is, our generators provide a basis for the vector space of traceless hermitian 3×3 matrices.

There are many ways we could pick our basis for the Lie algebra. However, it is generally wise to make as many as possible diagonal. In this case, there are three linearly-independent 3×3 diagonal hermitian matrices, but the traceless requirement reduces these to only two. The number of simultaneously diagonalizable generators is called the *rank* of the Lie algebra, hence $SU(3)$ is rank two.

A common choice for the generators, with two diagonal generators, is the ‘‘Gell-Mann matrices’’:

$$\lambda_1 = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \quad \lambda_2 = \begin{pmatrix} 0 & -i & 0 \\ i & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \quad \lambda_3 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \quad (4.68)$$

$$\lambda_4 = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix}, \quad \lambda_5 = \begin{pmatrix} 0 & 0 & -i \\ 0 & 0 & 0 \\ i & 0 & 0 \end{pmatrix}, \quad (4.69)$$

$$\lambda_6 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}, \quad \lambda_7 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & -i \\ 0 & i & 0 \end{pmatrix}, \quad \lambda_8 = \frac{1}{\sqrt{3}} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -2 \end{pmatrix}. \quad (4.70)$$

Notice the $SU(2)$ substructure. For example, the upper left 2×2 submatrices of λ_1, λ_2 , and λ_3 are just the Pauli matrices. The group $SU(3)$ contains subgroups isomorphic with $SU(2)$ (but not invariant subgroups).

One area where $SU(3)$ plays an important role is in the ‘‘Standard Model’’ – $SU(3)$ is the ‘‘gauge group’’ of the strong interaction (Quantum Chromodynamics). In this case, the group elements describe transformations in ‘‘color’’ space,

where color is the analog of charge in the strong interaction. Instead of the single dimension of electromagnetic charge, color space is three-dimensional. The $SU(3)$ symmetry reflects the fact that all colors couple with the same strength – there is no preferred “direction” in color space. In field theory, once the gauge symmetry is specified, the form of the interaction is determined.

There is another example in particle physics where $SU(3)$ enters. Instead of the color symmetry just discussed, there is a “flavor” symmetry. The three lightest quarks are called “up” (u), “down” (d), and “strange” (s). The quantum number that distinguishes these is called flavor. The strong interaction couples with the same strength to each flavor. Thus, we may make “rotations” in this three-dimensional flavor space without changing the interaction. These rotations are described by the elements of $SU(3)$. The symmetry is actually broken, because the u , d , and s quarks have different masses (also, the electromagnetic and weak interaction couplings depend on flavor), but it is still a useful approximation in many situations. We’ll develop this application somewhat further here.

We use the Gell-Mann representation, in which λ_3 and λ_8 are the diagonal generators. According to the assumption of $SU(3)$ flavor symmetry, our operators in flavor space commute with the Hamiltonian. We’ll label our quark flavor basis according to the eigenvalues of λ_3 and λ_8 . It is conventional to notice the $SU(2)$ substructure of $(\lambda_1, \lambda_2, \lambda_3)$ and refer to the two-dimensional operations of these generators as operations on “isospin” (short for isotopic spin) space. This is the ordinary nuclear isospin. It really doesn’t have anything to do with angular momentum, but gets its “spin” nomenclature from the analogy with angular momentum where $SU(2)$ also enters. By analogy with angular momentum, a two-dimensional representation gets “third-component” quantum numbers of $\pm 1/2$. That is, we define, in this representation,

$$I_3 = \frac{1}{2}\lambda_3 = \frac{1}{2} \begin{pmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 0 \end{pmatrix}. \quad (4.71)$$

The eigenstates with $I_3 = +1/2, -1/2$ are called the u quark and the d quark, respectively. The strange quark in this convention has $I_3 = 0$, it is an $I = 0$ state (an isospin “singlet”). Note that this three-dimensional representation of $SU(2)$ is reducible to two-dimensional and one-dimensional irreps.

For the other quantum number, we define the “hypercharge” operator, in this representation:

$$Y = \frac{1}{\sqrt{3}}\lambda_8 = \frac{1}{3} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -2 \end{pmatrix}. \quad (4.72)$$

Thus, the u and d quarks both have $Y = 1/3$, and the s quark has $Y = -2/3$. The basis for this three-dimensional representation of flavor $SU(3)$ is illustrated in Fig. 4.1.

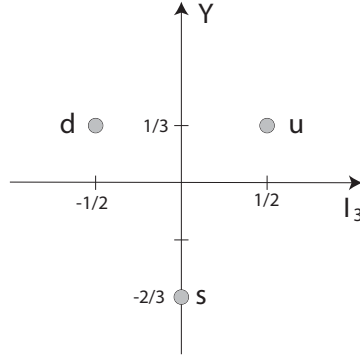


Figure 4.1: The $\mathbf{3}$ representation of $SU(3)$, in the context of quark flavors.

Now, we can also generate additional representations of $SU(3)$, and interpret in this physical context. Under complex conjugation of an element of $SU(3)$,

$$U = e^{i\epsilon^\alpha \lambda_\alpha} \rightarrow U^* = e^{-i\epsilon^\alpha \lambda_\alpha^*}. \quad (4.73)$$

This generates a new three-dimensional representation, called $\bar{\mathbf{3}}$. The I_3 and Y quantum numbers switch signs. Thus, the diagram for $\bar{\mathbf{3}}$ looks like the diagram for $\mathbf{3}$ reflected through the origin. We label the states $\bar{u}, \bar{d}, \bar{s}$, reflecting their interpretation as anti-quark states. Notice that the complex conjugate representation $\bar{\mathbf{3}}$ is not equivalent to the $\mathbf{3}$ representation. This is a difference from $SU(2)$, where the two representations ($\mathbf{2}$ and $\bar{\mathbf{2}}$) are equivalent.

We may also generate higher dimension representations of $SU(3)$ by forming direct product representations. Some of these have special interpretation in particle physics: Combining a quark with an anti-quark, that is, forming the $\mathbf{3} \otimes \bar{\mathbf{3}}$ representation, gives meson states. Combining three quarks, $\mathbf{3} \otimes \mathbf{3} \otimes \mathbf{3}$, gives baryons. As usual, these direct product representations may be expected to be reducible. For example, we have the reduction to irreducible representations: $\mathbf{3} \otimes \bar{\mathbf{3}} = \mathbf{8} \oplus \mathbf{1}$. We will discuss the graph in Fig. 4.2 in class.

4.4 Exercises

1. Show that $SU(n)$ requires $(n-1)(n+1)$ real parameters to specify an element.
2. Show that $C_{\alpha\beta\gamma}$ is antisymmetric under interchange of any pair of indices.
3. Show that the complex conjugate representation, $\bar{\mathbf{2}}$, of $SU(2)$ is equivalent to the original $\mathbf{2}$ representation.
4. Consider the Helmholtz equation in two dimensions:

$$\nabla^2 f + f = 0, \quad (4.74)$$

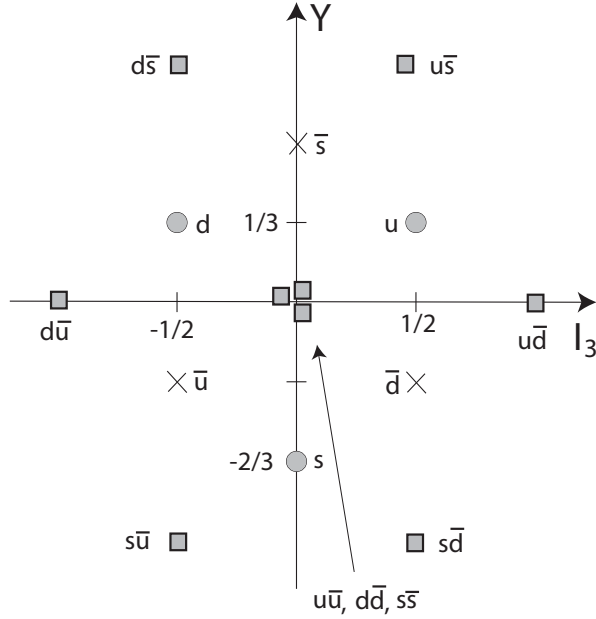


Figure 4.2: The $3 \otimes \bar{3} = 8 \oplus 1$ representation of $SU(3)$, in the context of quark flavors.

where

$$\nabla^2 \equiv \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2}. \quad (4.75)$$

- (a) Show that the equation is left invariant under the transformation:

$$\tau(\epsilon, \theta, \alpha, \beta) : \begin{pmatrix} x \\ y \end{pmatrix} \rightarrow \begin{pmatrix} x' \\ y' \end{pmatrix} = \begin{pmatrix} x \cos \theta - y \sin \theta + \alpha \\ x \epsilon \sin \theta + y \epsilon \cos \theta + \beta \end{pmatrix}, \quad (4.76)$$

where $\epsilon = \pm 1$, $-\pi \leq \theta < \pi$, and α and β are any real numbers (actually, α, β , and θ could be complex, but we'll restrict to real numbers here).

- (b) The set of transformations $\{\tau(\epsilon, \theta, \alpha, \beta)\}$ obviously forms a Lie group, where group multiplication is defined as the application of successive transformations. Is it a compact group? Is it connected? What is the identity element? The group multiplication table can be shown to be:

$$\tau(\epsilon_1, \theta_1, \alpha_1, \beta_1) \tau(\epsilon_2, \theta_2, \alpha_2, \beta_2) = \tau(\epsilon_3, \theta_3, \alpha_3, \beta_3), \quad (4.77)$$

where

$$\epsilon_3 = \epsilon_1 \epsilon_2 \quad (4.78)$$

$$\theta_3 = \epsilon_2\theta_1 + \theta_2 \pmod{(-\pi, \pi)} \quad (4.79)$$

$$\alpha_3 = \alpha_2 \cos \theta_1 - \beta_2 \sin \theta_1 + \alpha_1 \quad (4.80)$$

$$\beta_3 = \epsilon_1(\alpha_2 \sin \theta_1 + \beta_2 \cos \theta_1) + \beta_1. \quad (4.81)$$

What is the inverse $\tau^{-1}(\epsilon, \theta, \alpha, \beta)$?

5. We consider some properties of a group algebra which can be useful for obtaining characters: Let the elements of a class be denoted $\{a_1, a_2, \dots, a_{n_a}\}$, the elements of another class be denoted $\{b_1, b_2, \dots, b_{n_b}\}$, etc. Form element $A = \sum_{i=1}^{n_a} a_i$ of the group algebra, and similarly for B , etc.

Suppose D is an n -dimensional irreducible representation. You showed in problem 19 that

$$D(A) \equiv \sum_{i=1}^{n_a} D(a_i) = \frac{n_a}{n} \chi(A) I, \quad (4.82)$$

where $\chi(A)$ is the character of irrep D for class A .

- (a) Now consider the multiplication of two elements, A and B , of the group algebra. Show that the product consists of complete classes, i.e.,

$$AB = \sum_C s_C C, \quad (4.83)$$

where s_C are non-negative integers. You may find it helpful to show that $g^{-1}ABg = AB$ for all group elements g .

- (b) Finally, prove the potentially useful relation:

$$n_a \chi(A) n_b \chi(B) = n \sum_C s_C n_c \chi(C). \quad (4.84)$$

6. We have discussed Lie algebras (with Lie product given by the commutator) and Lie groups, in our attempt to deal with rotations. At one point, we asserted that the structure (multiplication table) of the Lie group in some neighborhood of the identity was completely determined by the structure (multiplication table) of the Lie algebra. We noted that, however intuitively pleasing this might sound, it was not actually a trivial statement, and that it followed from the ‘‘Baker-Campbell-Hausdorff’’ theorem. Let’s try to tidy this up a bit further here.

First, let’s set up some notation: Let \mathcal{L} be a Lie algebra, and \mathcal{G} be the Lie group generated by this algebra. Let $X, Y \in \mathcal{L}$ be two elements of the algebra. These generate the elements $e^X, e^Y \in \mathcal{G}$ of the Lie group. We assume the notion that if X and Y are close to the zero element of the Lie algebra, then e^X and e^Y will be close to the identity element of the Lie group.

What we want to show is that the group product $e^X e^Y$ may be expressed in the form e^Z , where $Z \in \mathcal{L}$, at least for X and Y not too ‘‘large’’. Note

that the non-trivial aspect of this problem is that, first, X and Y may not commute, and second, objects of the form XY may not be in the Lie algebra. Elements of \mathcal{L} generated by X and Y must be linear combinations of X, Y , and their repeated commutators.

- (a) Suppose X and Y commute. Show explicitly that the product $e^X e^Y$ is of the form e^Z , where Z is an element of \mathcal{L} . (If you think this is trivial, don't worry, it is!)
- (b) Now suppose that X and Y may not commute, but that they are very close to the zero element. Keeping terms to quadratic order in X and Y , show once again that the product $e^X e^Y$ is of the form e^Z , where Z is an element of \mathcal{L} . Give an explicit expression for Z .
- (c) Finally, for more of a challenge, let's do the general theorem: Show that $e^X e^Y$ is of the form e^Z , where Z is an element of \mathcal{L} , as long as X and Y are sufficiently "small". We won't concern ourselves here with how "small" X and Y need to be – you may investigate that at more leisure.

Here are some hints that may help you: First, we remark that the differential equation

$$\frac{df}{du} = Xf(u) + g(u), \quad (4.85)$$

where $X \in \mathcal{L}$, and letting $f(0) = f_0$, has the solution:

$$f(u) = e^{uX} f_0 + \int_0^u e^{(u-v)X} g(v) dv. \quad (4.86)$$

This can be readily verified by back-substitution. If g is independent of u , then the integral may be performed, with the result:

$$f(u) = e^{uX} f_0 + h(u, X)g, \quad (4.87)$$

Where, formally,

$$h(u, X) = \frac{e^{uX} - 1}{X}. \quad (4.88)$$

Second, if $X, Y \in \mathcal{L}$, then

$$e^X Y e^{-X} = e^{X_c} (Y), \quad (4.89)$$

where I have introduced the notation " X_c " to mean "take the commutator". That is, $X_c(Y) \equiv [X, Y]$. This fact may be demonstrated by taking the derivative of

$$A(u; Y) \equiv e^{uX} Y e^{-uX} \quad (4.90)$$

with respect to u , and comparing with our differential equation above to obtain the desired result.

Third, assuming $X = X(u)$ is differentiable, we have

$$e^{X(u)} \frac{d}{du} e^{-X(u)} = -h(1, X(u)_c) \frac{dX}{du}. \quad (4.91)$$

This fact may be verified by considering the object:

$$B(t, u) \equiv e^{tX(u)} \frac{\partial}{\partial u} e^{-tX(u)}, \quad (4.92)$$

and differentiating (carefully!) with respect to t , using the above two facts, and finally letting $t = 1$.

One final hint: Consider the quantity

$$Z(u) = \ln(e^{uX} e^Y). \quad (4.93)$$

The series:

$$\ell(z) = \frac{\ln z}{z-1} = 1 - \frac{z-1}{2} + \frac{(z-1)^2}{3} - \dots \quad (4.94)$$

plays a role in the explicit form for the result. Again, you are not asked to worry about convergence issues.

7. In the next few problems we'll pursue further the example we discussed in the notes and in class with $SU(3)$. We consider systems made from the u, d , and s quarks (for "up", "down", and "strange"). Except for the differences in masses, the strong interaction is supposed to be symmetric as far as these three different "flavors" of quarks are concerned. Thus, if we imagine our matter fields to be a triplet:

$$\psi = \begin{pmatrix} \psi_u \\ \psi_d \\ \psi_s \end{pmatrix}, \quad (4.95)$$

then we expect invariance (under the strong interaction) under the transformations

$$\psi \rightarrow \psi' = U\psi, \quad (4.96)$$

where U is any 3×3 matrix. Thus, U is any element of $SU(3)$, and the interaction possesses $SU(3)$ symmetry.

You have already shown that $SU(n)$ is an $(n^2 - 1)$ parameter group. Thus, $SU(3)$ has 8 parameters, and an arbitrary element in $SU(3)$ can be expressed in the form:

$$U = \exp \left\{ \frac{i}{2} \sum_{j=1}^8 a_j \lambda_j \right\}$$

where the $\{\lambda_j\}$ is a set of eight 3×3 traceless, hermitian matrices. One such set is the following: (Gell-Mann)

$$\lambda_1 = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \quad \lambda_2 = \begin{pmatrix} 0 & -i & 0 \\ i & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \quad \lambda_3 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 0 \end{pmatrix},$$

$$\lambda_4 = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix}, \quad \lambda_5 = \begin{pmatrix} 0 & 0 & -i \\ 0 & 0 & 0 \\ i & 0 & 0 \end{pmatrix},$$

$$\lambda_6 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}, \quad \lambda_7 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & -i \\ 0 & i & 0 \end{pmatrix}, \quad \lambda_8 = \frac{1}{\sqrt{3}} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -2 \end{pmatrix}.$$

If the a_j are infinitesimal numbers, we have

$$\psi' = \left(1 + \frac{i}{2} \sum a_j \lambda_j\right) \psi$$

and hence, the quantities $\Lambda_j = \frac{1}{2} \lambda_j$ are called the generators of the infinitesimal transformations, or, simply, the generators of the group. These generators satisfy the commutation relations: (and we have a Lie algebra)

$$[\Lambda_i, \Lambda_j] = i f_{ijk} \Lambda_k$$

Evaluate the structure constants, f_{ijk} , of $SU(3)$.

8. We may find ourselves interested in “states” consisting of more than one quark, thus we must consider (infinitesimal) transformations of the form

$$\psi \rightarrow \psi' = (1 + i \vec{\alpha} \cdot \vec{\Lambda}) \psi \quad (4.97)$$

$$\vec{\alpha} \cdot \vec{\Lambda} \equiv \sum_{j=1}^8 a_j \Lambda_j$$

where the Λ_j may be represented by matrices of dimension other than 3. Let us develop a simple graphical approach to dealing with this problem (We could also use less intuitive method of Young diagrams, as in the final problem of this problem set).

First, let us introduce the new operators (“canonical form”):

$$I_{\pm} = \Lambda_1 \pm i \Lambda_2 \quad (4.98)$$

$$U_{\pm} = \Lambda_6 \pm i \Lambda_7 \quad (4.99)$$

$$V_{\pm} = \Lambda_4 \pm i \Lambda_5 \quad (4.100)$$

$$I_3 = \Lambda_3 \quad (\text{“3rd component of isotopic spin”}) \quad (4.101)$$

$$Y = \frac{2}{\sqrt{3}} \Lambda_8 \quad (\text{“hypercharge”}) \quad (4.102)$$

Second, we remark that only two of the 8 generators of $SU(3)$ can be simultaneously diagonalized (e.g., see the explicit λ matrices I wrote down earlier). [Thus, $SU(3)$ is called a group of *rank 2* – in general, $SU(n)$ has rank $n - 1$.] We choose I_3 and Y to be the diagonalized generators. Thus, our states will be eigenstates of these operators, with eigenvalues which will denote by i_3 and y . With the structure constants, you may easily find, e.g.,

$$[I_3, I_{\pm}] = \pm I_{\pm}$$

Thus, if $\psi(i_s)$ is an eigenstate of I_3 with eigenvalue i_s :

$$\begin{aligned} I_3 I_+ \psi(i_s) &= I_+ (1 + I_3) \psi(i_s) = I_+ (1 + i_s) \psi(i_s) \\ &= (1 + i_s) I_+ \psi(i_s) \end{aligned} \quad (4.103)$$

So I_+ acts as a “raising” operator for i_3 , since $I_+ \psi(i_s)$ is again an eigenstate of I_3 , with eigenvalue $1 + i_s$. Likewise, we have other commutation relations, such as:

$$[I_3, U_{\pm}] = \mp \frac{1}{2} U_{\pm} \quad (4.104)$$

$$[I_3, V_{\pm}] = \pm \frac{1}{2} V_{\pm} \quad (4.105)$$

$$[Y, I_{\pm}] = 0 \quad (4.106)$$

$$[Y, U_{\pm}] = \pm U_{\pm} \quad (4.107)$$

$$[Y, V_{\pm}] = \pm V_{\pm} \quad (4.108)$$

$$[I_3, Y] = 0 \quad (4.109)$$

etc.

Thus, the action of the “raising and lowering” operators $I_{\pm}, U_{\pm}, V_{\pm}$ can be indicated graphically, as in Fig. 4.3.

Thus, we may generate all states of an irreducible representation starting with one state by judicious application of the raising and lowering operators. As a simplest example, and to keep the connection to quarks alive, we consider the 3-dimensional representation: Let’s start at the u – quark; it has $i_3 = \frac{1}{2}$ and $y = \frac{1}{3}$. See Fig. 4.4.

Why did we stop after we generated d and s , starting from u ? Well, clearly we can’t have more components (or “occupied sites”) than the dimensional-maximum allowed. In fact, since this a 3-dimensional representation, we can just look at the matrices we gave earlier and see that the eigenvalues of I_3 are going to be $\pm \frac{1}{2}$ and 0, and those of Y will be $\frac{1}{3}, \frac{1}{3}$, and $-\frac{2}{3}$. A little more consideration of the matrices convinces us that, e.g., $I_+ u = 0, I_+ s = 0, U_+ d = 0$, etc.,

We have given the $i_3 - y$ graph for the “3” representation of $SU(3)$. Now give the corresponding graph for the “3*” (or $\bar{3}$) representation, that is, the conjugate representation.

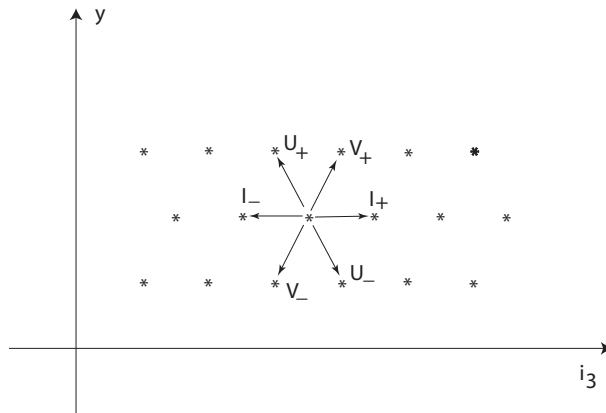


Figure 4.3: The actions of the $SU(3)$ raising and lowering operators $SU(3)$, in the $i_3 - y$ state space.

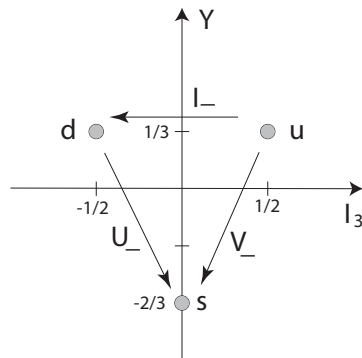


Figure 4.4: The 3 irreducible representation of $SU(3)$, in the $i_3 - y$ state space.

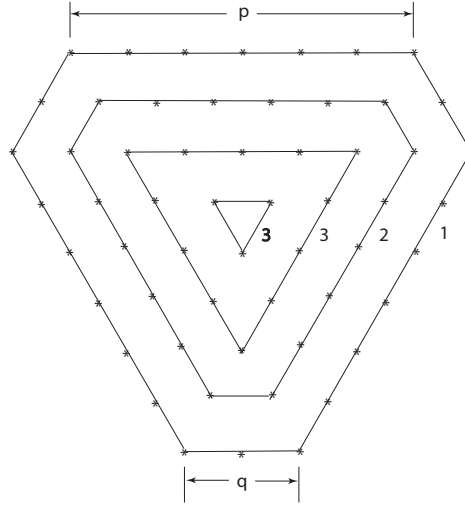


Figure 4.5: The graph of the $SU(3)$ irreducible representation $(p, q) = (6, 2)$. The numbers indicate the multiplicities at each site.

9. You are encouraged to develop the detailed arguments, using the commutation relations for the following observations: The graph for a given irreducible representation is a convex graph which is 6-sided in general (or three-sided if a side length goes to zero). A graph (of an irreducible rep.) is uniquely labelled by two numbers (p, q) . An example will suffice to get the idea across. Fig. 4.5 shows the graph for $(p, q) = (6, 2)$. The origin of the $I_3 - y$ coordinate system is inside the innermost triangle. The rule giving the multiplicity of states at each site is that *i*) the outermost ring has multiplicity of 1, *ii*) as one moves to inner rings, the multiplicity increases by one at each ring, until a triangular ring is reached, whereupon no further increases occur.

By counting the total number of states (*i.e.*, by counting sites, weighted according to multiplicity), we arrive at the dimensionality of the representation. The result, as you may wish to convince yourselves, is

$$\dim = N = \frac{1}{2}(p+1)(q+1)(p+q+2)$$

For the 3 and 3* representations, give the corresponding pairs (p, q) , and check that the dimensions come out correctly.

One more remark: If we have $p \geq q$, we denote the representation by its dimensionality N . If $p < q$, we call it a conjugate representation, and denote it by N^* [*e.g.*, $(2, 0)$ is the representation 6, but $(0, 2)$ is 6*.] An alternative notation is to use a “bar”, *e.g.*, \bar{N} to denote the conjugate representa-

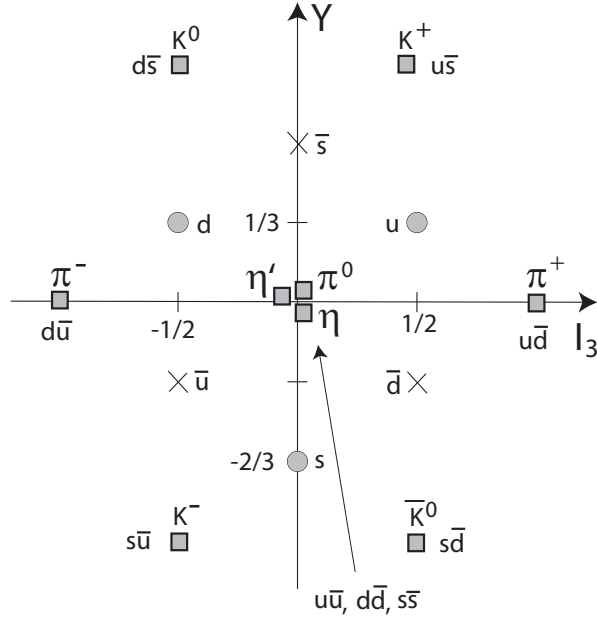


Figure 4.6: The graph of the $SU(3)$ representation for $3 \times \bar{3}$. Physical particle names for the lowest pseudoscalar mesons are indicated at each site.

tion (since for unitary representations the adjoint and complex conjugate representations are the same).

10. We know that the mesons are states of a quark and an antiquark. If you have done everything fine so far, you will see that we can thus generate the mesons by $3 \otimes 3^*$. The result is shown in Fig. 4.6 (don't worry about the particle names, unless you're interested) Using the rules given above concerning irreducible representations, we find, from this graph, the decomposition $3 \otimes 3^* = 8 \oplus 1$.

We know baryons are made of three quarks (no antiquarks). Make sure you understand how I did the mesons, and apply the same graphical approach to the baryons, and determine the decomposition of $3 \otimes 3 \otimes 3$ into a direct sum of irreducible reps. Do not use Young diagrams (next problem) to do this problem, although you are encouraged to check you answer with Young diagrams. You may find it amusing, if you know something about particle properties, to assign some known baryon names to the points on your graphs.

11. Go to the URL: <http://pdg.lbl.gov/2007/reviews/youngrpp.pdf>. Study the section on “ $SU(n)$ Multiplets and Young Diagrams,” and use the techniques described there to answer the following question: We consider

the special unitary group $SU(4)$. This is the group of unimodular unitary 4×4 matrices. We wish to consider the product representation of the irreducible representation given by the elements of the group itself with the irreducible representation formed by the isomorphism of taking the complex conjugate of every element. This turns out to yield a representation which is not equivalent to 4 . We could call this new representation 4^* , but it is perhaps more typical to use the notation $\bar{4}$. Note that, since we are dealing with unitary matrices, the complex conjugate and the adjoint representation are identical, so this notation is reasonable.

The question to be answered is: What are dimensions of the irreducible representations obtained in the decomposition of the product representation $4 \otimes \bar{4}$?

The principal purpose of this problem (which is mechanically very simple) is to alert you to the existence of convenient graphical techniques in group theory – most notably that of Young diagrams. We make no attempt yet to understand “why it works”.

A few more words are in order concerning the language on the web page: Since it is taken from the Particle Data Group’s “Review of Particle Properties,” it is concerned with the application to particle physics, and the language reflects this. However, it is easily understood once one realizes that the number of particles in a “multiplet” is just the dimension of a representation for the group. Effectively, the particles are labels for basis vectors in a space of dimension equal to the multiplet size. [The basic physics motivation for the application of $SU(n)$ to the classification and properties of mesons and baryons is that the strong interaction is supposed to be symmetric as far as the different flavors are concerned. The “ n ” in this $SU(n)$ is just the number of different flavors. Note that this (flavor) $SU(n)$ is a different application from the “color” $SU(3)$ symmetry in QCD.] Those of you who know something about particles may find it amusing to try to attach some known particle names to the $4 \otimes \bar{4}$ multiplets.

Chapter 5

The Permutation Group and Young Diagrams

5.1 Definitions

The permutation, or symmetric, group, S_n is interesting at least partly because it contains subgroups isomorphic to all groups of order $\leq n$. This result is known as “Cayley’s theorem”. It is also of great value in tensor analysis as the means to describe the tensor space in terms of symmetries under permutations of indicies. Here, we develop a diagrammatic approach to determining the irreducible representations of S_n , which will turn out to have applications beyond this immediate one.

Recall that we can express an element of S_n in cycle notation. For example, the element of S_5 :

$$p = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 4 & 1 & 5 & 2 & 3 \end{pmatrix} \quad (5.1)$$

is described in cycle notation as $(142)(35)$. We can make a useful correspondence between cycle structures and the “partitions” of integer n :

Def: A *partition* of a positive integer n is a set of integers $(\lambda_1, \lambda_2, \dots, \lambda_n)$ such that

$$\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_n \geq 0, \quad (5.2)$$

and

$$\lambda_1 + \lambda_2 + \dots + \lambda_n = n. \quad (5.3)$$

Consider the class structure of the symmetric group S_n . Classes are given by cycle structures, *i.e.*, a particular class is specified by giving the n numbers $\omega_1, \omega_2, \dots, \omega_n$, where ω_i is the number of i cycles in an element belonging to

the class. Thus, for $(142)(35) \in S_5$, $\omega_1 = 0$, $\omega_2 = 1$, $\omega_3 = 1$, $\omega_4 = \omega_5 = 0$. Noticing that $\sum_{i=1}^n i\omega_i = n$, we see that the specification of a class of the symmetric group corresponds to the specification of a partition of n , according to the construction:

$$\begin{aligned}\lambda_1 &= \omega_1 + \omega_2 + \dots + \omega_n \\ \lambda_2 &= \omega_2 + \dots + \omega_n \\ &\vdots \\ \lambda_n &= \omega_n.\end{aligned}$$

For our S_5 example:

$$\begin{aligned}\lambda_1 &= 2 \\ \lambda_2 &= 2 \\ \lambda_3 &= 1 \\ \lambda_4 &= 0 \\ \lambda_5 &= 0,\end{aligned}\tag{5.4}$$

and the sum of these numbers is five.

We use this correspondence in the invention of a graphical description known as Young Diagrams.

Def: A *Young Diagram* is a diagram with n boxes arranged in n rows corresponding to a partition of n , *i.e.*, with row i containing λ_i boxes.

For example, the diagram:

$$\begin{array}{|c|c|} \hline \square & \square \\ \hline \square & \square \\ \hline \square & \\ \hline \end{array},\tag{5.5}$$

for S_5 corresponds to $\lambda_1 = 2$, $\lambda_2 = 2$, $\lambda_3 = 1$, and $\lambda_4 = \lambda_5 = 0$. Because of the ordering of the λ 's, each row of a Young diagram has at most as many boxes as the row above it.

Note that giving all the Young diagrams for a given n classifies all of the classes of S_n . Since $n_c = n_r$ it may not be surprising that such diagrams are also useful in identifying irreducible representations of S_n . That is, there is a 1 : 1 correspondence between Young diagrams and irreducible representations of S_n . Furthermore, these diagrams are useful in decomposing products of irreducible representations.

Def: A *Young tableau* is a Young diagram in which the n boxes have been filled with the numbers $1, \dots, n$, each number used exactly once.

For example:

$$\begin{array}{|c|c|} \hline 4 & 1 \\ \hline 2 & 3 \\ \hline 5 & \\ \hline \end{array} . \quad (5.6)$$

There are $n!$ Young tableau for a given Young diagram.

Def: A *standard Young tableau* is a Young tableau in which the numbers appear in ascending order within each row or column from left to right and top to bottom.

For example, the following are the possible standard Young tableau with the given shape:

$$\begin{array}{|c|c|} \hline 1 & 2 \\ \hline 3 & 4 \\ \hline 5 & \\ \hline \end{array} , \quad \begin{array}{|c|c|} \hline 1 & 2 \\ \hline 3 & 5 \\ \hline 4 & \\ \hline \end{array} , \quad \begin{array}{|c|c|} \hline 1 & 3 \\ \hline 2 & 4 \\ \hline 5 & \\ \hline \end{array} , \quad \begin{array}{|c|c|} \hline 1 & 3 \\ \hline 2 & 5 \\ \hline 4 & \\ \hline \end{array} , \quad \begin{array}{|c|c|} \hline 1 & 4 \\ \hline 2 & 5 \\ \hline 3 & \\ \hline \end{array} . \quad (5.7)$$

Def: A *normal tableau* is a standard Young tableau in which the numbers are in order, left to right and top to bottom.

There is only one normal tableau of a given shape, e.g.,

$$\begin{array}{|c|c|} \hline 1 & 2 \\ \hline 3 & 4 \\ \hline 5 & \\ \hline \end{array} . \quad (5.8)$$

From a normal tableau, we may obtain all other standard tableau by suitable permutations, for example:

$$Y_1 : \begin{array}{|c|c|} \hline 1 & 2 \\ \hline 3 & 4 \\ \hline 5 & \\ \hline \end{array} \xrightarrow{(1)(2453)} Y_2 : \begin{array}{|c|c|} \hline 1 & 4 \\ \hline 2 & 5 \\ \hline 3 & \\ \hline \end{array} . \quad (5.9)$$

That is, $(2453)Y_1 = Y_2$.

5.2 Examples: System of particles

Suppose we have a system of n identical particles, in which the Hamiltonian, H , is invariant under permutations of the particles. Let x_i be the coordinate (position, spin, etc.) of particle i . Suppose $\psi(x_1, x_2, \dots, x_n)$ is an eigenfunction of H belonging to eigenvalue E . Then any permutation of the particles:

$$\begin{aligned} P_a \psi &= \begin{pmatrix} 1 & 2 & \cdots & n \\ a_1 & a_2 & \cdots & a_n \end{pmatrix} \psi(x_1, x_2, \dots, x_n) \\ &= \psi(x_{a_1}, x_{a_2}, \dots, x_{a_n}) \end{aligned} \quad (5.10)$$

is another eigenfunction belonging to the same eigenvalue.

In quantum mechanics, we have symmetric wave functions, under interchange of any pair of particle coordinates, for bosons, and anti-symmetric wave functions for fermions. Define a “symmetrizer operator” by:

$$S \equiv \frac{1}{n!} \sum_P P, \quad (5.11)$$

where \sum_P is short for $\sum_{P \in S_n}$, that is a sum over all permutations of the n particle coordinates. Likewise, define an “anti-symmetrizer operator” by:

$$A \equiv \frac{1}{n!} \sum_P \delta_P P, \quad (5.12)$$

where

$$\delta_P \equiv \begin{cases} +1 & \text{if } P \text{ is even} \\ -1 & \text{if } P \text{ is odd.} \end{cases} \quad (5.13)$$

We call δ_P the “parity” of the permutation. It is given by

$$\delta_P = (-1)^q, \quad (5.14)$$

where q is the number of transpositions required to produce permutation P starting from the normal tableau.

It is an exercise for the reader to show that a k -cycle has parity $(-1)^{k-1}$. Therefore, if a permutation, P , consists of ℓ cycles with structure $\{k_1, k_2, \dots, k_\ell\}$ then the parity of P is:

$$\begin{aligned} \delta_P &= (-1)^{\sum_{i=1}^{\ell} (k_i - 1)} \\ &= (-1)^{n - \ell}, \end{aligned} \quad (5.15)$$

where the second line follows because $\sum_{i=1}^{\ell} k_i = n$. The quantity $n - \ell$ is called the *decrement* of P .

We also leave it as an exercise for the reader to show that, for any $P_a \in S_n$:

$$P_a S = S \quad (5.16)$$

$$P_a A = A P_a = \delta_{P_a} A \quad (5.17)$$

$$S^2 = S \quad (5.18)$$

$$A^2 = A. \quad (5.19)$$

Thus, S and A act as projection operators.

Consider two-particle states. Let u and d be orthogonal single-particle states¹, and $\psi_N = u(x_1)d(x_2)$. We have symmetrizer:

$$S_{12} = \frac{1}{2}(e + P_{12}), \quad (5.20)$$

¹Alternatively, we could be talking about the two angular momentum states of a spin-1/2 system, with u corresponding, say to spin “up”, and d to spin “down”.

where e is the identity operator of S_2 . The operator S_{12} projects out the symmetric part of ψ_N :

$$\psi^S \equiv S_{12}\psi_N = \frac{1}{2}[u(x_1)d(x_2) + d(x_1)u(x_2)]. \quad (5.21)$$

Likewise, the anti-symmetrizer,

$$A_{12} = \frac{1}{2}(e - P_{12}), \quad (5.22)$$

projects out the antisymmetric piece:

$$\psi^A \equiv A_{12}\psi_N = \frac{1}{2}[u(x_1)d(x_2) - d(x_1)u(x_2)]. \quad (5.23)$$

Note that the combinations $u(x_1)u(x_2)$ and $d(x_1)d(x_2)$ are already symmetric.

Now we relate this discussion to our graphical formalism. The Young diagram \square corresponds to the class of two 1-cycles, that is, the identity of S_2 . The Young diagram \square corresponds to the class of one 2-cycle, that is transposition. Thus, we make the identification of tableau:

$$\begin{array}{|c|c|} \hline 1 & 2 \\ \hline \end{array} \text{ with } \begin{cases} u(x_1)u(x_2) \\ \frac{1}{2}[u(x_1)d(x_2) + d(x_1)u(x_2)] \\ d(x_1)d(x_2), \end{cases} \quad (5.24)$$

and

$$\begin{array}{|c|} \hline 1 \\ \hline 2 \\ \hline \end{array} \text{ with } \frac{1}{2}[u(x_1)d(x_2) - d(x_1)u(x_2)]. \quad (5.25)$$

That is, two boxes in a row correspond to a symmetric state, and two in a column to an antisymmetric state.

Let's try this with three-particle states, with u, d, s as orthonormal single particle states. We'll drop the x from our notation, and simply write $\psi_N = u(1)d(2)s(3)$. There are $3! = 6$ linearly independent functions obtained by permuting the 1, 2, 3 particle labels, or by permuting the state labels u, d, s . We'll do the latter, and also simplify our notation still further and drop the particle labels, with the understanding that they remain in the order 123.

We rewrite the six linearly independent functions obtained by permutations into a different set of six linearly independent functions, based on symmetry properties under interchange. First, the completely symmetric arrangement:

$$\begin{aligned} \psi^S &= S_{123}\psi_N = S_{123}uds \\ &= \frac{1}{3!}(e + P_{12} + P_{13} + P_{23} + P_{123} + P_{132})uds \\ &= \frac{1}{3!}(uds + dus + sdu + usd + dsu + sud). \end{aligned} \quad (5.26)$$

Once again, this corresponds to the identity class of three 1-cycles: $\begin{array}{|c|c|c|} \hline 1 & 2 & 3 \\ \hline \end{array}$. This symmetric state is invariant under the actions of S_3 , hence it generates the one-dimensional identity representation.

The completely antisymmetric arrangement is:

$$\begin{aligned}\psi^A &= A_{123}\psi_N = A_{123}uds \\ &= \frac{1}{3!}(e - P_{12} - P_{13} - P_{23} + P_{123} + P_{132})uds \\ &= \frac{1}{3!}(uds - dus - sdu - usd + dsu + sud),\end{aligned}\quad (5.27)$$

corresponding to $\begin{array}{|c|} \hline 1 \\ \hline 2 \\ \hline 3 \\ \hline \end{array}$. All actions of the S_3 group on ψ^A yield ± 1 times ψ^A . Thus, this function is a vector in another one-dimensional invariant subspace under the actions of the group elements and hence generates another one-dimensional irreducible representation of S_3 . Note that it is not equivalent to the identity representation.

There are four more functions to build; these must have mixed symmetry. We may proceed by symmetrizing uds with respect to two particles, and then antisymmetrizing with respect to two particles (or vice versa), with one particle in common between the two operations.² There is some arbitrariness in how we choose to carry out this program. Let us take:

$$\begin{aligned}\psi_1 &= A_{13}S_{12}\psi_N = A_{13}\frac{1}{2}(uds + dus) = \frac{1}{4}(uds - sdu + dus - sud) \\ \psi_2 &= A_{23}S_{12}\psi_N = \frac{1}{4}(uds - usd + dus - dsu) \\ \psi_3 &= S_{13}A_{12}\psi_N = \frac{1}{2}(e + P_{13})\frac{1}{2}(uds - dus) = \frac{1}{4}(uds + sdu - dus - sud) \\ \psi_4 &= S_{23}A_{12}\psi_N = \frac{1}{4}(uds + usd - dus - dsu).\end{aligned}\quad (5.29)$$

We note that ψ_1 and ψ_2 form an invariant subspace under S_3 :

$$\begin{aligned}(12)\psi_1 &= \frac{1}{4}(dus - dsu + uds - usd) = \psi_2 \\ (13)\psi_1 &= -\psi_1 \\ (12)\psi_2 &= \psi_1 \\ (13)\psi_2 &= \frac{1}{4}(sdu - dsu + sud - usd) = \psi_2 - \psi_1,\end{aligned}\quad (5.30)$$

with the other S_3 elements obtained by products of these. Likewise, ψ_3 and ψ_4 form an invariant subspace.

Typically, we want to form an orthogonal system. We may check whether our states are orthogonal. For example,

$$\begin{aligned}(\psi_1, \psi_3) &= (A_{13}S_{12}uds, S_{13}A_{12}uds) \\ &= (uds, S_{12}A_{13}S_{13}A_{12}uds) = 0,\end{aligned}\quad (5.31)$$

²Note that

$$S_{ij}A_{ij} = \frac{1}{2}(e + P_{ij})\frac{1}{2}(e - P_{ij}) = \frac{1}{4}(e + P_{ij} - P_{ij} - e) = 0. \quad (5.28)$$

That is, our projections project onto orthogonal subspaces.

since $A_{13}S_{13} = 0$. Likewise, we find that

$$(\psi_1, \psi_4) = (\psi_2, \psi_3) = (\psi_2, \psi_4) = 0. \quad (5.32)$$

However, we also find that $(\psi_1, \psi_2) \neq 0$ and $(\psi_3, \psi_4) \neq 0$, so our $\psi_1, \psi_2, \psi_3, \psi_4$ states do not yet form an orthogonal system. But we make take linear combinations (a, b, c, d are normalization constants):

$$\begin{aligned} \psi'_1 &= a(\psi_1 + \psi_2) \\ &= -\frac{1}{\sqrt{12}}(2uds + 2dus - sdu - sud - usd - dsu), \end{aligned} \quad (5.33)$$

$$\begin{aligned} \psi'_2 &= b(\psi_1 - \psi_2) \\ &= -\frac{1}{2}(usd + dsu - sdu - sud), \end{aligned} \quad (5.34)$$

$$\begin{aligned} \psi'_3 &= c(\psi_3 + \psi_4) \\ &= \frac{1}{\sqrt{12}}(2uds - 2dus + sdu - sud + usd - dsu), \end{aligned} \quad (5.35)$$

$$\begin{aligned} \psi'_4 &= d(\psi_3 - \psi_4) \\ &= \frac{1}{2}(-sdu + sud + usd - dsu), \end{aligned} \quad (5.36)$$

where we have normalized and adopted phase conventions.

Thus, we have a set of six orthonormal functions. Both ψ'_1 and ψ'_4 are symmetric under the transposition (12), hence both correspond to the Young

tableau $\begin{array}{|c|c|} \hline 1 & 2 \\ \hline 3 \\ \hline \end{array}$. Likewise, ψ'_2 and ψ'_3 are antisymmetric under (12), corresponding

to tableau $\begin{array}{|c|c|} \hline 1 & 3 \\ \hline 2 \\ \hline \end{array}$. The states $\{\psi'_1, \psi'_2\}$ form an invariant subspace under S_3 , and the states $\{\psi'_3, \psi'_4\}$ form another invariant subspace. Both subspaces lead to the same irreducible representation of S_3 , a ‘‘mixed’’ representation (that is, neither purely symmetric nor purely antisymmetric under transpositions), with Young diagram $\begin{array}{|c|c|} \hline & \\ \hline & \\ \hline \end{array}$. This is a two-dimensional representation, acting on either of the two-dimensional invariant subspaces. For example,

$$\begin{aligned} (12)\psi'_1 &= \psi'_1 \\ (12)\psi'_2 &= -\psi'_2 \end{aligned} \quad (5.37)$$

tells us that the (12) element is represented in this basis by:

$$D(12) = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}. \quad (5.38)$$

Now notice that we are also generating an orthonormal basis of states of the $3 \otimes 3 \otimes 3$ representation of $SU(3)$! We thus have a connection between $SU(3)$ and the permutation symmetry. Let us pursue this idea further in this example. We’ll make the example more concrete by interpreting that the particles u, d, s

as quark flavor eigenstates. In this case, we are generating the wavefunctions of the baryons, in terms of quark flavor content. Now $3 \times 3 \times 3 = 27$, so we are dealing with a 27-dimensional representation of $SU(3)$. We proceed to find the decomposition of this into irreducible representations, and obtain the baryon flavor wavefunctions:

First, we have,

$$\begin{array}{|c|} \hline 1 \\ \hline 2 \\ \hline 3 \\ \hline \end{array} = \psi^A = \frac{1}{6}(uds - dus + sud - usd + dsu - sdu) = \begin{array}{|c|} \hline u \\ \hline d \\ \hline s \\ \hline \end{array}, \quad (5.39)$$

where the graph on the left is our familiar Young tableau indicating complete antisymmetry under transposition of coordinates in S_3 . The graph on the right, called a ‘‘Weyl tableau’’, indicates that the wave function is also completely antisymmetric under interchange of *flavors* in $SU(3)$. This is the only completely antisymmetric state: Any ‘‘rotation’’ in $SU(3)$ gives back this state. Hence this generates a one-dimensional representation of $SU(3)$, the identity representation.

We have seen that we also have states with mixed symmetry under the actions of S_3 . We found four such states comprised of uds , two associated with $\begin{array}{|c|c|} \hline 1 & 2 \\ \hline 3 & \\ \hline \end{array}$ and two with $\begin{array}{|c|c|} \hline 1 & 3 \\ \hline 2 & \\ \hline \end{array}$. There are, in addition states with two identical quarks.

For example, we may obtain the $\begin{array}{|c|c|} \hline 1 & 2 \\ \hline 3 & \\ \hline \end{array}$ uud state by letting $s \rightarrow u$ in ψ'_1 :

$$\begin{aligned} \psi'_1 &= -\frac{1}{\sqrt{12}}(2uds + 2dus - sdu - sud - usd - dsu), \\ &\rightarrow -\frac{1}{\sqrt{12}}(2udu + 2duu - udu - uud - uud - duu), \\ &= -\frac{1}{\sqrt{12}}(udu + duu - 2uud), \\ &\rightarrow -\frac{1}{\sqrt{6}}(udu + duu - 2uud), \end{aligned} \quad (5.40)$$

where we have normalized to one in the last step.

Similarly, the $\begin{array}{|c|c|} \hline 1 & 3 \\ \hline 2 & \\ \hline \end{array}$ uud state is obtained from ψ'_2 :

$$\begin{aligned} \psi'_2 &= -\frac{1}{2}(usd + dsu - sdu - sud), \\ &\rightarrow -\frac{1}{2}(uud + duu - udu - uud), \\ &\rightarrow -\frac{1}{\sqrt{2}}(udu - duu). \end{aligned} \quad (5.41)$$

Notice that we get the same state by replacing the s quark in ψ'_3 with a u quark. Likewise, ψ'_4 gives the same state as ψ'_1 .

Let us summarize the mixed symmetry states of the baryons:³

Baryon name	$\begin{array}{ c c } \hline 1 & 2 \\ \hline 3 \\ \hline \end{array}$ (12) symmetric	$\begin{array}{ c c } \hline 1 & 3 \\ \hline 2 \\ \hline \end{array}$ (12) antisymmetric	Weyl tableau
N^+	$-\frac{1}{\sqrt{6}}(udu + duu - 2uud)$	$\frac{1}{\sqrt{2}}(udu - duu)$	$\begin{array}{ c c } \hline u & u \\ \hline d \\ \hline \end{array}$
N^0	$\frac{1}{\sqrt{6}}(udd + dud - 2ddu)$	$\frac{1}{\sqrt{2}}(udd - dud)$	$\begin{array}{ c c } \hline u & d \\ \hline d \\ \hline \end{array}$
Σ^+	$\frac{1}{\sqrt{6}}(usu + suu - 2uus)$	$\frac{1}{\sqrt{2}}(usu - suu)$	$\begin{array}{ c c } \hline u & u \\ \hline s \\ \hline \end{array}$
Σ^0	$-\frac{1}{\sqrt{12}}(2uds + 2dus - sdu - sud - usd - dsu)$	$-\frac{1}{2}(usd + dsu - sdu - sud)$	$\begin{array}{ c c } \hline u & d \\ \hline s \\ \hline \end{array}$
Σ^-	$\frac{1}{\sqrt{6}}(sdd + dsd - 2dds)$	$\frac{1}{\sqrt{2}}(sdd - dsd)$	$\begin{array}{ c c } \hline d & d \\ \hline s \\ \hline \end{array}$
Λ^0	$\frac{1}{2}(sud - sdu + usd - dsu)$	$\frac{1}{\sqrt{12}}(2uds - 2dus + sdu - sud + usd - dsu)$	$\begin{array}{ c c } \hline u & s \\ \hline d \\ \hline \end{array}$
Ξ^0	$-\frac{1}{\sqrt{6}}(uss + sus - 2ssu)$	$-\frac{1}{\sqrt{2}}(uss - sus)$	$\begin{array}{ c c } \hline u & s \\ \hline s \\ \hline \end{array}$
Ξ^-	$-\frac{1}{\sqrt{6}}(dss + sds - 2ssd)$	$-\frac{1}{\sqrt{2}}(dss - sds)$	$\begin{array}{ c c } \hline d & s \\ \hline s \\ \hline \end{array}$

Thus, we have two eight-dimensional irreducible representations of $SU(3)$ with mixed symmetry. Together with the completely antisymmetric state, we so far have irreducible representations of dimensions 1, 8, 8 in our 27-dimensional $3 \otimes 3 \otimes 3$ product representation. We next consider the representation generated by the completely symmetric states, corresponding to $\begin{array}{|c|c|c|} \hline 1 & 2 & 3 \\ \hline \end{array}$. We may start with:

$$\psi^S = \frac{1}{\sqrt{6}}(uds + dus + sdu + usd + dsu + sud). \quad (5.42)$$

The particle name attached to this state is Σ^{*0} . If we replace the s by a u , for example, we get

$$\frac{1}{\sqrt{3}}(udu + uud + duu), \quad (5.43)$$

known as Δ^+ .

We summarize the symmetric states in a table:

Baryon name	$\begin{array}{ c c c } \hline 1 & 2 & 3 \\ \hline \end{array}$ completely symmetric	Weyl tableau
Δ^{++}	uuu	$\begin{array}{ c c c } \hline u & u & u \\ \hline \end{array}$

³Note that the superscripts give the electric charges of the states, where the u has charge $\frac{2}{3}$ and the d and s both have charge $-\frac{1}{3}$. Thus the charge operator, Q is related to the I_3 and Y operators by $Q = I_3 + \frac{Y}{2}$.

Δ^+	$\frac{1}{\sqrt{3}}(udu + uud + duu)$	$\boxed{u} \boxed{u} \boxed{d}$
Δ^0	$\frac{1}{\sqrt{3}}(udd + dud + ddu)$	$\boxed{u} \boxed{d} \boxed{d}$
Δ^-	ddd	$\boxed{d} \boxed{d} \boxed{d}$
Σ^{*+}	$\frac{1}{\sqrt{3}}(uus + usu + suu)$	$\boxed{u} \boxed{u} \boxed{s}$
Σ^{*0}	$\frac{1}{\sqrt{6}}(uds + dus + sdu + usd + dsu + sud)$	$\boxed{u} \boxed{d} \boxed{s}$
Σ^{*-}	$\frac{1}{\sqrt{3}}(dds + dsd + sdd)$	$\boxed{d} \boxed{d} \boxed{s}$
Ξ^{*0}	$\frac{1}{\sqrt{3}}(uss + sus + ssu)$	$\boxed{u} \boxed{s} \boxed{s}$
Ξ^{*-}	$\frac{1}{\sqrt{3}}(dss + sds + ssd)$	$\boxed{d} \boxed{s} \boxed{s}$
Ω^-	sss	$\boxed{s} \boxed{s} \boxed{s}$

There are thus ten symmetric states, generating a ten-dimensional irreducible representation of $SU(3)$. We have once again found that $3 \otimes 3 \otimes 3 = 10 \oplus 8 \oplus 8 \oplus 1$ in $SU(3)$. Notice that we can generate all of the irreducible representations and bases from the “Weyl” diagrams, with two simple rules:

1. No column contains the same label twice.
2. Within each row or column, the state labels must be in non-decreasing order (according to whatever convention is chosen for the ordering of u, d, s).

Let us notice something now: When we formed the $3 \otimes 3 \otimes 3$ product representation of $SU(3)$, we obtained the Clebsch-Gordan series consisting of $SU(3)$ irreducible representations:

Number of irreps	Dimension of irrep	Young diagram
1	1	$\begin{array}{c} \square \\ \square \\ \square \end{array}$
2	8	$\begin{array}{cc} \square & \square \\ \square & \square \end{array}$
1	10	$\square \square \square$

But we also obtained irreducible representations of S_3 . That is, we obtained the decomposition of our 27-dimensional representation of S_3 , acting on our 27-dimensional state space, into the irreducible representations of S_3 :

Number of irreps	Dimension of irrep	Young diagram
1	1	$[1^3] = \begin{array}{c} \square \\ \square \\ \square \end{array}$
8	2	$[12] = \begin{array}{cc} \square & \square \\ \square & \square \end{array}$
10	1	$[3] = \square \square \square$

Here we have introduced the notation $[a^i b^j \dots]$ to stand for a partition of n with i occurrences of “ a ”, j occurrences of “ b ”, etc. The first one-dimensional representation acts on the completely antisymmetric basis vector, the eight two-dimensional representations act on the vectors of mixed symmetry, and the

final ten one-dimensional representations act on each of the ten symmetric basis vectors.

We notice a kind of “duality” between the number of irreducible representations of $SU(3)$ and the dimensions of the S_3 irreducible representations, and vice versa. This result holds more generally than this example. The general statement is:

Theorem: The multiplicity of the irreducible representation $[f]$ of S_n , denoted by $m_{[f]}(S_n)$ is equal to the dimension of the irreducible representation $[f]$ of $SU(N)$, denoted by $d_{[f]}(SU(N))$:

$$m_{[f]}(S_n) = d_{[f]}(SU(N)), \quad (5.44)$$

and vice versa:

$$m_{[f]}(SU(N)) = d_{[f]}(S_n), \quad (5.45)$$

in the same tensor space of dimension N^n .

We have introduced the language of a “tensor space” here, we’ll define and discuss this in the next section.

We conclude this section with an important theorem on the irreducible representations of S_n , generalizing the observations we have made for S_2 and S_3 . We introduce the notation Θ_λ to refer to the normal Young tableau associated with partition of n specified by $\lambda = \{\lambda_1, \lambda_2, \dots, \lambda_n\}$. We let Θ_λ^p refer to the standard tableau obtained by permutation p on Θ_λ .

Now define:

Def: The *irreducible symmetrizer*, or *Young symmetrizer*, e_λ^p associated with the Young tableau Θ_λ^p is

$$e_\lambda^p \equiv \sum_{h,v} \delta_v h v, \quad (5.46)$$

where h is a horizontal permutation of Θ_λ^p and v is a vertical permutation.

An example should help to make this clear. Consider S_3 . We have (up to a factor of $3!$ for S and A):

$$\Theta_3 = \begin{array}{|c|c|c|} \hline 1 & 2 & 3 \\ \hline \end{array} : e_3 = \sum_h h = \sum_{p \in S_3} p = S \quad (5.47)$$

$$\begin{aligned} \Theta_{21} &= \begin{array}{|c|c|} \hline 1 & 2 \\ \hline 3 \\ \hline \end{array} : e_{21} = [(e + (12))][e - (13)] \\ &= e + (12) - (13) - (132) \end{aligned} \quad (5.48)$$

$$\begin{aligned} \Theta_{21}^{(23)} &= \begin{array}{|c|c|} \hline 1 & 3 \\ \hline 2 \\ \hline \end{array} : e_{21}^{(23)} = [(e + (13))][e - (12)] \\ &= e + (13) - (12) - (123) \end{aligned} \quad (5.49)$$

$$\Theta_{1^3} = \begin{array}{|c|} \hline 1 \\ \hline 2 \\ \hline 3 \\ \hline \end{array} : e_{1^3} = \sum_v \delta_v v = \sum_{p \in S_3} \delta_p p = A. \quad (5.50)$$

This exhausts the standard tableau for S_3 .

We are ready for the theorem, which tells us that these irreducible symmetrizers generate the irreducible representations of S_n :

Theorem: The irreducible symmetrizers $\{e_\lambda\}$ associated with the normal Young tableau $\{\Theta_\lambda\}$ generate all of the inequivalent irreducible representations of S_n .

The general proof of this may be found in Tung and in Hamermesh. We'll make some observations here:

1. The number of inequivalent irreducible representations of S_n is given by the number of different Young diagrams, since they can be put into 1:1 correspondence with the classes.
2. There is one e_λ for each Young diagram, since there is one normal tableau for each diagram. Thus, the number of elements of $\{e_\lambda\}$ is the number of irreducible representations.
3. The remainder of the proof requires showing that each e_λ generates an inequivalent irreducible representation.

Notice that a corollary to this theorem is the fact that e_λ and e_λ^p generate equivalent irreducible representations. We may further notice that the dimension of an irreducible representation $[f]$ of S_n is equal to the number of standard Young tableaux associated with $[f] = [f_1 f_2 \dots f_n]$. For example, in S_3 , $\begin{array}{|c|c|c|} \hline 1 & 2 & 3 \\ \hline \end{array}$

generates a one-dimensional representation, $\begin{array}{|c|c|} \hline 1 & 2 \\ \hline 3 \\ \hline \end{array}$ and $\begin{array}{|c|c|} \hline 1 & 3 \\ \hline 2 \\ \hline \end{array}$ generate a two-

dimensional representation, and $\begin{array}{|c|} \hline 1 \\ \hline 2 \\ \hline 3 \\ \hline \end{array}$ generates a one-dimensional representation. We may check that $1^2 + 2^2 + 1^2 = 6$, the order of S_3 .

5.3 Tensors and tensor spaces

Def: Let V be an N -dimensional vector space:

$$x = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_N \end{pmatrix} \in V. \quad (5.51)$$

The product of n vectors: $x(1) \otimes x(2) \otimes \dots \otimes x(n)$ forms a *tensor of rank n* in a *tensor space* of N^n dimensions. That is the direct product space; $V \otimes V \otimes \dots \otimes V$ is called a tensor space.

We may denote the N^n tensor components by:

$$T_{i_1 i_2 \dots i_n} = x_{i_1}(1)x_{i_2}(2) \cdots x_{i_n}(n), \quad (5.52)$$

where the indices i_1, \dots, i_n range over $1, 2, \dots, N$.

Let G be a continuous group of linear transformations on V :

$$x \xrightarrow{a \in G} x' \Rightarrow x' = ax, \quad (5.53)$$

where $a \in G$ is an $N \times N$ matrix (depending on the parameters of group G). Under the action of $a \in G$, the tensor components transform according to:

$$T'_{i_1 i_2 \dots i_n} = a_{i_1 i'_1} a_{i_2 i'_2} \cdots a_{i_n i'_n} T_{i'_1 i'_2 \dots i'_n}, \quad (5.54)$$

where it is understood that repeated indices are summed over.

Notice the connection with direct product representations: In the tensor space, the transformation $a \in G$ is represented by $N^n \times N^n$ component matrix:

$$D(a) = a \otimes a \otimes a \cdots \otimes a, \quad (5.55)$$

with components

$$D(a)_{i_1 i_2 \dots i_n, i'_1 i'_2 \dots i'_n} = a_{i_1 i'_1} a_{i_2 i'_2} \cdots a_{i_n i'_n}. \quad (5.56)$$

This is a generalization of our earlier discussion on direct product matrices.

The representation $D(a)$ is generally reducible with respect to both G and S_n , the latter corresponding to symmetries with respect to permutations of the indices. For a tensor of rank $n = 1$ the relevant symmetric group is S_1 . Hence the components of a vector x which form a tensor of rank one correspond to the Young diagram \square .

Now consider the second rank tensor $T_{i_1 i_2}$. Permuting the indices gives $T_{i_2 i_1}$. We may form:

$$T_{i_1 i_2} \pm T_{i_2 i_1}, \quad (5.57)$$

forming the basis of the symmetric and antisymmetric product representations, described by the Young diagrams $\square\square$ and \square . The indices i_1 and i_2 run from 1 to N . The matrix $D(a) = a \otimes a$ may be reduced to the direct sum of an antisymmetric representation and a symmetric representation. The antisymmetric representation (of S_2) has dimension

$$d_A = \frac{N(N-1)}{2}. \quad (5.58)$$

We may see this as follows: The index i_1 takes on values $1, \dots, N$. For each i_1 , i_2 can take on $N-1$ values different from i_1 . But each $T_{i_1 i_2} - T_{i_2 i_1}$ occurs twice (with opposite sign) in this counting, hence the factor of $1/2$. This leaves a symmetric representation with dimension

$$d_S = N^2 - \frac{N(N-1)}{2} = \frac{N(N+1)}{2}. \quad (5.59)$$

Notice that the interchange of i_1 with i_2 corresponds to transposition $p = (12)$ on $T'_{i_1 i_2} = a_{i_1 i'_1} a_{i_2 i'_2} T'_{i'_1 i'_2}$, and therefore:

$$\begin{aligned} pT'_{i_1 i_2} = T'_{i_2 i_1} &= a_{i_2 i'_2} a_{i_1 i'_1} T'_{i'_2 i'_1} \\ &= a_{i_1 i'_1} a_{i_2 i'_2} T'_{i'_2 i'_1} \\ &= a_{i_1 i'_1} a_{i_2 i'_2} pT'_{i'_1 i'_2}. \end{aligned} \quad (5.60)$$

Thus, any $a \in G$ commutes with $p \in S_2$. This property remains valid for n^{th} rank tensors: Let $p \in S_n$, and

$$T_{i_1 i_2 \dots i_n} = T_{(i)} = x_{i_1}(1) x_{i_2}(2) \cdots x_{i_n}(n), \quad (5.61)$$

where we have introduced a shorter notation for the indices. Then

$$\begin{aligned} (pT)_{(i)} &= x_{i_1}(a_1) x_{i_2}(a_2) \cdots x_{i_n}(a_n) \\ &= T_{p(i)}, \end{aligned} \quad (5.62)$$

since the permutation of the n objects $1, 2, \dots, n$ is equivalent to the permutation of the indices i_1, i_2, \dots, i_n . Now,

$$\begin{aligned} (pT')_{(i)} = T'_{p(i)} &= D_{p(i)p(j)} T_{p(j)} \\ &= D_{p(i)p(j)} (pT)_{(j)} \\ &= D_{(i)(j)} (pT)_{(j)}, \end{aligned} \quad (5.63)$$

since $D(a)$ is bisymmetric, that is invariant under the simultaneous identical permutations of both the i 's and j 's.

Thus, any $p \in S_n$ commutes with any transformation of linear operator G on the tensor space. This is an important observation. It means that linear combinations which have a particular permutation symmetry transform among themselves, and can also be described by Young tableaux associated with the same Young diagram, generating an invariant subspace of S_n . The space of an n -rank tensor is reducible into subspaces of tensors of different symmetries. A tensor space can be reduced with respect to both G and S_n , and a kind of duality between a linear group G and a symmetric group S_n exists in a tensor space. We noted this earlier in our $3 \otimes 3 \otimes 3$ example under $SU(3)$. The 27-dimensional

because we cannot make a totally antisymmetric combination under S_4 from three distinct components (u, d, s) .

Under S_4 , the dimensions of the surviving irreducible representations are:

$$\boxed{1\ 2\ 3\ 4} \quad d_4^{S_4} = 1 \quad \Rightarrow \quad m_4^{SU(3)} = 1$$



$$d_{31}^{S_4} = 3 \quad \Rightarrow \quad m_{31}^{SU(3)} = 3$$



$$d_{22}^{S_4} = 2 \quad \Rightarrow \quad m_{22}^{SU(3)} = 2$$



$$d_{211}^{S_4} = 3 \quad \Rightarrow \quad m_{211}^{SU(3)} = 3$$



To determine the multiplicities under S_4 , or the dimensions under $SU(3)$, we could do the same sort of constructive analysis as we did for $2 \otimes 2$ under $SU(2)$ or $3 \otimes 3 \otimes 3$ under $SU(3)$. For example, the dimension $d_{211}^{SU(3)}$ is clearly 3, since $\begin{array}{|c|} \hline 1 \\ \hline \end{array}$ is completely antisymmetric in (u, d, s) , hence of dimension one, and adding one more $u, d,$ or s gets us to three dimensions. Likewise, for the diagram $\begin{array}{|c|c|c|} \hline & & \\ \hline & & \\ \hline & & \\ \hline \end{array}$ we have a 15-dimensional representation of $SU(3)$, with a set of linearly independent vectors:

- $uuuu$
- $dddd$
- $ssss$
- $uuud + uudu + uduu + duuu$
- $uus + uusu + usuu + suuu$
- $dddu + ddud + dudd + uddd$
- $ddds + ddsd + dsdd + sddd$
- $sssu + ssus + suss + usss$
- $sssd + ssds + sdss + dsss$
- $uudd + udud + uddu + duud + dudu + dduu$
- $uuss + usus + ussu + suus + susu + ssuu$

$ddss + dsds + dssd + sdds + sdsd + ssdd$
 $udsu + udus + uuds + usdu + usud + uUSD + sudu + suud + dusu + duus + dsuu + sduu$
 $udsd + udds + duds + usdd + dusd + sudd + sdud + dsud + ddus + dsdu + ddsu + sddu$
 $udss + usds + suds + ussd + susd + ssud + duss + dsus + sdus + dssu + sdsu + ssdu$

We could also use the general formula:

$$d_{[f]}^{SU(N)} = \prod_{i < j}^N \frac{f_i - f_j + j - i}{j - i}. \quad (5.66)$$

For example, for $\square\square\square\square$, $f = (4, 0, 0, 0)$ and

$$d_{[4]}^{SU(3)} = \binom{4+1}{1} \binom{4+2}{2} \binom{0+1}{1} = 15, \quad (5.67)$$

remembering that there is no $j = 4$ contribution since $N = 3$. Likewise,

$$d_{[31]}^{SU(3)} = \binom{3-1+1}{1} \binom{3+2}{2} \binom{1+1}{1} = 15, \quad (5.68)$$

$$d_{[22]}^{SU(3)} = \binom{2+1}{1} \binom{2+2}{2} = 6, \quad (5.69)$$

$$d_{[211]}^{SU(3)} = \binom{2}{1} \binom{3}{2} = 3. \quad (5.70)$$

Notice that

$$15 \times 1 + 15 \times 3 + 6 \times 2 + 3 \times 3 = 81 = 3^4, \quad (5.71)$$

so all dimensions in the representation are accounted for in our reduction to irreducible representations. We notice that there are no singlets in this decomposition. A physical application of this is in $SU(3)_{\text{color}}$, where we find that no colorless (i.e., color singlet) four-quark states are possible. Under the hypothesis that the physical hadron states are colorless, this implies that we should not observe any particles made of four quarks.

5.4 Exercises

1. How many transpositions are required to generate a k -cycle? Hence, what is the parity of a k -cycle?
2. Show that, for any $P_a \in S_n$:

$$\begin{aligned}
 P_a S &= S \\
 P_a A &= A P_a = \delta_{P_a} A \\
 S^2 &= S \\
 A^2 &= A.
 \end{aligned}$$

3. We gave the representation of one element of the two-dimensional irreducible representation of S_3 in basis $\{\psi'_1, \psi'_2\}$ in Eqn. 5.38. Find the other matrices in this representation.
4. Quarks are spin- $\frac{1}{2}$ particles, hence they are fermions. According to quantum mechanics, the wave function of a system of identical fermions must be antisymmetric under the interchange of the fermions (the celebrated “connection between spin and statistics”). To see the idea, first consider a system of two electrons (an electron is also a spin- $\frac{1}{2}$ particle). We put the “first” electron at position x_1 , with spin orientation s_1 , and the second at x_2 with spin orientation s_2 . The wave function is $\psi(x_1, s_1; x_2, s_2)$. This wave function must be antisymmetric under interchange of the two electrons:

$$\psi(x_2, s_2; x_1, s_1) = -\psi(x_1, s_1; x_2, s_2). \quad (5.72)$$

Suppose our two electrons are in an orbital angular momentum $L = 0$ state. The spin states may be described by the z components of the spins, $\pm\frac{1}{2}$, which we'll represent with arrows, \uparrow for spin “up” and \downarrow for spin “down”. But in making a system of two electrons (with $L = 0$), we are generating a product representation of $SU(2)$ in angular momentum: $2 \otimes 2 = 3 \oplus 1$. That is the irreducible representations of our total angular momentum state are three-dimensional, corresponding to total spin one, and one-dimensional or spin zero. We have already worked out the symmetries of these combinations in this note: the spin one system is symmetric under interchange, and the spin zero is antisymmetric. Note that, since we have specified $L = 0$ the wave function is symmetric under the interchange of the spatial coordinates. We may conclude that the only way we can put two electrons together in an $L = 0$ state is with total spin $S = 0$:

$$\psi(x_2, s_2; x_1, s_1) = \frac{1}{\sqrt{2}}(|e \uparrow; e \downarrow\rangle - |e \downarrow; e \uparrow\rangle), \quad (5.73)$$

where the symmetry under spatial interchange is not explicitly shown.

Now let us return to quarks, and consider baryons. To keep this simple, we'll also put our three quarks together in a state with no orbital angular momentum (S -wave). That is, the spatial state is symmetric under the interchange of any pair of quarks. We'll regard the “flavor” quantum number (“ u ”, “ d ”, or “ s ”, or equivalently, I_3, Y) as analogous to the spin projections, and regard them as additional quantum numbers that get interchanged when we act on a wave function with permutations of the quarks.

Treating the angular momentum, when we combine three quarks in S -wave, we build the $2 \otimes 2 \otimes 2 = 4 \oplus 2 \oplus 2$ representation of $SU(2)$. Thus, the three quarks could be in a total spin state of $1/2$ or $3/2$. The spin $3/2$ state is clearly symmetric under interchange of the spins. The two spin $1/2$ representations have mixed symmetry. We may choose a basis

for one of these representations that corresponds to symmetry under the interchange of the quarks at x_1 and x_2 (or, quarks 1 and 2, for short):

$$\begin{aligned}\chi_+^\lambda &= -\frac{1}{\sqrt{6}}(\uparrow\downarrow\uparrow + \downarrow\uparrow\uparrow - 2\uparrow\uparrow\downarrow) \\ \chi_-^\lambda &= -\frac{1}{\sqrt{6}}(\uparrow\downarrow\downarrow + \downarrow\downarrow\downarrow - 2\downarrow\downarrow\uparrow).\end{aligned}\quad (5.74)$$

Likewise, the spin basis wavefunctions for the other two dimension wave function, with antisymmetry under interchange of the first two quarks, may be chosen as:

$$\begin{aligned}\chi_+^\rho &= \frac{1}{\sqrt{2}}(\uparrow\downarrow\uparrow - \downarrow\uparrow\uparrow) \\ \chi_-^\rho &= \frac{1}{\sqrt{2}}(\uparrow\downarrow\downarrow - \downarrow\uparrow\downarrow).\end{aligned}\quad (5.75)$$

We must deal with a small (but extremely important in physical implication!) complication before we construct the (spin, flavor) wave functions of the S -wave baryons. Consider the Δ^{++} baryon. This is made of three u quarks, clearly in a symmetric flavor state. It is also a spin- $\frac{3}{2}$ particle, with all of the quark spins aligned, that is, in a spin symmetric state. Thus, the Δ^{++} is symmetric in spatial interchange (since it is S -wave), flavor interchange, and spin interchange. Combined, it appears that we have built a baryon which is symmetric under interchange of the constituent quarks. But this violates our fermion principle, which says it must be antisymmetric. This observation was historically one of the puzzles in the 1960's when this model was proposed. Eventually, we learned that the most promising way out was to give the quarks another quantum number, called "color". To combine three quarks with three different colors requires a minimum of three colors, hence the hypothesis that there are three colors, and the relevant group for rotations in color space is also the $SU(3)$ group. It is a hypothesis (perhaps justifiable with QCD) that the physical particles (such as baryons) we see are overall colorless. That is, the color basis wave function corresponds to a singlet representation of $SU(3)_{\text{color}}$. We have already seen that the one-dimensional representation in the decomposition $3 \otimes 3 \otimes 3 = 10 \oplus 8 \oplus 8 \oplus 1$ is antisymmetric under interchange. Thus, the introduction of color saves our fermi statistics. We simply assume that the color wavefunction of baryons is antisymmetric. Then the (space, spin, flavor) wave function must be overall symmetric.

Now consider the proton, a spin $\frac{1}{2}$ baryon made with two u 's and a d . In $SU(3)_{\text{flavor}}$, the proton wave function must be some linear combination of

the basis states corresponding to the $\begin{array}{|c|c|} \hline 1 & 2 \\ \hline 3 & \\ \hline \end{array}$ and $\begin{array}{|c|c|} \hline 1 & 3 \\ \hline 2 & \\ \hline \end{array}$ representations (you may wish to ponder why there is no $\begin{array}{|c|c|c|} \hline 1 & 2 & 3 \\ \hline \end{array}$ piece). Let us call the (12)-symmetric wave functions ϕ^λ and the (12)-antisymmetric wave functions ϕ^ρ :

$$\phi_{uud}^{\lambda} = -\frac{1}{\sqrt{6}}(udu + duu - 2uud) \quad (5.76)$$

$$\phi_{uud}^{\rho} = \frac{1}{\sqrt{2}}(udu - duu). \quad (5.77)$$

The problem you are asked to solve is: What is the wave function of a spin up proton? Assume that the spatial wave function is symmetric, and give the spin/ flavor wave function. It is perhaps easiest to use some notation such as kets, forming the wave function from kets of the form $|u \uparrow u \uparrow d \downarrow\rangle$, etc.

Note: I won't go into the physics further, but it should be remarked that this isn't just an idle exercise in mathematics – this wave function implies observable physical consequences on quantities such as the magnetic moment of the proton.