

Particles and Symmetries

Physics 226, Spring 2014

Stephen D. Ellis

Based on notes by Laurence G. Yaffe

University of Washington

Contents

-1 Preface	1
0 Introduction	4
1 Math Methods: A Quick Review	9
1.1 Power Series	9
1.2 The Exponential Function	10
1.3 Along the Real Axis: The Hyperbolic Functions	10
1.4 Along the Imaginary Axis: The Sinusoidal Functions	13
1.5 Rotations in 3-D	16
1.6 Examples	20
2 Special relativity	22
2.1 Galilean relativity	22
2.2 Constancy of c	24
2.3 Clocks and rulers	24
2.4 Observational tests	26
2.5 Superluminal motion?	28
2.6 Further resources	28
2.7 Example Problems	28
3 Minkowski spacetime	32
3.1 Events	32
3.2 Reference frames	32
3.3 Lightcones	34
3.4 Simultaneity	34
3.5 Lorentz transformations	36
3.6 Spacetime vectors	39
3.7 Units: What is large and what is small?	42

3.8	Minkowski spacetime	44
3.9	The pole and the barn	46
3.10	Causality	49
3.11	Example Problems	50
4	Relativistic dynamics	57
4.1	Proper time	57
4.2	4-velocity	58
4.3	Relativistic Addition of Velocities	59
4.4	4-momentum	60
4.5	4-force	62
4.6	Constant acceleration	63
4.7	Plane waves	65
4.8	Electromagnetism	67
4.9	Scattering	68
4.10	Example Problems	71
5	QM and Angular Momentum	73
5.1	Angular Momentum Operators	73
5.2	Spin 1/2 in Vector/Matrix Notation	78
5.3	Spin 1 in Vector/Matrix Notation	80
5.4	Examples	81
6	Known particles	85
6.1	Ordinary matter	85
6.2	Stability of particles	90
6.3	Nuclear decays	91
6.4	Photons	92
6.5	Antiparticles	93
6.6	Leptons	94
7	Quarks and hadrons	96
7.1	Quark flavors	97
7.2	Hadrons	97
7.3	Mesons	100
7.4	Baryons	105
7.5	Baryon wavefunctions	107
7.6	Baryon number	111

7.7	Hadronic decays	111
7.8	Sample calculations	117
8	Symmetries	121
8.1	Quantum dynamics	121
8.2	Symmetries	122
8.3	Continuous symmetries	126
8.4	Spacetime symmetries	127
8.5	Charge, lepton, and baryon number	129
8.6	Approximate symmetries	130
8.7	Flavor symmetries	131
8.8	Isospin	131
8.9	Parity	135
8.10	Charge conjugation	137
8.11	CP	139
8.12	Time reversal and CPT	140
8.13	Sample calculations	140
9	Weak interactions	143
9.1	Muon decay	144
9.2	Neutrino scattering	147
9.3	Weak gauge bosons	150
10	Introduction to Group Theory	154
10.1	Definitions	154
10.2	Finite Groups	155
10.3	Lie Groups	157
11	Multiplets and Young Diagrams	164
11.1	Basic Definitions	164
11.2	Combine Multiplets	165
	Index	168

Chapter -1

Preface

The preparation of these notes began in 2008 when my colleague Laurence Yaffe taught the first offering of the newly designed class, Particles and Symmetries. This class was created to give undergraduate physics students, early in their studies, an introduction to the fundamental constituents of matter and the symmetries which characterize their interactions. The presentation begins with an overview of special relativity, and then moves into an examination of the building blocks of the current Standard Model of particle physics. The material, by design, takes advantage of the fact that a remarkable amount of particle physics may be understood quantitatively using relatively few basic concepts. Students are assumed to have had introductory physics and at least one quarter of quantum mechanics introducing state vectors (bars and kets), quantum time evolution, observables and expectation values, spin-1/2 and related two-state systems, quantized angular momentum, and quantized harmonic oscillators. Facility with calculus, linear algebra, and basic mathematical methods is also assumed. Prior exposure to special relativity, or particle physics, is not required.

The spring 2014 version of these notes incorporates or adapts a number of subsequent changes to the original notes from both myself (I have taught Particles and Symmetries several times starting in 2011) and Professor Yaffe. His contributions are gratefully acknowledged. Note, in particular, that this spring there are explicitly Chapters (1 and 5) reviewing the most relevant concepts from Physics 227 and 225 (respectively). Note also that there is no required textbook for this course, only some suggestions of books that you may find useful. On the other hand, it is essential that you read the notes according to the reading assignments on the class webpage and come to class prepared to discuss the covered topics, especially those that you might find confusing. We will have frequent in-class “clicker quizzes” to encourage you to do the reading and also help me identify those concepts that are causing difficulty.

Next, here are some comments regarding conventions. Arrows are used to indicate three-dimensional spatial vectors, such as \vec{x} . Components of spatial vectors are written as x^j , with a Latin index (such as j), which runs from 1 to 3 (corresponding to x to z). Four-dimensional spacetime vectors, which are introduced in chapter 3, will not be marked with a vector sign, but their meaning should be clear from context. Components of a spacetime vector are written as x^μ , with a Greek index (such as μ) running from 0 to 3. Sadly, there are two different conventions in common use in the physics community for defining the dot product of spacetime vectors, differing by an overall minus sign. These notes use the so-called West Coast metric (+,-,-,-) that renders time-like invariants positive, but the dot product of spacetime vectors having vanishing time components the negative of the usual three-dimensional dot product. The notes from Prof. Yaffe use the so-called East Coast metric

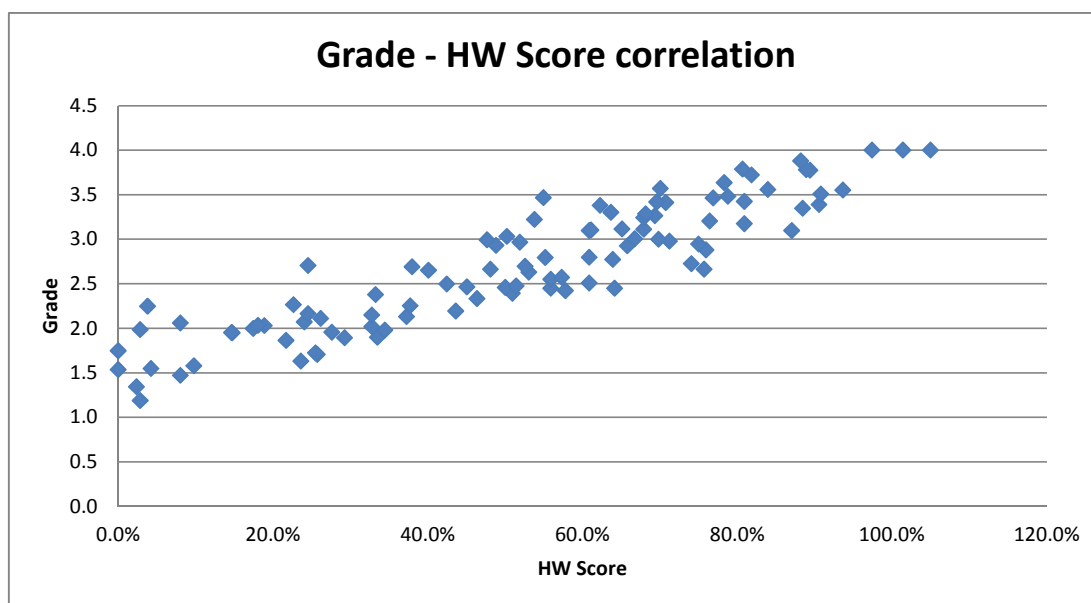


Figure 1: Correlation between HW scores and Final Grade for Spring 2013

(-,+,+,+) (that he labels "the only sensible choice"), which makes the dot product of spacetime vectors having vanishing time components the same as the usual three-dimensional dot product, but time-like quantities are negative.

Finally we include a few words concerning the role of the Homework (HW) assignments. Since this is a physics course with a focus on learning to quantitatively analyze the properties of physical systems, the HW provides the core of this course. It is essential that you do the HW in order to achieve command of the various concept and techniques covered in this course. This connection is confirmed by the fact that the HW and the exams, for which you should consider the HW as practice, constitute the basis of the course grade.

An illustration of the correlation between HW scores and Final Grade for the class last Spring is provided in Fig. 1, which clearly indicates a fairly direct correlation between performance on the HW and in the class. In particular, students who put little effort into the HW, invariably received a poor grade in the course. There is, in fact, an important further dimension to the question of learning from the HW in this class. Solutions to all HW exercises are posted on the web (Catalyst) as soon as the HW is turned in. It is important to look through these solutions, especially if you had difficulty solving a particular exercise (or all of them), in order to identify and remedy any misunderstandings that may be causing you difficulty. As part of the "postmortem" last spring, I looked through the not-picked-up HW papers still in my office at the end of the quarter. Again there was a strong correlation between the fraction of HW (turned-in but) not picked up and the final grade, as illustrated in Fig. 2. Several students turned in most (or all) of the HW but never picked it up. In many such cases, these students were apparently unaware of the mistakes they made on the HW and repeated these mistakes on the exams (and received poor final grades). You are encouraged to make the HW a central part of your study process. Do the HW (or at least attempt it), then study the solutions to identify the issues that are causing you trouble and then rectify that

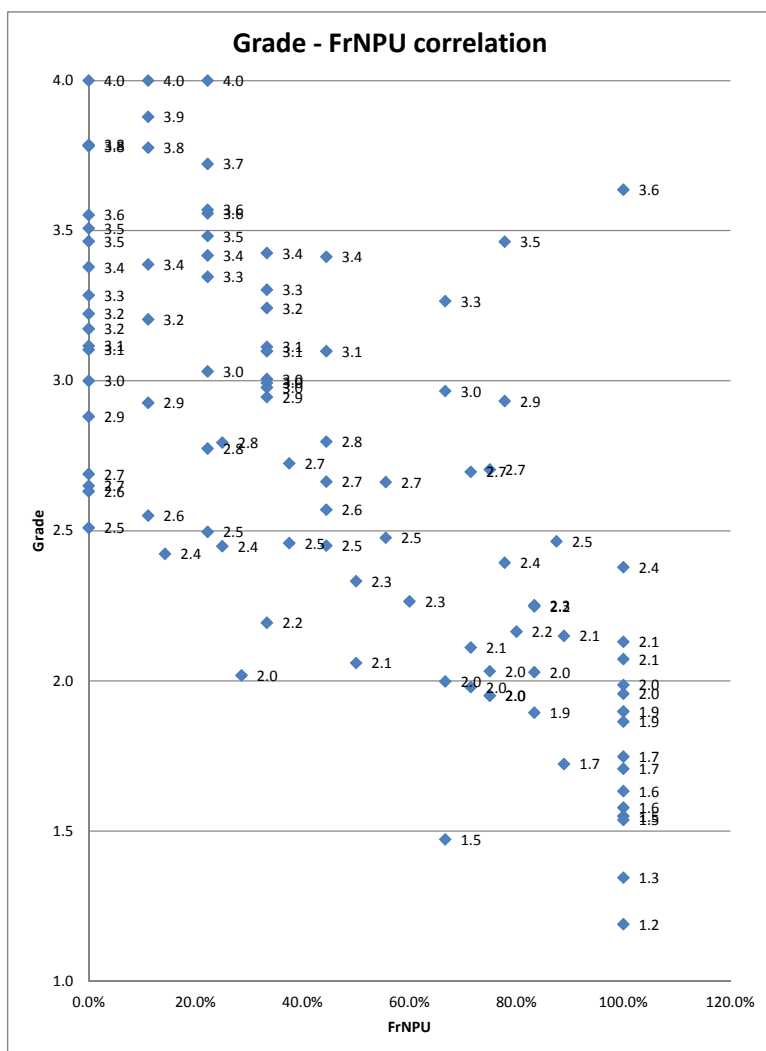


Figure 2: Correlation between Final Grade and the fraction of HW not picked up for Spring 2013

misunderstanding.

Note that there is a late HW option (after solutions are posted) for partial credit. I am also hoping to have sufficient TA resources to allow a re-submittal option this Spring so that you can correct your HW and turn it in again for more credit.

Stephen D. Ellis

April 2014

Chapter 0

Introduction

As we start this study of Particles and Symmetries it seems appropriate to try to provide an overview, *i.e.*, some version of the big picture goals for this course. As the title of the course implies, our goal is to provide an introduction to an area of physics that has seen dramatic progress in the last 50 years - particle physics. A central tool during this progress has been the exploitation of the underlying symmetries, the other subject in the title, of the interactions of these particles. The short version summary of this progress is encoded in the so-called Standard Model of Particle Physics (typically denoted the SM), which identifies the particles (degrees of freedom) and interactions between the particles relevant to the understanding of nearly all of the physical universe. When we include collective behavior (quarks bound in nuclei, electrons bound in atoms, atoms bound in solid matter) plus classical gravity, we have a nearly complete explanation¹ for the physics of the very large, *e.g.*, the evolution of the universe from very early times, down to the physics at the shortest distances now observable at particle accelerators. To have full quantitative command of this fundamental understanding requires a tool not at our disposal - quantum field theory. However, we will be able to outline a “semi-classical” (building block) picture of particle physics using only special relativity, quantum mechanics and symmetries, which is remarkably complete and relatively quantitative. The most recent addition to the SM is the so-called Higgs boson (named after the British theoretical physicist Peter Higgs), whose initial discovery at the Large Hadron Collider (the LHC) at CERN (in Geneva, Switzerland) was announced on the 4th of July, 2012, and whose detailed properties have by now been largely confirmed to match those expected. Indeed the Nobel Prize in physics was awarded to Peter Higgs and Francois Englert last autumn for their work (50 years ago) that led to the prediction of this particle. Interestingly the existence and interactions of this particle were *predicted* based on theoretical (*i.e.*, mathematical) considerations, and the confirmation of the expected properties constitutes a major step forward in particle physics research.

From a pedagogical perspective, this study of particle physics will help us learn about two of the pillars of twentieth century physics, special relativity and quantum mechanics. The corresponding dimension-full constants c and \hbar serve to set the “scale” for most of what we observe. The “uncertainty principle” of quantum mechanics, which says, for example, that the product the uncertainties

¹The primary *missing* pieces are “dark matter” that serves to gravitationally binds galaxies but does not form stars, and “dark energy” that is apparently causing an *acceleration* in the expansion rate of the universe. The former category includes approximately 25% of the energy content of the universe, while the latter is about 70%. Stuff like us is a 5% effect.

in where we are (Δx) and where we are going (Δp , with p the momentum) is bounded below, $\Delta x \Delta p \geq \hbar/2$, guarantees that particles confined to small volumes must have large momenta. Thus, since the masses of individual particles (of the variety discussed in this course) are so small, they are very often (*i.e.*, in most reference frames of interest) moving with velocities approaching the speed of light. This situation will provide us with the opportunity to discuss special relativity in detail. The exercises will allow us to practice using special relativity to describe the kinematics of particle collisions at high energy, especially the role of 4-dimensional momentum conservation (which is itself associated with the symmetry associated with the invariance of physics under translations in space *and* time), and the speed of light as the *universal* speed limit. We will want to develop facility with 4-vector notation and the transformations (boosts) that take us between different inertial reference frames. Similarly, since the total angular momentum of an individual particle (*i.e.*, its “spin”) is of order \hbar , the fact that angular momentum is quantized on this scale means that quantum mechanics will play a central role. We will make use of (and practice using) the uncertainty principle and the important role of the eigenstates of (commuting) operators. In particular, we will want to become efficient at using operator notation to relate different states within the degenerate multiplets that arise due to symmetries. You should have seen this structure in the context of states of *definite* total angular momentum, but *varying* angular momentum component along 1 spatial axis (*e.g.*, $|J, J_z\rangle$). Transformations between these states are accomplished using the (hopefully familiar) raising/lowering (ladder) operators, which are just a special form of a rotation (*i.e.*, a particle with its spin pointing “up” will look like its spin is pointing “down” if you stand on your head). Finally we will discuss how to use symmetries (and the underlying mathematics of group theory) to tie this all together and keep the mathematics simple. This approach will include the use of “approximate” symmetries - where there is no true (exact) invariance under certain transformations, but rather the transformations induce only numerically “small” changes. This will allow us to simplify complex computations in terms of *perturbative* expansions organized in terms of powers of these small changes (such expansions are an essential tool for your physics toolbox). All during the quarter we should be honing our skills for making order-of-magnitude estimates, *i.e.*, being able to estimate the numerical value of a given quantity even when we do not know (or do not understand) all of the details.

Do not be concerned if all of these concepts are not clear at the outset. Also, you should expect that initially portions of our discussions may seem more “abstract” than you are accustomed to. However, you should become concerned if clarity does not develop quickly over the next 10 weeks. Finally do not be surprised if our approach seems somewhat circular. We will try to introduce the relevant vocabulary and concepts in the early lectures, and then return to the same concepts in the context of a more complete formalism in the latter lectures.

We end this Introduction with a brief summary of the ideas, definitions, mathematical tools that you have already (hopefully) mastered in the prerequisite courses in mathematics and physics (especially Phys. 225, 227 and 228, although these classes do not always cover the same material with different instructors). If parts of the following do not seem familiar, you are encouraged to perform a more thorough review, and/or come chat with me.

From Phys. 227-8:

1. Power series expansion: We very often want to expand a mathematical expression in terms of some small parameter, which here we call $|\delta| \ll 1$ and note that δ is *necessarily* dimensionless. For this purpose it is useful to recall the simple power series expansion (really the usual Taylor series expansion),

$$(1 + \delta)^\alpha \approx 1 + \alpha\delta + (\alpha(\alpha - 1)/2!) \delta^2 + \mathcal{O}(\delta^3).$$

This expansion is valid independent of the signs of α and δ , but actually requires that the product $|\alpha\delta| \ll 1$ to be useful. However, the exponent α is typically of order unity.

2. Complex numbers:

$$z = x + iy = re^{i\phi} = r \cos \phi + ir \sin \phi, \operatorname{Re} z = x, \operatorname{Im} z = y, |z| = r = \sqrt{x^2 + y^2}, \phi = \arctan \frac{y}{x}.$$

3. Exponential function e^z : Defined by

$$e^z = \sum_{n=0}^{\infty} \frac{z^n}{n!} = \frac{d}{dz} e^z$$

4. Sinusoidal functions $\sin z, \cos z$: Defined by

$$\sin z = \frac{e^{iz} - e^{-iz}}{2i} = \sum_{n=0}^{\infty} \frac{(-1)^n z^{2n+1}}{(2n+1)!}, \frac{d}{dz} \sin z = \cos z.$$

$$\cos z = \frac{e^{iz} + e^{-iz}}{2} = \sum_{n=0}^{\infty} \frac{(-1)^n z^{2n}}{(2n)!}, \frac{d}{dz} \cos z = -\sin z.$$

$$\cos^2 z + \sin^2 z = 1.$$

5. Hyperbolic functions $\sinh z, \cosh z$: Defined by

$$\sinh z = \frac{e^z - e^{-z}}{2} = \sum_{n=0}^{\infty} \frac{z^{2n+1}}{(2n+1)!}, \frac{d}{dz} \sinh z = \cosh z.$$

$$\cosh z = \frac{e^z + e^{-z}}{2} = \sum_{n=0}^{\infty} \frac{z^{2n}}{(2n)!}, \frac{d}{dz} \cosh z = \sinh z.$$

$$\cosh^2 z - \sinh^2 z = 1.$$

6. Vectors and Matrices: Facility with vector descriptions of both configuration space (“where the particle is”) and momentum space (“where the particle is going”) are central to this course. We will use vector notation in two dimensions (labeled 1 and 2 or x and y), three dimensions (labeled 123 or xyz) and four dimensions (labeled 0, 123 or t, xyz), where the role of t will be distinguished from xyz . This facility with vector notation should include some familiarity with the representations of transformations like rotations as matrices. In particular, we would like to be able to think of transformations in four dimensions - Lorentz transformations - as generalizations (with some “funny” minus signs) of rotations in three dimensions.

7. Matrices, operators and commutation relations: A (hopefully) familiar feature of rotations in three dimensions is that, unlike rotations in two dimensions, 3D rotations do *not* commute in general. This essential property is explicitly represented by the corresponding rotation matrices, *i.e.*, $M_1 M_2 - M_2 M_1 \equiv [M_1, M_2] \neq 0$.

8. Another important set of functions of mathematical physics, which hopefully you were introduced to in Physics 227 or 228 (or somewhere that you learned about physics in 3 spatial dimensions), are the spherical harmonics, $Y_{\ell,m}(\theta, \phi)$. These functions form a *complete* (and orthonormal) set of functions on the surface of a sphere. This means that *any* function of θ and ϕ , $0 \leq \theta \leq \pi$ and $0 \leq \phi < 2\pi$, can be represented as a linear combination of the $Y_{\ell,m}(\theta, \phi)$ with ℓ integer valued and $-\ell \leq m \leq \ell$, also integer valued. We can interpret the $Y_{\ell,m}(\theta, \phi)$ as the eigenstates of the orbital angular momentum L and z projection L_z . The corresponding eigenvalues are $\ell\hbar$ and $m\hbar$.

Quite generally we find it useful to have complete (and orthogonal) sets of functions of a certain class. For example, the functions $\sin(2\pi tn/T)$ and $\cos(2\pi tn/T)$ with n an integer comprise a complete (and orthogonal) set of functions of t that are periodic with period T ($f(t) = f(t+T)$). The $Y_{\ell,m}(\theta, \phi)$ above are such a set for functions that are well behaved (continuous) on the surface of a sphere. Such functions are almost always defined in terms of solving an eigenvalue/eigenfunction problem. This is especially true in the context of quantum mechanics.

From Phys. 225:

1. The uncertainty principle, $\Delta x \Delta p \geq \hbar/2$, requires that a small uncertainty in the location of a particle must be matched by a large uncertainty in the particle's momentum. We understand this point “physically” through that the fact that fine spatial resolution (*i.e.*, small Δx) corresponds to interactions with particles with short wavelengths (λ) and thus large momenta, where $p = h/\lambda$, with λ the de Broglie wavelength. This last point emphasizes the essential quantum mechanical point that real particles simultaneously exhibit *both* point-like (classical) and wave-like (quantum mechanical) behavior. As you may recall this feature is often used to motivate that fact that only certain energies and momenta are present for a particle confined to a box or an orbit, *i.e.*, the wavelength must be such as to “fit” in the box.

The uncertainty principle encodes the fact that the processes corresponding to the measurement of location and the measurement of momentum “interact” with each other, *i.e.* the corresponding operators do *not* commute. Nonzero commutators play an essential role in quantum mechanics and in this course. Commutation relations (equations involving commutators) are at the heart of defining symmetries in terms of the underlying group theory.

2. The possible states of quantum mechanical (QM) systems are typically labeled in terms of the eigenvalues of a (complete) set of commuting operators. An example is the labeling of states in terms of the total energy and the total angular momentum J , or “spin” S for a single particle. Further, these quantities are “quantized” in the sense that only certain discrete (*not* continuous) values are allowed: $E_n = (n + 1/2)\hbar\omega$ (for quantized harmonic oscillators with n an integer), and $S = (n + 1/2)\hbar$ (fermions) or $S = n\hbar$ (bosons). (We are unaware of this quantization in the context of “classical physics” simply because n is so large for classical systems, and a change by unity (*i.e.*, one unit of \hbar) is very difficult to detect. For example, recall that the number of atoms in a mole is of order 6×10^{23} and 10^{-23} is a *very* small number.)

To the extent that the underlying physics is invariant under translations in time and rotations in space total energy and total angular momentum are constants of the motion. *i.e.*, are “conserved” quantities (a concept from Introductory physics).

The simplest spin system is spin zero with the corresponding state vector (or “bra”) labeled as $|S, S_z\rangle = |0, 0\rangle$. This representation of the rotation group has only this *one* element and it follows that it *cannot* be changed by a rotation, *i.e.*, the spin structure is the same in all reference frames related by rotations and is typically referred to as the *singlet* representation. The next simplest case is spin 1/2 with two elements in the representation, $|S, S_3\rangle = |1/2, 1/2\rangle$ and $|S, S_3\rangle = |1/2, -1/2\rangle$ (the doublet representation). This fact leads to the *pair* of possible outcomes in the Stern-Gerlach experiment you discussed in Phys. 225. An appropriate rotation (or boost) can turn one of these states into the other. In general, a representation of spin S (with S an integer or half-integer) corresponds to $2S + 1$ elements (distinct states) labeled by $S_3 = S, S_3 = S - 1 \dots S_3 = -S$ (for a fixed choice of the $\hat{3}$ direction). Since the shift in S_3 at each step is unity (this is the quantization of spin in terms of \hbar), this arithmetic only works for S an integer or a half-integer, *i.e.*, for $2S + 1$ an integer. It is these states (same S but different S_3) that are transformed into each other by rotations (or the ladder-operator) and thus constitute a representation of the rotation group.

3. Quantum Mechanics as a description of “small” systems is characterized by “wave functions” for the energy eigenstates, which are *complex* valued solutions of differential equations, *e.g.*, Schrödinger’s Equation. The amplitudes for something to happen in QM are expressed in terms of “matrix elements” of operators, *i.e.*, an operator between a “bra” ($\langle|$) and a “ket” ($| \rangle$) representing the

“outgoing” and “incoming” states, respectively. The underlying arithmetic is greatly simplified by writing the wave functions or state vectors in terms of the eigenstates of the relevant operators, which typically means in terms of representations of the relevant symmetries. Hopefully the linear algebra describing the arithmetic of the corresponding state vectors and matrix elements (matrices) is a familiar concept for simple (low dimensionality) systems. (If not, we will work to make it familiar.) Finally note that probabilities, rates, cross sections are proportional to the absolute squares of amplitudes, *i.e.*, real numbers.

We will explicitly discuss some of the most useful of these topics in Chapters 1 and 5. The reader is also strongly encouraged to review the lecture notes from the last time I taught Phys. 227-228 (2008-2009), which are available [here](#). The content of essentially all of the first ten lectures has application to our studies in Physics 226. Finally note that these 227/8 notes include worked examples and samples of how the computer program *Mathematica* can be used to both think about (make plots, *etc.*) and solve relevant exercises. That will remain true in Physics 226.

Chapter 1

Math Methods: A Quick Review

In your Elementary Mathematical Physics course (227/8) you learned about a variety of functions and techniques that will be useful in Physics 226. We will attempt to review the most relevant of those, especially for the analysis of Special Relativity, in these notes. The reader is also strongly encouraged to review the lecture notes from the last time I taught Phys. 227-228 (2008-2009), which are available here. The content of essentially all of the first ten lectures has application to our studies in Physics 226. Also note that these 227/8 notes include worked examples and samples of how the computer program *Mathematica* can be used to both think about (make plots, *etc.*) and solve relevant exercises.

1.1 Power Series

A extremely powerful tool for both understanding and evaluating functions is the power series expansion, typically the Taylor series expansion,

$$f(z) = \sum_{n=0}^{\infty} c_n z^n, \quad (1.1.1)$$

where, for the expansion about the origin, the coefficients are the derivatives at the origin divided by $n!$,

$$c_n = \frac{d^n f(z)}{d^n z} \Big|_{z=0} \frac{1}{n!}. \quad (1.1.2)$$

The analytic properties of the function $f(z)$ in the complex z -plane are then characterized by the convergence properties of this series. Note that this power series, where it converges, serves to define the function whether z is a real number, a complex number or matrix valued. We will use this last point later in this Chapter.

A particularly useful approximate result arises when we have a small parameter, say $|\delta| \ll 1$, so that, for example,

$$(1 + \delta)^\alpha \approx 1 + \alpha\delta + (\alpha(\alpha - 1)/2!) \delta^2 + \mathcal{O}(\delta^3). \quad (1.1.3)$$

This expansion is valid independent of the signs of α and δ , but actually requires that the product $|\alpha\delta| \ll 1$ to be useful. However, the exponent α is typically of order unity.

1.2 The Exponential Function

One of the most useful functions of mathematical physics is the exponential function. It is effectively defined as the solution of the following (trivial) first order differential equation plus boundary/initial condition,

$$\frac{df(z)}{dz} = f(z); \quad f(0) = 1 \Rightarrow f(z) = \exp(z) = e^z. \quad (1.2.1)$$

So the exponential function is the eigenfunction of the derivative operator with a specific boundary condition, *i.e.*, specific normalization. Iterating the form of Eq. (1.2.1) leads to the conclusion that all derivatives of the exponential function at the origin have unit value,

$$\left. \frac{d^n f(z)}{d^n z} \right|_{z=0} = 1. \quad (1.2.2)$$

Thus the Taylor series expansion of this function about the origin is given by

$$e^z = \sum_{n=0}^{\infty} \frac{z^n}{n!}. \quad (1.2.3)$$

Since only the $n = 0$ term contributes at the origin, this sum clearly satisfies the boundary condition,

$$e^0 = 1, \quad (1.2.4)$$

as long as you know that, by convention, $0! = 1$ (but note that $(-1)! = \infty$). Likewise taking a derivative yields

$$\frac{de^z}{dz} = \sum_{n=1}^{\infty} \frac{z^{n-1}}{(n-1)!} = \sum_{n=0}^{\infty} \frac{z^n}{n!} = e^z. \quad (1.2.5)$$

A little 227 style analysis confirms that this sum is convergent in the entire z plane (*i.e.*, it is singular only at ∞), and serves to define the exponential function as an analytic function everywhere in the finite (complex) z plane. To explore this function we first focus on the behavior of the exponential function separately along both the real and imaginary axes.

1.3 Along the Real Axis: The Hyperbolic Functions

Focusing on the case $z = x + iy \rightarrow x$ with x (and y) purely real, we can easily confirm that the series in Eq. (1.2.3) increases quickly with increasing positive x and diverges to ∞ as $x \rightarrow +\infty$. On the hand, for negative x , there is substantial cancellation between the terms, which alternate in sign, and the exponential function decreases rapidly as x becomes more negative. In the limit $x \rightarrow -\infty$ the exponential function vanishes ($e^{-\infty} \equiv 0$). This behavior is illustrated in Fig. 1.1. Since the logarithm function is the inverse of the exponential function, a semi-log plot, as on the right in Fig. 1.1, yields “linear” behavior in the plot, *i.e.*, $\ln e^x = x$.

It is useful to define even and odd functions in terms of the real exponential (remember that symmetries are important), which yields the so-called hyperbolic functions:

$$\begin{aligned} \cosh(x) &\equiv \frac{e^x + e^{-x}}{2}, & \cosh(x) &= \cosh(-x), \\ \sinh(x) &\equiv \frac{e^x - e^{-x}}{2}, & \sinh(x) &= -\sinh(-x). \end{aligned} \quad (1.3.1)$$

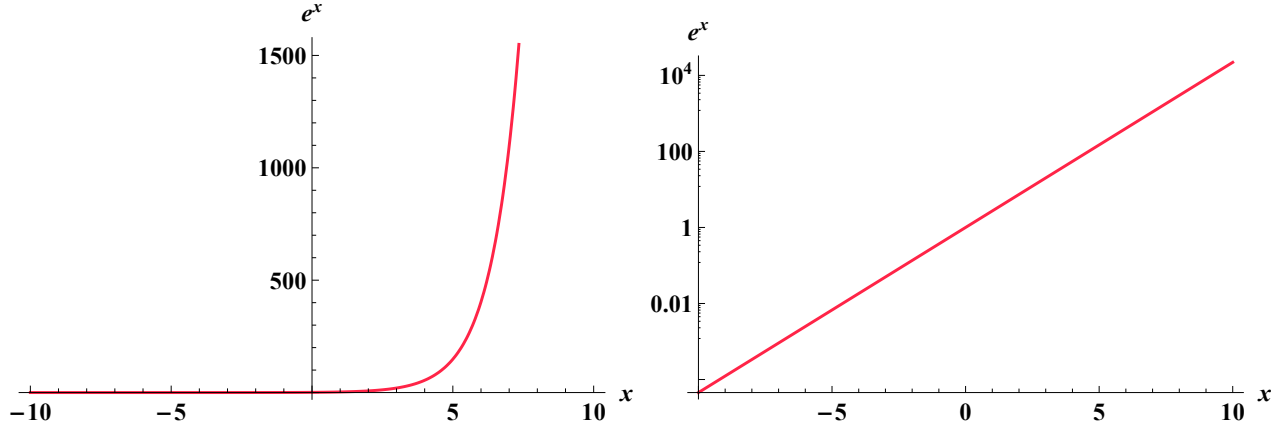


Figure 1.1: The exponential function along the real axis: LEFT- linear scale, RIGHT- logarithmic scale.

Using Eq. (1.2.2) the hyperbolic functions have the expected series expansions in terms of even and odd powers;

$$\begin{aligned}\cosh(x) &= \sum_{n=0}^{\infty} \frac{x^{2n}}{(2n)!}, \\ \sinh(x) &= \sum_{n=0}^{\infty} \frac{x^{2n+1}}{(2n+1)!}.\end{aligned}\tag{1.3.2}$$

Other useful properties follow from these definitions and Eq. (1.2.1). In particular, we have

$$\begin{aligned}\frac{d \cosh(x)}{dx} &= \frac{e^x - e^{-x}}{2} = \sinh(x), \\ \frac{d \sinh(x)}{dx} &= \frac{e^x + e^{-x}}{2} = \cosh(x).\end{aligned}\tag{1.3.3}$$

So it follows that the hyperbolic functions are eigenfunctions of the second order derivative (with eigenvalue +1),

$$\begin{aligned}\frac{d^2 \cosh(x)}{d^2 x} &= \cosh(x), \\ \frac{d^2 \sinh(x)}{d^2 x} &= \sinh(x).\end{aligned}\tag{1.3.4}$$

We also have

$$\begin{aligned}\cosh^2(x) &= \frac{e^{2x} + e^{-2x} + 2}{4}, \\ \sinh^2(x) &= \frac{e^{2x} + e^{-2x} - 2}{4},\end{aligned}\tag{1.3.5}$$

so that

$$\cosh^2(x) - \sinh^2(x) = 1.\tag{1.3.6}$$

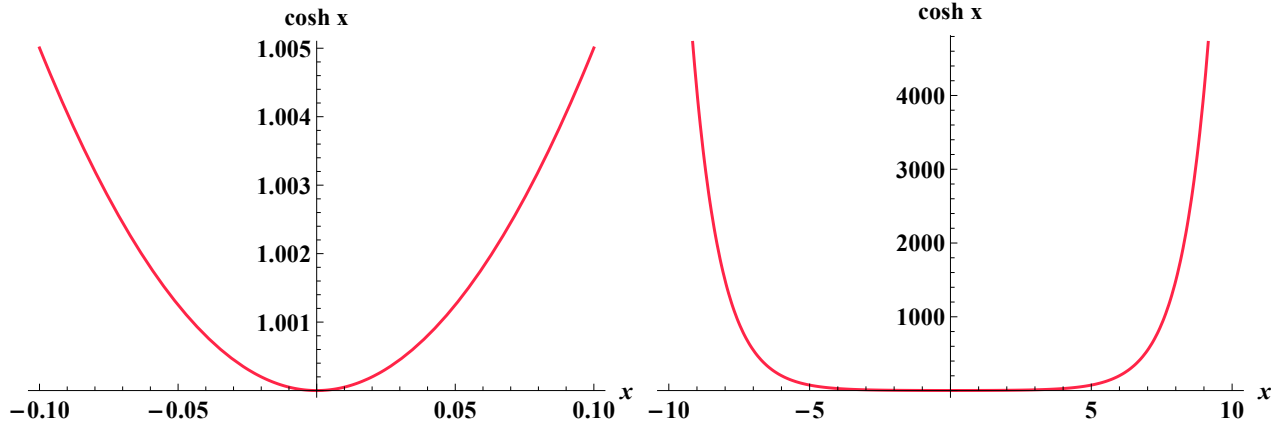


Figure 1.2: The hyperbolic function $\cosh x$: LEFT- small x , RIGHT- large x .

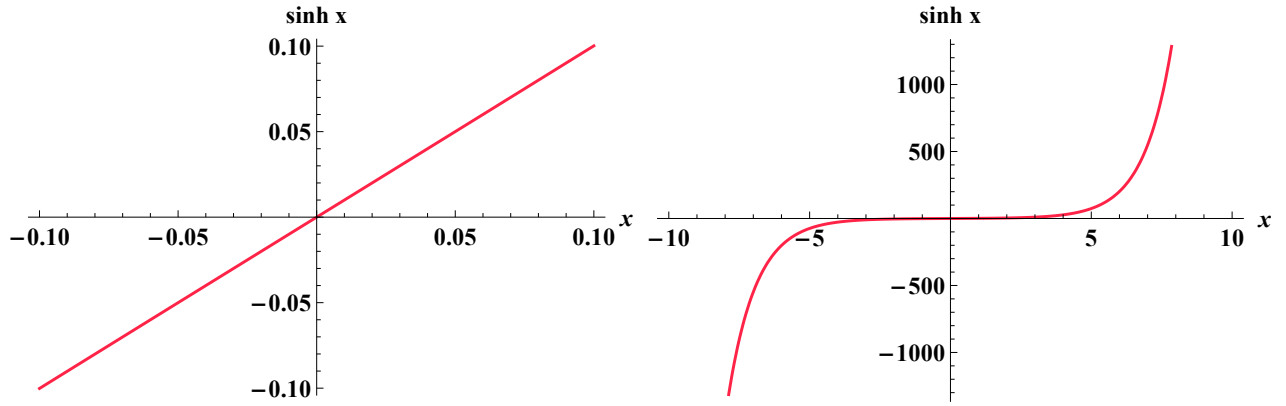


Figure 1.3: The hyperbolic function $\sinh x$: LEFT- small x , RIGHT- large x .

It follows from the behavior of the exponential function that

$$\begin{aligned} \cosh(x \rightarrow 0) &\rightarrow 1 + \frac{x^2}{2} \rightarrow 1, \\ \cosh(x \rightarrow \pm\infty) &\rightarrow +\infty, \end{aligned} \quad (1.3.7)$$

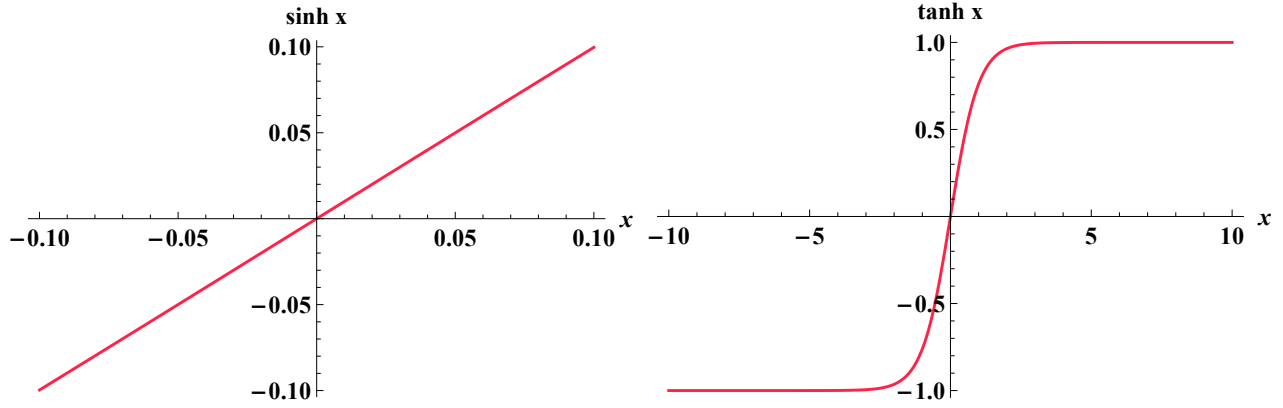
while

$$\begin{aligned} \sinh(x \rightarrow 0) &\rightarrow x \rightarrow 0, \\ \sinh(x \rightarrow \pm\infty) &\rightarrow \pm\infty. \end{aligned} \quad (1.3.8)$$

The behavior of the hyperbolic function $\cosh x$ is illustrated in Fig. 1.2 for both small (left) and large (right) x . The corresponding plots for $\sinh x$ appear in Fig. 1.3.

A related (and useful) function is the hyperbolic tangent defined by the ratio,

$$\tanh(x) \equiv \frac{\sinh(x)}{\cosh(x)}. \quad (1.3.9)$$


 Figure 1.4: The hyperbolic function $\tanh x$: LEFT- small x , RIGHT- large x .

It follows from the above properties of the hyperbolic sine and cosine that for small x values ($x \ll 1$), $\tanh(x)$ behaves like $\sinh(x)$, *i.e.*, like x , while its magnitude is bounded by 1 for large x . We have

$$\begin{aligned} \tanh(x \ll 1) &\rightarrow x, \\ \tanh(x \rightarrow \pm\infty) &\rightarrow \pm 1, \end{aligned} \quad (1.3.10)$$

as illustrated in Fig. 1.4.

As we will see when we discuss Special Relativity in detail, the hyperbolic functions play an essential role in explicitly representing transformations between reference frames which are moving with a fixed velocity with respect to each other. Next we consider the exponential function along the *imaginary* axis.

1.4 Along the Imaginary Axis: The Sinusoidal Functions

Consider the exponential function with a purely imaginary argument, $z = iy$ (with y real). We can write the series form in terms of separate even, real and odd, imaginary parts ($i^2 = -1$, $i^3 = -i$, $i^4 = 1$, *etc.*) ,

$$e^{iy} \equiv \sum_{n=0}^{\infty} \frac{(iy)^n}{n!} = \sum_{n=0}^{\infty} \frac{(-1)^n y^{2n}}{(2n)!} + i \sum_{n=0}^{\infty} \frac{(-1)^n y^{2n+1}}{(2n+1)!}. \quad (1.4.1)$$

This immediately suggests the usual series definitions of the sinusoidal functions:

$$\begin{aligned} \cos(y) &\equiv \sum_{n=0}^{\infty} \frac{(-1)^n y^{2n}}{(2n)!}, \\ \sin(y) &\equiv \sum_{n=0}^{\infty} \frac{(-1)^n y^{2n+1}}{(2n+1)!}. \end{aligned} \quad (1.4.2)$$

Note that these expressions are very similar to the hyperbolic forms in Eq. (1.3.2) *except* for the alternating signs in the sums. Thus Eq. (1.4.1) can be written as (called Euler's formula)

$$e^{iy} = \cos(y) + i \sin(y), \quad \cos(y) = \operatorname{Re}(e^{iy}), \quad \sin(y) = \operatorname{Im}(e^{iy}), \quad (1.4.3)$$

which leads to the analogs of Eq. (1.3.1) illustrating similar symmetry properties,

$$\begin{aligned}\cos(y) &= \frac{e^{iy} + e^{-iy}}{2}, & \cos(y) &= \cos(-y), \\ \sin(y) &= \frac{e^{iy} - e^{-iy}}{2i}, & \sin(y) &= -\sin(-y).\end{aligned}\tag{1.4.4}$$

We can also immediately obtain the analogs of Eq. (1.3.3)

$$\begin{aligned}\frac{d \cos(y)}{dy} &= \frac{ie^{iy} - ie^{-iy}}{2} = -\sin(y), \\ \frac{d \sin(y)}{dy} &= \frac{ie^{iy} + ie^{-iy}}{2i} = \cos(y).\end{aligned}\tag{1.4.5}$$

So it follows that the sinusoidal functions are eigenfunctions of the second derivative operator (with eigenvalue -1),

$$\begin{aligned}\frac{d^2 \cos(y)}{d^2 y} &= -\cos(y), \\ \frac{d^2 \sin(y)}{d^2 y} &= -\sin(y).\end{aligned}\tag{1.4.6}$$

The next step is to determine the analogs of Eqs. (1.3.5) and (1.3.6),

$$\begin{aligned}\cos^2(y) &= \frac{e^{2iy} + e^{-2iy} + 2}{4}, \\ \sin^2(y) &= \frac{e^{2iy} + e^{-2iy} - 2}{-4},\end{aligned}\tag{1.4.7}$$

so that

$$\cos^2(y) + \sin^2(y) = 1.\tag{1.4.8}$$

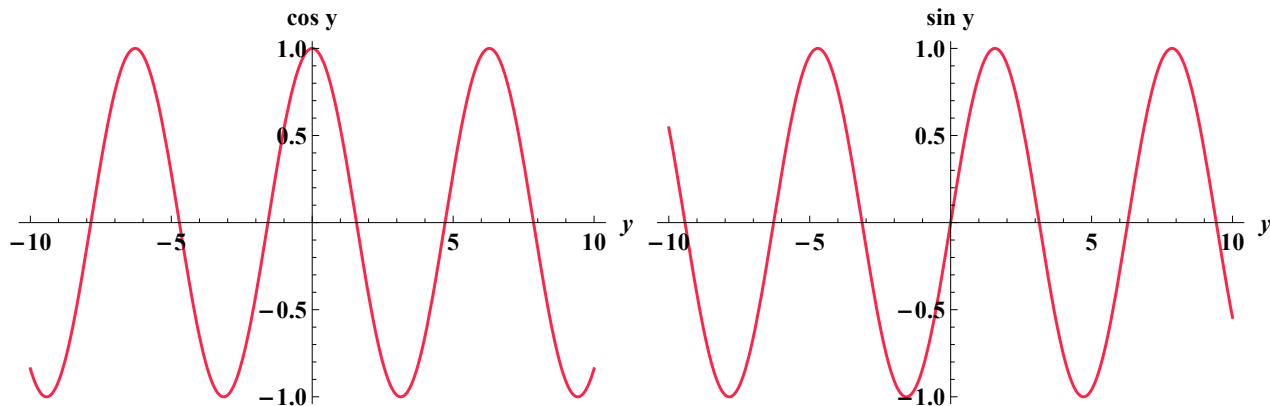
Note, in particular, the similarities and *differences* from the hyperbolic case.

Now consider the values of these functions. From the series expansions, it follows that at the origin, like the hyperbolic functions, we have

$$\begin{aligned}\cos(0) &= 1, \\ \sin(0) &= 0.\end{aligned}\tag{1.4.9}$$

However, away from the origin (along the imaginary axis for e^z), things are more interesting. The alternating signs in the series expansions suggests that any growth will be smaller than that of the hyperbolic functions. In fact, since the expression e^{iy} is a pure phase with modulus 1, $|e^{iy}| = 1$, it is not surprising that the sinusoidal functions have magnitudes bounded above by 1, as is already suggested by Eq. (1.4.8). The most straightforward approach is to simply evaluate the series in Eq. (1.4.2). It is perhaps surprising, based on simply looking at the series expression, that the numerical results lead to the conclusion that both of these functions are periodic with period 2π ,

$$\begin{aligned}\cos(y + 2\pi) &= \cos(y), \\ \sin(y + 2\pi) &= \sin(y), \\ \sin(y) &= \cos\left(\frac{\pi}{2} - y\right).\end{aligned}\tag{1.4.10}$$


 Figure 1.5: The sinusoidal functions: LEFT- $\cos y$, RIGHT- $\sin y$.

In fact, the transcendental number π can be determined numerically by solving for the smallest positive real number for which the series expression for the sine function vanishes. In this path to π no reference is made to circles or trigonometry, which is how this quantity is usually first introduced. We are also led directly to the special cases,

$$\begin{aligned} \cos(\pi) &= -1 = e^{i\pi}, \quad \cos(2\pi) = e^{i2\pi} = 1, \quad \cos\left(\frac{\pi}{2}\right) = \cos\left(\frac{3\pi}{2}\right) = 0, \\ \sin(\pi) &= \sin(2\pi) = 0, \quad \sin\left(\frac{\pi}{2}\right) = 1, \quad \sin\left(\frac{3\pi}{2}\right) = -1. \end{aligned} \quad (1.4.11)$$

This behavior is illustrated in Fig. 1.5. The analogue to Eq. (1.3.9) is the sinusoidal tangent function defined by

$$\tan y \equiv \frac{\sin y}{\cos y}, \quad (1.4.12)$$

and illustrated in Fig. 1.6. Note that, since the numerator of this expression is maximum when the denominator vanishes, this function has regularly spaced singularities along the real axis (separated by zeros).

Another way to approach the discussion of periodicity is to recall that, as we already noted, the series expansion for the exponential function in Eq. (1.2.3) defines an *entire* function with no singularities in the finite complex z plane. We also need to recall (from Phys 227) that a complex number can be written in terms of its real and imaginary parts, $z = x + iy$, or in terms of its magnitude and phase, $z = |z|e^{i\phi}$ with $\phi = \tan^{-1}(y/x)$. Thus, if we evaluate the exponential function just *above* the real axis, $z_+ = 1e^{i\delta}$ $\delta \ll 1$, we must get the *same* value if we go around the unit circle and approach the real axis from below, $z_- = 1e^{i(2\pi-\delta)}$. It is consistent for the exponential function to be branch cut *free* if and only if the sinusoidal functions are periodic, $e^{iy} = e^{iy+i2\pi}$.

We close this discussion by exploring the situation in the complex plane more generally. With the definitions above we have

$$\begin{aligned} \cos(z) &= \cos(x + iy) = \cos(x) \cosh(y) - i \sin(x) \sinh(y), \\ \sin(z) &= \sin(x + iy) = \sin(x) \cosh(y) + i \cos(x) \sinh(y). \end{aligned} \quad (1.4.13)$$

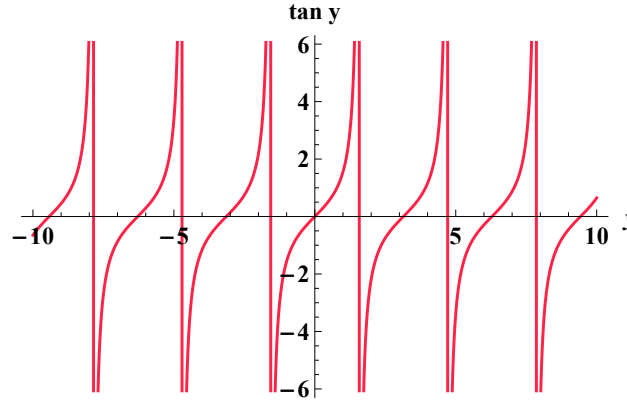


Figure 1.6: The sinusoidal tangent function.

This should look familiar to the addition of angles formulas you learned in high school,

$$\begin{aligned}\cos(\alpha + \beta) &= \cos(\alpha) \cos(\beta) - \sin(\alpha) \sin(\beta), \\ \sin(\alpha + \beta) &= \sin(\alpha) \cos(\beta) + \cos(\alpha) \sin(\beta),\end{aligned}\tag{1.4.14}$$

plus

$$\cos(iy) = \cosh(y), \quad \sin(iy) = i \sinh(y),\tag{1.4.15}$$

which follow directly from the series expressions above.

1.5 Rotations in 3-D

As a final mathematical physics subject to review we turn to rotations in 3 dimensions (3-D), which will allow us to use the sinusoidal functions from above and prepare for the idea of Lorentz transformations (in 4-D). For a discussion of rotations as a specific set of transformations described by the mathematics of Group Theory see Chapter 10. For the current discussion consider vectors in 3-D, for example, the position vector \vec{r} (or the velocity $\vec{v} = \dot{\vec{r}}$, where the dot signifies the derivative with respect to time), measured with respect to a chosen origin. To use specific vector and matrix notation we introduce 3 orthonormal unit basis vectors, \hat{e}_1 , \hat{e}_2 and \hat{e}_3 to obtain a completely defined reference frame. (Note that you may be more familiar with the \hat{x} , \hat{y} and \hat{z} notation, but the 1-2-3 notation is more common in the 4-D world to which we are headed.) These are represented by

$$\hat{e}_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \quad \hat{e}_2 = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, \quad \hat{e}_3 = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}.\tag{1.5.1}$$

A general 3-D position vector is then represented by

$$\vec{r} = x^1 \hat{e}_1 + x^2 \hat{e}_2 + x^3 \hat{e}_3 = \begin{pmatrix} x^1 \\ x^2 \\ x^3 \end{pmatrix}.\tag{1.5.2}$$

With the usual definition of spherical coordinates with polar angle θ (measured from the \hat{e}_3 direction) and azimuthal angle ϕ (measured from the \hat{e}_1 direction in the \hat{e}_1 - \hat{e}_2 plane), along with the usual

trigonometric definitions of the sinusoidal functions, we have the familiar vector components

$$x^1 = |\vec{r}| \sin \theta \cos \phi, \quad x^2 = |\vec{r}| \sin \theta \sin \phi, \quad x^3 = |\vec{r}| \cos \theta. \quad (1.5.3)$$

Now we want to consider performing a rotation. There are actually two types of rotations possible. We could choose to rotate the location vector \vec{r} with the basis vectors *fixed*, called an “active” rotation, or we could instead rotate the basis vectors with the location vector fixed, *i.e.*, rotate the reference frame, called a “passive” rotation. The mathematics is the same if the two angles of rotation differ by a sign. In this class we will be concentrating on transformations between different reference frames and thus on passive rotations.

We can think of generating a rotation through a specified angle α as being represented by the exponentiation of the appropriate “generator” of an *infinitesimal* rotation (see Chapter 10 for more details). As you may have learned in your Quantum Mechanics (or Classical Mechanics, or 227/8) class and we will discuss more later, the generators of an infinitesimal rotation about any of the 3 axes (*i.e.*, a rotations in the plane orthogonal to that axis, are given by the 3 components of the angular momentum operator modulo a factor of \hbar that carries the appropriate units). These 3 operators obey the commutation relation (here we include the explicit factor of \hbar),

$$[J_k, J_l] \equiv J_k J_l - J_l J_k = i\hbar \epsilon_{klm} J_m \quad [k, l, m = 1, 2, 3], \quad (1.5.4)$$

which is the *algebra* of the Rotation Group, $SO(3)$.¹

A useful 3-D representation of these operators is given by

$$J_1 = \hbar \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & -i \\ 0 & i & 0 \end{pmatrix}, \quad J_2 = \hbar \begin{pmatrix} 0 & 0 & i \\ 0 & 0 & 0 \\ -i & 0 & 0 \end{pmatrix}, \quad J_3 = \hbar \begin{pmatrix} 0 & -i & 0 \\ i & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}. \quad (1.5.5)$$

The reader is encouraged to verify that the matrices in Eq. (1.5.5) satisfy Eq. (1.5.4). Note further that these matrices are traceless and Hermitian ($J_k^\dagger = J_k$) (see also the discussion in Chapter 10). While the precise form of these matrices may not be intuitively obvious, the general form should be clear from our understanding of how ordinary rotations work. For example, a rotation about the 1-axis is a rotation in the 2-3 plane. It should serve to mix the 2 and 3 components of an ordinary 3-vector. Modulo the issue of phases, this is precisely what the form of the J_1 matrix in Eq. (1.5.5) does. It transforms a 3-component into a 2-component and a 2-component into a 3-component (*i.e.*, the only non-zero elements of the matrix perform this transformation).²

As an explicit example we consider a rotation by an angle α about the 3-axis. This is obtained by exponentiating the corresponding generator times the angle of rotation. How do we evaluate the exponentiation of a matrix? We simply recall that the exponential is *defined* by the power series of Eq. (1.2.3) and proceed. In particular, we want to evaluate the expression

$$R_3(\alpha) \equiv e^{i\alpha J_3/\hbar} = \sum_{n=0}^{\infty} \frac{(i\alpha J_3/\hbar)^n}{n!} = \mathbf{1} + \sum_{n=1}^{\infty} \frac{(i\alpha J_3/\hbar)^n}{n!}, \quad (1.5.6)$$

¹The symbol ϵ_{klm} (called the Levi-Civita symbol) represents the unique completely *antisymmetric* $3 \times 3 \times 3$ matrix, which is normalized so that $\epsilon_{123} = 1$. Cyclical permutations of the indices 123 also yield unity, *e.g.*, $\epsilon_{231} = 1$, while non-cyclical permutations give -1, *e.g.*, $\epsilon_{213} = -1$. Repeated indices yield zero, *e.g.*, $\epsilon_{11k} = 0$.

²Being able to think of rotations as *either* occurring in a 2-D plane or about the direction orthogonal to that plane is an accident of (apparently) living where there are precisely 3 spatial dimensions. The “rotation in a plane” interpretation is the one that generalizes to a larger number of spatial dimensions.

with $\mathbf{1}$ the unit matrix,

$$\mathbf{1} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}. \quad (1.5.7)$$

Note that, since the quantity in the exponent must be dimension-LESS, we have divided out the factor of \hbar . To evaluate the rest of the sum we note that each term is a 3×3 matrix determined by the following properties of J_3 ,

$$(J_3/\hbar)^{2n} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \quad (J_3/\hbar)^{2n+1} = J_3/\hbar. \quad (1.5.8)$$

So, by “pulling apart” the unit matrix and using Eq. (1.5.8), we can rewrite Eq. (1.5.6) as

$$\begin{aligned} R_3(\alpha) &= e^{i\alpha J_3/\hbar} = \mathbf{1} + \sum_{n=1}^{\infty} \frac{(i\alpha J_3/\hbar)^n}{n!} \\ &= \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix} + \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix} \sum_{n=0}^{\infty} \frac{(i\alpha)^{2n}}{(2n)!} + (J_3/\hbar) \sum_{n=0}^{\infty} \frac{(i\alpha)^{2n+1}}{(2n+1)!} \\ &= \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix} + \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix} \sum_{n=0}^{\infty} \frac{(-1)^n (\alpha)^{2n}}{(2n)!} + i(J_3/\hbar) \sum_{n=0}^{\infty} \frac{(-1)^n (\alpha)^{2n+1}}{(2n+1)!}. \end{aligned} \quad (1.5.9)$$

We recognize the series from Eq. (1.4.2) and can write

$$\begin{aligned} R_3(\alpha) &= e^{i\alpha J_3/\hbar} = \mathbf{1} + \sum_{n=1}^{\infty} \frac{(i\alpha J_3/\hbar)^n}{n!} \\ &= \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix} + \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix} \cos \alpha + i(J_3/\hbar) \sin \alpha \\ &= \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix} + \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix} \cos \alpha + \begin{pmatrix} 0 & 1 & 0 \\ -1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \sin \alpha \\ &= \begin{pmatrix} \cos \alpha & \sin \alpha & 0 \\ -\sin \alpha & \cos \alpha & 0 \\ 0 & 0 & 1 \end{pmatrix}. \end{aligned} \quad (1.5.10)$$

As expected for a rotation about the 3-axis, the 3-component is unchanged (the 1 in the lower right corner of the rotation matrix), while the 1 and 2-components are mixed via the sinusoidal functions. Note that the full rotation matrix is an Orthogonal matrix ($R_3(\alpha)^{-1} = R_3(\alpha)^T$, the transpose is the inverse) as we would expect for the representation of an element of the 3-D Orthogonal group, $SO(3)$, and has determinant 1 ($1 \times (\cos^2 \alpha - (-\sin^2 \alpha)) = 1$).

If we explicitly apply this rotation to a general location vector, we obtain (recall Eq. (1.5.2), recall also that we are performing a “passive” rotation where the location we are describing remains *fixed*

while the axes of the reference frame are rotated)

$$\vec{r}' = R_3(\alpha)\vec{r} = \begin{pmatrix} \cos \alpha x^1 + \sin \alpha x^2 \\ -\sin \alpha x^1 + \cos \alpha x^2 \\ x^3 \end{pmatrix}. \quad (1.5.11)$$

Returning to the spherical coordinate notation of Eq. (1.5.3), we have

$$\begin{aligned} x'^1 &= |\vec{r}| \sin \theta (\cos \phi \cos \alpha + \sin \phi \sin \alpha) = |\vec{r}| \sin \theta \cos(\phi - \alpha), \\ x'^2 &= |\vec{r}| \sin \theta (\sin \phi \cos \alpha - \cos \phi \sin \alpha) = |\vec{r}| \sin \theta \sin(\phi - \alpha), \\ x'^3 &= x^3 = |r| \cos \theta, \end{aligned} \quad (1.5.12)$$

where the last steps in the first two lines use the expressions in Eq. (1.4.14). This result should be intuitively reasonable. First, as already noted, a rotation about the 3-axis does *not* change the 3-component, *i.e.*, the polar angle θ is unchanged. Such a rotation does, however, mix the 1 and 2-components. Since the rotation of the axes by angle α is in the same sense as the definition of the azimuthal angle ϕ , the azimuthal angle of the (unrotated) location vector as measured in the rotated reference frame is *reduced* by α , $\phi' = \phi - \alpha$. This explicit example also serves to illustrate the again intuitive result that such a rotation does *not* change the length of the rotation vector, which is given by the “scalar” product of the vector with itself denoted by $|\vec{r}|^2 = \vec{r} \cdot \vec{r}$. This is called a scalar product precisely because it is *not* changed by rotations. (The label scalar is to be contrasted with objects labeled as vectors that are changed by rotations.) In detail we have

$$\begin{aligned} \vec{r} \cdot \vec{r} &= (x^1)^2 + (x^2)^2 + (x^3)^2 = |\vec{r}|^2 (\sin^2 \theta \cos^2 \phi + \sin^2 \theta \sin^2 \phi + \cos^2 \theta) \\ &= |\vec{r}|^2 (\sin^2 \theta + \cos^2 \theta) = |\vec{r}|^2, \end{aligned} \quad (1.5.13)$$

and

$$\vec{r}' \cdot \vec{r}' = (x'^1)^2 + (x'^2)^2 + (x'^3)^2 = |\vec{r}|^2 (\sin^2 \theta \cos^2(\phi - \alpha) + \sin^2 \theta \sin^2(\phi - \alpha) + \cos^2 \theta) = |\vec{r}|^2. \quad (1.5.14)$$

The scalar product of two *different* 3-vectors is also unchanged by rotations, since it depends only on the polar angle between the directions of the two vectors. Again in detail we have

$$\begin{aligned} \vec{r}_1 \cdot \vec{r}_2 &= (x_1^1)(x_2^1) + (x_1^2)(x_2^2) + (x_1^3)(x_2^3) \\ &= |\vec{r}_1| |\vec{r}_2| (\sin \theta_1 \sin \theta_2 (\cos \phi_1 \cos \phi_2 + \sin \phi_1 \sin \phi_2) + \cos \theta_1 \cos \theta_2) \\ &= |\vec{r}_1| |\vec{r}_2| (\sin \theta_1 \sin \theta_2 \cos(\phi_1 - \phi_2) + \cos \theta_1 \cos \theta_2) \\ &= |\vec{r}_1| |\vec{r}_2| \cos \Delta\theta_{12} = \vec{r}_1' \cdot \vec{r}_2', \end{aligned} \quad (1.5.15)$$

where the last step arises from recognizing the expression for the cosine of the angle between two directions, $\cos \Delta\theta_{12}$, expressed in spherical coordinates $((\theta_1, \phi_1), (\theta_2, \phi_2))$ (which you learned about in Phys 227/8). Since this expression involves only *differences* between the spherical coordinate angles, it will be unchanged when these angles are changed in identical ways by a rotation. The corresponding group of transformations (rotations) is labeled the Orthogonal Group since orthogonal vectors remain orthogonal after the transformation.

The reader is encouraged to evaluate more general 3-D rotations and prepare to consider 4-D transformations - the Lorentz transformations. These will be expressed in terms of 4×4 matrices similar, but not identical, to the rotation matrices above. 4-D Lorentz *scalar* products of a pair of 4-vectors (invariant under Lorentz transformations) will play a central role in our discussion.

1.6 Examples

To complete this discussion we present the results of rotation about the 1 and 2 axes, results the reader should check. It follows from the explicit expressions in Eq. (1.5.5) that we have the analogues of Eq. (1.5.8) as

$$(J_1/\hbar)^{2n} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, (J_1/\hbar)^{2n+1} = J_1/\hbar, \quad (1.6.1)$$

and

$$(J_2/\hbar)^{2n} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}, (J_2/\hbar)^{2n+1} = J_2/\hbar. \quad (1.6.2)$$

Thus we can nearly guess the corresponding rotations by an angle α about these two axes. For the rotation about the 1 axis we have

$$\begin{aligned} R_1(\alpha) &\equiv e^{i\alpha J_1/\hbar} = \mathbf{1} + \sum_{n=1}^{\infty} \frac{(i\alpha J_1/\hbar)^n}{n!} \\ &= \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} + \begin{pmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \cos \alpha + i(J_1/\hbar) \sin \alpha \\ &= \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} + \begin{pmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \cos \alpha + \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & -1 & 0 \end{pmatrix} \sin \alpha \\ &= \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos \alpha & \sin \alpha \\ 0 & -\sin \alpha & \cos \alpha \end{pmatrix}. \end{aligned} \quad (1.6.3)$$

To obtain an intuitive picture (or check) of the “signs in front of the sines” consider a vector in the 2-3 plane where both the 2- and the 3-components are positive ($x^2 > 0$ and $x^3 > 0$) and the vector lies in the 2-3 plane oriented between the positive 2- and 3-axes. Then a positive rotation about the 1 axis (your right hand is essential here) rotates the 2-axis *towards* the fixed location vector and the 3-axis *away* from the fixed location vector. Hence we expect that the 2-component in the new, rotated frame to be *larger* (than in the old frame), while the 3-component will be smaller. This result is precisely what the “signs” in Eq. (1.6.3) tell us, *i.e.*, addition occurs for the 2-component while subtraction occurs for the 3-component.

Finally the result for a rotation about the 2 axis is clear except, perhaps, the signs. We have

$$\begin{aligned}
 R_2(\alpha) &\equiv e^{i\alpha J_2/\hbar} = \mathbf{1} + \sum_{n=1}^{\infty} \frac{(i\alpha J_2/\hbar)^n}{n!} \\
 &= \begin{pmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix} + \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix} \cos \alpha + i(J_2/\hbar) \sin \alpha \\
 &= \begin{pmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix} + \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix} \cos \alpha + \begin{pmatrix} 0 & 0 & -1 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix} \sin \alpha \\
 &= \begin{pmatrix} \cos \alpha & 0 & -\sin \alpha \\ 0 & 1 & 0 \\ \sin \alpha & 0 & \cos \alpha \end{pmatrix}.
 \end{aligned} \tag{1.6.4}$$

Again these are the expected Orthogonal matrices with unit determinant. To check the signs we can perform a similar exercise to that above, but now in the 1-3 plane. A positive rotation about the 2-axis moves the 3-axis towards the fixed location vector and the 1-axis away. Hence we expect the “+” sign in the 3-row of the matrix and the “-” sign in the 1-row as seen in Eq. (1.6.4).

Chapter 2

Special relativity

2.1 Galilean relativity

We start by considering an important example of an invariance, *i.e.* an invariance of the equations of motion under a change of the coordinate system. In particular, Newton's laws of motion,

$$\frac{d\vec{p}}{dt} = \vec{F}, \quad \frac{d\vec{x}}{dt} = \frac{\vec{p}}{m}, \quad (2.1.1)$$

retain the same form if one substitutes, *i.e.*, changes coordinates via,

$$\vec{x} \rightarrow \vec{x}' + \vec{u}t, \quad \vec{p} \rightarrow \vec{p}' + m\vec{u}, \quad (2.1.2)$$

for any velocity \vec{u} which is constant, *i.e.*, independent of time, $d\vec{u}/dt = 0$. In other words, equations (2.1.1) and (2.1.2) imply that

$$\frac{d\vec{p}'}{dt} = \vec{F}, \quad \frac{d\vec{x}'}{dt} = \frac{\vec{p}'}{m}. \quad (2.1.3)$$

This argument indicates that changing coordinates to those of a (relatively) moving (inertial) reference frame does not affect the form of Newton's equations. In other words, there is no *preferred* inertial frame in which Newton's equations are valid; if they hold in one frame, then they hold in *all* inertial frames. This is referred to as *Galilean* relativity. Note that an intrinsic feature (assumption) of Galilean relativity is that clocks, *i.e.*, time, are the same in all inertial frames. Once they are synchronized between two different inertial frames, they will remain synchronized.

Consider a particle, or wave moving with velocity \vec{v} when viewed in the unprimed frame. In that frame the position of the particle (or crest of the wave) is given by $\vec{x}(t) = \vec{x}_0 + \vec{v}t$. In the primed frame, using (2.1.2), the location of the same particle or wave-crest is given by $\vec{x}'(t) = \vec{x}_0 + (\vec{v} - \vec{u})t$. Hence, when viewed in the primed frame, the velocity of the particle or wave is given by

$$\vec{v}' = \vec{v} - \vec{u}. \quad (2.1.4)$$

This *shift* in velocities upon transformation to a moving frame is completely in accord with everyday experience. For example, as illustrated in Figure 2.1 if a person standing on the ground sees a car moving at 100 kph (kilometers per hour) parallel to a train moving at 80 kph, then a person

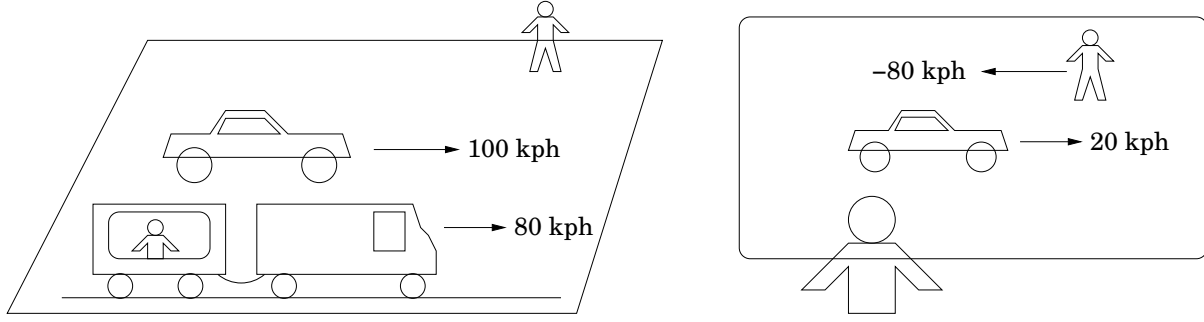


Figure 2.1: A moving train and car, as seen from the ground (left), and from the train (right).

sitting in the train will see that car moving with a *relative* velocity of $20 \text{ kph} = (100 - 80) \text{ kph}$, while the person on the ground recedes from view at a velocity of -80 kph . Similarly, a sound wave propagating at the speed of sound v_s (in a medium), as seen by an observer at rest with respect to the medium, will be seen (or heard) as propagating with speed $v' = v_s - u$ by an observer moving in the same direction as the sound wave with speed u (with respect to the medium). Consequently, the frequency $f' = v'/\lambda$ heard by the moving observer (*i.e.*, the number of wave fronts passing the observer per unit time) will differ from the frequency $f = v_s/\lambda$ heard by the stationary observer,

$$f' = \frac{v_s - u}{\lambda} = f \left(1 - \frac{u}{v_s} \right). \quad (2.1.5)$$

This is the familiar Doppler shift for the case of a moving observer and stationary source, where both velocities are measured with respect to the medium. Recall from introductory physics that the *medium* plays an important role here. If, with respect to medium, it is instead the observer who is stationary and the source that is moving (with the two still separating), the resulting Doppler shift is now

$$f' = \frac{f}{\left(1 + \frac{u}{v_s} \right)}. \quad (2.1.6)$$

Of course, to first order in u/v_s the results are identical, *i.e.*, (2.1.6) approaches (2.1.5), for $u/v_s \ll 1$,

$$f' = \frac{f}{\left(1 + \frac{u}{v_s} \right)} \Big|_{u/v_s \ll 1} \simeq f \left(1 - \frac{u}{v_s} \right). \quad (2.1.7)$$

However, as u approaches v_s ($u/v_s \rightarrow 1$) the limits of (2.1.6) and (2.1.5) are quite different,

$$f' = f \left(1 - \frac{u}{v_s} \right) \Big|_{u/v_s \rightarrow 1} \rightarrow 0, \quad (2.1.8a)$$

$$f' = \frac{f}{\left(1 + \frac{u}{v_s} \right)} \Big|_{u/v_s \rightarrow 1} \rightarrow \frac{f}{2}. \quad (2.1.8b)$$

Thus there is a *special* reference frame for sound, the rest frame of the medium in which the sound propagates.

2.2 Constancy of c

When applied to light (*i.e.*, electromagnetic radiation) the Galilean relativity velocity transformation (2.1.4) predicts that observers moving at different speeds will measure different propagation velocities for light coming from a given source (perhaps a distant star). This conclusion is *wrong*. Many experiments, including the famous Michelson-Morley experiment, have looked for, and failed to find, any variation in the speed of light as a function of the velocity of the observer. It has been unequivocally demonstrated that (2.1.4) does *not* apply to light. Note also that, unlike sound, light requires *no* medium to propagate.

Newton's laws, and the associated Galilean relativity relations (2.1.2) and (2.1.4), provide extremely accurate descriptions for the dynamics of particles and waves which move slowly compared to the speed of light c . But Newtonian dynamics does not correctly describe the behavior of light or (as we will see) any other particle or wave moving at speeds which are not very small compared to c . Our goal is to find a formulation of the dynamics which does not have this limitation.

We will provisionally adopt two postulates:

Postulate 1 *The speed of light (in a vacuum) is the same in all inertial reference frames.*

Postulate 2 *There is no preferred reference frame: the laws of physics take the same form in all inertial reference frames.*

We will see that these postulates lead to a fundamentally different view of space and time, as well as to many predictions which have been experimentally tested — successfully. In particular, we *must* view the world as intrinsically 4-dimensional. Not only do velocities change as we move between different inertial frames, but time does also. Yet, as required to match experiment, this new description of space-time must reduce to the familiar Galilean result in the limit $v \ll c$.

2.3 Clocks and rulers

A clock is some construct which produces regular “ticks” that may be counted to quantify the passing of time. An ideal clock is one whose period is perfectly regular and reproducible. Real clocks must be based on some physical phenomenon which is nearly periodic — as close to periodic as possible. Examples include pendula, vibrating crystals, and sun-dials. All of these have limitations. The period of a pendulum depends on its length and the acceleration of Earth's gravity. Changes in temperature will change the length of a pendulum. Moreover, the Earth is not totally rigid: tides, seismic noise, and even changes in weather produce (small) changes in the gravitational acceleration at a given point on the Earth's surface. The frequency (or period) of vibration of a crystal is affected by changes in temperature and changes in mass due to adsorption of impurities on its surface. In addition to practical problems (weather), the length of days as measured by a sun-dial changes with the season and, on much longer time scales, changes due to slowing of the Earth's rotation caused by tidal friction. On the other hand, if we observe the behavior of quantum mechanical systems such as individual atoms, we see much more robust periodic behavior.¹

¹Atomic clocks can now provide a very high time standard indeed as described in this Wikipedia article.

An idealized clock, which is particularly simple to analyze, is shown in Fig. 2.2. A short pulse of light repeatedly bounces back and forth (in a vacuum) between two parallel mirrors. Each time the light pulse reflects off one of the mirrors constitutes a “tick” of this clock.² If L is the distance between the mirrors, then the period (round-trip light travel time) of this clock is $\Delta t = 2L/c$.

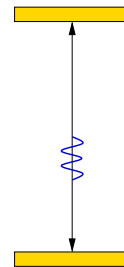


Figure 2.2: An idealized clock in which a pulse of light repeatedly bounces between two mirrors.

Now consider this same clock as seen by an observer moving to the left (perpendicular to the direction of the bouncing light) at velocity $-u$. In the observer’s frame, the clock moves to the right at velocity u , as shown in Fig. 2.3. Let $\Delta t'$ be the period of the clock as viewed in this frame, so that the pulse of light travels from the lower mirror to the upper mirror and back to the lower mirror in time $\Delta t'$. The upper reflection takes place halfway through this interval, when the upper mirror has moved a distance $u \Delta t' / 2$ to the right, and the light returns to the lower mirror after it has moved a distance $u \Delta t'$. Hence the light must follow the oblique path shown in the figure. The distance the light travels in one period is twice the hypotenuse, $D = 2\sqrt{L^2 + (u \Delta t' / 2)^2} = \sqrt{4L^2 + (u \Delta t')^2}$. Now use the first postulate: the speed of light in this frame is c , exactly the same as in the original frame. This means that the distance D and the period $\Delta t'$ must be related via $D = c \Delta t'$. Combining these two expressions gives $c \Delta t' = \sqrt{4L^2 + (u \Delta t')^2}$ and solving for $\Delta t'$ yields $\Delta t' = 2L / \sqrt{c^2 - u^2}$. Inserting $2L = c \Delta t$ and simplifying produces

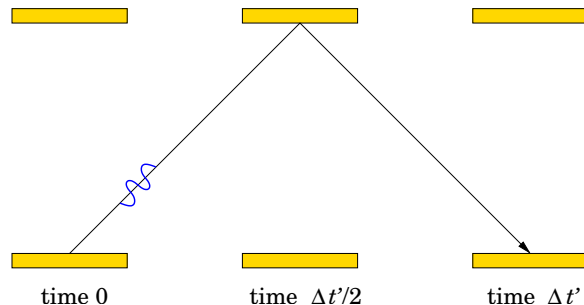


Figure 2.3: Three snapshots of the same clock viewed from a moving frame.

$$\Delta t' = \frac{\Delta t}{\sqrt{1 - (u/c)^2}}. \quad (2.3.1)$$

This is a remarkable result. It shows that the period of a clock, when viewed in a frame in which the clock is moving, is different, and *longer*, than the period of the clock as viewed in its rest frame. This phenomena is known as *time dilation*. It is an inescapable consequence of the constancy of the speed of light. Although we have analyzed a particularly simple model of a clock to deduce the existence of time dilation, the result is equally valid for *any* good clock.³ In other words, moving clocks run *slower* than when at rest, by a factor of

$$\gamma \equiv \frac{1}{\sqrt{1 - (u/c)^2}}, \quad (2.3.2)$$

where u is the speed with which the clock is moving. Note that $\gamma > 1$ for any non-zero speed u which is less (in magnitude) than c .

²To actually build such a clock, one would make one of the mirrors partially reflecting so that a tiny part of each light pulse is transmitted and measured by a photo-detector. These practical aspects are inessential for our purposes.

³After all, if some other good clock remains synchronized with our idealized clock when viewed in their common rest frame, then the same synchronization between the two clocks must also be present when the two clocks are viewed in a moving frame.

In the above discussion, we examined the case where the axis of our idealized clock was perpendicular to the direction of motion. What if the axis of the clock is parallel to the direction of motion? This situation is shown in Fig. 2.4. Analyzing this case is also instructive.

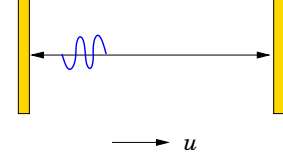


Figure 2.4: Our idealized clock, now rotated so that its axis is parallel to the direction of motion.

The round-trip light travel time (or period) must again be $\Delta t' = \gamma \Delta t$, because time dilation applies to *any* clock.⁴ Let L' be the distance between the mirrors, as viewed in the primed frame. The mirrors are moving to the right at velocity u , as shown in the figure. Suppose the light reflects off the right-hand mirror at time $\delta t'$ after leaving the left-hand mirror. During this time the right-hand mirror will have moved a distance $u \delta t'$ and therefore the distance light travels on this leg is $L' + u \delta t'$, longer than L' due to the motion of the mirror. Since $\Delta t'$ is the round-trip time, the light travel time for the return leg must be $\Delta t' - \delta t'$. On the way back, the light travel distance is $L' - u(\Delta t' - \delta t')$, since the motion of the left-hand mirror is *decreasing* the distance the light must travel.

Now use Postulate 1. For the first leg, the light travel distance $L' + u \delta t'$ must equal $c \delta t'$, since the speed of light in any (inertial) frame is c . Hence $\delta t' = L'/(c - u)$. And for the second leg, equating the distance $L' - u(\Delta t' - \delta t')$ with $c(\Delta t' - \delta t')$ implies that $\Delta t' - \delta t' = L'/(c + u)$. Substituting in $\delta t'$ gives

$$\Delta t' = \frac{L'}{c + u} + \frac{L'}{c - u} = \frac{2cL'}{c^2 - u^2} = \gamma^2 (2L'/c). \quad (2.3.3)$$

But we already know that $\Delta t' = \gamma \Delta t = \gamma (2L/c)$. The only way these two results for $\Delta t'$ can be consistent is if the distance L' between the mirrors, as seen in the frame in which the clock is moving parallel to its axis, is *smaller* than L by a factor of γ ,

$$L' = \frac{L}{\gamma} = L \sqrt{1 - (u/c)^2}. \quad (2.3.4)$$

This phenomena is known as *Lorentz contraction*. We have deduced it by using an ideal clock to convert a measurement of distance (the separation between mirrors) into a measurement of time. But the same result must apply to the measurement of any length which is parallel to the direction of motion. In other words, a ruler whose length is L , as measured in its rest frame, will have a length of $L' = L/\gamma$ when viewed in a frame in which the ruler is moving with a velocity parallel to itself (*i.e.*, parallel to the long axis of the ruler).

2.4 Observational tests

As we have seen, both time dilation and Lorentz contraction are direct, logical consequences of the frame-*independence* of the speed of light. Therefore every experimental test of the frame independence of c is a test of the existence of both time dilation and Lorentz contraction. Nevertheless, it is interesting to ask how these effects can be directly observed.

⁴To expand on this, imagine constructing two identical copies of our idealized clock. In their common rest frame, orient the axis of one clock perpendicular to the axis of the other clock. Since these two ideal clocks remain synchronized when viewed in their rest frame, they must also be synchronized when viewed from a moving frame whose velocity is parallel to one clock and perpendicular to the other.

One place where time dilation has a “real world” impact is in the functioning of the global positioning system (GPS). Time-dilation, due to the orbital motion of GPS satellites, slows the atomic clocks carried in these satellites by about 7 microseconds per day. This is easily measurable, and is a huge effect compared to the tens of nanosecond (per day) timing accuracy which can be achieved using GPS signals.⁵

A different observable phenomena where time dilation plays a key role involves muons produced in cosmic ray showers. When a high energy cosmic ray (typically a proton or atomic nucleus) strikes an air molecule in the upper reaches of the atmosphere (typically above 50 km = 5×10^4 m), this can create a particle shower containing many elementary particles of various types (which we will be discussing later) including electrons, positrons, pions, and muons. Muons are unstable particles; their lifetime τ is 2.2 microseconds. Moving at almost the speed of light, a high energy muon will travel a distance of about $c\tau \approx (3 \times 10^8 \text{ m/s}) \times (2 \times 10^{-6} \text{ s}) = 600 \text{ m}$ in time τ . This is small compared to the height of the atmosphere, and yet muons produced in showers originating in the upper atmosphere are easily observed on the ground. How can this be, if muons decay after merely a couple of microseconds?

The resolution of this apparent paradox is time dilation. Two microseconds is the lifetime of a muon *in its rest frame*. One may view a muon, or a bunch of muons moving together, as a type of clock. If there are N_0 muons initially, then after some time t (as measured in the rest frame of the muons) on average all but $N_1 = N_0 e^{-t/\tau}$ muons will have decayed. Turning this around, if all but some fraction N_1/N_0 of muons decay after some interval of time, then the length of this interval equals $\tau \ln(N_0/N_1)$ — as measured in the muons’ rest frame. But as we have seen above, a moving clock (any moving clock!), runs slower by a factor of γ . Therefore, fast moving muons decay more slowly than do muons at rest. This means that muons produced in the upper atmosphere at a height H (typically tens of kilometers) will have a substantial probability of reaching the ground before decaying provided they are moving fast enough so that $\gamma c\tau > H$.

Muons produced in the upper atmosphere and reaching the earth before decaying also illustrate Lorentz contraction — if one considers what’s happening from the muon’s perspective. Imagine riding along with a muon produced in an atmospheric shower. Or, as one says more formally, consider the co-moving reference frame of the muon. In this frame, the muon is at rest but the Earth is racing toward the muon at nearly the speed of light. The muon will decay, on average, in two microseconds. But the thickness of the atmosphere, in this frame, is reduced by Lorentz contraction. Therefore, the surface of the Earth will reach the muon before it (typically) decays if $(H/\gamma)/c < \tau$. This is the same condition obtained above by considering physics in the frame of an observer on the ground. This example nicely illustrates the second relativity postulate: because the laws of physics are frame independent, one may use whatever frame is most convenient to analyze some particular phenomena. In this example, whether one regards time dilation or Lorentz contraction as being responsible for allowing muons produced in the upper atmosphere to reach the ground depends on the frame one chooses to use. However, both approaches agree with the fact that high energy muons *can* reach the ground from the upper atmosphere.

⁵However, this is only part of the story regarding relative clock rates in GPS satellites. The difference in gravitational potential between the satellites’ orbits and the Earth’s surface also produces a change in clock rates due to a general relativistic effect known as *gravitational redshift*. This effect goes in the opposite direction (speeding orbiting clocks relative to Earth-bound ones) and is larger in magnitude, 45 microseconds per day. So GPS clocks actually run faster than clocks on the ground by $45 - 7 = 38$ microseconds per day.

2.5 Superluminal motion?

The expressions for time dilation (2.3.1) and length contraction (2.3.4) make sense (*i.e.*, yield real, not imaginary, results) only for $u \leq c$. As we will discuss more explicitly in Chapter 3, a basic feature of Special Relativity is that nothing (no signal, no particle, no information) can travel *faster* than the speed of light c . Thus there was considerable excitement in autumn 2011 when the OPERA neutrino detector at the Gran Sasso Laboratory in Italy (see link below) reported that neutrinos (which are thought to have a very small but nonzero mass) had seemed to travel to the detector from CERN in Geneva, Switzerland at a speed that exceeded c . (see, *e.g.*, this Science Daily story.) Such a measurement requires the realization of the typical (introductory physics) picture of a reference frame densely populated by synchronized identical clocks. In particular, the clocks in Geneva and in Gran Sasso need to be synchronized with a precision of better than 50 nanoseconds. As the discussion above suggests, this is a daunting challenge indeed, but possible using the GPS system.

Hence the OPERA result fundamentally conflicted with Special Relativity. Either our postulates, or the experimental measurement, must be in error. All indications are that the original measurement was in error. The OPERA team reported in early 2012 that this original measurement likely suffered from a synchronization error caused by a loose connection in a cable relaying the GPS signals to the experiment's clocks. Subsequent results reported by the companion experiment ICARUS confirmed that the speed of neutrinos is indeed bounded above by the speed of light.

2.6 Further resources

Michelson-Morley experiment, Wikipedia

GPS and Relativity, R. Pogge

Relativity in the Global Positioning System, N. Ashby

GPS, Wikipedia

Introduction to Cosmic Rays, VVC SLAC

Do-it-yourself Cosmic Ray Muon Detector

Cosmic ray, Wikipedia

OPERA, Wikipedia

ICARUS, Wikipedia

2.7 Example Problems

Kogut 2-1

This is a typical “Star Trek” style multi-frame problem. We have 2 frames of reference: the Earth (frame S) and the spaceship (frame S') moving with velocity $v = 0.6c$ with respect to each other. Everything is synchronized at event 1 ($t_1 = t'_1 = 0$, $x_1 = x'_1 = 0$), as the (small) spaceship passes the earth. Event 2 marks the emission of a pulse of light from the earth towards the spaceship at $t_2 = 10$ minutes = 600 seconds. Event 3 marks the detection of the light pulse by our friends on the spaceship. We want to use the concept of *proper* time, which is the time measured by a clock for

events occurring *at* the location of the clock, *i.e.*, for events occurring at the *same* point in the (rest) frame of the clock.

(a) Q: Is the time interval between events 1 and 2 a proper time interval in the spaceship frame? In the Earth frame?

A: Events 1 and 2 occur at the same point in frame S (*i.e.*, on the Earth), but not at the same point in frame S' on the spaceship. Hence the time interval between events 1 and 2 is a *proper* time interval on the Earth, but not on the spaceship.

(b) Q: Is the time interval between events 2 and 3 a proper time interval in the spaceship frame? In the Earth frame?

A: Events 2 and 3 occur at different points in both frames. Hence the time interval between events 2 and 3 is *not* a *proper* time interval in *either* frame.

(c) Q: Is the time interval between events 1 and 3 a proper time interval in the spaceship frame? In the Earth frame?

A: Events 1 and 3 occur at the same point on the spaceship (frame S') (*e.g.*, at the center of the small ship), but not at the same point on the Earth. Hence the time interval between events 1 and 3 is a *proper* time interval on the spaceship but not on the Earth.

(d) Q: What is the time of event 2 as measured on the spaceship?

A: We want to determine the time of the light emission in frame S' , t'_2 . This time interval (from event 1) is not a proper time interval in the S' frame and we must account for time dilation (with respect to the proper time interval in frame S). We have

$$\gamma = \frac{1}{\sqrt{1 - (v/c)^2}} = \frac{1}{\sqrt{1 - (0.6)^2}} = 1.25, \quad (2.7.1a)$$

$$t'_2 = \gamma t_2 = 1.25(600 \text{ seconds}) = 750 \text{ seconds} = 12.5 \text{ minutes}. \quad (2.7.1b)$$

(e) Q: According to the spaceship, how far away is the Earth when the light pulse is emitted?

A: We want to determine the distance to the earth from the spaceship at the time of the emission in frame S' . This is just the distance traveled by the earth as viewed by spaceship at velocity v during the time interval determined in part (d). We have

$$l'_2 = vt'_2 = (0.6) \times 3.0 \times 10^8 \text{ m/s} \times 750 \text{ s} = 1.35 \times 10^{11} \text{ m}. \quad (2.7.2)$$

(f) Q: From your answers in parts (d) and (e), what does the spaceship clock read when the light pulse arrives?

A: So we want to determine the time of event 3 in frame S' on the spaceship. We have determined both the time of emission and the distance to earth at emission in this frame, and we know that light travels at speed c in *all* frames (plus the fact that the spaceship is not moving in its own rest frame, S'). Thus we need only calculate the time for light to travel from the earth to the spaceship and add it to the time of the emission (all in the S' frame)

$$t'_3 = t'_2 + l'_2/c = 750 \text{ s} + 1.35 \times 10^{11} \text{ m} / (3.0 \times 10^8 \text{ m/s}) = (750 + 450) \text{ s} = 1200 \text{ s} = 20 \text{ minutes}. \quad (2.7.3)$$

(g) Q: Find the time of event 3 according to the Earth's clock by analyzing everything from the Earth's perspective.

A: We return to frame S and find the time of event 3. On the Earth we have that, at the time of event 2, the distance to the rocket ship is

$$l_2 = vt_2 = (0.6) \times 3.0 \times 10^8 \text{ m/s} \times 600 \text{ s} = 1.08 \times 10^{11} \text{ m}. \quad (2.7.4)$$

So the time interval (in this frame) between events 2 and 3 is given by (note that the spaceship continues to move in the S frame)

$$c(t_3 - t_2) = l_2 + v(t_3 - t_2) \rightarrow t_3 - t_2 = \frac{l_2}{c - v} = \frac{1.08 \times 10^{11} \text{ m}}{1.2 \times 10^8 \text{ m/s}} = 900 \text{ s}. \quad (2.7.5)$$

So finally we obtain

$$t_3 = t_2 + (t_3 - t_2) = (600 + 900) \text{ s} = 1500 \text{ s} = 25 \text{ minutes}. \quad (2.7.6)$$

(h) Q: Are your answers to parts (f) and (g) consistent with your conclusions from parts (a), (b) and (c)?

A: We learned in (c) that the time interval between events 1 and 3 is a proper time interval in frame S' , but not S , where the interval is dilated. We can check this point via

$$t'_3 \gamma = (20 \text{ minutes})(1.25) = 25 \text{ minutes} = t_3, \quad (2.7.7)$$

which checks with our result in (g).

Kogut 2-2

Here we consider two rockets, A and B , to define two reference frames, and let the rockets have identical proper lengths (*i.e.*, lengths in their respective rest frames) of 100 m. The two rockets pass each other moving in opposite directions and we consider two events defined in frame A by the passing of the front of rocket B . Event 1 is when the front of B passes the front end of A and event 2 is when the front of B passes the back end of A . The time interval in frame A between the two events is 1.5×10^{-6} s.

(a) Q: What is the relative velocity of the rockets?

A: Since we know the length of rocket A (in its rest frame) and the time interval for the front of rocket B to travel the length of A , all measured in frame A , we can find the relative velocity from

$$v_{\text{rel}} = \frac{100 \text{ m}}{1.5 \times 10^{-6} \text{ s}} = 6.67 \times 10^7 \text{ m/s}. \quad (2.7.8)$$

(b) Q: According to the clocks on rocket B , how long does the front end of A take to pass the entire length of rocket B ?

A: The passing of rocket A viewed from B will be exactly equivalent to the passing of B viewed in A (by symmetry, the relative speed is the same of both frames). Hence the time interval for these 2 events in B is again 1.5×10^{-6} s,

$$t_B = \frac{100 \text{ m}}{v_{\text{rel}}} = \frac{100 \text{ m}}{6.67 \times 10^7 \text{ m/s}} = 1.5 \times 10^{-6} \text{ s} = t_A. \quad (2.7.9)$$

(c) Q: According to the clocks on rocket B , how much time passes between the time when the front end of A passes the front end of B and the time when the rear end of A passes the front end of B ? Does this time interval agree with your answer to (b)? Should it?

A: Now consider a third event in B (event 1 was the front of A passing the front of B and event 2 was the front of A passing the back of B) defined by when the back of A passes the front of B . The important point now is that in frame B the length of rocket A is (relativistically) contracted. We have

$$\gamma = \frac{1}{\sqrt{1 - (v_{\text{rel}}/c)^2}} = 1.0257, \quad (2.7.10a)$$

$$L_{A \text{ in } B} = \frac{100 \text{ m}}{\gamma} = 97.50 \text{ m}, \quad (2.7.10b)$$

$$t_C = \frac{L_{A \text{ in } B}}{v_{\text{rel}}} = \frac{t_B}{\gamma} = 1.46 \times 10^{-6} \text{ seconds}. \quad (2.7.10c)$$

This is a time interval defined by a *new* pair of events, not equivalent to the previous pair, and the new time interval does not and should not agree with the result in (b).

Chapter 3

Minkowski spacetime

3.1 Events

An *event* is some occurrence which takes place at some instant in time at some particular point in (3-D) space. Your birth was an event. JFK's assassination was an event. Each downbeat of a butterfly's wingtip is an event. Every collision between air molecules is an event. Snap your fingers right now — that was an event. The set of all possible events is called *spacetime*. A point particle, or any stable object of negligible size, will follow some trajectory through spacetime which is called the *worldline* of the object. The set of all spacetime trajectories of the points comprising an extended object will fill some region of spacetime which is called the *worldvolume* of the object.

3.2 Reference frames

To label points in space, it is convenient to introduce spatial coordinates so that every point is uniquely associated with some triplet of numbers (x^1, x^2, x^3) . Similarly, to label events in spacetime, it is convenient to introduce spacetime coordinates so that every event is uniquely associated with a set of *four* numbers. The resulting spacetime coordinate system is called a (4-D) *reference frame*. Particularly convenient are *inertial reference frames*, in which coordinates have the form (t, x^1, x^2, x^3) where the superscripts here are coordinate labels and *not* powers. The set of events in which x^1, x^2 , and x^3 have arbitrary *fixed* (real) values while t ranges from $-\infty$ to $+\infty$ represent the worldline of a particle, or hypothetical observer, which is subject to no external forces and is at rest in this particular reference frame with no acceleration. This is illustrated in Figure 3.1.

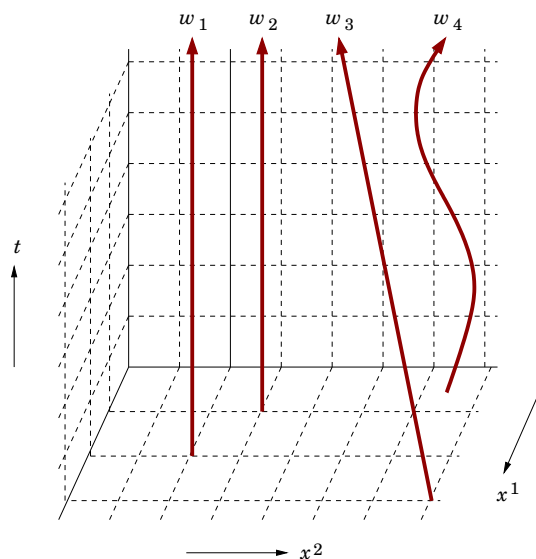


Figure 3.1: An inertial reference frame. Worldlines w_1 and w_2 represent observers at rest in this reference frame, w_3 is the spacetime trajectory of an inertial observer who is moving in this frame, and w_4 is the spacetime trajectory of a non-inertial object whose velocity and acceleration fluctuates.

As Figure 3.2 tries to suggest, one may view an inertial reference frame as being defined by an infinite set of inertial observers, one sitting at every point in space, all of whom carry *synchronized* (ideal) clocks and all of whom are at rest with respect to each other (but recall that this situation is a challenge to realize in practice - see the discussion of the OPERA experiment at the end of Chapter 2). You can imagine every observer carrying a notebook (or these days a tablet computer) and recording the time, according to his clock, of events of interest.

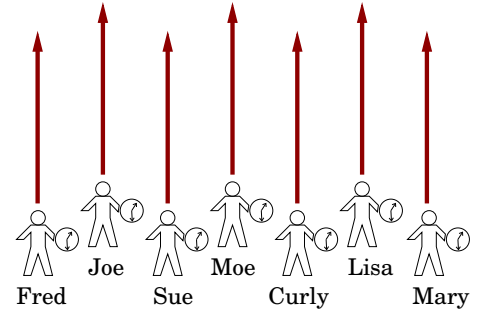


Figure 3.2: A family of inertial observers, all with synchronized watches and mutually at rest, defines an inertial reference frame.

For example, consider a statement like “a moving rod has length L ”. Suppose that the worldline of the left end of the rod intersects the worldline of some observer A at the event labeled A^* whose time, according to observer A’s clock, is t_1 . The worldline of the right end of the rod intersects the worldline of observer B at the event labeled B^* whose time (according to B) is also t_1 , and then intersects the worldline of observer C at event C^* at the later time t_2 (according to C). The interior of the rod sweeps out a flat two-dimensional surface in spacetime — the shaded “ribbon” bounded by the endpoint worldlines shown in Figure 3.3.

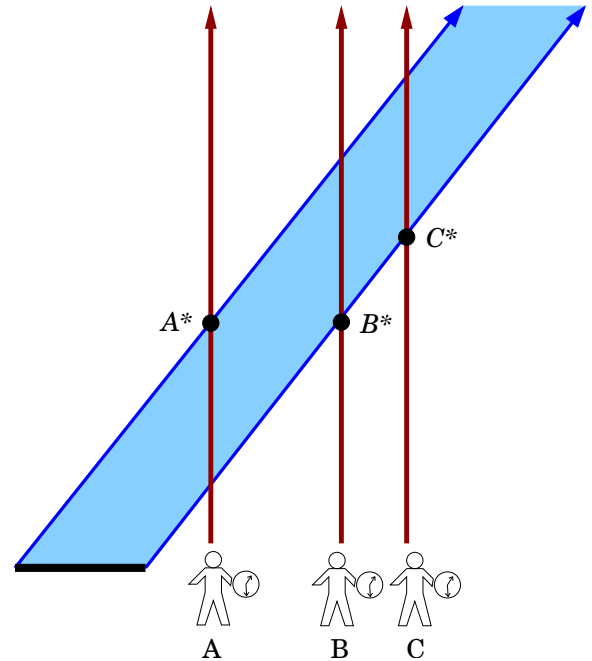


Figure 3.3: A moving rod which passes by three inertial observers who are at rest.

The *surface of simultaneity* of event A^* , in the reference frame in which observer A is at rest, is the set of all events whose time coordinates in this frame coincide with the time of event A^* . So event B^* is on the surface of simultaneity of event A^* (it is displaced precisely horizontally), while event C^* is *not* on the surface of simultaneity of event A^* .

The length of the rod, in this reference frame, is defined as the spatial distance between observers A and B, *i.e.*, the spatial distance between the ends of the rod *at the same time* in this frame (on a surface of simultaneity). As usual, it is convenient to choose Cartesian spatial coordinates, so that, if observers A and B have spatial coordinates (x_A^1, x_A^2, x_A^3) and (x_B^1, x_B^2, x_B^3) , then their relative spatial separation is given by

$$d_{AB} = [(x_B^1 - x_A^1)^2 + (x_B^2 - x_A^2)^2 + (x_B^3 - x_A^3)^2]^{1/2}. \quad (3.2.1)$$

One should stop and ask how the observers defining an inertial reference frame could, in principle, test whether their clocks are synchronized, and whether they are all mutually at rest. The simplest approach is to use the propagation of light. Suppose observer A flashes a light, momentarily, while observer B holds a mirror which will reflect light coming from observer A back to its source. If the light is emitted at time t_A , according to A’s clock, it will be reflected at time t_B , according to B’s

clock, and the reflected pulse will then be detected by A at some time $t_A + \Delta t$. If A and B's clocks are synchronized, then the time t_B at which B records the reflection must equal $t_A + \frac{1}{2}\Delta t$. Any deviation from this indicates that the clocks are not synchronized. If this experiment is repeated, then any change in the value of Δt indicates that the two observers are not mutually at rest.

3.3 Lightcones

Before proceeding further, it will be helpful to introduce a useful convention for spacetime coordinates. When one does dimensional analysis, it is customary to regard time and space as having *different* dimensions. If we define the spacetime coordinates of an event as the time and spatial coordinates in a chosen inertial frame, (t, x^1, x^2, x^3) , then the differing dimensions of the time and space coordinates will be a nuisance. Because the value of the speed of light, c , is universal — *independent* of reference frame — we can use it as a simple conversion factor which relates units of time to units of distance. Namely, we define the new coordinate (with dimensions of length)

$$x^0 \equiv ct, \quad (3.3.1)$$

which is the distance light can travel in time t . Henceforth we will use x^0 in place of the time t as the first entry in the spacetime coordinates of an event, (x^0, x^1, x^2, x^3) .

Now consider a flash of light which is emitted from the event with coordinates $x^0 = x^1 = x^2 = x^3 = 0$ — *i.e.*, from the spacetime origin in this coordinate system. The light will propagate outward in a spherical shell whose radius at time t equals ct , which is x^0 . Therefore, the set of events which form the entire history of this light flash are those events for which $[(x^1)^2 + (x^2)^2 + (x^3)^2]^{1/2} = x^0$. We can think of these events as forming a “cone” as illustrated in Figure 3.4. The intersection of this cone with the (2-D) x^0 – x^1 plane is comprised of the two half-lines at $\pm 45^\circ$, for which $x^0 = \pm x^1$ and $x^0 > 0$. These 45° lines describe the path of light which is emitted from the origin traveling in the $\pm x^1$ directions.

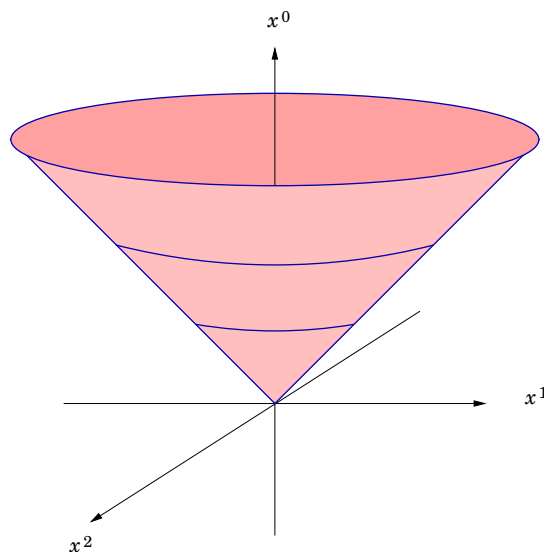


Figure 3.4: The “lightcone” of a flash of light emitted from the origin.

3.4 Simultaneity

Next consider the reference frames of two different inertial (non-accelerating) observers, A and B, who are *not* at rest with respect to each other. As viewed in A's reference frame, suppose that observer B is moving with speed v in the x^1 direction (with respect to A), so that B's position satisfies

$$x^1 = vt = (v/c)x^0 \quad (\text{in frame A}).$$

Figure 3.5 depicts this situation graphically. (We have chosen the origin of time to be when A and B are at the same point.) In reference frame A, the worldline of observer A is the vertical axis (labeled w_A), since this corresponds to all events with $x^1 = x^2 = x^3 = 0$ and x^0 arbitrary. The worldline of observer B (in reference frame A and labeled w_B) is a tilted line with a slope of c/v (slope here is defined as $\Delta x^0 / \Delta x^1$, *i.e.*, the tangent of the angle with respect to the x^1 axis), since this corresponds to all events with $x^0 = (c/v)x^1$ (and vanishing x^2 and x^3). As expected $v \rightarrow 0$ corresponds to a vertical line (infinite slope), while $v \rightarrow c$ is the line at 45° (corresponding to unit slope and the light cone in frame A).

Surfaces of simultaneity for observer A correspond to horizontal planes in this diagram, because such planes represent all events with a common value of time (or x^0) according to A's clock. But what are surfaces of simultaneity for observer B? In other words, what set of events share a common value of time according to B's clock? These turn out to be tilted planes with slope v/c (not c/v), as shown in the figure by the red lines labeled $x'^0 = 1$, $x'^0 = 0$ and $x'^0 = -1$.

A quick way to see that this must be the case is to note that the 45° path of a light ray traveling from the origin in the $+x^1$ direction (the dashed line in Fig. 3.5) bisects the angle between observer A's worldline (the x^0 axis in Fig. 3.5) and his surface of simultaneity defined by $x^0 = 0$. Exactly the same statement must also be true for observer B — she will also

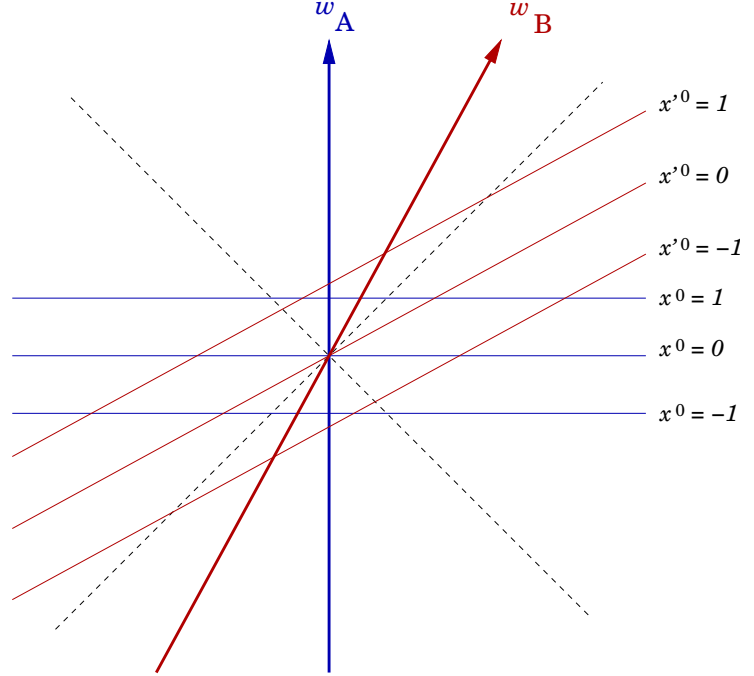


Figure 3.5: Worldline of two observers, and corresponding surfaces of simultaneity. The dashed lines show the lightcone of the origin.

describe the path of the light as bisecting the angle between her worldline and her surface of simultaneity which contains the origin (the red $x'^0 = 0$ line). This is an application of our second postulate (the physics looks the same in all inertial reference frames). Therefore, when plotted in A's reference frame, as in Figure 3.5, observer B's worldline and surfaces of simultaneity must have *complementary* slopes (c/v versus v/c), so that they form *equal* angles with the lightcone at 45° .

The essential point, which is our most important result so far, is that the concept of simultaneity is observer *dependent*. Events that one observer views as occurring simultaneously will *not* be simultaneous when viewed by a different observer moving at a non-zero relative velocity (as long as the events occur at spatial points separated by a nonzero distance).

Because this is a key point, it may be helpful to go through the logic leading to this conclusion in a more explicit fashion. To do so, consider the experiment depicted in Figure 3.6. Two flashes of light (the black lines) are emitted at events R and S and meet at event T . In observer B's frame, shown in the left panel of Figure 3.6, the emission events are *simultaneous* and separated by some distance L' . The reception event T is necessarily equi-distant between R and S . Lines w_B , $w_{B'}$, and $w_{B''}$ show the worldlines of observers who are at *rest* in this frame and who witness events R , T , and S , respectively. (In other words, w_B is the worldline of observer B, sitting at the origin in this

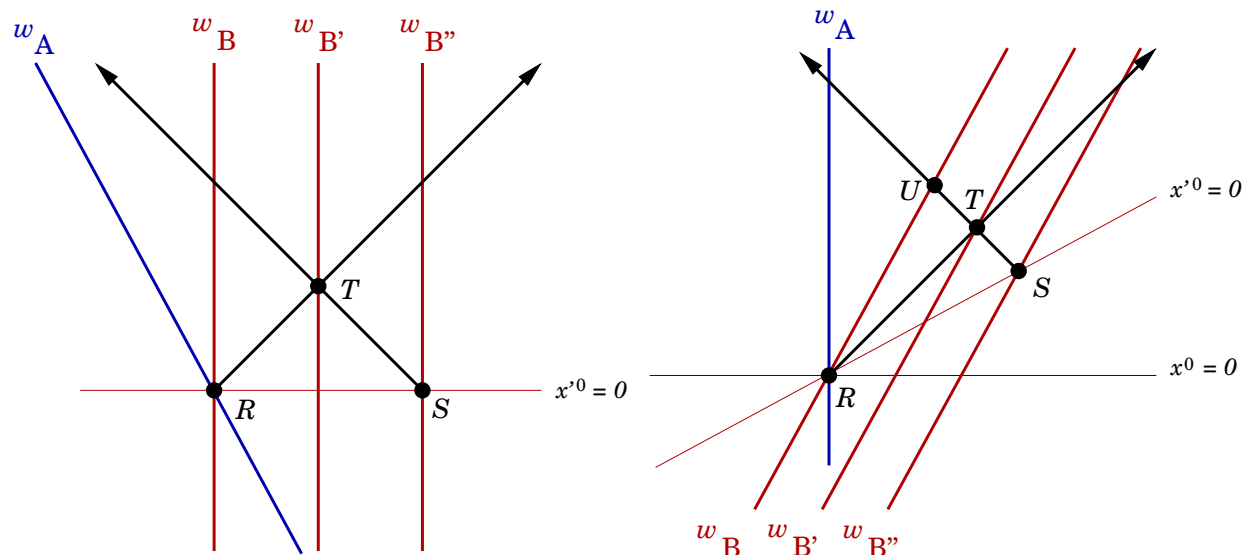


Figure 3.6: Two flashes of light emitted at events R and S which meet at event T , as described in two different frames.

frame, while $w_{B'}$ is the worldline of an observer sitting at rest a distance $L'/2$ away, and $w_{B''}$ is the worldline of an observer at rest a distance L' away, with all distances in the same direction.)

In observer A's frame, shown in the right panel of Figure 3.6, the worldlines of observers at rest in frame B are now *tilted* lines all with slope c/v . But the paths of the light rays (propagating within the plane shown) lie at $\pm 45^\circ$ in *both* frames, because the speed of light is universal. The emission event S , which lies on B's surface of simultaneity, is the intersection between the leftward propagating light ray and the worldline $w_{B''}$ of an observer who is at rest in B's frame and twice as far from the origin as the worldline, $w_{B'}$, which contains the reception event T . Since events R and S are simultaneous, as seen in frame B (and the distance L' in this construction is arbitrary), the frame B surface of simultaneity containing events R and S must, in frame A, appear as a straight line connecting these events. From the geometry of the figure, one can see that the triangles RTU and RTS are similar, and hence the angle between the simultaneity line RS and the 45° lightcone is the same as the angle between the worldline w_B and the lightcone. This implies that the slope of the simultaneity line is the inverse of the slope of worldline w_B , as asserted above. (As an exercise determine where the point U lies in the left panel and whether the triangles RTU and RTS are again similar - they are.)

3.5 Lorentz transformations

Just as many problems in ordinary spatial geometry are easier when one introduces coordinates and uses analytic geometry, spacetime geometry problems of the type just discussed are also simpler if one introduces and uses analytic formulas relating coordinates in different reference frames. These relations are referred to as *Lorentz transformations*. Recall that in Chapter 1 we considered the transformation of coordinates between two reference frames related by a rotation.

Using the two frames discussed above, let (x^0, x^1, x^2, x^3) denote spacetime coordinates in the inertial

reference frame of observer A, and let (x'^0, x'^1, x'^2, x'^3) denote spacetime coordinates in the inertial reference frame of observer B, who is moving in the x^1 direction with velocity v relative to observer A. How are these coordinates related?

Assume, for simplicity, that the spacetime origins of both frames coincide. Then there must be some linear transformation which relates coordinates in the two frames,

$$\begin{pmatrix} x^0 \\ x^1 \\ x^2 \\ x^3 \end{pmatrix} = \Lambda \begin{pmatrix} x'^0 \\ x'^1 \\ x'^2 \\ x'^3 \end{pmatrix}, \quad (3.5.1)$$

where Λ is some 4×4 (real) matrix. (This is the 4-D analog of the 3×3 rotation matrix in Eq. (1.5.10).) Since the transformation Λ describes the effect of switching to a moving frame, it is referred to as a *Lorentz boost*, or simply a ‘boost’.

If the spatial coordinates of frame B are not rotated with respect to the axes of frame A, so that observer B describes observer A as moving in the $-x'^1$ direction with velocity $-v$, then the Lorentz boost will only affect lengths in the 1-direction, leaving the 2 and 3 directions unaffected. Therefore, we should have

$$x^2 = x'^2, \quad x^3 = x'^3 \quad (\text{for a boost along } x^1), \quad (3.5.2)$$

implying that the boost matrix Λ has the block diagonal form

$$\Lambda = \begin{pmatrix} \alpha & \beta & 0 & 0 \\ \Gamma & \Delta & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}, \quad (3.5.3)$$

with an identity matrix in the lower-right 2×2 block, and some non-trivial 2×2 matrix in the upper-left block, which we need to determine.

Now the coordinates of events on the worldline of observer B, in frame B coordinates, satisfy $x'^1 = x'^2 = x'^3 = 0$ since observer B is sitting at the spatial origin of her coordinate system. Specializing to this worldline, the transformation (3.5.3) gives

$$x^0 = \alpha x'^0, \quad x^1 = \Gamma x'^0, \quad (3.5.4)$$

implying that $x^1 = (\Gamma/\alpha)x^0$. But we already know that this worldline, in frame A coordinates, should satisfy $x^1 = (v/c)x^0$ since observer B moves with velocity v in the 1-direction relative to observer A. Therefore, we must have $\Gamma/\alpha = v/c$. We also know that from observer A’s perspective, clocks at rest in frame B run slower than clocks at rest in frame A by a factor of $\gamma = 1/\sqrt{1 - (v/c)^2}$. In other words,

$$\gamma = \frac{\Delta t_A}{\Delta t_B} = \frac{dx^0}{dx'^0} = \alpha. \quad (3.5.5)$$

Combining this with the required value of Γ/α implies that $\Gamma = \gamma(v/c)$. This determines the first column of the Lorentz boost matrix (3.5.3).

To fix the second column, consider the events comprising the x'^1 axis in frame B, or those events with $x'^0 = x'^2 = x'^3 = 0$ and x'^1 arbitrary. These events lie on the surface of simultaneity of the spacetime origin in frame B. Above we learned that this surface, as viewed in reference frame A, is

the tilted plane with slope v/c , whose events satisfy $x^0 = (v/c)x^1$. But applied to the x'^1 axis in frame B, the transformation (3.5.3) gives

$$x^0 = \beta x'^1, \quad x^1 = \Delta x'^1, \quad (3.5.6)$$

or $x^0 = (\beta/\Delta)x^1$. Therefore, we must have $\beta/\Delta = v/c$. Finally, we can use the fact that events on the path of a light ray emitted from the spacetime origin and moving in the 1-direction must satisfy both $x'^1 = x'^0$ and $x^1 = x^0$, since observers in both frames will agree that the light moves with speed c . But if $x'^1 = x'^0$, then the transformation (3.5.3) gives $x^0 = (\alpha + \beta)x'^0$, and $x^1 = (\Gamma + \Delta)x'^0$. Therefore, we must have $\alpha + \beta = \Gamma + \Delta$. Inserting $\alpha = \gamma$, $\Gamma = (v/c)\gamma$, $\beta = (v/c)\Delta$ and solving for Δ yields $\Delta = \gamma$. Putting it all together, we have

$$\Lambda = \begin{pmatrix} \gamma & \gamma(v/c) & 0 & 0 \\ \gamma(v/c) & \gamma & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}, \quad (3.5.7)$$

for a boost along the 1-direction with velocity v . The mixing of the 0 and 1 components of four-vectors provided by this matrix is the direct analogue of the usual mixing of 2 spatial components under an ordinary spatial rotation (recall Eq. (1.5.10)). In some sense the difference when mixing with the 0 (or time) component is that the rotation “angle” is now *imaginary* and we obtain hyperbolic functions (instead of sinusoidal functions - recall the discussion in Chapter 1), and no minus sign. To see this point explicitly, a useful notation is

$$\gamma \equiv \cosh y, \quad \gamma \frac{v}{c} \equiv \sinh y, \quad \tanh y = \frac{v}{c}, \quad (3.5.8)$$

so that Eq. (3.5.7) can be written in the evocative form

$$\Lambda = \begin{pmatrix} \cosh y & \sinh y & 0 & 0 \\ \sinh y & \cosh y & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}. \quad (3.5.9)$$

The quantity y is called the “rapidity” and is a useful kinematic variable at particle colliders like the LHC. This notation efficiently encodes the fact that (following from the definition of γ and the hyperbolic functions)

$$\gamma^2 - \gamma^2 \left(\frac{v}{c}\right)^2 = \cosh^2 y - \sinh^2 y = 1, \quad (3.5.10)$$

which ensures that $\det \Lambda = 1$ as was also true for ordinary rotations.

Using the matrix (3.5.7) or (3.5.9) and multiplying out the transformation (3.5.1) yields

$$x^0 = \gamma(x'^0 + \frac{v}{c}x'^1) = (\cosh y x'^0 + \sinh y x'^1), \quad x^2 = x'^2, \quad (3.5.11a)$$

$$x^1 = \gamma(\frac{v}{c}x'^0 + x'^1) = (\sinh y x'^0 + \cosh y x'^1), \quad x^3 = x'^3. \quad (3.5.11b)$$

With a little more work, one may show that the general Lorentz transformation matrix for a boost with speed v in an arbitrary direction specified by a unit vector $\hat{n} = (n_1, n_2, n_3)$, $n_1^2 + n_2^2 + n_3^2 = 1$ is given by

$$\Lambda = \begin{pmatrix} \gamma & \gamma(v/c)n_1 & \gamma(v/c)n_2 & \gamma(v/c)n_3 \\ \gamma(v/c)n_1 & 1 + (\gamma-1)n_1^2 & (\gamma-1)n_1n_2 & (\gamma-1)n_1n_3 \\ \gamma(v/c)n_2 & (\gamma-1)n_1n_2 & 1 + (\gamma-1)n_2^2 & (\gamma-1)n_2n_3 \\ \gamma(v/c)n_3 & (\gamma-1)n_1n_3 & (\gamma-1)n_2n_3 & 1 + (\gamma-1)n_3^2 \end{pmatrix}. \quad (3.5.12)$$

Finally, it is always possible for two inertial reference frames to differ by a spatial rotation (of the axes), in addition to a boost. The coordinate transformation corresponding to a spatial rotation may also be written in the form (3.5.1), but with a transformation matrix which has the block-diagonal form

$$\Lambda = \begin{pmatrix} 1 & \\ & R \end{pmatrix} \quad (\text{spatial rotation}), \quad (3.5.13)$$

where R is some 3×3 rotation matrix (an orthogonal matrix with determinant one as, for example, in Eq. (1.5.10) in our discussion in Chapter 1, a representation of an element of the Special Orthogonal Group $\text{SO}(3)$ ¹). In other words, for such transformations the time coordinates are not affected, $x^0 = x'^0$, while the spatial coordinates are transformed by the rotation matrix R . The most general Lorentz transformation is a product of a rotation of the form (3.5.13) and a boost of the form (3.5.12),

$$\Lambda = \Lambda_{\text{boost}} \times \Lambda_{\text{rotation}}, \quad (3.5.14)$$

and is an element of the group $\text{SO}(3,1)$, where the $3,1$ notation reminds us of the difference (in the signs in the metric, see below) between the 3 spatial dimensions and the 1 time dimension.

3.6 Spacetime vectors

In ordinary three-dimensional (Euclidean) space, if one designates some point O as the spatial origin then one may associate every other point X with a vector which extends from O to X . One can, and should, regard vectors as geometric objects, independent of any specific coordinate system. However, it is very often convenient to introduce a set of basis vectors $\{\hat{e}_1, \hat{e}_2, \hat{e}_3\}$ (normally chosen to point along orthogonal, right-handed coordinate axes), and then express arbitrary vectors as linear combinations of the chosen basis vectors,

$$\vec{v} = \sum_{i=1}^3 \hat{e}_i v^i. \quad (3.6.1)$$

The components $\{v^i\}$ of the vector depend on the choice of basis vectors, but the geometric vector \vec{v} itself does not.

In exactly the same fashion, once some event O in spacetime is designated as the spacetime origin, one may associate every other event X with a *spacetime* vector which extends from O to X . Spacetime vectors (also called “4-vectors”) are geometric objects, whose meaning is independent of any specific reference frame. However, once one chooses a reference frame, one may introduce an associated set of spacetime basis vectors, $\{\hat{e}_0, \hat{e}_1, \hat{e}_2, \hat{e}_3\}$, which point along the corresponding coordinate axes. And, as in any vector space, one may then express an arbitrary spacetime vector v as a linear combination of these basis vectors,

$$v = \sum_{\mu=0}^3 \hat{e}_\mu v^\mu. \quad (3.6.2)$$

We will use Greek letters (most commonly α and β , or μ and ν) to represent spacetime indices which run from 0 to 3. And typically we will use Latin letters i, j, k to represent spatial (only) indices

¹If you are not familiar with the concepts and language of group theory, which will be useful in much of our discussion this quarter, you are encouraged to look at the brief introduction to group theory in Chapter 10 of the (supplementary) lecture notes for this class.

which run from 1 to 3. We will often use an *implied* summation convention in which the sum sign is omitted, but is implied by the presence of *repeated* indices:

$$\hat{e}_\mu v^\mu \equiv \sum_{\mu=0}^3 \hat{e}_\mu v^\mu. \quad (3.6.3)$$

We will generally *not* put vector signs over spacetime vectors, instead relying on the context to make clear whether some object is a (4-)vector. But we will put vector signs over three-dimensional spatial vectors, to distinguish them from spacetime vectors.

The spacetime coordinates of an event are the components of the spacetime vector x associated with this event in the chosen reference frame,

$$x = \hat{e}_\mu x^\mu \equiv \hat{e}_0 x^0 + \hat{e}_1 x^1 + \hat{e}_2 x^2 + \hat{e}_3 x^3. \quad (3.6.4)$$

A different reference frame will have basis vectors which are linear combinations of the basis vectors in the original frame. Consider a ‘primed’ frame whose coordinates $\{x'^\mu\}$ are related to the coordinates $\{x^\nu\}$ of the original frame via a Lorentz transformation (3.5.1). It is convenient to write the components of the transformation matrix as $\Lambda^\mu{}_\nu$ (where the first index labels the row and the second labels the column, as usual for matrix components). Then the linear transformation (3.5.1) may be compactly rewritten as

$$x^\mu = \Lambda^\mu{}_\nu x'^\nu. \quad (3.6.5)$$

The inverse transformation, expressing primed coordinates in terms of unprimed ones, is

$$x'^\mu = (\Lambda^{-1})^\mu{}_\nu x^\nu, \quad (3.6.6)$$

where $(\Lambda^{-1})^\mu{}_\nu$ are the components of the inverse matrix Λ^{-1} .² The components of *any* 4-vector transform in exactly the same fashion when one transforms between two given reference frames.

The Lorentz transformation matrix also relates the basis vectors in the two frames (note the indices),

$$\hat{e}'_\nu = \hat{e}_\mu \Lambda^\mu{}_\nu. \quad (3.6.7)$$

In other words, if you view the list $(\hat{e}_0, \hat{e}_1, \hat{e}_2, \hat{e}_3)$ as a row-vector, then it is multiplied on the *right* by a Lorentz transformation matrix Λ . The transformation of basis vectors must have precisely this form so that the complete spacetime vector is frame independent, as initially asserted,

$$x = \hat{e}'_\mu x'^\mu = \hat{e}_\nu \Lambda^\nu{}_\mu (\Lambda^{-1})^\mu{}_\alpha x^\alpha = \hat{e}_\nu x^\nu. \quad (3.6.8)$$

Next recall that the dot or scalar product of two spatial vectors, $\vec{a} \cdot \vec{b}$, is defined geometrically, without reference to any coordinate system, as the product of the length of each vector times the cosine of the angle between them. One can then show that this is the same as the component-based definition, $\vec{a} \cdot \vec{b} = \sum_i a^i b^i$, for *any* choice of Cartesian coordinates. It is this frame or rotation *independence* that ensures this product is a scalar.

²For boost matrices of the form (3.5.7) or (3.5.12), changing the sign of v (or y) converts Λ into its inverse. Note that this changes the sign of the off-diagonal components in the first row and column, leaving all other components unchanged. For transformations which also include spatial rotations, to convert the transformation to its inverse one must transpose the matrix in addition to flipping the sign of these “time-space” components, corresponding to changing the sign of the rotation angle.

What is the appropriate generalization of the dot or scalar product for spacetime vectors? This should be some operation which, given two 4-vectors a and b , produces a single number. The operation should be *symmetric*, so that $a \cdot b = b \cdot a$, and *linear*, so that $a \cdot (b + c) = a \cdot b + a \cdot c$. The result should also be *independent* of the choice of (inertial) reference frame one uses to specify the components of these vectors, *i.e.*, be a scalar (unchanged) under Lorentz transformations. Finally it should essentially reduce to the usual spatial dot product if both a and b lie within a common surface of simultaneity. There is a (nearly) unique solution to these requirements.

However, there is a *sign* ambiguity when satisfying the above constraints (except the last) and you will see two definitions of the Lorentz scalar product in common usage (and it is important to recognize this fact in order to avoid confusion). The one typically labeled the “East Coast” choice is given by $a \cdot b \equiv -a^0 b^0 + a^1 b^1 + a^2 b^2 + a^3 b^3$. This definition of the dot product differs from the four dimensional Euclidean space definition of a dot product merely by the change in sign of the time-component term. It satisfies the required linearity and reduces to the usual spatial dot product if the time components of both four vectors vanish. The alternative “West Coast” scalar product, which is used in the the text by Kogut and will be used in this class, is given by

$$a \cdot b \equiv +a^0 b^0 - a^1 b^1 - a^2 b^2 - a^3 b^3, \quad (3.6.9)$$

or with an implied summation on spatial indices, $a \cdot b = +a^0 b^0 - a^i b^i$. Only the overall sign of the scalar or dot product changes between the two definitions, and not the underlying symmetry properties or the physics. Note that with the “West Coast” scalar product it is the 3 spatial components that differ in sign from the Euclidean space scalar product, but with the advantage that typical physically interesting scalar quantities will have positive values. (But be warned that the East Coast definition is used when Prof. Yaffe teaches this course. As a result his lecture notes exhibit sign differences in several places.)

To see that this definition of the scalar product is frame-independent (*i.e.*, is really a scalar), it is sufficient to check the effect of a boost of the form (3.5.7) (since we already know that a rotation of coordinates does not affect the three-dimensional dot product). Transforming the components of the 4-vectors a and b to a primed frame, as in Eq. (3.6.6), using the boost (3.5.7) gives

$$a'^0 = \gamma \left(a^0 - \frac{v}{c} a^1 \right), \quad a'^1 = \gamma \left(a^1 - \frac{v}{c} a^0 \right), \quad a'^2 = a^2, \quad a'^3 = a^3, \quad (3.6.10a)$$

$$b'^0 = \gamma \left(b^0 - \frac{v}{c} b^1 \right), \quad b'^1 = \gamma \left(b^1 - \frac{v}{c} b^0 \right), \quad b'^2 = b^2, \quad b'^3 = b^3. \quad (3.6.10b)$$

Hence

$$\begin{aligned} a'^0 b'^0 - a'^1 b'^1 &= \gamma^2 \left[\left(a^0 - \frac{v}{c} a^1 \right) \left(b^0 - \frac{v}{c} b^1 \right) - \left(a^1 - \frac{v}{c} a^0 \right) \left(b^1 - \frac{v}{c} b^0 \right) \right] \\ &= \gamma^2 \left[1 - (v/c)^2 \right] (+a^0 b^0 - a^1 b^1) \\ &= +a^0 b^0 - a^1 b^1, \end{aligned} \quad (3.6.11)$$

where the last step used $\gamma^2 \equiv 1/[1 - (v/c)^2]$. Therefore, as claimed, the value of the dot product (3.6.9) (or with the alternative definition) is *independent* of the specific inertial frame one uses to define the vector coefficients and, in that sense, is a scalar.

The spacetime dot product (3.6.9) is a useful construct in many applications (since the underlying physics is Lorentz invariant and thus expressible in terms of Lorentz scalars). As a preview of things to come, consider some plane wave (acoustic, electromagnetic, or any other wave type) propagating with frequency ω and wave-vector \vec{k} . One normally writes the complex amplitude for such a wave

as some overall coefficient times $e^{-i\omega t + i\vec{k} \cdot \vec{x}}$. Having already defined the spacetime position vector x whose time component $x^0 \equiv ct$, if we also define a spacetime wave-vector k whose time component $k^0 \equiv \omega/c$ then this ubiquitous phase factor may be written as a spacetime dot product,

$$e^{-i\omega t + i\vec{k} \cdot \vec{x}} = e^{-ik \cdot x}. \quad (3.6.12)$$

Similarly, in quantum mechanics the wave function of a particle with definite momentum \vec{p} and energy E moving in empty space is proportional to $e^{-iEt/\hbar + i\vec{p} \cdot \vec{x}/\hbar}$. If we define a 4-momentum p with time component $p^0 = E/c$, then this phase factor may also be written as a spacetime dot product,

$$e^{-iEt/\hbar + i\vec{p} \cdot \vec{x}/\hbar} = e^{-ip \cdot x}. \quad (3.6.13)$$

(Note that in the East Coast definition the minus sign in the exponent becomes a plus sign.)

3.7 Units: What is large and what is small?

It may be helpful at this point to say a few words about units and the size of things. Recall from freshman physics that one of the most confusing issues in the introductory course is the question of units. For quantities with units (which we will call “dimensionfull” quantities) the specific size will depend on the choice of units. For example, in (old) English units a typical student is approximately 6 feet tall, while in by now standard (except in the US) MKS units that means just 2 meters tall. This is clearly a confusing situation. A (single) dimensionfull quantity has no intrinsic “size” as its numerical value depends on the (arbitrary) choice of units. However, a dimensionfull quantity can be (meaningfully) large or small *compared* to another dimensionfull quantity with the *same* units. We often say that non-relativistic kinematics apply for small velocities. What we really mean is for velocities small *compared* to the velocity of light c . Thus in the equations above the relevant measure of relativistic effects is the ratio v/c (often labeled $\beta = v/c$) as in $\gamma = 1/\sqrt{1 - v^2/c^2} = 1/\sqrt{1 - \beta^2}$. When β is small compared to one (the “natural” separator between large and small), non-relativistic approximations are accurate, while as $\beta \rightarrow 1$ we must use the full relativistic description.

Similarly when quantities like $p \cdot x$ in Eq. (3.6.13) are large *compared* to \hbar (many “quanta”) the effects of interference are numerically small and “classical mechanics” pertains. Yet when $p \cdot x/\hbar$ is of order unity or smaller even bullets can display “wavy” (*i.e.*, quantum mechanical) behavior.

A related issue is that the MKS system exhibits *three* fundamental varieties of dimensionfull quantities, length (m), mass (kg) and time (s). Yet in the relativistic and quantum mechanical world of particle physics that we want to discuss here, we clearly want to employ 4-vectors, which relate time with space (and energies with momentum) as in Eqs. (3.3.1) and (3.6.4) (and Eq. (3.6.13)). In order to make the units of the different components match-up (as they must in order that we can Lorentz transform the components into one another), we had to introduce all those factors of c . We also introduced the factor of $1/\hbar$ in Eq. (3.6.13) to ensure that the argument of the exponential is dimensionless. Since the exponential is defined by a power series and each term in the series must have the *same* units, the only possibility is that the exponent (the argument of the exponential) has *no* units, *i.e.*, is dimensionless.

Further, as noted above, the actual magnitudes of the standard units were chosen to correspond to human scales (*e.g.*, the size of a king). These choices are, of course, unnatural for particle physics applications. For example, the mass of a proton is 1.67×10^{-27} kg while the spatial “size” of a proton is measured in fermi’s (1 femtometer = 1 fm = 10^{-15} m), not meters. Likewise the lifetime of a

typical particle that decays via the strong interactions is of order 10^{-23} s, which is the time for light to travel across a particle of size 1 fm. Since the particle physics we will discuss later in this course is “naturally” relativistic and quantum mechanical, we would like to make a different choice of scales so that the speed of light c and \hbar are both of order 1. It turns out we can address all of the above issues by defining a new set of “particle physics units” such that both c and \hbar are exactly equal to 1!!!! In the process we have reduced the number of types of dimensionfull quantities to 1. In these rather surprising “natural” units we have

$$c = 2.9979 \times 10^8 \text{ m/s} = 1, \quad (3.7.1a)$$

$$\hbar = 1.055 \times 10^{-34} \text{ Js} = 6.58 \times 10^{-22} \text{ MeVs} = 1. \quad (3.7.1b)$$

Thus time now has the same units as distance. Likewise mass and energy have the same units and both go like 1/distance or 1/time. In these new units the mass of the proton is essentially 1 GeV ($0.938 \text{ GeV}/c^2$) ($1 \text{ GeV} = 1 \text{ gigaelectronvolt} = 10^9 \text{ electronvolts}$). We also have one fm equal to $1/(197 \text{ MeV}) \sim 1/(200 \text{ MeV}) = 1/(0.2 \text{ GeV})$ ($1 \text{ MeV} = 10^6 \text{ electronvolts}$). It is typical in particle physics to express (nearly) all dimensionfull quantities in terms of the “natural” (particle physics) unit of GeV. A list of useful values is provided in the following table, where the “old” units are indicated in the [] brackets.

Units
$1 \text{ kg} = 5.61 \times 10^{26} \text{ GeV} [\text{GeV}/c^2]$
$1 \text{ m} = 5.07 \times 10^{15} \text{ GeV}^{-1} [\hbar c/\text{GeV}]$
$1 \text{ s} = 1.52 \times 10^{24} \text{ GeV}^{-1} [\hbar/\text{GeV}]$
$1 \text{ TeV} = 10^{12} \text{ eV} = 10^3 \text{ GeV} [\text{GeV}/c^2]$
$1 \text{ fm} = 1 \text{ F} = 10^{-13} \text{ cm} = 5.07 \text{ GeV}^{-1}$
$(1 \text{ fm})^2 = 10 \text{ mb} = 10^{-26} \text{ cm}^2 = 25.7 \text{ GeV}^{-2}$
$(1 \text{ GeV})^{-2} = 0.389 \text{ mb}$

As suggested by the last 2 lines, the “areas” of particles (*i.e.*, the cross sections for scattering) are typically measured in millibarns (mb). Masses, energies and momenta are measured in GeV, while distances and times are in GeV^{-1} . Thus the product of distance and momenta (time and energy) is

dimensionless, as desired. In these units the sizes of various “objects” become:

$$\begin{aligned}
 &\text{Sizes } (\sim \text{ means ignore factors of } 2) \\
 &\text{Universe } \sim 10^{26} \text{ m} = 5 \times 10^{41} \text{ GeV}^{-1} (\sim 10^{11} \text{ galaxies}) \\
 &\text{Galaxy Supercluster } \sim 10^{24} \text{ m} = 5 \times 10^{39} \text{ GeV}^{-1} \\
 &\text{Galaxy } \sim 10^{21} \text{ m} = 5 \times 10^{36} \text{ GeV}^{-1} (\sim 10^{11} \text{ stars}) \\
 &\text{Star } \sim 10^9 \text{ m} = 5 \times 10^{24} \text{ GeV}^{-1} \\
 &\text{Earth } \sim 10^7 \text{ m} = 5 \times 10^{22} \text{ GeV}^{-1} \\
 &\text{Human } \sim 10^0 \text{ m} = 5 \times 10^{15} \text{ GeV}^{-1} \\
 &\text{Atom } \sim 10^{-10} \text{ m} = 5 \times 10^5 \text{ GeV}^{-1} \\
 &\text{Nucleus } \sim 10^{-14} \text{ m} = 5 \times 10^1 \text{ GeV}^{-1} \\
 &\text{Proton } \sim 10^{-15} \text{ m} = 5 \times 10^0 \text{ GeV}^{-1} \\
 &\text{Present observational limit } \sim 10^{-19} \text{ m} = 5 \times 10^{-4} \text{ GeV}^{-1} \\
 &\text{Planck length } \sim 10^{-35} \text{ m} = 5 \times 10^{-20} \text{ GeV}^{-1}
 \end{aligned}$$

This last quantity is the length scale (inverse mass scale) set by the (very weak) gravitational interactions. Note the huge range of sizes that characterize our universe.

You will not be surprised to learn that with only one fundamental type of dimensionfull unit it is easy to define dimensionless ratios. In many instances these are the simplest quantities to understand in particle physics. On the other hand, the really interesting (and more difficult to explain) quantities are the small number of dimensionfull quantities. Examples include $\Lambda_{\text{QCD}} (\approx 0.2 \text{ GeV})$, the fundamental dimensionfull parameter characterizing the strong interaction, G_F (the Fermi constant, $\approx 1.2 \times 10^{-5} \text{ GeV}^{-2}$) or M_W (the mass of the W boson, $\approx 80 \text{ GeV}$), the dimensionfull parameters that characterize the weak interactions and G_N , Newton’s constant ($\approx 6.7 \times 10^{-35} \text{ GeV}^{-2}$), that characterizes the gravitational interaction.

For now in this course, we will keep the explicit factors of c and \hbar , but our goal is to become comfortable with the natural units of particle physics where $c = \hbar = 1$.

3.8 Minkowski spacetime

In Euclidean space, the dot product of a vector with itself gives the square of the norm (or length) of the vector, $\vec{v} \cdot \vec{v} \equiv |\vec{v}|^2$. This is the familar situation for three dimensional spatial vectors. Proceeding by analogy, we will *define* the square of a spacetime vector using the dot product (3.6.9), so that

$$(a)^2 \equiv a \cdot a = (a^0)^2 - (a^1)^2 - (a^2)^2 - (a^3)^2. \quad (3.8.1)$$

If Δx is a spacetime vector representing the separation between two events, then the square of Δx is called the *invariant interval* separating these events. This is usually denoted by s^2 , so that

$$s^2 \equiv (\Delta x^0)^2 - (\Delta x^1)^2 - (\Delta x^2)^2 - (\Delta x^3)^2. \quad (3.8.2)$$

Spacetime in which the “distance” between events is defined by this expression is called *Minkowski spacetime*.³

The definition of the invariant interval (3.8.2), or the square of a vector (3.8.1), differ from the usual Euclidean space relations due to the minus signs in front of the spatial component terms (or in front of the time components in the other definition). But this is a fundamental change. Unlike Euclidean distance, the spacetime interval s^2 can be positive, negative, or zero (and this is true for either definition of where the minus signs go). Let Δx be the spacetime displacement from some event X to another event Y . If the interval $s^2 = (\Delta x)^2$ vanishes, then the spatial separation between these events equals their separation in time multiplied by c ,

$$s^2 = 0 \implies (\Delta \vec{x})^2 = (\Delta x^0)^2 = (c \Delta t)^2 \quad (\text{lightlike separation}). \quad (3.8.3)$$

This means that light could propagate from X to Y (if $\Delta t > 0$), or from Y to X (if $\Delta t < 0$). In other words, event Y is on the lightcone of X , or vice-versa. In this case, one says that the separation between X and Y is *lightlike*.

If the interval s^2 is positive (in our metric), then the spatial separation is less than the time separation (times c),

$$s^2 > 0 \implies (\Delta \vec{x})^2 < (\Delta x^0)^2 = (c \Delta t)^2 \quad (\text{timelike separation}). \quad (3.8.4)$$

This means that some particle moving *slower* than light could propagate from X to Y (if $\Delta t > 0$), or from Y to X (if $\Delta t < 0$). In other words, event Y is in the *interior* of the lightcone of X , or vice-versa. In this case, one says that the separation between X and Y is *timelike*.

Finally, if the interval s^2 is negative (in our metric), then the spatial separation is greater than the time separation (times c),

$$s^2 < 0 \implies (\Delta \vec{x})^2 > (\Delta x^0)^2 = (c \Delta t)^2 \quad (\text{spacelike separation}). \quad (3.8.5)$$

In other words, event Y is *outside* the lightcone of X , and vice-versa. In this case, one says that the separation between X and Y is *spacelike*. These possibilities are shown pictorially in Figure 3.7.

With the alternate definition of the scalar product, *i.e.*, the extra overall minus sign, spacelike separations correspond to positive values of s^2 while timelike separations are negative. This is the most confusing feature of having two definitions in wide usage - you need to know what definition is being used to distinguish spacelike from timelike from the sign alone.⁴

³Minkowski spacetime is the domain of special relativity, in which gravity is neglected. Correctly describing gravitational dynamics leads to general relativity, in which spacetime can have curvature and the interval between two arbitrary events need not have the simple form (3.8.2). We will largely ignore gravity.

A further word about index conventions may also be appropriate. It is standard in modern physics to write the components of 4-vectors with superscripts, like a^μ or x^ν , as we have been doing. Although we will not need this, it is also conventional to define subscripted components which, in Minkowski space (with our choice of the scalar product), differ by flipping the sign of the space components, so that $a_k \equiv -a^k$ ($k = 1, 2, 3$) for any 4-vector a . This allows one to write the dot product of two 4-vectors a and b as $a_\mu b^\mu$ (with the usual implied sum). More generally, in curved space one defines a metric tensor $g_{\mu\nu}$ via a differential relation of the form $ds^2 = g_{\mu\nu} dx^\mu dx^\nu$, and then defines $a_\mu \equiv g_{\mu\nu} a^\nu$ so that $a \cdot b = a_\mu b^\mu = a^\mu b_\mu = g_{\mu\nu} a^\mu b^\nu$. In flat spacetime the metric tensor is diagonal: “West Coast” $g_{\mu\nu} = \text{diag}[1, -1, -1, -1]$, “East Coast” $g_{\mu\nu} = \text{diag}[-1, 1, 1, 1]$.

⁴When you begin a physics conversation with another physicist, the first question should be to establish what sign convention to use (an essential part of the “secret” physicist’s handshake). The choice is typically correlated with where the physicist went to graduate school and this explains the coast-based labels. Those trained on the East Coast, *e.g.*, Larry Yaffe, use the sign convention used, *e.g.*, in the autumn 2013 version of this class, while those trained on the West Coast, *e.g.*, John Kogut and Steve Ellis, tend to use the convention used in this class this quarter. You should learn to be fluent in both.

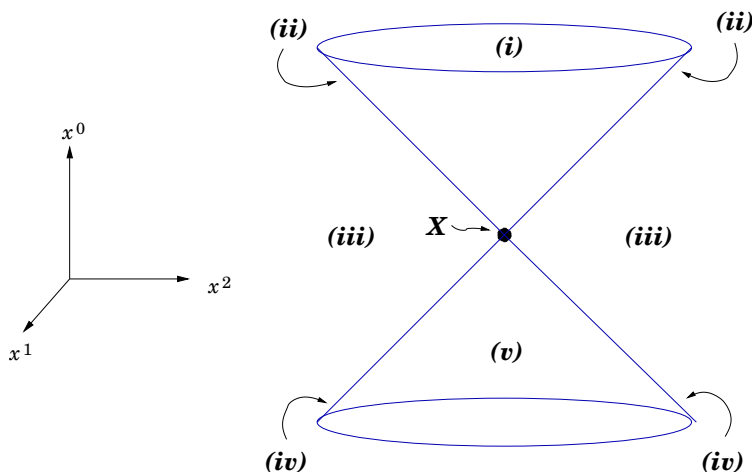


Figure 3.7: The past and future lightcones of an event X separate spacetime into those events which are: (i) timelike separated and in the future of X , (ii) lightlike separated and in the future of X , (iii) spacelike separated, (iv) lightlike separated and in the past of X , and (v) timelike separated and in the past of X .

3.9 The pole and the barn

A classic puzzle illustrating basic aspects of special relativity is the pole and the barn, sketched in Figure 3.8. You are standing outside a barn whose front and back doors are open. A (very fast!) runner carrying a long horizontal pole is approaching the barn. The (proper) length of the barn (measured in its rest frame) is 10 meters. The length of the pole, when measured at rest, is 20 meters. But the relativistic runner is moving at a speed of $\frac{\sqrt{3}}{2}c \approx 0.866c$, and hence the

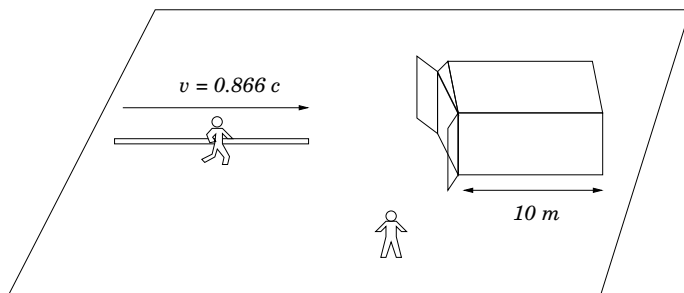


Figure 3.8: A relativistic runner, carrying a long pole, approaches and passes through a barn. Does the pole fit within the barn?

pole (in your frame) is Lorentz contracted by a factor of $1/\gamma = \sqrt{1 - (v/c)^2} = 1/2$. Consequently, from your standpoint, the pole just “fits” within the barn; when the front of the pole emerges from one end of the barn, the back of the pole will have just passed into the barn through the other door.

But now consider this situation from the runner’s perspective. In his (or her) co-moving frame, the pole is 20 meters long. The barn is coming toward the runner at a speed of $-\frac{\sqrt{3}}{2}c$, and hence the barn which is 10 meters long in its rest frame is Lorentz contracted to a length of only 5 meters. The pole cannot possibly fit within the barn!

Surely the pole either does, or does not, fit within the barn. Right? Which description is correct?

This puzzle, like all apparent paradoxes in special relativity, is most easily resolved by drawing a spacetime diagram which clearly displays the relevant worldlines and events of interest. It is often also helpful to draw contour lines on which the invariant interval s^2 (relative to some key event) is constant. For events within the x^0 – x^1 plane, the invariant interval from the origin is just $s^2 = (x^0)^2 - (x^1)^2$. Therefore, the set of events in the x^0 – x^1 plane which are at some fixed interval

s^2 from the origin lie on a hyperbola.⁵

Let us create a spacetime diagram for this puzzle working in the reference frame of the barn. (This is an arbitrary choice. We could just as easily work in the runner's frame.) Try doing this yourself before reading the following step-by-step description of Figure 3.9.

Orient coordinates so that the ends of the barn are at $x^1 = 0$ and $x^1 = 10\text{ m}$. Therefore, the worldline of the left end of the barn (w_L) is a vertical line at $x^1 = 0$, while the worldline of the right end of the barn (w_R) is a vertical line at $x^1 = 10\text{ m}$. Since the pole is moving at velocity $\frac{\sqrt{3}}{2}c$ (in the x^1 direction), the worldlines of the ends of the pole are straight lines in the x^0 - x^1 plane with a slope of $c/v = 2/\sqrt{3} \simeq 1.155$. Call the moment when the back end of the pole passes into the barn time zero. So the worldline of the back end of the pole (w'_B) crosses the worldline of the left end of the barn at event A with coordinates $(x^0, x^1) = (0, 0)$. In the frame in which we're working the pole is Lorentz contracted to a length of 10 meters. Hence, the worldline of the front end of the pole (w'_F) must cross the x^1 axis at event B with coordinates $(x^0, x^1) = (0, 10\text{ m})$. This event lies on the worldline w_R of the right end of the barn, showing that in this reference frame, at time $t = 0$, the Lorentz contracted pole just fits within the barn.

Now add to the diagram the surface of simultaneity of event A in the runner's frame. From section 3.4 we know that this surface, in the frame in which we are drawing our diagram, is tilted upward so that its slope is $v/c \simeq 0.866$ (and the 45° lightcone of event A bisects the angle between this surface and the worldline w'_B). The worldline w_R of the right end of the barn intersects this surface of simultaneity at event C , while the worldline w'_F of the front of the pole intersects this surface at event D . This surface of simultaneity contains events which, in the runner's frame, occur at the same instant in time. From the diagram it is obvious that event C lies between events A and D . In other words, in the runner's frame, at the moment when the back end of the pole passes into the barn, the front end of the pole is far outside the other end of the barn — the pole does *not* fit in the barn.

The essential point of this discussion, and the spacetime diagram in Figure 3.9, is the distinction between events which are simultaneous in the runner's frame (events A , C , and D), and events which are simultaneous in the barn's frame (A and B). *Both* descriptions given initially were correct. The only fallacy was thinking that it was meaningful to ask whether the pole does (or does not) fit within the barn, without first specifying a reference frame. The answer *depends* on the choice of frame.

To complete our discussion of this spacetime diagram, consider the invariant interval between event A (which is our spacetime origin) and each of the events B , C , and D . Within the two-dimensional plane of the figure, the invariant interval from the origin is $s^2 = (x^0)^2 - (x^1)^2$. We know that event B has coordinates $(x^0, x^1) = (0, 10\text{ m})$ so it is immediate that $s_{AB}^2 = -(10\text{ m})^2$. We could work out the (x^0, x^1) coordinates of events C and D , and from those coordinates evaluate their interval from event A . But this is not necessary since we can use the fact that events C and D lie on the runner's frame surface of simultaneity of event A . We are free to evaluate the interval from event A using the runner's frame coordinates, instead of barn frame coordinates. Within the two-dimensional plane of the figure, $s^2 = (x'^0)^2 - (x'^1)^2$. Events A , C , and D are simultaneous in the runner's frame, so all their x'^0 coordinates vanish. And in this frame (the rest frame of the pole) we know that the pole's length is 20 m , while the barn's length is Lorentz contracted to 5 m . Hence $s_{AC}^2 = -(5\text{ m})^2$ and $s_{AD}^2 = -(20\text{ m})^2$. Therefore, event C must lie on the hyperbola whose intersection with the x^1 axis

⁵Recall that the equation $y^2 - x^2 = s^2$ defines a hyperbola in the (x, y) plane whose asymptotes are the 45° lines $y = \pm x$. If $s^2 < 0$ then one branch opens toward the right and the other opens toward the left. If $s^2 > 0$ then one branch opens upward and one opens downward.

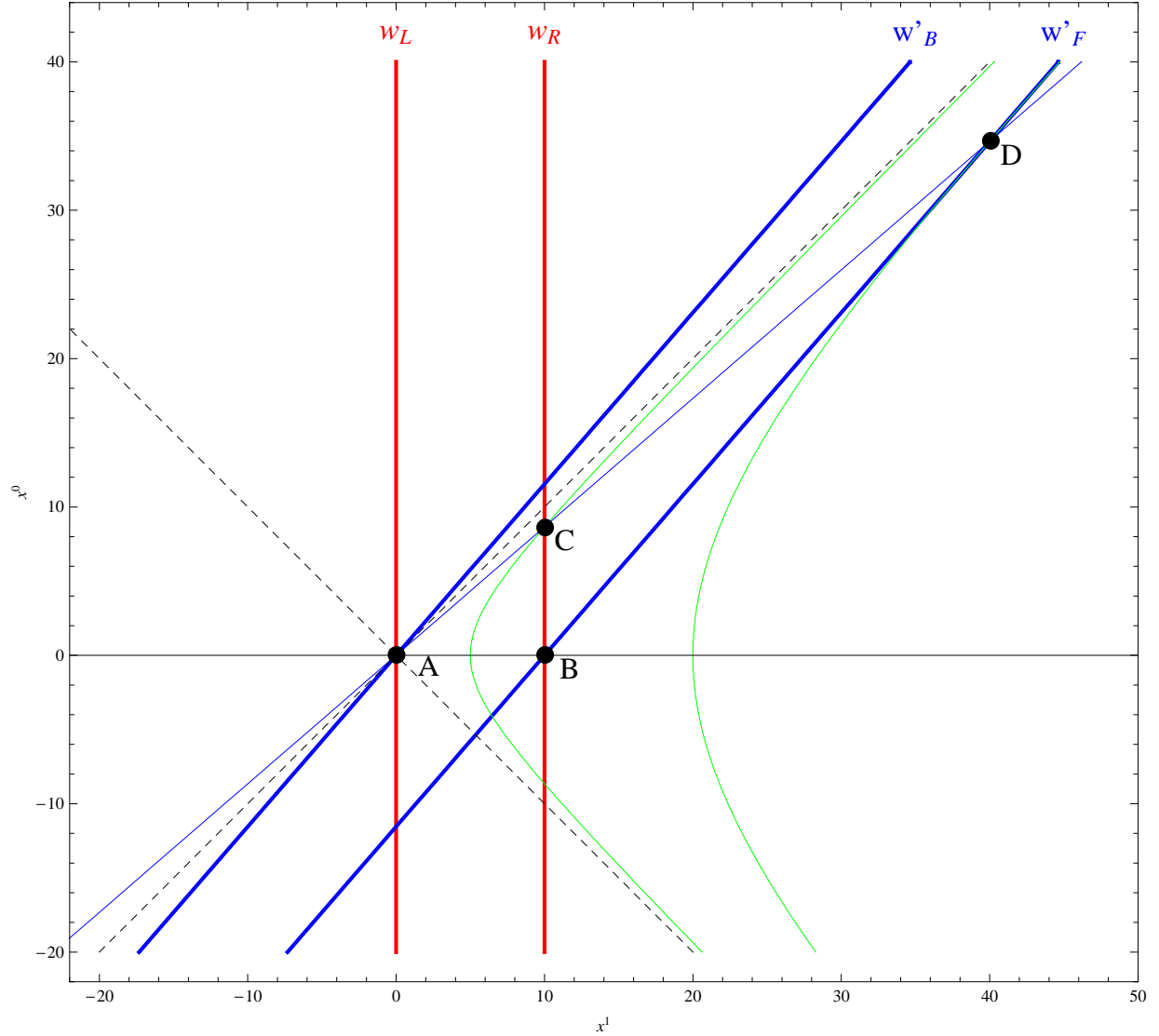


Figure 3.9: A spacetime diagram of the pole and the barn, showing events in the rest frame of the barn. The red vertical lines are the worldlines w_L and w_R of the left and right ends of the barn. The blue lines labeled w'_B and w'_F are the worldlines of the front and back of the pole, respectively. The thin blue line passing through events A, C, and D is a surface of simultaneity in the runner's reference frame. The hyperbola passing through event C shows events at invariant interval $s^2 = -(5m)^2$ relative to event A. Note that this hyperbola intercepts the x^1 axis at $5m$. The hyperbola passing through event D shows events at invariant interval $s^2 = -(20m)^2$ relative to event A. Note that this hyperbola intercepts the x^1 axis at $20m$.

is at $5m$, while event D must lie on the hyperbola whose intersection with the x^1 axis is at $20m$, as indicated by the 2 green curves.

3.10 Causality

Consider any two spacetime events A and B which are spacelike separated. A basic consequence of the fact that surfaces of simultaneity are observer dependent is that different observers can *disagree* about the temporal ordering of spacelike separated events. For example, in the unprimed reference frame illustrated in Fig. 3.10, event B lies in the *future* of event A — its x^0 coordinate is bigger. But event B lies below the $x'^0 = 0$ surface of simultaneity which passes through event A . This means that event B lies in the *past* of event A in the primed reference frame.

This should seem bizarre. If observers at rest in the unprimed frame were to see some particle or signal travel from event A to event B , then this signal would be traveling backwards in time from the perspective of observers at rest in the primed frame. This is inconsistent with causality — the fundamental idea that events in the past influence the future, but *not* vice-versa.

An idealized view of the goal of physics is the prediction of future events based on knowledge of the past state of a system. But if different observers disagree about what events are in the future and what events are in the past, how can the laws of physics possibly take the same form in all reference frames? Are our two relativity postulates fundamentally inconsistent?

If it is possible for some type of signal to travel between events A and B then, because these two events are outside each other's lightcones, this would be superluminal propagation of information. The only way that our postulates can be consistent is if it is simply *not possible* for any signal to travel between spacelike separated events. In other words, a necessary consequence of our postulates is that *no* signal whatsoever can travel faster than light. For fans of science fiction this is a sad state of affairs, but it is an inescapable conclusion. (Read again the discussion at the end of Chapter 2 of the recent, apparently wrong, observation of neutrinos traveling faster than the speed light.)

The situation is different if events A and B are *timelike* separated. First, A and B will be timelike separated in *all* frames. Further, if B occurs after A in some reference frame (so that a signal could propagate from A to B), then this same temporal ordering will obtain in *all* frames. To see this last point, first note that, for a timelike separation, we have $(t_A - t_B)^2 - (\vec{x}_A - \vec{x}_B)^2 > 0$ in all frames. The temporal ordering statement means that $t_B > t_A$ in some frame. In order to switch this temporal ordering to $t'_B < t'_A$ in a different reference frame, there must be an intermediate reference frame where $t''_B = t''_A$, since this quantity changes smoothly with the intervening boosts. But the

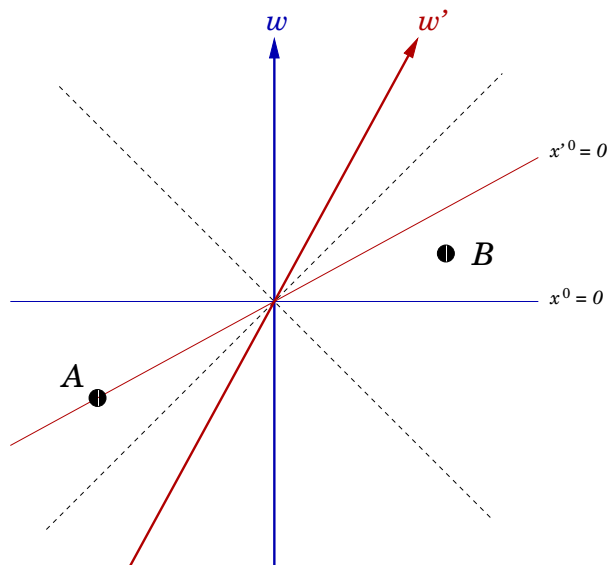


Figure 3.10: Two spacelike separated events A and B . In the unprimed frame, B is in the future of A , but in the primed frame B is in the past of A .

temporal separation can *never* vanish for a timelike separated pair of events (*i.e.*, $(t_A - t_B)^2 > 0$ in all frames).

3.11 Example Problems

Kogut 2-6

The emission and the absorption of a light ray are two distinct spacetime events, which are separated by a distance ℓ in the common rest frame of the emitter and the absorber. This question asks for the spatial and temporal separation of these events as observed in a boosted reference frame traveling with velocity v parallel to the direction from the emitter to the absorber. It is very similar to Kogut problem 2-5. Three different methods for solving the problem (each of which are instructive) are presented below.

Method #1: Thought-experiment

(a) In the original frame, the light ray travels a distance $x_2 - x_1 = \ell$ in a time $t = \ell/c$. Now consider the light ray emission/absorption process in a frame moving with speed v along the x^1 direction of the original frame. Without loss of generality, assume that the origin of the boosted frame coincides with the emission event. As seen in the boosted frame, the original frame is moving with velocity $-v$ along the x^1 direction. Call the time between emission and absorption events (in the boosted frame) t' , so in this frame the light ray travels a distance ct' . Since the distance between x_1 and x_2 was ℓ in the original frame, it is now ℓ/γ in the boosted frame due to Lorentz contraction. But it is also *essential* to realize that while x_1 and x_2 are fixed in the original frame, they are moving as viewed in the boosted frame. In particular, x_2 moves a distance $-vt'$ while the light is traveling, which we must add on to ℓ/γ to obtain the net distance traveled by the light in this frame. Therefore, $ct' = \ell/\gamma - vt'$. Write this as $ct' = \ell/\gamma - (v/c)ct'$, and solve for ct' ,

$$ct' = \frac{\ell}{\gamma(1 + v/c)} = \ell \sqrt{\frac{1 - v/c}{1 + v/c}}.$$

(b) The time between events in the boosted frame is just

$$t' = ct'/c = \frac{\ell}{c} \sqrt{\frac{1 - v/c}{1 + v/c}},$$

(since the speed of light is frame-independent). Notice that this result is not a simple time dilation. For positive v , the time interval between emission and absorption as measured in the boosted frame is less than in the original frame. For negative v , that time interval is greater.

ASIDE: This result allows us to make a connection to our discussion of clocks in Chapter 2. Imagine that, instead of being absorbed, the light ray is reflected back and detected at the emitter. The corresponding time interval (in the original frame) between emission and detection,

$$\Delta t = \frac{2\ell}{c},$$

is just the time between ticks of the clock we discussed in Chapter 2 ($L \rightarrow \ell$). As observed in the moving frame (moving in the configuration of Figure 2.4), the time interval is (note the different

direction of motion in the two segments)

$$\Delta t' = \frac{\ell}{c} \sqrt{\frac{1-v/c}{1+v/c}} + \frac{\ell}{c} \sqrt{\frac{1+v/c}{1-v/c}} = \frac{2\ell}{c} \frac{1}{\sqrt{1-(v/c)^2}} = \gamma \Delta t,$$

which is the usual time dilation result.

Method #2: Lorentz transformation

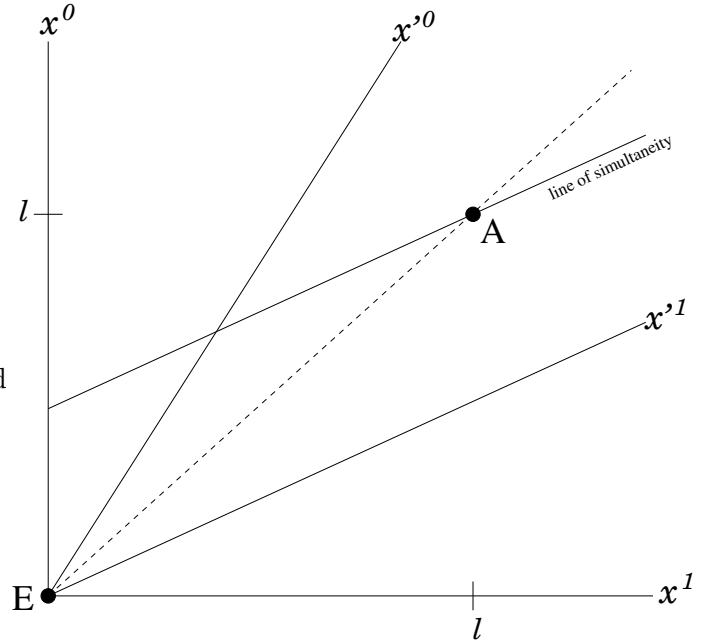
In the original frame, the emission event may be placed at the origin of the Minkowski diagram of spacetime. The absorption event then has coordinates $(x^0, x^1) = (\ell, \ell)$ which lies on the lightcone (since it describes the motion of light!). Under a boost, the origin is mapped to the origin so the emission event also occurs at the origin of the boosted frame (since we assumed that this was the synchronizing event). The absorption event has coordinates (x'^0, x'^1) given by

$$\begin{pmatrix} x'^0 \\ x'^1 \end{pmatrix} = \begin{pmatrix} \gamma & -\gamma \frac{v}{c} \\ -\gamma \frac{v}{c} & \gamma \end{pmatrix} \begin{pmatrix} \ell \\ \ell \end{pmatrix}.$$

The spatial separation is given by $x'^1 = \gamma \ell (1 - v/c)$, which reduces to the same answer given above for ct' , i.e., $\ell \sqrt{(1 - v/c)/(1 + v/c)}$. Since the events lie on the lightcone, the time separation (times c) and spatial separation are equivalent.

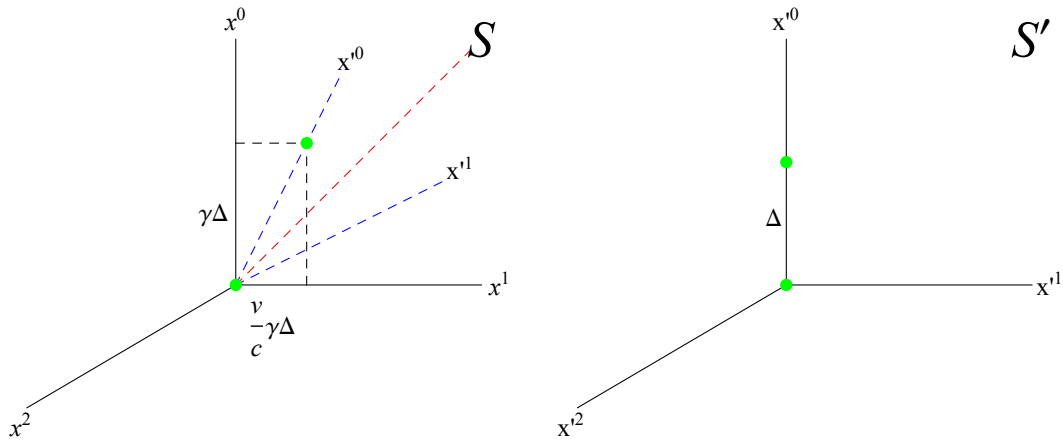
Method #3: Spacetime diagram

In the diagram to the right we have drawn the lines of simultaneity for the boosted observer that intersect the emission and absorption events, E and A. The upper line of simultaneity is described by the equation $(x^0 - \ell)/(x^1 - \ell) = v/c$, which when written in the more familiar slope-intercept form is $x^0 = (v/c)x^1 + \ell(1 - v/c)$.⁶ The x^0 -intercept is $\ell(1 - v/c)$ and as you can see from the diagram it gives the time (times c) between emission and absorption events for the boosted observer. Well, almost. We must realize that the orthogonal axes of the diagram are drawn in the original frame, not the boosted one. So the time we have just extracted is the time measured in the original frame, not the boosted one. But we already know how to convert time intervals between inertial frames in relative motion—use time dilation. A clock carried by the boosted observer will run slower than that carried by the observer at rest. So we again obtain the same result $x'^0 = \gamma x^0 = \gamma \ell (1 - v/c) = \ell \sqrt{(1 - v/c)/(1 + v/c)}$.



More spacetime separation examples.

⁶You should keep in mind that the line of simultaneity is merely the intersection of the three-dimensional hyperplane of simultaneity with the $x^0 - x^1$ plane, so the complete equation is $x^0 - (v/c)x^1 + x^2 + x^3 - \ell(1 - v/c) = 0$.


 Figure 3.11: Timelike separation in S' frame.

Let us make use of the specific Lorentz transformation in Eq. (3.5.7) and the (West Coast) metric to look explicitly at an illustrative variety of pairs of events in the 2 reference frames defined by the boost. As usual, we assume that the two frames have a common origin and that the spatial directions are aligned (*i.e.*, there is no rotation in the transformation, as should be clear from its form).

Timelike separation

Consider the situation suggested in Fig. (3.11). In the S' frame (the right-hand figure) two events (the green dots) occur at the spatial origin, but separated in time (*i.e.*, in x'^0) by a distance Δ . In the S frame (the left-hand figure) the lightcone (red dashed line) and the boosted x'^0 and x'^1 directions (blue dashed lines) are indicated. Note that the two events lie along the x'^0 direction in *both* frames. The specific components of the 4-vector separations of the two events in the two frames are given by (the reader is encouraged to explicitly evaluate the matrix multiplication to find Δx)

$$\Delta x' = \begin{pmatrix} \Delta \\ 0 \\ 0 \\ 0 \end{pmatrix}, \quad \Delta x = \Lambda \Delta x' = \begin{pmatrix} \gamma \Delta \\ (v/c) \gamma \Delta \\ 0 \\ 0 \end{pmatrix}, \quad (3.11.1)$$

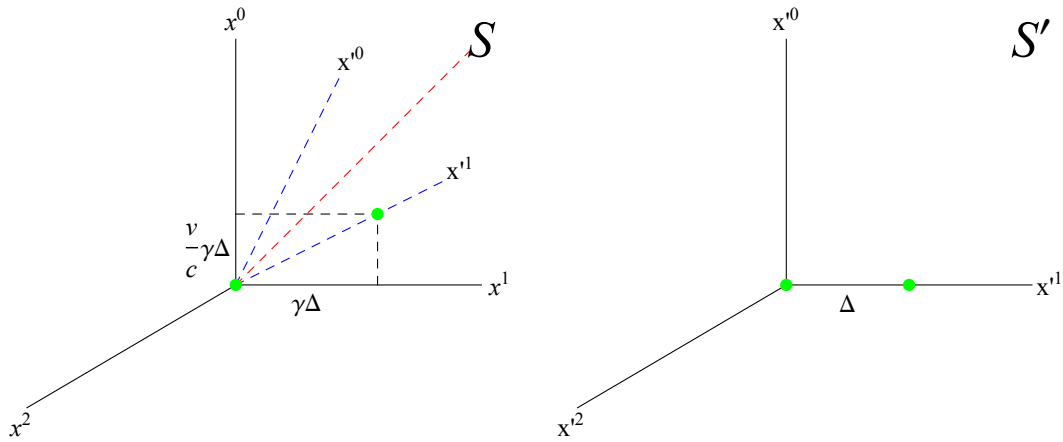
in agreement with the figure. The *invariant* separation squared is given by

$$(\Delta x')^2 = +\Delta^2 = (\Delta x)^2 = \Delta^2 \gamma^2 (1 - (v/c)^2) = +\Delta^2. \quad (3.11.2)$$

The factor of γ in the zeroth component of Δx is the usual time *dilation* factor, but note that the two events occur at *different* spatial points in the S frame. However, since both the zeroth and first components of the separation change between the two frames in just the correct fashion, the invariant separation (squared) is unchanged, *i.e.*, is invariant, under the Lorentz transformation.

Spacelike separation, \hat{e}^1 direction

Next consider two *simultaneous* events in the S' frame, one at the origin and one translated by Δ in the \hat{e}^1 direction as indicated in Fig. (3.12). Note that this corresponds to the usual (simultaneous)


 Figure 3.12: Spacelike separation (in x^1 direction) in S' frame.

definition of a length in the S' frame. Now the 4-vector separations in the two frames are

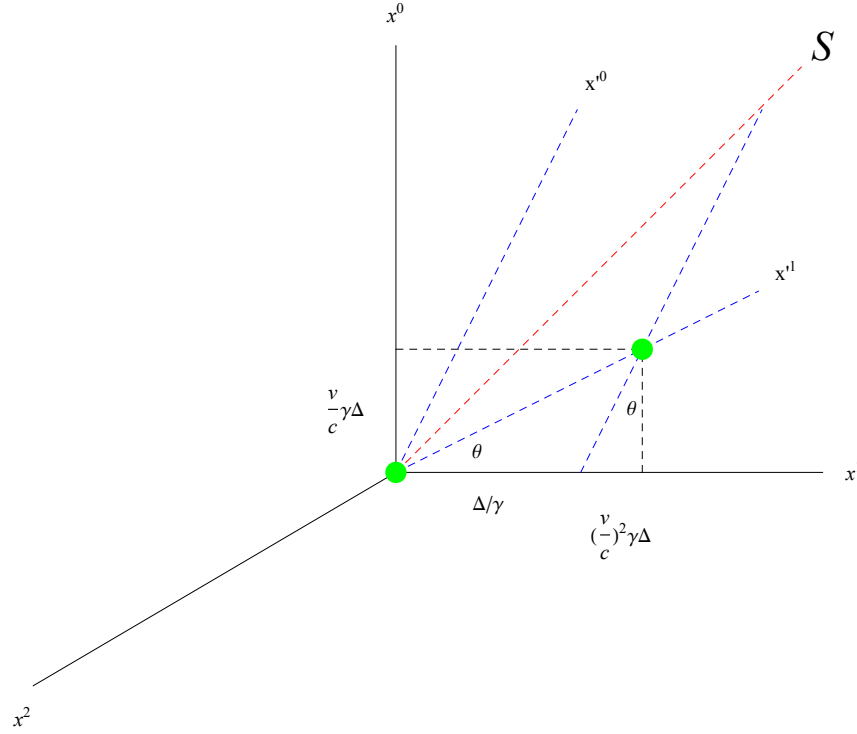
$$\Delta x' = \begin{pmatrix} 0 \\ \Delta \\ 0 \\ 0 \end{pmatrix}, \quad \Delta x = \Lambda \Delta x' = \begin{pmatrix} (v/c)\gamma\Delta \\ \gamma\Delta \\ 0 \\ 0 \end{pmatrix}, \quad (3.11.3)$$

while the invariant separation is

$$(\Delta x')^2 = -\Delta^2 = (\Delta x)^2 = \Delta^2 \gamma^2 ((v/c)^2 - 1) = -\Delta^2. \quad (3.11.4)$$

As expected for a spacelike separation the invariant has a negative value (in our West Coast metric). The astute reader may be concerned by the fact that the *spatial* component of the separation in the S frame is $\gamma\Delta$, and not the “expected” *contracted* length. All readers are encouraged to think about this issue, and, in particular, how to measure lengths in different reference frames. The essential point is that a length is defined by the spatial separation of two events that occur at the *same* time in the given frame. If we think of the green dots as defined by the ends of a “ Δ -stick”, at rest in the S' frame, we can measure the length of the same Δ -stick by determining the location of the right-hand end of the Δ -stick when the left-hand end is at the origin, *i.e.*, at $x^0 = 0$. This requires a little bit of trigonometry as indicated in Fig. (3.13). In particular, we can use the fact that the motion of the right-hand end of the Δ -stick in the S frame (recall the Δ -stick is at rest in the S' frame) will be along a line parallel to the x'^0 direction (as indicated in the figure). Next we use the two *similar* triangles (indicated by the identical angles θ , where $\tan \theta = v/c$) to determine the length of the lower side of the smaller triangle to be $(v/c)\gamma\Delta \times (v/c) = (v/c)^2\gamma\Delta$, as noted in the figure. To find the measured length of the Δ -stick in the S frame we need the location of the two ends measured simultaneously at $x^0 = 0$ (or any other shared x^0 value). Thus the length we want in Fig. (3.13) is the lower side of the larger triangle ($\gamma\Delta$) *minus* the side of the smaller triangle $((v/c)^2\gamma\Delta)$. Thus the *length* of the Δ -stick in the S frame is

$$\text{Length} = \gamma\Delta - (v/c)^2\gamma\Delta = (1 - (v/c)^2)\gamma\Delta = \Delta/\gamma, \quad (3.11.5)$$


 Figure 3.13: Spacelike separation (in x^1 direction) in S' frame, viewed in S frame.

which is just the expected *contracted* length.

Spacelike separation, \hat{e}^2 (or \hat{e}^3) direction

Next consider two simultaneous events in the S' frame, one at the origin and one now translated by Δ in the \hat{e}^2 (or \hat{e}^3) direction. Note that again this corresponds to the usual (simultaneous) definition of a length in the S' frame. Since the boost is not along the (spatial) direction of the separation, the separations in the two frames are identical as indicated in Fig. (3.14),

$$\Delta x' = \begin{pmatrix} 0 \\ 0 \\ \Delta \\ 0 \end{pmatrix}, \quad \Delta x = \Lambda \Delta x' = \begin{pmatrix} 0 \\ 0 \\ \Delta \\ 0 \end{pmatrix}. \quad (3.11.6)$$

Hence in this case the 4-vector separation is unchanged by the boost (as is its invariant square). This is an illustration of the fact that spatial separations *orthogonal* to the direction of a boost are unchanged by the boost.

Lightlike separation, \hat{e}^1 direction

Next we consider two events separated by a lightlike displacement in the S' frame, one at the origin and one translated by $\Delta/\sqrt{2}$ in *both* the \hat{e}^0 and \hat{e}^1 directions (*i.e.*, separated by a distance Δ along

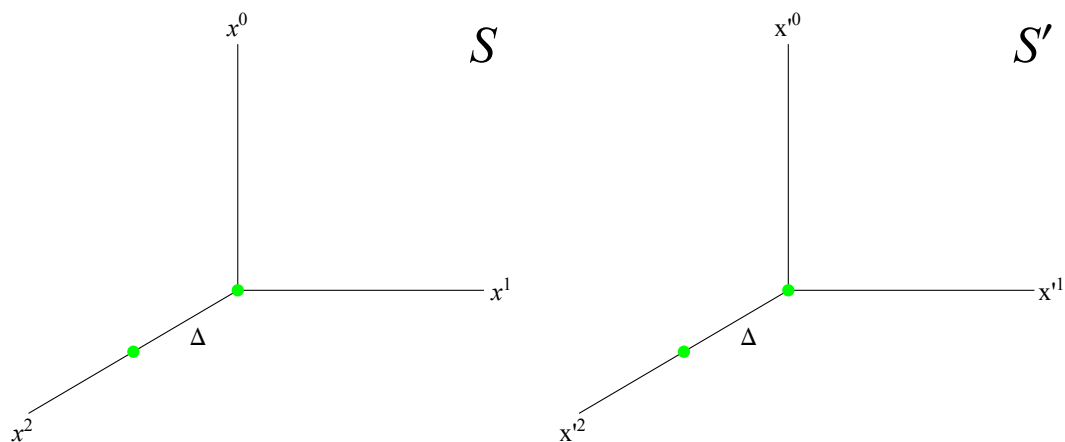


Figure 3.14: Spacelike separation (in x^2 direction) in S' frame (and S frame).

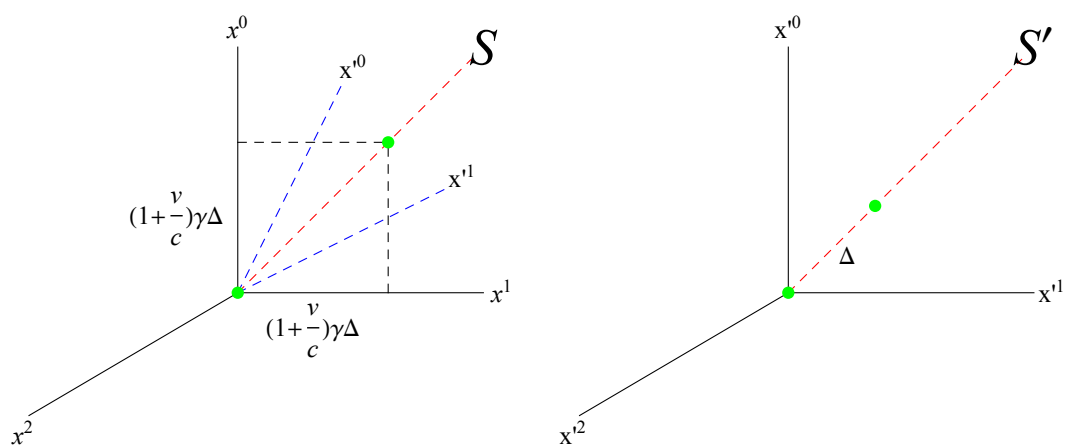


Figure 3.15: Lightlike separation (in x'^0 and x^1 direction) in S' frame.

the lightcone) as indicated in Fig. (3.15). Now the separations in the two frames are

$$\Delta x' = \begin{pmatrix} \Delta/\sqrt{2} \\ \Delta/\sqrt{2} \\ 0 \\ 0 \end{pmatrix}, \quad \Delta x = \Lambda \Delta x' = \begin{pmatrix} (1 + (v/c)) \gamma \Delta/\sqrt{2} \\ (1 + (v/c)) \gamma \Delta/\sqrt{2} \\ 0 \\ 0 \end{pmatrix}. \quad (3.11.7)$$

Thus, although there is a dilation by the factor $(1 + v/c)\gamma$ for both components, the separation remains lightlike (and along the light cone),

$$(\Delta x')^2 = 0 = (\Delta x)^2. \quad (3.11.8)$$

Lightlike separation, \hat{e}^2 direction

Finally we consider two events separated by a lightlike displacement in the S' frame, one at the origin and one translated by $\Delta/\sqrt{2}$ in *both* the \hat{e}^0 and \hat{e}^2 directions (*i.e.*, separated by a distance Δ along the lightcone, but not parallel to the boost). Now the separations in the two frames are

$$\Delta x' = \begin{pmatrix} \Delta/\sqrt{2} \\ 0 \\ \Delta/\sqrt{2} \\ 0 \end{pmatrix}, \quad \Delta x = \Lambda \Delta x' = \begin{pmatrix} \gamma \Delta/\sqrt{2} \\ (v/c) \gamma \Delta/\sqrt{2} \\ \Delta/\sqrt{2} \\ 0 \end{pmatrix}. \quad (3.11.9)$$

Thus in this case the impact of the boost is more complicated, dilating the zeroth component and changing the direction of the spatial component (*i.e.*, in the S frame the separation is no longer in just the $\hat{e}^0 - \hat{e}^2$ plane), but the resulting separation is still lightlike,

$$(\Delta x')^2 = 0 = (\Delta x)^2 = (\Delta^2/2) (\gamma^2(1 - (v/c)^2) - 1) = (\Delta^2/2) (1 - 1). \quad (3.11.10)$$

The reader is encouraged to invest the time necessary to ensure that the differences between these various examples are clear.

Chapter 4

Relativistic dynamics

We have seen in the previous lectures that our relativity postulates suggest that the most efficient (lazy but smart) approach to relativistic physics is in terms of 4-vectors, and that velocities never exceed c in magnitude. In this chapter we will see how this 4-vector approach works for dynamics, *i.e.*, for the interplay between motion and forces.

A particle subject to forces will undergo non-inertial motion. According to Newton, there is a simple (3-vector) relation between force and acceleration,

$$\vec{f} = m \vec{a}, \quad (4.0.1)$$

where acceleration is the second time derivative of position,

$$\vec{a} = \frac{d\vec{v}}{dt} = \frac{d^2\vec{x}}{dt^2}. \quad (4.0.2)$$

There is just one problem with these relations — they are *wrong*! Newtonian dynamics is a good *approximation* when velocities are very small compared to c , but outside of this regime the relation (4.0.1) is simply incorrect. In particular, these relations are inconsistent with our relativity postulates. To see this, it is sufficient to note that Newton's equations (4.0.1) and (4.0.2) predict that a particle subject to a constant force (and initially at rest) will acquire a velocity which can become arbitrarily large,

$$\vec{v}(t) = \int_0^t \frac{d\vec{v}}{dt'} dt' = \frac{\vec{f}}{m} t \rightarrow \infty \quad \text{as } t \rightarrow \infty. \quad (4.0.3)$$

This flatly contradicts the prediction of special relativity (and causality) that no signal can propagate faster than c . Our task is to understand how to formulate the dynamics of non-inertial particles in a manner which is consistent with our relativity postulates (and then verify that it matches observation, including in the non-relativistic regime).

4.1 Proper time

The result of solving for the dynamics of some object subject to known forces should be a prediction for its position as a function of time. But whose time? One can adopt a particular reference frame, and then ask to find the spacetime position of the object as a function of coordinate time t in

the chosen frame, $x^\mu(t)$, where, as always, $x^0 \equiv ct$. There is nothing wrong with this, but it is a *frame-dependent* description of the object's motion.

For many purposes, a more useful description of the object's motion is provided by using a choice of time which is directly associated with the object in a *frame-independent* manner. Simply imagine that the object carries with it its own (good) clock. Time as measured by a clock whose worldline is the same as the worldline of the object of interest is called the *proper time* of the object. To distinguish proper time from coordinate time in some inertial reference frame, proper time is usually denoted as τ (instead of t).

Imagine drawing ticks on the worldline of the object at equal intervals of proper time, as illustrated in Figure 4.1. In the limit of a very fine proper time spacing $\Delta\tau$, the invariant interval between neighboring ticks is constant, $s^2 = (c\Delta\tau)^2$. In the figure, note how the tick spacing, as measured by the coordinate time x^0 , varies depending on the instantaneous velocity of the particle. When the particle is nearly at rest in the chosen reference frame (*i.e.*, when the worldline is nearly vertical), then the proper time clock runs at nearly the same rate as coordinate time clocks, but when the particle is moving fast then its proper time clock runs more slowly than coordinate time clocks due to time dilation.

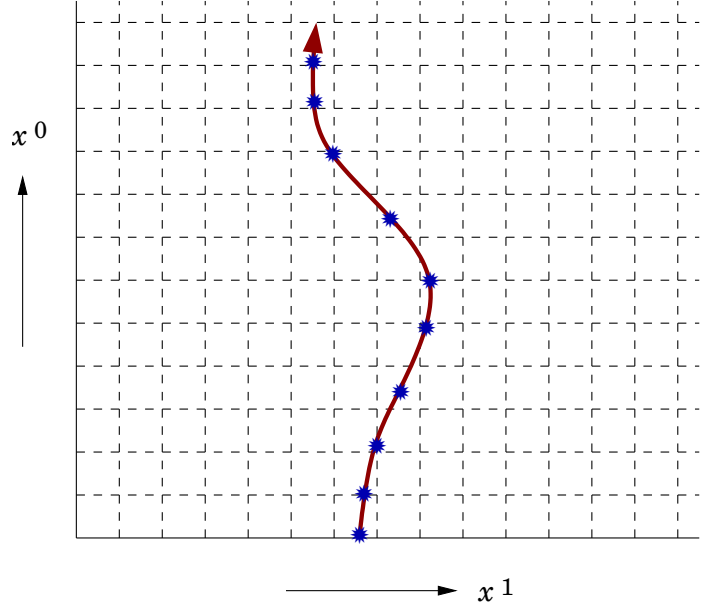


Figure 4.1: The worldline of a non-inertial particle, with tick marks at constant intervals of proper time.

4.2 4-velocity

Using the proper time to label points on the spacetime trajectory of a particle means that its spacetime position is some function of proper time, $x(\tau)$. The time component of x (in a chosen reference frame) gives the relation between coordinate time and proper time of events on the worldline,

$$ct = x^0(\tau). \quad (4.2.1)$$

The corresponding *four-velocity* of a particle is the derivative of its spacetime position with respect to proper time, (note that both u and x are 4-vectors)

$$u \equiv \frac{dx(\tau)}{d\tau}. \quad (4.2.2)$$

Since $x^0 = ct$, the time component of the 4-velocity gives the rate of change of coordinate time with respect to proper time,

$$u^0 = c \frac{dt}{d\tau}. \quad (4.2.3)$$

The spatial components of the 4-velocity give the rate of change of the spatial position with respect to proper time, $u^k = dx^k/d\tau$. This is *not* the same as the ordinary 3-velocity \vec{v} , which is the rate of

change of position with respect to coordinate time, $v^k = dx^k/dt$. But we can relate the two using calculus,

$$u^k = \frac{dx^k}{d\tau} = \frac{dt}{d\tau} \frac{dx^k}{dt} = \frac{u^0}{c} v^k. \quad (4.2.4)$$

From our discussion of time dilation, we already know that moving clocks run slower than clocks at rest in the chosen reference frame by a factor of γ . In other words, it must be the case that

$$\frac{u^0}{c} = \frac{dt}{d\tau} = \gamma = \left[1 - \frac{\vec{v}^2}{c^2} \right]^{-1/2}. \quad (4.2.5)$$

Combined with Eq. (4.2.4), this shows that the spatial components of the 4-velocity equal the 3-velocity times a factor of γ ,

$$u^k = \gamma v^k = \frac{v^k}{\sqrt{1 - \vec{v}^2/c^2}}. \quad (4.2.6)$$

We can now use Eqs. (4.2.5) and (4.2.6) to evaluate the square of the 4-velocity,

$$u^2 = (u^0)^2 - (u^k)^2 = \gamma^2 (c^2 - \vec{v}^2) = c^2. \quad (4.2.7)$$

So a 4-velocity vector *always* squares to $+c^2$, regardless of the value of the 3-velocity. (Recall that the plus sign here corresponds to our choice of metric; the East Coast metric yields $u^2 = -c^2$, but still a constant.)

Let's summarize what we've learned a bit more geometrically. The worldline $x(\tau)$ describes some trajectory through spacetime. At every event along this worldline, the four-velocity $u = dx/d\tau$ is a 4-vector which is *tangent* to the worldline. When one uses proper time to parametrize the worldline, the tangent vector u has a constant square, $u^2 = c^2$. So you can think of u/c as a tangent 4-vector which has unit “length” everywhere along the worldline. The fact that u^2 is positive (in our metric choice) shows that the 4-velocity is always a timelike vector. (Note that it is a timelike vector in both metrics, but with appropriately differing signs for the square.)

Having picked a specific reference frame in which to evaluate the components of the 4-velocity u , Eqs. (4.2.5) and (4.2.6) show that the components of u are completely determined by the ordinary 3-velocity \vec{v} , so the information contained in u is precisely the same as the information contained in \vec{v} . You might then ask “why bother with 4-velocity?” The answer is that the 4-velocity u is a more natural quantity to use — it has geometric meaning which is *independent* of any specific choice of reference frame. Moreover, the components u^μ of the 4-velocity transform linearly under a Lorentz boost in exactly the *same* fashion as any other 4-vector. [See Eq. (3.6.5)]. In contrast, under a Lorentz boost the components of the 3-velocity v transform in a somewhat messy fashion, but we can use the four-velocity to analyze this question.

4.3 Relativistic Addition of Velocities

Consider a point particle moving with 3-velocity v' in the x^1 direction in the S' frame (to match our previous convention) such that $(u')^T = (\gamma_{v'} c, \gamma_{v'} v', 0, 0)$, with $\gamma_{v'} = 1/\sqrt{1 - v'^2/c^2}$. Now view the motion of this particle in the S frame, which is defined such that, in the S frame, the S' frame

is moving in the $+x^1$ direction with velocity v_0 . Thus the boost between the two frames is (recall Eq. (3.5.7))

$$\Lambda(v_0) = \begin{pmatrix} \gamma_0 & \gamma_0(v_0/c) & 0 & 0 \\ \gamma_0(v_0/c) & \gamma_0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}, \gamma_0 = \frac{1}{\sqrt{1 - v_0^2/c^2}}, \quad (4.3.1)$$

so that

$$u = \Lambda(v_0)u' = \begin{pmatrix} c\gamma_0\gamma_{v'}(1 + v'v_0/c^2) \\ \gamma_0\gamma_{v'}(v' + v_0) \\ 0 \\ 0 \end{pmatrix}. \quad (4.3.2)$$

These results allow us to obtain the ordinary 3-velocity in the S frame from the 4-velocity via

$$v = |\vec{v}| = v^1 = c \frac{u^1}{u^0} = \frac{v_0 + v'}{1 + v_0 v' / c^2}, \quad (v^2 = v^3 = 0). \quad (4.3.3)$$

While the numerator is the familiar Galilean result for velocity addition (and reduces to this result for velocities small compared to c), the denominator is new to the *relativistic* addition of 3-velocities. (Note that the plus signs in the numerator and denominator correspond to the two 3-velocities being in the *same* direction. The signs would be negative for velocities in opposite directions.) This expression has the interesting feature, required by our relativistic Postulates, that, if either (or both) of the initial three-velocities approach c , v also approaches but *never* exceeds c . The reader is encouraged to complete this analysis and carry out the algebra necessary to obtain the following results,

$$(u')^2 = u^2 = c^2, \quad \gamma_0\gamma_{v'}(1 + v'v_0/c^2) = \gamma_v = \frac{1}{\sqrt{1 - v^2/c^2}}. \quad (4.3.4)$$

The first equation confirms the Lorentz invariance of the four-velocity squared, while the second equation confirms that, in the S frame, the four-velocity can be written in the standard form $u^T = (\gamma_v c, \gamma_v v, 0, 0)$.

4.4 4-momentum

To discuss momentum we should first be explicit concerning what we mean by the symbol m . The *rest mass* m of any object is the mass of the object as measured in its rest frame. The *4-momentum* of a particle (or any other object) with rest mass m is defined to be m times the object's 4-velocity,

$$p = m u. \quad (4.4.1)$$

For systems of interacting particles, this is the quantity to which conservation of momentum will apply. Spatial momentum components (in a given reference frame) are just the spatial components of the 4-momentum. The definition of 3-momentum which you learned in introductory physics, $\vec{p} = m \vec{v}$, is, at best, a non-relativistic approximation. This is important, so let us repeat,

$$\vec{p} \neq m \vec{v}. \quad (4.4.2)$$

From now on do *not* think of momentum as mass times 3-velocity. Rather, think 4-dimensionally with momentum as mass times 4-velocity.¹

If the spatial components of the 4-momentum are the (properly defined) spatial momentum, what is the time component p^0 ? There is only one possible answer — it must be related to energy.² In fact, the total energy E of an object equals the time component of its four-momentum times c , or

$$p^0 = E/c. \quad (4.4.3)$$

Using the relation (4.4.1) between 4-momentum and 4-velocity, plus the result (4.2.5) for u^0 , allows one to express the the total energy E of an object in terms of its rest mass and its velocity,

$$E = cp^0 = mcu^0 = mc^2\gamma = \frac{mc^2}{\sqrt{1 - \vec{v}^2/c^2}} = mc^2 \cosh y. \quad (4.4.4)$$

In other words, the relativistic gamma factor of any object is equal to the ratio of its total energy to its rest energy (recall Eq. (3.5.8)),

$$\gamma = \frac{E}{mc^2} = \cosh y. \quad (4.4.5)$$

When the object is at rest, its kinetic energy (or energy due to motion) vanishes, but its *rest energy*, given by Einstein's famous expression mc^2 , remains. If the object is moving slowly (compared to c), then it is appropriate to expand the relativistic energy (4.4.4) in powers of \vec{v}^2/c^2 . This gives

$$E = mc^2 + \frac{1}{2}m\vec{v}^2 + \dots. \quad (4.4.6)$$

In other words, for velocities small compared to c , the total energy E equals the rest energy mc^2 plus the usual non-relativistic kinetic energy, $\frac{1}{2}m\vec{v}^2$, up to higher order corrections which, relative to the non-relativistic kinetic energy, are suppressed by additional powers of \vec{v}^2/c^2 . We can, of course, define the *relativistic* kinetic energy via $K = E - mc^2 = mc^2(\gamma - 1)$, which reduces to the non-relativistic form $\frac{1}{2}m\vec{v}^2$ for $\vec{v}^2/c^2 \ll 1$.

The corresponding spatial component of the four-momentum is then (recall Eqs. (4.4.2) and (3.5.8))

$$\vec{p} = m\vec{v}\gamma = \frac{m\vec{v}}{\sqrt{1 - \vec{v}^2/c^2}} = mc\hat{v} \sinh y, \quad (4.4.7)$$

(where \hat{v} is the spatial unit vector in the direction of \vec{v}) which, as advertised, approaches the non-relativistic definition for $v \ll c$, $\gamma \rightarrow 1$ and $\sinh y \rightarrow v/c$.

We saw above that (with our choice of metric) 4-velocities square to c^2 . Since 4-momentum is just mass times 4-velocity, the 4-momentum of any object with (rest) mass m satisfies

$$p^2 = m^2c^2 (\cosh^2 y - \sinh^2 y) = m^2c^2. \quad (4.4.8)$$

¹Many introductory relativity books introduce a velocity-dependent mass $m(v) \equiv m\gamma$, in order to write $\vec{p} = m(v)\vec{v}$, and thereby avoid ever introducing four-velocity (or any other 4-vector). This is pedagogically terrible and offers no benefit whatsoever. If you have previously seen this use of a velocity-dependent mass, erase it from your memory banks!

²To see why, recall from mechanics (quantum or classical) that translation invariance in space is related to the existence of conserved spatial momentum, and translation invariance in time is related to the existence of a conserved energy. We will discuss this in more detail later. Since Lorentz transformations mix space and time, it should be no surprise that the four-momentum, which transforms linearly under Lorentz transformations, must characterize both the energy *and* the spatial momentum.

Since $p^2 = (p^0)^2 - (p^k)^2$, and $p^0 = E/c$, this may be rewritten (in any chosen inertial reference frame) as

$$E^2 = c^2 \vec{p}^2 + (mc^2)^2. \quad (4.4.9)$$

So, if you know the spatial momentum \vec{p} and mass m of some object, you can directly compute its energy E without first having to evaluate the object's velocity. Note, in particular, that Eq. (4.4.9) is true for *either* choice of metric!

But what if you want to find the ordinary 3-velocity? Return to the relation $u^k = \gamma v^k$ [Eq. (4.2.6)] between 3-velocity and 4-velocity, and multiply both sides by m to rewrite this result in terms of four-momentum. Since spatial momentum $p^k = m u^k$, and total energy $E = mc^2 \gamma$, we have $p^k = (E/c^2) v^k$ or

$$v^k = \frac{p^k}{E/c^2}, \quad |\vec{v}| = c \tanh y. \quad (4.4.10)$$

Three-velocity is *not* equal to momentum divided by mass. Rather, the ordinary 3-velocity equals the spatial momentum divided by the total energy (over c^2). And its magnitude *never* exceeds c .

4.5 4-force

In the absence of any forces, the momentum of an object remains constant. In the presence of forces, an object's momentum will change. In fact, force is just the time rate of change of momentum. But what time and what momentum? Newtonian (non-relativistic) dynamics says that $d\vec{p}/dt = \vec{F}$ along with $d\vec{x}/dt = \vec{p}/m$, where \vec{p} is 3-momentum and t is coordinate time. This is wrong — inconsistent with our relativity postulates. A frame-independent formulation of dynamics must involve quantities which have intrinsic frame-independent meaning — such as 4-momentum and proper time. The appropriate generalization of Newtonian dynamics which is consistent with our relativity postulates is

$$\frac{dx}{d\tau} = \frac{p}{m}, \quad (4.5.1a)$$

$$\frac{dp}{d\tau} = f. \quad (4.5.1b)$$

Eq. (4.5.1a) is just the definition (4.2.2) of 4-velocity rewritten in terms of 4-momentum, while Eq. (4.5.1b) is the *definition* of force as a four-vector. The only difference in these equations, relative to Newtonian dynamics, is the replacement of 3-vectors by 4-vectors and coordinate time by proper time.

Equations (4.5.1) are written in a form which emphasizes the role of momentum. If you prefer, you can work with 4-velocity instead of 4-momentum and rewrite these equations as $dx/d\tau = u$ and $du/d\tau = f/m$. Defining the *4-acceleration* $a \equiv du/d\tau = d^2x/d\tau^2$, this last equation is just $f = m a$. This is the relativistic generalization of Newton's $\vec{f} = m \vec{a}$, with force and acceleration now defined as spacetime vectors.³

In non-relativistic dynamics, if you know the initial position and velocity of a particle, and you know the force $\vec{f}(t)$ which subsequently acts on the particle, you can integrate Newton's equations to find

³Eq. (4.5.1b) is equivalent to $f = m a$ *provided* the mass m of the object is constant. For problems involving objects whose mass can change, such as a rocket which loses mass as it burns fuel, these two equations are not equivalent and one must use the more fundamental $dp/d\tau = f$.

the trajectory $\vec{x}(t)$ of the particle. Initial conditions plus a three-vector $\vec{f}(t)$ completely determine the resulting motion. To integrate the relativistic equations (4.5.1), you need initial conditions plus a four-vector force $f(\tau)$. This would appear to be more information (four components instead of three), and yet relativistic dynamics must reduce to non-relativistic dynamics when velocities are small compared to c .

The resolution of this apparent puzzle is that the four-force *cannot* be a completely arbitrary four-vector. We already know that, for any object with mass m , its four-momentum must satisfy $p^2 = (mc)^2$ [Eq. (4.4.8)]. Take the derivative of both sides with respect to proper time. The right hand side is constant in time (provided that the object in question is some stable entity with a fixed rest mass), so its proper time derivative vanishes. The derivative of the left hand side gives twice the dot product of p with f , and hence the 4-force must always be *orthogonal* to the 4-momentum,

$$p \cdot f = 0. \quad (4.5.2)$$

Written out in components, this says that $p^0 f^0 = p^i f^i$, or

$$f^0 = \frac{p^i f^i}{p^0} = \frac{\vec{v}}{c} \cdot \vec{f}, \quad (4.5.3)$$

showing that the time component of the force is just a particular linear combination of the spatial components, *i.e.*, the four components of the force cannot vary freely, but rather must satisfy this constraint.

4.6 Constant acceleration

Let us put this formalism into action by examining the case of motion under the influence of a constant force. But what is a “constant” force? We have just seen that the 4-force must always be orthogonal to the 4-momentum. So it is *impossible* for the 4-force $f(\tau)$ to be a fixed four-vector in an arbitrary frame, independent of τ . However, it is possible for the force to be constant when viewed in a frame which is *instantaneously* co-moving with the accelerating object.

Suppose a particle begins at the spacetime origin with vanishing 3-velocity (or 3-momentum) at proper time $\tau = 0$, and a (3-)force of magnitude F , pointing in the x^1 direction, acts on the particle. Hence the components of the initial spacetime position, four-velocity, and four-force are $x_0^\mu = (0, 0, 0, 0)$, $u_0^\mu = (c, 0, 0, 0)$, and $f_0^\mu = (0, F, 0, 0)$, respectively. The 4-velocity at later times may be written as some time-dependent Lorentz boost acting on the initial 4-velocity,

$$u(\tau) = \Lambda_{\text{boost}}(\tau) u_0. \quad (4.6.1)$$

The condition that the force is constant (f_0) in a co-moving frame amounts to the statement that the *same* Lorentz boost (as in Eq. 4.6.1)) relates the 4-force at any time τ (in the frame where the velocity is $u(\tau)$) to the initial force,

$$f(\tau) = \Lambda_{\text{boost}}(\tau) f_0. \quad (4.6.2)$$

At all times, $u^2 = c^2$ (because u is a 4-velocity), and $f^2 = -F^2$, because the magnitude of the force is assumed to be constant.

Since the initial force points in the x^1 direction, the particle will acquire some velocity in this direction, but the x^2 and x^3 components of the velocity will always remain zero. Hence the boost

$\Lambda_{\text{boost}}(\tau)$ will always be some boost in the x^1 direction, and the force $f(\tau)$ will likewise always have vanishing x^2 and x^3 components. In other words, the 4-velocity and 4-force will have the form

$$u^\mu(\tau) = (u^0(\tau), u^1(\tau), 0, 0), \quad f^\mu(\tau) = (f^0(\tau), f^1(\tau), 0, 0), \quad (4.6.3)$$

with $u^0(0) = c$, $u^1(0) = 0$ and $f^0(0) = 0$, $f^1(0) = F$. From (4.5.2) the dot product $f \cdot u = f^0 u^0 - f^1 u^1$ must vanish, implying that $f^0/f^1 = u^1/u^0$. So the components of the 4-force must be given by

$$f^\mu(\tau) = \frac{F}{c} (u^1(\tau), u^0(\tau), 0, 0). \quad (4.6.4)$$

(Do you see why? This is the *only* form for which $f \cdot u = 0$ and $f^2 = -F^2$.)

Now we want to solve $m du/d\tau = f(\tau)$. Writing out the components explicitly (and dividing by m) gives

$$\frac{du^0(\tau)}{d\tau} = \frac{F}{mc} u^1(\tau), \quad \frac{du^1(\tau)}{d\tau} = \frac{F}{mc} u^0(\tau). \quad (4.6.5)$$

This is easy to solve if you remember some basic mathematical physics (from Chapter 1) - $\frac{d}{dz} \sinh z = \cosh z$ and $\frac{d}{dz} \cosh z = \sinh z$ (recall Eq. (1.3.3)). To satisfy Eq. (4.6.5), and our initial conditions, we simply choose

$$u^0(\tau) = c \cosh \frac{F\tau}{mc}, \quad u^1(\tau) = c \sinh \frac{F\tau}{mc}. \quad (4.6.6)$$

The ordinary velocity is given by $v^k = u^k (c/u^0)$ [Eq. (4.2.4)], so the speed of this particle subject to a constant force is

$$v(\tau) = c \tanh \frac{F\tau}{mc}. \quad (4.6.7)$$

Since $\tanh z \sim z$ for small values of the argument (recall Eq. (1.3.10)), the speed grows linearly with time at early times, $v(\tau) \sim (F/m)\tau$. This is precisely the expected *non-relativistic* behavior. But this approximation is only valid when $\tau \ll mc/F$ and the speed is small compared to c . The argument of the \tanh becomes large compared to unity when $\tau \gg mc/F$, and $\tanh z \rightarrow 1$ as $z \rightarrow \infty$. So the speed of the accelerating particle asymptotically approaches, but never reaches, the speed of light. In fact, we see from our previous definitions of the 4-momentum in terms of the rapidity y that it is the rapidity that grows linearly with τ in the case of “constant acceleration”,

$$\tanh y = \frac{v(\tau)}{c} = \tanh \frac{F\tau}{mc} \Rightarrow y = \frac{F\tau}{mc}. \quad (4.6.8)$$

At this point, we have determined how the velocity of the particle grows with time, but we need to integrate $dx/d\tau = u$ to find its spacetime position. Due to the properties of the hyperbolic functions the integrals are elementary,

$$x^0(\tau) = \int_0^\tau d\tau' u^0(\tau') = c \int_0^\tau d\tau' \cosh \frac{F\tau'}{mc} = \frac{mc^2}{F} \sinh \frac{F\tau}{mc}, \quad (4.6.9a)$$

$$x^1(\tau) = \int_0^\tau d\tau' u^1(\tau') = c \int_0^\tau d\tau' \sinh \frac{F\tau'}{mc} = \frac{mc^2}{F} \left[\cosh \frac{F\tau}{mc} - 1 \right]. \quad (4.6.9b)$$

Note that the hyperbolic sines and cosines grow exponentially for large arguments, $\sinh z \sim \cosh z \sim \frac{1}{2} e^z$ when $z \gg 1$. Hence, when $\tau \gg mc/F$ the coordinates $x^0(\tau)$ and $x^1(\tau)$ both grow like $e^{F\tau/mc}$ with increasing proper time. But the accelerating particle becomes ever more time-dilated; the rate of change of proper time with respect to coordinate time, $d\tau/dt = c/u^0 = 1/\cosh \frac{F\tau}{mc}$, behaves as $2e^{-F\tau/mc} \sim mc/(Ft)$.

4.7 Plane waves

Next we want to discuss the very important question of how *waves* are described in relativistic notation. Consider some wave (any type of wave) with wave-vector \vec{k} and frequency ω (3-vector notation), as measured in some inertial frame. The amplitude of the wave is described by a complex exponential, $\mathcal{A} e^{i\vec{k}\cdot\vec{x}-i\omega t}$, with the usual understanding that it is the *real* part of this function which describes the physical amplitude. Such a wave has a wavelength $\lambda = 2\pi/|\vec{k}|$ and planar wave-fronts orthogonal to the wave-vector that move at speed $v = \omega/|\vec{k}|$ in the direction of \vec{k} .

As mentioned earlier (Eq. (3.6.12)), it is natural to combine ω and \vec{k} into a spacetime wave-vector (*i.e.*, 4-vector) k with components

$$k^\mu = (\omega/c, k^1, k^2, k^3), \quad (4.7.1)$$

so that $\omega = ck^0$ and we have simply $e^{i\vec{k}\cdot\vec{x}-i\omega t} = e^{-ik\cdot x}$ (with our choice of metric).

The virtue of this formulation is that it is *frame-independent*. The spacetime position x and wave-vector k are geometric entities which you should think of as existing independent of any particular choice of coordinates. The value of the amplitude, $\mathcal{A} e^{-ik\cdot x}$, depends on the position x and the wave-vector k , but one may use whatever reference frame is most convenient to evaluate the dot product of these 4-vectors since $k\cdot x$ is the *same* in all frames. (This fact gives us the opportunity to be both lazy and smart, and is the real *power* of the 4-vector notation.)

Just as surfaces of simultaneity are observer-dependent, so is the frequency of a wave. After all, measuring the frequency of a wave involves counting the number of wave crests which pass some detector (or observer) in a given length of time. The time component of the wave-vector gives (by construction) the frequency of the wave as measured by observers who are at rest in the frame in which the components k^μ are defined. Such observers have 4-velocities whose components are just $(c, 0, 0, 0)$ (in that frame, *i.e.*, in their rest frame). Consequently, for these observers the frequency of the wave may be written as a dot product of the observer's 4-velocity and the wave-vector,

$$\omega_{\text{obs}} = u_{\text{obs}} \cdot k. \quad (4.7.2)$$

This expression is now written in a completely general fashion that is observer-*dependent* but frame-*independent*. That is, the expression (4.7.2) depends explicitly on the observer's 4-velocity u , but is independent of the frame used to evaluate the dot product between u and k (*i.e.*, the dot product must be the same in every frame). Therefore, the frequency which is measured by *any* observer will be given by the dot product of the observer's 4-velocity u and the wave-vector k . The dot product can be evaluated in *any* convenient (lazy but smart) frame, but, of course, u and k must *both* be evaluated in that *same* frame.

As we will discuss a bit more below, it should be no surprise that this approach is particularly useful for the discussion of light waves, where $v = c$, $\omega = c|\vec{k}|$ and

$$k_{\text{light}}^\mu = \frac{\omega}{c} (1, \hat{k}), \quad k_{\text{light}}^2 = 0. \quad (4.7.3)$$

As a simple first application of Eqs. (4.7.2) and (4.7.3) we can derive the basic form of the relativistic Doppler shift of light. Consider a source of plane wave light, which emits light with frequency $\nu_0 = \omega_0/2\pi$ as measured in the rest frame of the source (note that ω has units of radians per second, while ν is measured in cycles per second - NOT the same units!). Further we take the light waves

to be moving in the x^1 direction ($\hat{k} = \hat{e}_1$). Consider an observer moving *away* from the source (the *receding* case) also in the x^1 direction with velocity v , as measured in the source rest frame. Clearly the lazy but smart choice of frame is the source frame, where we have

$$k^\mu = \frac{\omega_0}{c} (1, 1, 0, 0), \quad u_{\text{obs}} = \frac{c}{\sqrt{1 - v^2/c^2}} (1, v/c, 0, 0), \quad (4.7.4a)$$

$$\frac{\nu_{\text{obs}}}{\nu_0} = \frac{\omega_{\text{obs}}}{\omega_0} = \frac{u_{\text{obs}} \cdot k}{\omega_0} = \sqrt{\frac{1 - v/c}{1 + v/c}} \leq 1 \text{ [receding]}. \quad (4.7.4b)$$

In this case (see Eq. (3.3.7) in Kogut), with the source and observer *receding* from each other, the observer sees a *smaller* frequency than the source emits (the light is red-shifted, $\nu_{\text{obs}}/\nu_0 = \sqrt{(1 - v/c)/(1 + v/c)} \leq 1$). For an observer *approaching* a source, we simply change the sign of v in Eq. (4.7.4) and the light is blue-shifted to a larger frequency,⁴

$$\frac{\nu_{\text{obs}}}{\nu_0} = \frac{\omega_{\text{obs}}}{\omega_0} = \frac{u_{\text{obs}} \cdot k}{\omega_0} = \sqrt{\frac{1 + v/c}{1 - v/c}} \geq 1 \text{ [approaching]}. \quad (4.7.5)$$

A more sophisticated application of Eq. (4.7.2), demonstrating the value of writing physical quantities in frame independent form, is illustrated in Figure 4.2.⁵ Mounted on the inner surface of a centrifuge, which is rotating at angular frequency Ω , is an emitter of light at one point, and a receiver at a different point. Let ϕ be the angle between emitter and receiver, relative to the center of the centrifuge, as measured in the inertial lab frame. The (inner) radius of the centrifuge is R . The frequency of the light as measured by an observer who is instantaneously at rest relative to the emitter is ν_e . The frequency of the light as measured by an observer who is instantaneously at rest relative to the receiver is ν_r . What is the fractional difference $(\nu_r - \nu_e)/\nu_e$? How does this frequency shift depend on the angle ϕ and the rotation frequency Ω ?

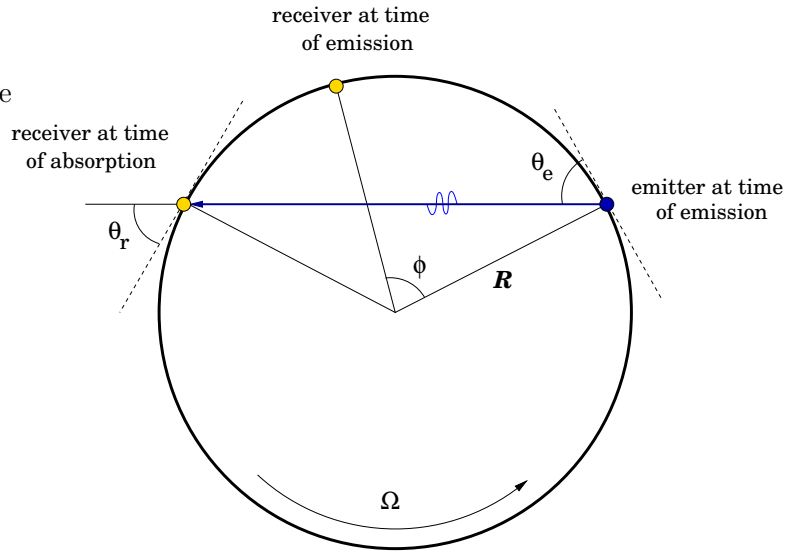


Figure 4.2: Inside a rotating centrifuge, light is emitted at one point and later received at another point. Is there a Doppler shift between the frequencies of emission and reception?

One approach for solving this problem would involve explicitly constructing the Lorentz transformations which relate the lab frame to the instantaneous rest frames of the emitter and receiver, and then combining these two transformations to determine the net transformation which directly connects

⁴A familiar application of this result in the context of astronomy and cosmology is the redshift z , defined as the fractional change in the frequency (or wavelength). For a radially expanding (receding) universe (with radial velocity v), this leads to the relation $1 + z = \sqrt{(1 + v/c)/(1 - v/c)}$.

⁵This discussion is an adaptation of an example in *Gravitation* by Misner, Thorne and Wheeler.

emitter and receiver. Given the three-dimensional geometry involved, this is rather involved (and would not correspond to the lazy but smart approach).

A much better approach is to choose a convenient single frame, namely the lab frame, in which to evaluate the components of the four-vectors appearing in the frame-independent expression (4.7.2) for the frequency. We need to compute

$$\frac{\nu_r}{\nu_e} = \frac{u_r \cdot k}{u_e \cdot k} = \frac{u_r^0 k^0 - \vec{u}_r \cdot \vec{k}}{u_e^0 k^0 - \vec{u}_e \cdot \vec{k}}. \quad (4.7.6)$$

Here u_e is the four-velocity of the emitter at the moment it emits light, and u_r is the four-velocity of the receiver at the moment when it receives the light.

If θ_e denotes the angle between the spatial wavevector and the direction of motion of the emitter (at the time of emission), and θ_r denotes the angle between \vec{k} and receiver's direction (at the time of reception) (all as indicated in Figure 4.2), then we can express the spatial dot products in terms of cosines of these angles,

$$\frac{\nu_r}{\nu_e} = \frac{u_r^0 k^0 - |\vec{u}_r| |\vec{k}| \cos \theta_r}{u_e^0 k^0 - |\vec{u}_e| |\vec{k}| \cos \theta_e}. \quad (4.7.7)$$

The speed of the inner surface of the centrifuge is constant, $v = \Omega R$, and hence the *speeds* of the emitter and receiver, as measured in the lab frame, are identical — even though their velocity vectors are different. The time component of a 4-velocity, $u^0/c = (1 - v^2/c^2)^{-1/2}$, only depends on the *magnitude* of the velocity \vec{v} , and hence $u_r^0 = u_e^0$. The equality of the emitter and receiver speeds also implies that the magnitudes of the spatial parts of the 4-velocities coincide, $|\vec{u}_r| = |\vec{u}_e|$. So using expression (4.7.7) for the frequency ratio, the only remaining question is how does θ_r compare to θ_e ?

This just involves ordinary geometry. Looking at the figure, notice that θ_e and θ_r are the angles between the path of the light, which is a chord of the circle, and tangents to the circle at the endpoints of the chord. But the angle a chord makes with these tangents is the same at either end, implying that $\theta_e = \theta_r$. And this means $\nu_r = \nu_e$ — there is *no* Doppler shift no matter how fast the centrifuge rotates (or what the values of ϕ , θ_e or R are)! (Try obtaining this result directly using boosts and more complex trigonometry.)

4.8 Electromagnetism

As noted earlier, it should be no surprise that the technology we are developing is especially useful for “objects” that travel at the speed of light, such as light itself. Unfortunately we do not have time here for an extensive explorations of the relativistic aspects of electromagnetism, which will be left for other classes. But one aspect, how to represent the Lorentz force in the framework we have been discussing, is natural to describe here.

As we have seen above, generalizations from non-relativistic to relativistic dynamics are mostly a matter of replacing 3-vectors by 4-vectors (and coordinate time by proper time). But what about electric and magnetic fields? Both are (apparently) 3-vectors, and there is no sensible way to turn them into 4-vectors. Recall that \vec{E} is an ordinary “polar” 3-vector, which changes sign under reflection, while \vec{B} is an “axial” 3-vector, which does *not* change sign under reflection. It turns out that what *is* sensible (and natural) is to package the components of \vec{E} and \vec{B} together into a 4×4 matrix

(a 4-tensor) called the *field strength tensor*, whose components are

$$\|F^\mu{}_\nu\| = \begin{pmatrix} 0 & E_x & E_y & E_z \\ E_x & 0 & cB_z & -cB_y \\ E_y & -cB_z & 0 & cB_x \\ E_z & cB_y & -cB_x & 0 \end{pmatrix}. \quad (4.8.1)$$

It is important to realize that this explicit form for the field strength tensor (*i.e.*, the explicit signs and factors of c) corresponds to our specific choices of units (here SI, with E measured in Newtons/Coulomb and B in Tesla) and the choice to write down the form with one superscript and one subscript.⁶ Different choices for these details yield slightly different expressions.

With this repackaging of electric and magnetic fields, the Lorentz force (as a 4-vector) has a remarkably simple form,

$$f_{\text{Lorentz}}^\mu = \frac{q}{c} F^\mu{}_\nu u^\nu. \quad (4.8.2)$$

Verifying that this 4-force leads to exactly the same rate of change of energy and momentum as does the traditional form of writing the Lorentz force, $\vec{f} = q(\vec{E} + \vec{v} \times \vec{B})$, is an instructive and recommended exercise.

4.9 Scattering

When objects (elementary particles, molecules, automobiles, ...) collide, the results of the collision can differ markedly from the initial objects. Composite objects can fall apart or change form (leading to large insurance premiums). Interestingly, dramatic changes during collisions can also occur for elementary particles. In fact, studying the collisions of elementary particles is a primary method used to investigate fundamental interactions and explains the existence of large energy particle colliders like the LHC. These machines are really just (large!) microscopes with *very fine* resolution and with the capability to produce particles, like the Higgs boson, that we do not observe in everyday life.

A complete description of what emerges from a collision (or ‘scattering event’) depends on microscopic details of the interaction between the incident objects. But certain general principles constrain the possibilities, most importantly, the conservation of energy and momentum. As discussed in section 4.4, the total energy E and spatial momentum \vec{p} of any object may be combined to form the 4-momentum $p^\mu = (E/c, \vec{p})$. Consequently, energy and momentum conservation may be rephrased as the conservation of 4-momentum: in the absence of any external forces, the total 4-momentum of any system cannot change,

$$\frac{d}{dt} p_{\text{tot}}(t) = 0. \quad (4.9.1)$$

⁶To appreciate the field strength tensor in its full 4-glory, we recognize that it is defined in terms of a 4-vector potential, $A^\mu = (\phi, c\vec{A})$, where ϕ is the usual electric scalar potential (in volts) and \vec{A} is the usual 3-vector potential, which you may have seen in previous courses. Then the field strength tensor is the 4-curl, $F^\mu{}_\nu = \partial^\mu A_\nu - \partial_\nu A^\mu$, with $\partial^\mu = \partial/\partial x_\mu = (\partial_t/c, -\vec{\nabla})$ and $\partial_\mu = \partial/\partial x^\mu = (\partial_t/c, \vec{\nabla})$ (in our metric). Finally we need the 3-vector definitions $\vec{E} = -\partial\vec{A}/\partial t - \vec{\nabla}\phi$ and $\vec{B} = \vec{\nabla} \times \vec{A}$, where this cross product explains why \vec{B} is an axial or pseudo-vector. The interested reader is encouraged to work out the terms in Eq. (4.8.1) using these definitions. Note that with both indices either up or down the resulting tensor is fully anti-symmetric instead of the mixed symmetry (metric independent) form in Eq. (4.8.1).

In a scattering process two or more objects, initially far apart, come together and interact in some manner (which may be very complicated), thereby producing some number of objects that subsequently fly apart. When the incoming objects are far apart and not yet interacting, the total 4-momentum is just the sum of the 4-momentum of each object,

$$p_{\text{in}} = \sum_{a=1}^{N_{\text{in}}} p_a, \quad (4.9.2)$$

where N_{in} is the number of incoming objects (and the index a labels particles, not spacetime directions). Similarly, when the outgoing objects are arbitrarily well separated, they are no longer interacting and the total 4-momentum is (again) the sum of the individual 4-momenta of all of the outgoing objects,

$$p_{\text{out}} = \sum_{b=1}^{N_{\text{out}}} p_b. \quad (4.9.3)$$

Hence, for any scattering processes, conservation of energy and momentum implies that the total incident 4-momentum equals the total outgoing 4-momentum (independent of the values of N_{in} and N_{out}),

$$p_{\text{in}} = p_{\text{out}}. \quad (4.9.4)$$

As with any 4-vector equation, one may choose to write out the components of this equation in whatever reference frame is most convenient (as long as we use the *same* frame for both p_{in} and p_{out}). For analyzing scattering processes, sometimes it is natural to work in the rest frame of one of the initial objects (the ‘target’); this is commonly called the *lab frame* and experiments of this variety are called “fixed target” experiments (the frame of the actual lab is the target frame). Alternatively, one may choose to work in the reference frame in which the total spatial momentum vanishes. In this frame, commonly called the *CM frame*,⁷ the components of the total 4-momentum are

$$p_{\text{CM}}^\mu = (E_{\text{CM}}/c, 0, 0, 0), \quad (4.9.5)$$

where E_{CM} is the total energy of the system in the CM frame. In the early days of particle physics, where only a single beam of accelerated particles was available, fixed target experiments were the norm and the CM frame was an intellectual construct. For the kinematic reasons we are about to discuss, it became clear that moving the actual lab to the CM frame would provide an enormous increase in efficiency for particle production. As a result, we now live in the era of particle *colliders*, where two beams of accelerated particles, moving in opposite directions, are caused to collide essentially head-on.

As an application of these ideas, consider first the scattering of protons of energy $E_{\text{in}} = 1 \text{ TeV}$ on protons at rest (in ordinary matter). The proton rest energy $m_p c^2$ is a bit less than 1 GeV. Using Eq. (4.4.5), one sees that a proton with 1 TeV energy is ultrarelativistic, $\gamma = E_{\text{in}}/(m_p c^2) \approx 10^3$. When an ultrarelativistic proton strikes a target proton at rest, both protons can be disrupted and new particles may be created. Schematically,

$$p + p \rightarrow X,$$

⁷‘CM’ means ‘center of mass’, but this historical name is really quite inappropriate for relativistic systems, which may include massless particles that carry momentum but have no rest mass. The widely used ‘CM’ label should always be understood as referring to the zero (spatial) momentum frame.

where X stands for one or more outgoing particles. What is the largest mass of a particle which could be produced in such a collision?

The total energy of the incident particles (in the rest frame of the target) is $E_{\text{tot}} = E_{\text{in}} + m_p c^2 \approx 1.001 \text{ TeV}$. If all of this energy is converted into the rest energy of one or more outgoing particles, then you might conclude that these collisions could produce particles with mass up to $E_{\text{tot}}/c^2 \approx 10^3 m_p$. This would be consistent with conservation of energy. But this is *wrong*, as it completely ignores conservation of 3-momentum. In the rest frame of the target, the total spatial momentum \vec{p}_{tot} is non-zero (and equal to the momentum \vec{p}_{in} of the projectile proton). If there is a single outgoing particle X , it cannot be produced at rest — it must emerge from the collision with a non-zero spatial momentum equal to \vec{p}_{tot} . That means its energy will be greater than its rest energy.

To determine the largest mass of a particle which can be produced in this collision, one must simultaneously take into account conservation of both energy and momentum. That is, one must satisfy the 4-vector conservation equation (4.9.4). In the lab frame, if we orient coordinates so that the 3-axis is the collision axis, then

$$p_{\text{in}} = p_{\text{projectile}} + p_{\text{target}} = \begin{pmatrix} E_{\text{in}}/c \\ 0 \\ 0 \\ p_{\text{in}} \end{pmatrix} + \begin{pmatrix} m_p c \\ 0 \\ 0 \\ 0 \end{pmatrix}. \quad (4.9.6)$$

If a single particle X emerges, then its four-momentum is the total outgoing four-momentum,

$$p_{\text{out}} = p_X = \begin{pmatrix} E_X \\ p_X^1 \\ p_X^2 \\ p_X^3 \end{pmatrix}. \quad (4.9.7)$$

Demanding that p_{in} coincide with p_{out} determines $\vec{p}_X = p_{\text{in}} \hat{e}_3$ and $E_X = E_{\text{in}} + m_p c^2$. Eq. (4.4.9), applied to the projectile proton (with known mass), may be used to relate the incident spatial momentum and energy, $\vec{p}_{\text{in}}^2 = (E_{\text{in}}/c)^2 - (m_p c)^2$. The same relation (4.4.9), applied to the outgoing particle X , connects its energy E_X and momentum \vec{p}_X to the desired maximum mass m_X , $(m_X c^2)^2 = E_X^2 - (c \vec{p}_X)^2$. Inserting numbers and computing E_X , $|\vec{p}_X| = p_{\text{in}}$, and finally m_X is straightforward. But even less work is required if one recalls [from Eq. (4.4.8)] that the square of any four-momentum directly gives the rest mass of the object, $p^2 = m^2 c^2$. Hence

$$\begin{aligned} m_X^2 c^2 &= p_X^2 = p_{\text{out}}^2 = p_{\text{in}}^2 = (p_{\text{projectile}} + p_{\text{target}})^2 \\ &= p_{\text{projectile}}^2 + p_{\text{target}}^2 + 2 p_{\text{projectile}} \cdot p_{\text{target}} \\ &= 2 m_p^2 c^2 + 2 E_{\text{in}} m_p. \end{aligned} \quad (4.9.8)$$

Consequently, $m_X = \sqrt{2 m_p (m_p + E_{\text{in}}/c^2)} = m_p \sqrt{2 + 2 E_{\text{in}}/(m_p c^2)} \approx \sqrt{2002} m_p \approx 45 m_p$. Even though the projectile proton has an energy a thousand times greater than its rest energy, the maximum mass particle which can be created in this collision is only 45 times heavier than a proton. The rest of the energy must provide the kinetic energy associated with the conserved spatial momentum. More generally, the maximum mass that can be produced grows (only) like the *square root* of the lab frame energy, $m_X \sim \sqrt{2 E_{\text{in}} m_p / c^2}$, when $E_{\text{in}} \gg m_p c^2$. This is why “colliders”, where the lab and CM frames coincide with both the “beam” and “target” particles racing towards each other, are most efficient when hunting for new particles. In particular, if we collide two particles with the same

mass (*e.g.*, either identical particles or particle and antiparticle), the same energy (E_{in}) but opposite momenta, the largest rest mass (particle) we can produce is $m_X = 2E_{\text{in}}$, which in this case increases *linearly* with the beam energy E_{in} .

4.10 Example Problems

Kogut 4-3

In the S' frame we have an event at the 4-vector point $x' = (c \times 9 \times 10^{-8} \text{ s}, 100 \text{ m}, 0, 0)$. We want to determine the location of this event in the S frame, where the S' frame moves with velocity $v/c = 4/5$ along the x axis with respect to the S frame and, for convenience (we are free to be lazy but smart), we assume that the origins (in space and time) of the 2 frames are synchronized. The boost factor between the two frames is $\gamma = 1/\sqrt{1 - (v/c)^2} = 5/3$. Thus the corresponding Lorentz boost gives us

$$x = \Lambda(v)x' = \begin{pmatrix} 5/3 & 4/3 & 0 & 0 \\ 4/3 & 5/3 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 27 \text{ m} \\ 100 \text{ m} \\ 0 \\ 0 \end{pmatrix} = \begin{pmatrix} 178.33 \text{ m} \\ 202.67 \text{ m} \\ 0 \\ 0 \end{pmatrix} = \begin{pmatrix} c \times 59.44 \times 10^{-8} \text{ s} \\ 202.67 \text{ m} \\ 0 \\ 0 \end{pmatrix}.$$

Kogut 4-4

In frame S we are given two events defined by the 4-vectors $x_1 = (L, L, 0, 0)$ and $x_2 = (L/2, 2L, 0, 0)$, or $\Delta x^0 = L/2$ and $\Delta x^1 = -L$. We want to boost to a frame S' where the events (appear to) occur at the same time. Thus we want to solve (note v is the velocity of S' in S so using Λ^{-1} or $\Lambda(-v)$)

$$\Delta x'^0 = \gamma (\Delta x^0 + (-v/c)\Delta x^1) = 0 \Rightarrow L/2 + (-v/c)(-L) = 0 \Rightarrow v/c = -1/2,$$

corresponding to the S' frame moving towards negative x values. So we find the common time in the S' frame is

$$t' = \gamma (ct_1 + (-v/c)x_1) / c = (2/\sqrt{3}) (L + (1/2)L) / c = \sqrt{3}L/c.$$

As a check, note that we obtain the same result if we use instead t_2 and x_2 .

Kogut 6-11

Consider a relativistic particle whose (relativistic) kinetic energy is twice its rest energy, *i.e.*, its total energy is *three* times its rest energy. Thus we have

$$K = 2mc^2 \Rightarrow E = 3mc^2 \Rightarrow \gamma = 3 \Rightarrow \frac{v}{c} = \sqrt{1 - \frac{1}{9}} = 0.943.$$

Thus the magnitude of this particle's momentum is

$$p = \gamma mv = 3(v/c)mc = 2.83 mc.$$

If the kinetic energy is $5mc^2$, we have instead

$$\gamma = 6 \Rightarrow \frac{v}{c} = \sqrt{1 - \frac{1}{36}} = 0.986, p = 5.92 mc.$$

Kogut 6-16

Here we have the opportunity to consider the most fuel-efficient rocket exhaust - photons (the fastest exit velocity for any given energy) in a problem with an explicitly time *dependent* mass for the rocket. We are given only the rocket's initial and final masses, M_i and M_f , and want to calculate its final velocity (starting at rest). Being smart but lazy we do NOT integrate Newton's law! Instead we simply use 4-momentum (energy-momentum) conservation. We image that a certain fraction of the initial mass of the rocket, $\Delta M = M_i - M_f$, is instantaneously converted into a photon (or several collinear photons). To conserve momentum the rocket must recoil in the direction opposite to the photon(s). We have (in the initial rest frame of the rocket)

$$E_f = E_i \Rightarrow M_i c^2 = E_{\text{photon(s)}} + M_f \gamma c^2,$$

$$\vec{p}_{\text{total}} = 0 \Rightarrow M_f v \gamma = p_{\text{photon(s)}} = E_{\text{photon(s)}}/c = M_i c - M_f c \gamma \Rightarrow M_i = M_f \gamma \frac{c+v}{c} \quad (4.10.1)$$

$$\Rightarrow \frac{M_i}{M_f} = \gamma(1 + v/c) = \sqrt{\frac{1 + v/c}{1 - v/c}} \quad \text{or} \quad \frac{v}{c} = \frac{(M_i/M_f)^2 - 1}{(M_i/M_f)^2 + 1}. \quad (4.10.2)$$

So the price of going very fast, $v \rightarrow c$, is that the final mass (including the astronaut) must be very much smaller than the initial mass ($M_f \ll M_i$), *i.e.*, accelerating rockets to near light-speed is an expensive activity (as has always been known to NASA).

Chapter 5

QM and Angular Momentum

5.1 Angular Momentum Operators

In your Introductory Quantum Mechanics (QM) course you learned about the basic properties of low spin systems. Here we want to review that knowledge and indicate in more detail how it arises from the basic principles of QM, *i.e.*, that we work with operators, in particular with Hermitian operators, and those operators obey simple commutation relations. To illustrate these ideas let us review the formalism of eigenstates of definite (total) angular momentum. The following analysis applies as well to the spin operator S and, except in one detail noted at the end, to orbital angular momentum L . (Of course, these operators are related via $J = L + S$.) If you have not seen this sort of analysis before, consider it as an introduction to the power of symmetries as expressed in terms of quantum mechanical operators and states. The analysis involves a large number of steps, but, in a very real sense, each of those steps is quite small.

Also be aware of the larger picture, as discussed in Chapter 10, that we are actually discussing the properties of *representations* of the rotation group, $SO(3)$ (for integer angular momentum) and $SU(2)$ (for half-integer angular momentum). The different states in the representation are what you see when you perform rotations on the reference frame. The apparently different states are, in some sense, the same given the underlying symmetry, *i.e.*, we are simply labeling them differently as we change (rotate) the directions of the “axes”. Note, in particular, that the possible states of the system must always appear in *complete* representations of the underlying symmetries. So our understanding of symmetries and the associated representations will provide tools to organize our description of physical systems, *e.g.*, the particles of the Standard Model.

We want to work with the (hopefully) familiar (3-vector) total angular momentum operator \vec{J} with three components J_1 , J_2 and J_3 (or J_x , J_y and J_z , see, for example, Chapters 7 and 11 in McIntyre). We take all 3 to be Hermitian operators ($J_k^\dagger = J_k$, where $^\dagger = {}^{*T}$, *i.e.*, take complex conjugate and transpose) and thus to have *real* eigenvalues. An essential feature of the operator nature of J (and of QM) is the fact that these three operators obey the nontrivial commutation relation (*i.e.*, the *algebra* corresponding to $SO(3)$ and $SU(2)$)

$$[J_k, J_l] \equiv J_k J_l - J_l J_k = i\hbar \epsilon_{klm} J_m \quad [k, l, m = 1, 2, 3], \quad (5.1.1)$$

where ϵ_{klm} is the *unique* $3 \times 3 \times 3$ anti-symmetric tensor.

ASIDE The algebra serves to completely define the properties of the group elements, the transformations, near the identity - no change - operator. However, there may still be ambiguity about the properties of the elements “far” from the identity operator, and this point is related to the difference between $SO(3)$ and $SU(2)$ - see Chapter 10.

Next we define J^2 , the total angular momentum squared operator,

$$J^2 = J_1^2 + J_2^2 + J_3^2, \quad (5.1.2)$$

which is also a Hermitian operator ($J^{2\dagger} = J^2$) again with real eigenvalues. Actually, since the operators on the RHS of Eq. (5.1.2) are all the squares of Hermitian operators, the corresponding eigenvalues are all positive semi-definite (≥ 0), *i.e.*, the squares of real numbers. By the same token the related operator expression

$$J^2 - J_3^2 = J_1^2 + J_2^2 \quad (5.1.3)$$

tells us that the eigenvalues of $J^2 - J_3^2$ are also positive semi-definite, or that the eigenvalues of J^2 are greater than or equal to the eigenvalues of J_3^2 . (Actually, as we will see shortly, equality will only occur for the special case of *zero* total angular momentum.)

The next essential fact, following from Eq. (5.1.1), is that J^2 commutes with the individual J_k . For example, we have¹

$$\begin{aligned} [J^2, J_3] &= [J_1^2, J_3] + [J_2^2, J_3] + [J_3^2, J_3] \\ &= [J_1, J_3]J_1 + J_1[J_1, J_3] + [J_2, J_3]J_2 + J_2[J_2, J_3] + 0 \\ &= -i\hbar J_2 J_1 - i\hbar J_1 J_2 + i\hbar J_1 J_2 + i\hbar J_2 J_1 = 0. \end{aligned} \quad (5.1.4)$$

Clearly a similar result holds for $[J^2, J_1]$ and $[J^2, J_2]$.

Finally we make the conventional *choice* that our basis states be the simultaneousness eigenstates of J^2 and J_3 (possible because they commute), $|j, m\rangle$, with eigenvalues, j, m , and defined by

$$J^2|j, m\rangle \equiv (J_1^2 + J_2^2 + J_3^2)|j, m\rangle = j(j+1)\hbar^2|j, m\rangle, \quad J_3|j, m\rangle = m\hbar|j, m\rangle, \quad (5.1.5)$$

where we take these eigenstates to be normalized

$$\langle j, m|j, m\rangle = 1.0. \quad (5.1.6)$$

We are encouraged to think of j as labeling the total angular momentum of the state *independent* of any choice of reference frame, while m labels the component of the angular momentum along the 3-axis in a *specific* choice of reference frame. As we will see in detail below, when we rotate the reference frame (or the state), the value of m changes, but j does not. Thus the states corresponding to a given j value and the possible m values, $-j \leq m \leq j$ comprise a *representation* of the rotation group, *i.e.*, these states are transformed into each other in a specific fashion by the rotations. In Group Theory language (see Chapter 10) the operator J^2 is formally labeled a Casimir operator. It is not an element of the algebra or the group, but does commute with the generators (and thus the group elements) and, as noted, its eigenvalues serve to label the specific representation of the group, while m labels the specific element of the $2j+1$ elements in the representation.

These results are presumably familiar from your QM course, including the fact that the allowed values of j are either integer or half-integer and the allowed values of m are the $2j+1$ values in the

¹The *signs* in Eq. (5.1.4) follow from the definition of ϵ_{klm} , *i.e.*, $\epsilon_{132} = -1$ while $\epsilon_{231} = +1$.

range $m = -j, -j+1, \dots, j-1, j$. Here we will see how these results follow from the basic properties of the operators noted above (and the following will serve as an introduction if this was not covered in your 225 class). To that end let us for now define the eigenvalues instead by

$$J^2|j, m\rangle \equiv N^2(j)\hbar^2|j, m\rangle, \quad J_3|j, m\rangle = m\hbar|j, m\rangle, \quad (5.1.7)$$

where we will *derive* below the specific form of $N^2(j)$ and the constraints on the possible values of m .

From the standpoint of the underlying Group theory, we label the J_k as the *generators* of the unitary rotation group ($SO(3)$ and $SU(2)$) in the sense that they “generate” an infinitesimal rotation. Since we want a rotation through a finite angle to be a unitary transformation (*i.e.*, it should conserve probability), the generators are necessarily Hermitian operators, $J_k^\dagger = J_k$ (so that the group elements arising from their exponentiation are Unitary). As already mentioned in Chapter 1 and described in some detail in Chapter 10, the finite rotation corresponds to exponentiating these generators times a continuous parameter (i times the rotation angle over \hbar). For example, $e^{iJ_3\theta/\hbar}$ corresponds to a rotation around the 3-axis by an angle θ and is a member of the rotation group. Note that, since

$$\left(e^{iJ_3\theta/\hbar}\right)^\dagger = e^{-iJ_3^\dagger\theta/\hbar} = e^{-iJ_3\theta/\hbar} = \left(e^{iJ_3\theta/\hbar}\right)^{-1}, \quad (5.1.8)$$

the operator $e^{iJ_3\theta/\hbar}$ is Unitary (*i.e.*, the Hermitian conjugate is the inverse) if (and only if) J_3 is Hermitian.

At this point we can also demonstrate that the eigenstates in Eq. (5.1.5) are orthogonal as desired (for different m values). We have (recall that $1 = e^{i0} = e^{-iJ_3\theta/\hbar}e^{+iJ_3\theta/\hbar}$)

$$\langle j, m'|j, m\rangle = \langle j, m'|e^{-iJ_3\theta/\hbar}e^{+iJ_3\theta/\hbar}|j, m\rangle = e^{i(m-m')\theta/\hbar}\langle j, m'|j, m\rangle. \quad (5.1.9)$$

There are two ways to satisfy this equation. *Either* the exponential factor is unity because $m = m'$ and these are really both the *same* state, *or* they are different states and the matrix element vanishes,

$$\langle j, m'|j, m\rangle = 0, \quad m' \neq m. \quad (5.1.10)$$

To proceed we want to make use of the two remaining generators (J_1 and J_2) that do not define our basis eigenstates, and will change the states when they operate. This is the part of the analysis that may not be familiar (it appears in Chapter 11 of McIntyre’s QM text), but it is illustrative of how we can prove useful results using only the properties of the operators. In particular, we can define the so-called “ladder” (or raising and lowering) operators by

$$J_\pm \equiv J_1 \pm iJ_2. \quad (5.1.11)$$

Since the J_k are Hermitian, it follows that

$$J_\pm^\dagger = J_1^\dagger \mp iJ_2^\dagger = J_1 \mp iJ_2 = J_\mp. \quad (5.1.12)$$

Using Eqs. (5.1.1) and (5.1.11), and some straightforward algebra, we can evaluate the new commutators

$$\begin{aligned} [J_+, J_-] &= [J_1, J_1] + [J_1, -iJ_2] + [iJ_2, J_1] + [iJ_2, -iJ_2] \\ &= 0 - i(iJ_3\hbar) + i(-iJ_3\hbar) + 0 = 2\hbar J_3 \end{aligned} \quad (5.1.13)$$

and

$$\begin{aligned} [J_3, J_{\pm}] &= [J_3, J_1] + [J_3, \pm iJ_2] = i\hbar J_2 + (\pm i)(-i\hbar J_1) \\ &= \hbar(iJ_2 \pm J_1) = \pm \hbar J_{\pm}. \end{aligned} \quad (5.1.14)$$

With a little more algebra we can demonstrate (and you should try this at home) that, since the total spin operator J^2 commutes with each of the components (recall Eq. (5.1.4)), it also commutes with the ladder operators,

$$[J^2, J_k] = 0 \Rightarrow [J^2, J_{\pm}] = 0. \quad (5.1.15)$$

Since we eventually want to be able to evaluate the result of operating on the eigenstates with the ladder operators, we want to first evaluate the products in terms of the eigen-operators J^2 and J_3 . This may seem unmotivated at first, but the usefulness of this step will be clear shortly. By explicit calculation it follows that

$$J_+ J_- = (J_1 + iJ_2)(J_1 - iJ_2) = J_1^2 + J_2^2 - i[J_1, J_2] = J^2 - J_3^2 + \hbar J_3, \quad (5.1.16)$$

and

$$J_- J_+ = (J_1 - iJ_2)(J_1 + iJ_2) = J_1^2 + J_2^2 + i[J_1, J_2] = J^2 - J_3^2 - \hbar J_3, \quad (5.1.17)$$

Note that the difference between these two equations is $2\hbar J_3$ as expected from Eq. (5.1.13). The content of these relations is that the ladder operators move us around in a given representation of $SO(3)$ or $SU(2)$ (hence the label), but do *not* change the representation, *i.e.*, do not change the eigenvalue of J^2 (recall Eq. (5.1.2)). To see this explicitly we first note the operator relation (recall Eq. (5.1.14))

$$J_3 J_{\pm} = J_{\pm} J_3 + [J_3, J_{\pm}] = J_{\pm} J_3 \pm \hbar J_{\pm}. \quad (5.1.18)$$

Thus, when we apply this operator to an eigenstate, we obtain

$$J_3 J_{\pm} |j, m\rangle = J_{\pm} J_3 |j, m\rangle + [J_3, J_{\pm}] |j, m\rangle = m\hbar J_{\pm} |j, m\rangle \pm \hbar J_{\pm} |j, m\rangle = (m \pm 1)\hbar J_{\pm} |j, m\rangle, \quad (5.1.19)$$

clearly indicating that the operator J_{\pm} raises/lowers the J_3 eigenvalue by one (explaining the “ladder” label),

$$J_{\pm} |j, m\rangle \propto |j, m \pm 1\rangle, \quad (5.1.20)$$

or, including an explicit coefficient,

$$J_{\pm} |j, m\rangle \equiv A_{\pm}(j, m \pm 1) |j, m \pm 1\rangle. \quad (5.1.21)$$

We will determine the coefficient $A_{\pm}(j, m \pm 1)$ shortly. From Eqs. (5.1.15) and (5.1.7) we have

$$J^2 J_{\pm} |j, m\rangle = J_{\pm} J^2 |j, m\rangle = N^2(j) \hbar^2 J_{\pm} |j, m\rangle = N^2(j) \hbar^2 A_{\pm}(j, m \pm 1) |j, m \pm 1\rangle, \quad (5.1.22)$$

confirming that J_{\pm} does *not* change the J^2 eigenvalue.

So, as already noted, these last equations tell us to interpret the operator J_{\pm} as stepping us through the 1-D representation labeled by total angular momentum j .

To determine the coefficient $A_{\pm}(j, m)$ we perform the following manipulations, which follow from the definitions above. First, from the definition of the coefficient in Eq. (5.1.21) and the unit normalization of the eigenstates, we have

$$\langle j, m | J_+ J_- | j, m \rangle = \langle j, m | J_-^{\dagger} J_- | j, m \rangle = |A_-(j, m-1)|^2 \langle j, m-1 | j, m-1 \rangle = |A_-(j, m-1)|^2, \quad (5.1.23)$$

and

$$\langle j, m | J_- J_+ | j, m \rangle = \langle j, m | J_+^\dagger J_+ | j, m \rangle = |A_+(j, m+1)|^2 \langle j, m+1 | j, m+1 \rangle = |A_+(j, m+1)|^2. \quad (5.1.24)$$

Now we can use Eqs. (5.1.16) and (5.1.17) to explicitly evaluate these matrix elements and find

$$\begin{aligned} \langle j, m | J_+ J_- | j, m \rangle &= \langle j, m | J^2 - J_3^2 + \hbar J_3 | j, m \rangle = N^2(j) \hbar^2 - m^2 \hbar^2 + m \hbar^2 \\ &= \hbar^2 (N^2(j) - m^2 + m) = |A_-(j, m-1)|^2, \end{aligned} \quad (5.1.25)$$

and

$$\begin{aligned} \langle j, m | J_- J_+ | j, m \rangle &= \langle j, m | J^2 - J_3^2 - \hbar J_3 | j, m \rangle = N^2(j) \hbar^2 - m^2 \hbar^2 - m \hbar^2 \\ &= \hbar^2 (N^2(j) - m^2 - m) = |A_+(j, m+1)|^2. \end{aligned} \quad (5.1.26)$$

We can choose the phases of the eigenstates so that both coefficients are positive, real (without any impact on the quantum physics) and define the coefficients in the operation of the ladder operators to be (keeping the still to be evaluated parameter $N^2(j)$)

$$A_\pm(j, m \pm 1) = \hbar \sqrt{N^2(j) - m^2 \mp m}, \quad (5.1.27)$$

Now we return to the discussion surrounding Eq. (5.1.3). We have seen that the ladder operators raise and lower the eigenvalue m without changing the eigenvalue $N^2(j)$. However, Eq. (5.1.3) tells us that $N^2(j) - m^2 \geq 0$ for all allowed values of $N^2(j)$ (*i.e.*, allowed values of j) and m . These two results can *both* be true if and only if the raising and lowering process truncates, since otherwise we will eventually obtain an m^2 value greater than any (fixed) $N^2(j)$ value. Thus there must be maximum and minimum values of m , m_{max} and m_{min} , such that

$$J_+ |j, m_{max}\rangle = 0, \quad J_- |j, m_{min}\rangle = 0. \quad (5.1.28)$$

These results can be rewritten as the statements that

$$A_+(j, m_{max} + 1) = 0, \quad A_-(j, m_{min} - 1) = 0. \quad (5.1.29)$$

Combining with Eq. (5.1.27) we have

$$N^2(j) - m_{max}^2 - m_{max} = 0, \quad N^2(j) - m_{min}^2 + m_{min} = 0. \quad (5.1.30)$$

Since the raising and lowering is *always* by a unit step (of \hbar , this is QM after all), we know that $m_{max} - m_{min} = n$ is an integer, which we label n . Using the difference of the two results in Eq. (5.1.30) to eliminate $N^2(j)$ and substituting $m_{max} = m_{min} + n$, we find

$$m_{min} = -\frac{n}{2}, \quad m_{max} = \frac{n}{2}. \quad (5.1.31)$$

Returning to Eq. (5.1.30) we find also that

$$N^2(j) = \frac{n}{2} \left(\frac{n}{2} + 1 \right). \quad (5.1.32)$$

So now we can make the standard identification for the eigenvalue j ,

$$\begin{aligned} j &\equiv \frac{n}{2} \Rightarrow -j \leq m \leq j \\ N^2(j) &= j(j+1). \end{aligned} \quad (5.1.33)$$

Substituting in Eq. (5.1.27) we have

$$A_{\pm}(j, m \pm 1) = \hbar \sqrt{j(j+1) - m(m \pm 1)} = \hbar \sqrt{(j \mp m)(j \pm m + 1)}, \quad (5.1.34)$$

These expressions for the coefficients explicitly verify the truncation results of Eq. (5.1.29), *i.e.*, $A_{+}(j, m_{\max} + 1 = j + 1) = 0$ and $A_{-}(j, m_{\min} - 1 = -j - 1) = 0$.

Since n is an integer, there are two possibilities corresponding to odd or even n . If n is an odd integer, then the “total angular momentum” eigenvalue $j = n/2$ is half-integer, while if n is an even integer, then $j = n/2$ is integer. In either case the number of distinct values of m is the familiar $2j + 1$ corresponding to the values $m = -j$ to $m = j$.² Since the same arithmetic applies to the spin of an individual particle, we see that both integer spin particles, *i.e.*, bosons, and half-integer spin particles, *i.e.*, fermions, are possible. On the other hand, *orbital* angular momentum arises from the $\vec{L} = (\vec{r} \times \vec{p})$ operator (with a clear classical connection) and assumes *only* integer values.

To recap, we have used only the facts that the 3 components of the total angular momentum operator are Hermitian operators (and so have *real* eigenvalues) and that these operators satisfy the commutation relation of Eq. (5.1.1), to derive that the possible eigenvalues of J^2 and J_3 are specified by a single parameter j . Further this parameter is either half-integer ($1/2, 3/2, \dots$) or integer ($0, 1, 2, \dots$). The eigenvalue of J^2 is given by $j(j+1)\hbar^2$ corresponding to the $2j+1$ m values in the range $-j \leq m \leq j$ with J_3 eigenvalues $m\hbar$ in the range $-j\hbar$ to $+j\hbar$. Note that only for the “trivial” case $j = 0$ is the eigenvalue of J^2 equal to the eigenvalue of J_3 (both are 0). For j greater than 0 we have $j(j+1)$ greater m^2 , as we should expect for QMical systems where the other components besides J_3 , which do *not* commute with J_3 , will exhibit nonzero, if indeterminate, values in an eigenstate of J_3 .

5.2 Spin 1/2 in Vector/Matrix Notation

Here we will study a spin 1/2 system as an example of a 2-state system as you studied in your QM class. We represent the corresponding states as

$$|\frac{1}{2}, \frac{1}{2}\rangle = |\uparrow\rangle = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \quad |\frac{1}{2}, -\frac{1}{2}\rangle = |\downarrow\rangle = \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \quad (5.2.1)$$

where the up and down arrow notation is specific to the spin 1/2 interpretation and corresponds to the component of spin along the 3 axis. In this basis we can use the so-called Pauli matrices, which you likely learned about in Physics 225 or 227 as a basis set of 2×2 matrices. They are defined by

$$\sigma_1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \sigma_2 = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \sigma_3 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}, \quad (5.2.2)$$

and obey the commutation relation (note the factors of $1/2$)

$$\left[\frac{\sigma_j}{2}, \frac{\sigma_k}{2} \right] = i\epsilon_{jkl} \frac{\sigma_l}{2}. \quad (5.2.3)$$

²The underlying physics point here is the *quantization* of angular momentum in units of \hbar , *i.e.*, changes in angular momentum (spin) can only occur in *integer* steps with magnitude \hbar . We don’t notice this in ordinary life because a typical (classical) angular momentum has magnitude of order $\text{kg m}^2/\text{s}$ or Joule * second. In units of \hbar this is of order $10^{34} \hbar$ and changes in the angular momentum of magnitude \hbar will appear to be continuous changes.

Thus we can define the representation of the spin operator for our spin 1/2 system as (recall Eqs. (1.5.4) and (1.5.5))

$$S_k = \hbar \frac{\sigma_k}{2}, \quad [S_j, S_k] = i\epsilon_{jkl}\hbar S_l. \quad (5.2.4)$$

Note, in particular, that

$$S_3 = \frac{\hbar}{2} \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}, \quad (5.2.5)$$

as is appropriate for basis states that are eigenstates of S_3 with eigenvalues $\pm\hbar/2$. It follows that

$$S_3|\frac{1}{2}, \frac{1}{2}\rangle = \frac{\hbar}{2}|\frac{1}{2}, \frac{1}{2}\rangle, \quad S_3|\frac{1}{2}, -\frac{1}{2}\rangle = -\frac{\hbar}{2}|\frac{1}{2}, -\frac{1}{2}\rangle, \quad S^2|\frac{1}{2}, \pm\frac{1}{2}\rangle = \frac{3}{4}\hbar^2|\frac{1}{2}, \pm\frac{1}{2}\rangle. \quad (5.2.6)$$

Now consider the raising and lowering (ladder) operators. The representations for the raising and lowering operators are

$$S_+ = S_1 + iS_2 = \hbar \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}, \quad S_- = S_1 - iS_2 = \hbar \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}, \quad (5.2.7)$$

which clearly perform the following transformations

$$S_+|\uparrow\rangle = 0, \quad S_-|\uparrow\rangle = \hbar|\downarrow\rangle, \quad S_+|\downarrow\rangle = \hbar|\uparrow\rangle, \quad S_-|\downarrow\rangle = 0. \quad (5.2.8)$$

This checks with Eq. (5.1.34) that yields for the spin 1/2 coefficients

$$\begin{aligned} A_+ \left(\frac{1}{2}, \frac{3}{2} \right) &= \hbar \sqrt{\frac{1}{2} \left(\frac{3}{2} \right) - \frac{1}{2} \left(\frac{3}{2} \right)} = 0, \\ A_- \left(\frac{1}{2}, -\frac{1}{2} \right) &= \hbar \sqrt{\frac{1}{2} \left(\frac{3}{2} \right) - \frac{1}{2} \left(-\frac{1}{2} \right)} = \hbar, \\ A_+ \left(\frac{1}{2}, \frac{1}{2} \right) &= \hbar \sqrt{\frac{1}{2} \left(\frac{3}{2} \right) - \left(-\frac{1}{2} \right) \frac{1}{2}} = \hbar, \\ A_- \left(\frac{1}{2}, -\frac{3}{2} \right) &= \hbar \sqrt{\frac{1}{2} \left(\frac{3}{2} \right) - \left(-\frac{1}{2} \right) \left(-\frac{3}{2} \right)} = 0. \end{aligned} \quad (5.2.9)$$

(Be certain to verify that you understand how these results arise.) The raising and lowering operators are simply flipping the spin component along the 3 axis, or producing zero if this spin component cannot be raised or lowered further.

Next we look at finite transformations in the underlying (Unitary Group) $SU(2)$. We proceed much as we did when we studied the group of rotations in 3-D in Chapter 1 (*i.e.*, the group $SO(3)$). To proceed it is useful to have the analogue of Eq. (1.5.8) for the Pauli matrices,

$$\sigma_k^{2n} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = \mathbf{1}, \quad \sigma_k^{2n+1} = \sigma_k. \quad (5.2.10)$$

ASIDE This result allows us to easily verify the last result in Eq. (5.2.6),

$$S_k^2 = \frac{\hbar^2}{4} \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad S^2 = S_1^2 + S_2^2 + S_3^2 = \frac{3\hbar^2}{4} \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}. \quad (5.2.11)$$

Proceeding as we did in Chapter 1, we find results similar to, but simpler than Eqs. (1.5.9) and (1.5.10) (see also Chapter 10). For the finite transformation generated by S_3 or σ_3 we have

$$\begin{aligned} T_3(\alpha) &\equiv e^{i\alpha S_3/\hbar} = e^{i\sigma_3\alpha/2} = \mathbf{1} + \sum_{n=1}^{\infty} \frac{(i\alpha\sigma_3)^n}{2^n n!} \\ &= \mathbf{1} \sum_{n=0}^{\infty} \frac{(-1)^n (\alpha)^{2n}}{2^{2n} (2n)!} + i\sigma_3 \sum_{n=0}^{\infty} \frac{(-1)^n (\alpha)^{2n+1}}{2^{2n+1} (2n+1)!}. \end{aligned} \quad (5.2.12)$$

Using what we know about the sinusoidal functions, especially Eqs. (1.3.2) and (1.3.3), we can write this transformation in the compact form

$$T_3(\alpha) = \begin{pmatrix} \cos \frac{\alpha}{2} + i \sin \frac{\alpha}{2} & 0 \\ 0 & \cos \frac{\alpha}{2} - i \sin \frac{\alpha}{2} \end{pmatrix} = \begin{pmatrix} e^{i\alpha/2} & 0 \\ 0 & e^{-i\alpha/2} \end{pmatrix} \quad (5.2.13)$$

(see also Eq. (10.3.23)). This $SU(2)$ transformation generated by S_3 or σ_3 in the basis of eigenstates of S_3 is simply a change of phase by $\pm\alpha/2$ where the sign depends on the sign of the eigenvalue. In particular, the transformation is diagonal and does *not* involve any mixing of the two eigenstates. In fact, since our basis states are eigenstates of S_3 , we could have evaluated this transformation directly,

$$T_3(\alpha) \left| \frac{1}{2}, m \right\rangle = e^{i\alpha S_3/\hbar} \left| \frac{1}{2}, m \right\rangle = e^{i\alpha m} \left| \frac{1}{2}, m \right\rangle, \quad (5.2.14)$$

where $m = \pm 1/2$. This is the *same* result as the matrix form in Eq. (5.2.13).

5.3 Spin 1 in Vector/Matrix Notation

To further strengthen our understanding of spin systems, we want to consider a spin 1 system³ where again we use the simultaneous eigenstates of S^2 and S_3 as the basis states. The possible eigenvalues of S_3 are now $+1, 0, -1$. So our vectors will have 3 components similar to the discussion of ordinary rotations, $SO(3)$, in Chapter 1, but is important to remember that here we are talking about $SU(2)$ Unitary transformations of a 3 state QMical system (where phases can matter) and not ordinary location vectors in 3 dimensions. For this case the basis states are

$$|s, m\rangle = |1, 1\rangle = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \quad |1, 0\rangle = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, \quad |1, -1\rangle = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}. \quad (5.3.1)$$

Next we want representations of S_1 , S_2 and S_3 in this basis. As just noted, this is not the same basis as in Chapter 1.5 (*i.e.*, not ordinary location 3-vectors) and we do not expect the same representation as in Eq. (1.5.5). In particular, in the basis of its *own* eigenstates, S_3 should be represented by

$$S_3 = \hbar \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & -1 \end{pmatrix}, \quad (5.3.2)$$

³In the individual particle language we will use shortly spin 1 particles are labeled vector particles in an obvious reference to the more familiar, but distinct, 3-component location vectors in 3-D space.

so that

$$S_3|1, 1\rangle = \hbar|1, 1\rangle, \quad S_3|1, 0\rangle = 0, \quad S_3|1, -1\rangle = -\hbar|1, -1\rangle. \quad (5.3.3)$$

A little algebra (or a good book) yields the corresponding representations of S_1 and S_2 to be

$$S_1 = \frac{\hbar}{\sqrt{2}} \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}, \quad S_2 = \frac{\hbar}{\sqrt{2}} \begin{pmatrix} 0 & -i & 0 \\ i & 0 & -i \\ 0 & i & 0 \end{pmatrix}. \quad (5.3.4)$$

The reader is *strongly* encouraged to verify that this set of representations (matrices) are Hermitian and satisfy the required commutator,

$$[S_j, S_k] = i\epsilon_{jkl}\hbar S_l. \quad (5.3.5)$$

With a little more algebra we find that

$$S_1^2 = \frac{\hbar^2}{2} \begin{pmatrix} 1 & 0 & 1 \\ 0 & 2 & 0 \\ 1 & 0 & 1 \end{pmatrix}, \quad S_2^2 = \frac{\hbar^2}{2} \begin{pmatrix} 1 & 0 & -1 \\ 0 & 2 & 0 \\ -1 & 0 & 1 \end{pmatrix}, \quad S_3^2 = \hbar^2 \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \quad (5.3.6)$$

so that

$$S^2 = S_1^2 + S_2^2 + S_3^2 = 2\hbar^2 \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \quad (5.3.7)$$

So we have confirmed that in a spin $s = 1$ system the shared eigenvalue of S^2 is $\hbar^2 s(s+1) = 2\hbar^2$.

Finally with this representation we can construct the corresponding raising and lowering operators,

$$\begin{aligned} S_+ &= S_1 + iS_2 = \frac{\hbar}{\sqrt{2}} \begin{pmatrix} 0 & 2 & 0 \\ 0 & 0 & 2 \\ 0 & 0 & 0 \end{pmatrix} = \sqrt{2}\hbar \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix}, \\ S_- &= S_1 - iS_2 = \frac{\hbar}{\sqrt{2}} \begin{pmatrix} 0 & 0 & 0 \\ 2 & 0 & 0 \\ 0 & 2 & 0 \end{pmatrix} = \sqrt{2}\hbar \begin{pmatrix} 0 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}. \end{aligned} \quad (5.3.8)$$

These matrices perform the expected transformations,

$$\begin{aligned} S_+|1, 1\rangle &= 0, \quad S_+|1, 0\rangle = \sqrt{2}\hbar|1, 1\rangle, \quad S_+|1, -1\rangle = \sqrt{2}\hbar|1, 0\rangle \\ S_-|1, 1\rangle &= \sqrt{2}\hbar|1, 0\rangle, \quad S_-|1, 0\rangle = \sqrt{2}\hbar|1, -1\rangle, \quad S_-|1, -1\rangle = 0. \end{aligned} \quad (5.3.9)$$

Recall from Eq. (5.1.34) that $\sqrt{2}\hbar$ is the expected nonzero coefficient.

5.4 Examples

To practice using the techniques described above let us consider the form of the $SU(2)$ transformation generated by S_1 and S_2 when operating on the eigenstates of S_3 as the basis states. Since the structure of Eq. (5.2.3) still obtains, the procedure follows much as it did for S_3 . For S_1 we have

$$\begin{aligned} T_1(\alpha) &\equiv e^{i\alpha S_1/\hbar} = e^{i\sigma_1\alpha/2} = \mathbf{1} + \sum_{n=1}^{\infty} \frac{(i\alpha\sigma_1)^n}{2^n n!} \\ &= \mathbf{1} \sum_{n=0}^{\infty} \frac{(-1)^n (\alpha)^{2n}}{2^{2n} (2n)!} + i\sigma_1 \sum_{n=0}^{\infty} \frac{(-1)^n (\alpha)^{2n+1}}{2^{2n+1} (2n+1)!}. \end{aligned} \quad (5.4.1)$$

Again we recognize the sums and rewrite in the compact form

$$T_1(\alpha) = \begin{pmatrix} \cos \frac{\alpha}{2} & i \sin \frac{\alpha}{2} \\ i \sin \frac{\alpha}{2} & \cos \frac{\alpha}{2} \end{pmatrix}. \quad (5.4.2)$$

Note that in this case the transformation is no longer diagonal, *i.e.*, this transformation does “rotate” the eigenstates of S_3 into one another (with some extra phases). This is intuitively reasonable as the transformation is about an axis orthogonal to the direction along which we quantized the spin component. Note also that

$$T_1(\pi) = \begin{pmatrix} \cos \frac{\pi}{2} & i \sin \frac{\pi}{2} \\ i \sin \frac{\pi}{2} & \cos \frac{\pi}{2} \end{pmatrix} = \begin{pmatrix} 0 & i \\ i & 0 \end{pmatrix}, \quad (5.4.3)$$

i.e., a “rotation” by π flips spin up to spin down and conversely (along with adding a phase of $\pi/2$), which is intuitively expected.

The story for S_2 is quite similar leading to

$$\begin{aligned} T_2(\alpha) &\equiv e^{i\alpha S_2/\hbar} = e^{i\sigma_2\alpha/2} = \mathbf{1} + \sum_{n=1}^{\infty} \frac{(i\alpha\sigma_2)^n}{2^n n!} \\ &= \mathbf{1} \sum_{n=0}^{\infty} \frac{(-1)^n (\alpha)^{2n}}{2^{2n} (2n)!} + i\sigma_2 \sum_{n=0}^{\infty} \frac{(-1)^n (\alpha)^{2n+1}}{2^{2n+1} (2n+1)!}. \end{aligned} \quad (5.4.4)$$

Again we recognize the sums and rewrite in the compact form

$$T_2(\alpha) = \begin{pmatrix} \cos \frac{\alpha}{2} & \sin \frac{\alpha}{2} \\ -\sin \frac{\alpha}{2} & \cos \frac{\alpha}{2} \end{pmatrix}. \quad (5.4.5)$$

This really does look like an ordinary rotation (recall Eq. (1.5.10)) except for the factor of $1/2$ in the arguments of the sines and cosines, which is the residue of spin $1/2$. It is also important to remember that this 2-state system is in terms of the eigenstates of S_3 and not in terms of the two components of an ordinary location 2-vector.

Note that a rotation through 2π about any of the 3 axes, which you might naively expect to bring us back to where we started (*i.e.*, yield the unit matrix), is given instead by the negative of the unit matrix (see Eqs. (5.4.2), (5.4.5) and (5.2.13))

$$T_1(2\pi) = T_2(2\pi) = T_3(2\pi) = \begin{pmatrix} -1 & 0 \\ 0 & -1 \end{pmatrix}. \quad (5.4.6)$$

This illustrates a fundamental difference between half-integer spin particles (fermions) and the more classically behaved integer-spin particles (bosons - see below). This difference plays an essential role in our understanding of how the fundamental particles behave.

As a final example let us evaluate finite $SU(2)$ transformations for vector particles, *i.e.*, in the basis of the previous section. The products of the (representations of the) generators satisfy slightly different relations than the more familiar form in Eq. (5.2.10) (also recall Eq. (5.3.6))

$$S_1^{2n} = \frac{\hbar^{2n}}{2} \begin{pmatrix} 1 & 0 & 1 \\ 0 & 2 & 0 \\ 1 & 0 & 1 \end{pmatrix}, \quad S_1^{2n+1} = \hbar^{2n} S_1, \quad (5.4.7)$$

$$S_2^{2n} = \frac{\hbar^{2n}}{2} \begin{pmatrix} 1 & 0 & -1 \\ 0 & 2 & 0 \\ -1 & 0 & 1 \end{pmatrix}, \quad S_2^{2n+1} = \hbar^{2n} S_2, \quad (5.4.8)$$

and

$$S_3^{2n} = \hbar^{2n} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \quad S_3^{2n+1} = \hbar^{2n} S_3. \quad (5.4.9)$$

Thus, following a path similar to the one above, we find that, in the vector $SU(2)$ representation, the finite transformation generated by S_1 is given by

$$\begin{aligned} T_1(\alpha) &\equiv e^{i\alpha S_1/\hbar} = \mathbf{1} + \sum_{n=1}^{\infty} \frac{(i\alpha S_1/\hbar)^n}{n!} \\ &= \frac{1}{2} \begin{pmatrix} 1 & 0 & -1 \\ 0 & 0 & 0 \\ -1 & 0 & 1 \end{pmatrix} + \frac{1}{2} \begin{pmatrix} 1 & 0 & 1 \\ 0 & 2 & 0 \\ 1 & 0 & 1 \end{pmatrix} \sum_{n=0}^{\infty} \frac{(-1)^n (\alpha)^{2n}}{(2n)!} + iS_1 \sum_{n=0}^{\infty} \frac{(-1)^n (\alpha)^{2n+1}}{(2n+1)!}, \end{aligned} \quad (5.4.10)$$

where in the second step we split apart the unit matrix to provide the $n = 0$ term in the first sum. As in the earlier analyses we can now rewrite the two sums as the cosine and sine functions. We have

$$T_1(\alpha) = \frac{1}{2} \begin{pmatrix} 1 + \cos \alpha & i\sqrt{2} \sin \alpha & -1 + \cos \alpha \\ i\sqrt{2} \sin \alpha & 2 \cos \alpha & i\sqrt{2} \sin \alpha \\ -1 + \cos \alpha & i\sqrt{2} \sin \alpha & 1 + \cos \alpha \end{pmatrix}. \quad (5.4.11)$$

This matrix clearly reduces to the unit matrix for $\alpha = 0$ and it is straight forward to demonstrate that it is a Unitary matrix ($T_1^{-1} = T_1^\dagger$) for real α . The form of the matrix for general α values is less intuitive. However, for $\alpha = \pi$, which we discussed above for the spin 1/2 case, and where we expect to be exchanging the spin up and spin down states we find

$$T_1(\pi) = \begin{pmatrix} 0 & 0 & -1 \\ 0 & -1 & 0 \\ -1 & 0 & 0 \end{pmatrix}. \quad (5.4.12)$$

This matrix does indeed exchange the $m = \pm 1$ states with each other, and introduces a phase of -1 everywhere.

The same analysis for the transformation generated by S_2 yields

$$\begin{aligned} T_2(\alpha) &\equiv e^{i\alpha S_2/\hbar} = \mathbf{1} + \sum_{n=1}^{\infty} \frac{(i\alpha S_2/\hbar)^n}{n!} \\ &= \frac{1}{2} \begin{pmatrix} 1 & 0 & 1 \\ 0 & 0 & 0 \\ 1 & 0 & 1 \end{pmatrix} + \frac{1}{2} \begin{pmatrix} 1 & 0 & -1 \\ 0 & 2 & 0 \\ -1 & 0 & 1 \end{pmatrix} \sum_{n=0}^{\infty} \frac{(-1)^n (\alpha)^{2n}}{(2n)!} + iS_2 \sum_{n=0}^{\infty} \frac{(-1)^n (\alpha)^{2n+1}}{(2n+1)!}, \end{aligned} \quad (5.4.13)$$

where in the second step we again split apart the unit matrix to provide the $n = 0$ term in the first sum. As in the previous analyses we can now rewrite the two sums as the cosine and sine functions. We have

$$T_2(\alpha) = \frac{1}{2} \begin{pmatrix} 1 + \cos \alpha & \sqrt{2} \sin \alpha & 1 - \cos \alpha \\ -\sqrt{2} \sin \alpha & 2 \cos \alpha & \sqrt{2} \sin \alpha \\ 1 - \cos \alpha & -\sqrt{2} \sin \alpha & 1 + \cos \alpha \end{pmatrix}. \quad (5.4.14)$$

Except for the factors of i this transformation is very similar to T_1 . T_2 is also a Unitary matrix. $T_2(0)$ is again the unit matrix, while

$$T_2(\pi) = \begin{pmatrix} 0 & 0 & 1 \\ 0 & -1 & 0 \\ 1 & 0 & 0 \end{pmatrix}, \quad (5.4.15)$$

which again exchanges the $m = \pm 1$ states, but with somewhat different phases.

Finally consider the transformation generated by S_3 , which, as noted earlier for the spin 1/2 case, is particularly simple to evaluate when the basis states are eigenstates of S_3 as here. We have

$$\begin{aligned} T_3(\alpha) &\equiv e^{i\alpha S_3/\hbar} = \mathbf{1} + \sum_{n=1}^{\infty} \frac{(i\alpha S_3/\hbar)^n}{n!} \\ &= \begin{pmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix} + \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix} \sum_{n=0}^{\infty} \frac{(-1)^n (\alpha)^{2n}}{(2n)!} + iS_3 \sum_{n=0}^{\infty} \frac{(-1)^n (\alpha)^{2n+1}}{(2n+1)!} \\ &= \begin{pmatrix} \cos \alpha + i \sin \alpha & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & \cos \alpha - i \sin \alpha \end{pmatrix} = \begin{pmatrix} e^{i\alpha} & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & e^{-i\alpha} \end{pmatrix}, \end{aligned} \quad (5.4.16)$$

which is, as expected, quite similar to Eq. (5.2.13)

To contrast with the result for spin 1/2 in Eq. (5.4.6) we can evaluate Eqs. (5.4.16), (5.4.11) and (5.4.14) for a rotation through 2π applied to a spin 1 system,

$$T_1(2\pi) = T_2(2\pi) = T_3(2\pi) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}. \quad (5.4.17)$$

For integer spin we obtain the expected “classical” result of returning to where we started.

The reader is encouraged to practice using the concepts described above by reproducing the above expressions for the various transformations.

Chapter 6

Known particles

6.1 Ordinary matter

What are you made of? Blood and guts and bone and muscle is a little more accurate than the traditional mother goose rhyme. Your tissues are made of cells, which are little bags of chemicals: proteins, nucleic acids, lipids, water and other molecules. Each molecule is a specific assembly of atoms. And each atom contains an atomic nucleus surrounded by some number of electrons.

This should all sound familiar. But stop for a minute and ask how this is known. You can see cells in a microscope. But for objects smaller than cells direct observation gets more difficult. How do you know that atoms and molecules, or electrons and nuclei, exist? Is it just because someone told you so? What's the evidence?

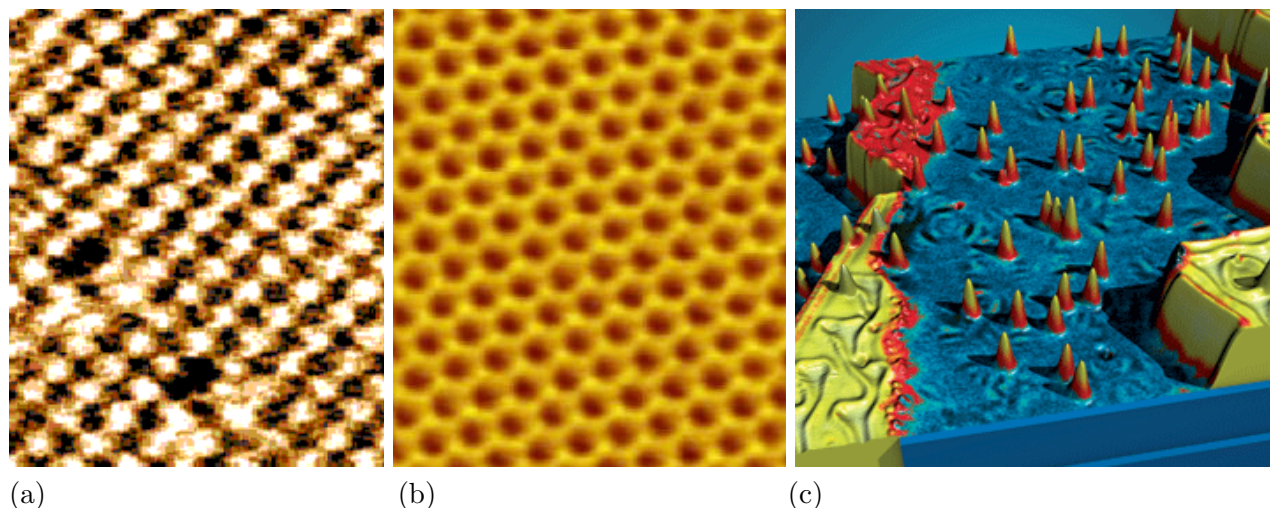


Figure 6.1: Three examples of modern atomic scale imaging. Image (a) shows the surface of sodium chloride, imaged by atomic force microscopy (AFM). Note the two surface defects. Image (b) (courtesy of E. Andrei) is a scanning tunneling microscope (STM) image of a freely suspended graphene sheet — a single atomic layer of graphite. The hexagonal structure, reflecting the sp^2 hybridization of valence electrons in the carbon atoms, is obvious. Image (c) (from the cover of the April 4, 2008 issue of *Science*) shows single cobalt atoms on a platinum surface with steps, imaged with spin-polarized scanning tunneling microscopy. Blue areas show the platinum substrate; red and yellow regions in front of the steps show adsorbed cobalt monolayer stripes with magnetization up (yellow) or down (red).

The historical basis for the atomic structure of matter owes much to the development of the kinetic theory of gases, the understanding of Brownian motion, and chemistry. From a more modern perspective, two compelling types of experimental evidence for the existence of atoms can be summed up as (i) chemistry works, and (ii) individual atoms and molecules can be imaged using a variety of modern techniques, such as scanning tunneling microscopy and atomic force microscopy. A few examples of atomic scale imaging are shown in Figure 6.1.

In the following discussion we will introduce the known particles essentially in historical order. Individual particles are characterized by their (rest) mass, their spin (*i.e.*, how they transform under ordinary spatial rotations) and their participation in the known types of interactions, which comprise the Standard Model (labeled the SM).¹ All particles participate in gravitational interactions, but for our purposes this interaction is extremely weak and will be largely ignored in this class. Nearly all particles participate in the Weak Interactions (all except photons and gluons), which are stronger than gravity but weaker than the Electromagnetic and Strong interactions. Particles with a nonzero electric charge (plus the photon) participate in the Electromagnetic interactions. Finally the hadrons (made from quarks and gluons) participate in the Strong interactions. The relative strengths and ranges of the known “fundamental” forces are characterized in the following table.

Force	Relative Strength	Range
Strong	1	$\approx \text{fm} = 10^{-15} \text{ meter}$
Electromagnetic	10^{-2}	“Infinite” ($\propto 1/r^2$)
Weak	10^{-6}	$\approx 10^{-3} \text{ fm}$
Gravitational	10^{-43}	“Infinite” ($\propto 1/r^2$)

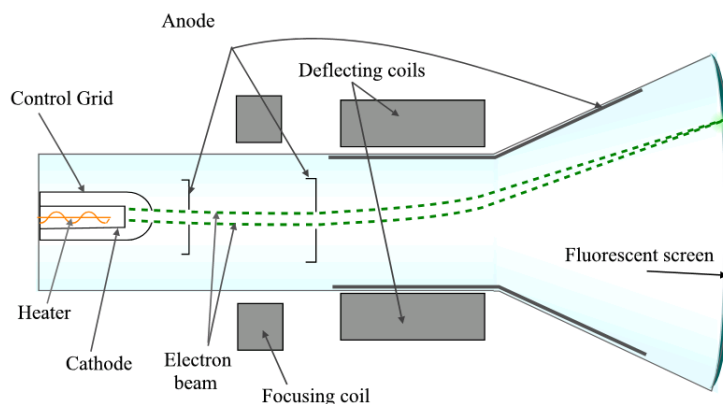


Figure 6.2: Simple sketch of cathode ray tube.

The existence of electrons has been known experimentally since the work of J.J. Thomson who, in 1897, studied the behavior of particles that pass through a cathode ray tube (which is essentially a *really old* tv) when a suitably large voltage is applied between the anode and cathode as illustrated in Figure 6.2.

¹We will eventually characterize this participation in terms of a variety of “charges”, which themselves characterize how the particles transform under more abstract “rotations.”

Thomson found that these particles have a mass to (electric) charge ratio which is *independent* of the type of material forming the cathode or the gas in the tube. Further the magnitude of this ratio is about 2000 times smaller than the corresponding mass to charge ratio of a hydrogen ion (*i.e.*, a proton). Measurement of the charge-to-mass ratio involves observing the deflection of a moving particle produced by a magnetic field (generated by the deflecting coils in Figure 6.2). The charge of a single electron can be measured using the approach of Millikan and Fletcher's famous oil drop experiment pictured in Figure 6.3. Based on refinements of such measurements, the magnitude of

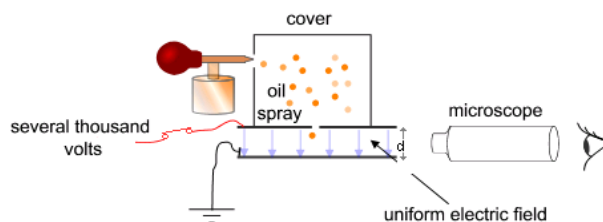


Figure 6.3: Simple sketch of Millikan and Fletcher's oil drop experiment.

the electron charge is now known to a precision of a few parts in 10^8 ,

$$|-e| = 1.602\,176\,565\,(35) \times 10^{-19} \text{ C}. \quad (6.1.1)$$

(The number in parentheses indicates the uncertainty in the last two digits.) In other words, a Coulomb, whose definition is based on macroscopic measurements of current plus the definition of a second, is equal in magnitude to $6.241\,509\,34\,(13) \times 10^{18}$ electron charges. The mass of the electron is also known to a similar precision,

$$m_e = 0.510\,998\,928\,(11) \text{ MeV}/c^2 = 9.109\,382\,91(40) \times 10^{-31} \text{ kg}. \quad (6.1.2)$$

One MeV ($= 10^6$ eV) is the energy acquired by an electron passing through a potential difference of one million volts. (The interested student is encouraged to become familiar with the vast amount of precision data available to you at the PDG website, which is linked at the bottom of the class web page.)

A few angstroms ($1 \text{ \AA} = 10^{-10} \text{ m} = 5 \times 10^5 \text{ GeV}^{-1}$) is the size of individual atoms, whereas nuclear sizes are naturally measured in units of the fermi (or femtometer), where $1 \text{ fm} = 10^{-15} \text{ m} \simeq 5 \text{ GeV}^{-1}$. Direct evidence of the size of atomic nuclei comes from scattering experiments, specifically measurements of the momentum dependence of the scattering cross section. This topic will be discussed more fully in a later chapter. For now, it suffices to note that, in order to learn about the structure of some object like an atomic nucleus, one must use some probe [such as photons (light), electrons, or other nuclei] whose wavelength is *smaller* than the size of the object of interest. To this end particle physics, starting around the middle of the 20th century, has been marked by the development of particle accelerators of ever increasing energy.

Atomic nuclei are known to be bound states of more fundamental particles, protons and neutrons (except for the lightest nucleus of hydrogen, which is just a single proton). This information again comes from scattering experiments: one can bombard nuclei with various projectiles, such as electrons or other nuclei, and observe individual protons or neutrons knocked out of the target nucleus. Just as atoms come in different types, which are usefully organized in the traditional periodic table

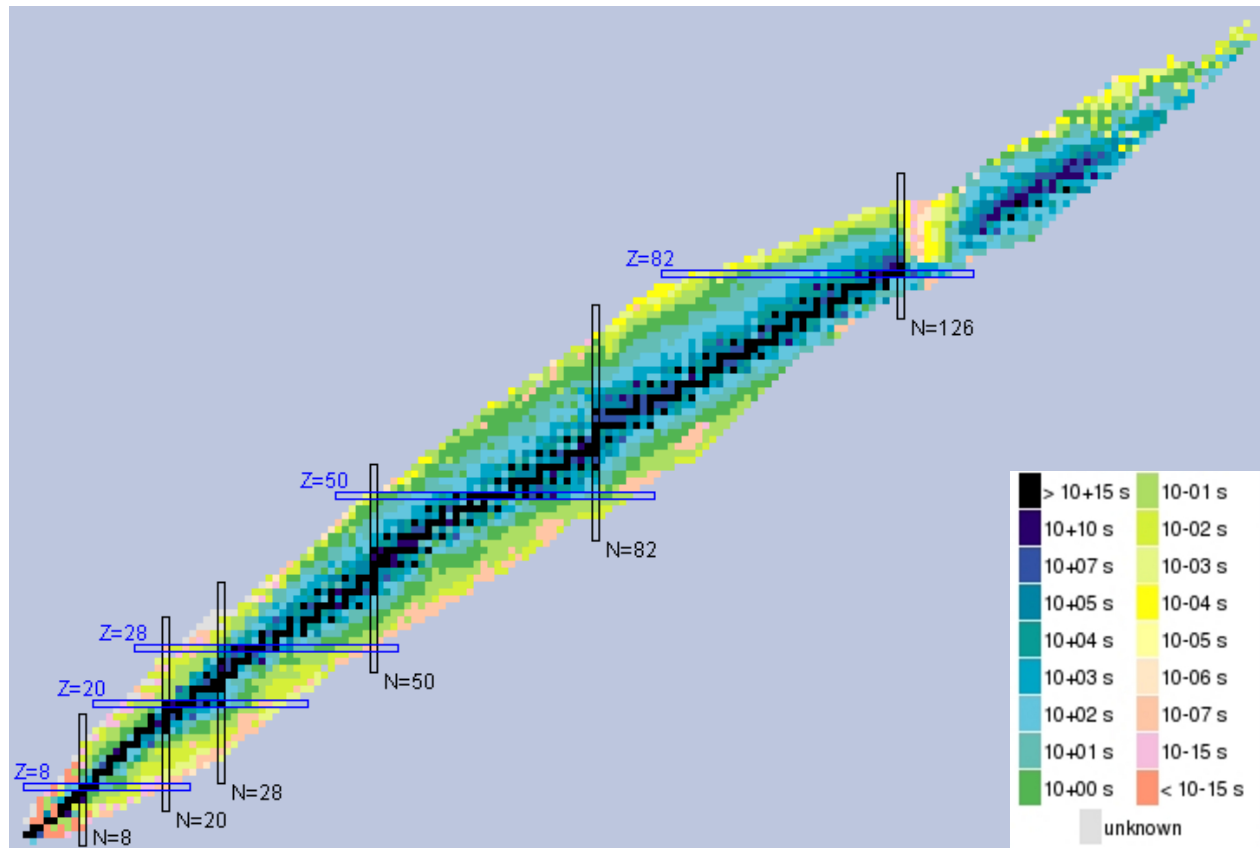


Figure 6.4: Chart of the nuclides (from the National Nuclear Data Center at Brookhaven National Laboratory). The number of protons, Z , is plotted vertically and the number of neutrons, N , horizontally. The color coding indicates the lifetime, with stable nuclei in black and lighter colors corresponding, as shown in the legend, to progressively shorter lifetimes. The rows and columns labeled with specific values of Z and N are so-called “magic” numbers where nuclei have enhanced stability.

and characterized by their differing chemical interactions, there are many different atomic nuclei distinguished by the numbers of neutrons and protons that they contain. It is conventional to label nuclei with the atomic symbol for the corresponding element, with a preceding *superscript* indicating the atomic number A , equal to the number of protons plus neutrons, and a preceding *subscript* Z indicating the number of protons. For example, the lithium-7 nucleus, ${}^7_3\text{Li}$, is a bound state of three protons and four neutrons. Figure 6.4 shows a plot of known nuclear species (or *nuclides*), color coded according to their stability (see the lifetimes in the legend). Useful interactive online versions may be found at www.nndc.bnl.gov and atom.kaeri.re.kr.²

Protons have charge $+e$, *precisely* equal in magnitude but opposite in sign to the electron. Note that this apparent *exact* equality (except for the sign) between the proton’s electric charge and that of the electron is a very important feature of our universe. We are built from atoms with zero net charge (to a very good approximation, *i.e.* to better than one part in 10^{21}), and thus electromagnetic repulsion

²Note that most tables of nuclides, including the one at atom.kaeri.re.kr, list *atomic* masses, not *nuclear* masses, the distinction being that the atomic mass is the mass of the neutral atom. In other words, the atomic mass includes the rest mass of all the electrons plus the mass of the nucleus, as well as the (negative) atomic binding energy.

does *not* exclude us from sitting next to our fellow students. This situation clearly calls for some fundamental explanation. So-called “GUTS”, or Grand Unified Theories, are intended to do just that by postulating a “grand” underlying symmetry that relates quarks and leptons. This “relationship” can then explain the exact equality of the magnitudes of the electric charges of electrons and protons. The existence of such an underlying symmetry, even if badly broken at the energy scales we are used to, would lead to interactions allowing the proton to decay (slowly, see below) into leptons. The search goes on, but for now proton decay had not been observed and we have no real explanation of why there is a single fundamental unit of electric charge.

The mass of a proton is measured to be

$$m_p = 0.938\,272\,046(21) \text{ GeV}/c^2. \quad (6.1.3)$$

This is about 2000 times larger than the mass of an electron. Neutrons, which are neutral (zero electric charge) particles, are slightly heavier than protons,

$$m_n = 0.939\,565\,379(21) \text{ GeV}/c^2. \quad (6.1.4)$$

Neutrons, protons, and electrons are all spin 1/2 particles, where spin is measured in terms of the fundamental quantum of spin, \hbar . Looking ahead, it is essential (and we will discuss and use this at several points) to recall that one way to classify particles is in terms of whether their spin is one-half integer (as here) or integer valued (in units of \hbar). The former are labeled “fermions” and the later “bosons”. One way to think about the difference between these classes of particles is how they transform under rotations. The familiar behavior (*i.e.*, behaving like you do) is that of bosons. After a rotation of 360 degrees, or 2π radians, about any axis a boson is *unchanged* (*i.e.*, comes back to where it started, $e^{2\pi i} = 1$). Less familiar is the behavior of a fermion under such a rotation; it does not come back to where it started but instead differs by a minus sign (*i.e.*, its phase has changed by π radians, $e^{2\pi i/2} = e^{\pi i} = -1$).³ Of course, we are sensitive to such phases only in the context of quantum mechanics. Thus the intrinsic difference between these two varieties of particles becomes absolutely clear in the context of quantum field theory (QFT) where we must define operators to represent these particles (and the fields that describe them). For fermions these operators must *anti*-commute ($\{A, B\} = AB + BA = 0$ or $AB = -BA$), while for bosons the corresponding operators must commute ($[C, D] = CD - DC = 0$ or $CD = DC$). Thus, when we build states out of identical fermions, the states must be *anti*-symmetric under the interchange of any pair of fermions. This immediately leads to the Pauli Exclusion principle - *no* two identical fermions can reside in the *same* state, since such a situation would necessarily be symmetric. On the other hand, a state constructed of identical bosons must be symmetric under the interchange of any two of the bosons. Thus bosons are “happy” to be in the same state as that guarantees symmetry (and lasers really do produce beams of coherent photons, all in the same state). You will often see this connection between spin and the interchange symmetry referred to as the Spin-Statistics Theorem.

Protons and neutrons are collectively referred to as *nucleons*. Recall that the masses of the neutron and proton are *very* similar (see Eqs. (6.1.3) and (6.1.4)), suggesting that these two particles should be related somehow (by a slightly broken symmetry?). Nucleons are known to have internal structure: they may be regarded as bound states of three quarks. We will later be discussing quarks, and

³Recall that in our discussion in Chapter 5 of transformations about the 3-axis by an “angle” α spin 1/2 systems picked up a phase proportional to $\alpha/2$ while spin 1 systems (vector systems) picked up a phase proportional to α . We further noted that this result is true for a 2π “rotation” about any of the axes, Eq. (5.4.6).

their possible bound states, in much greater detail. For now, we simply note that the observational evidence for quarks is necessarily somewhat indirect. It turns out that scattering experiments with nucleons *cannot* liberate free quarks. Why is that? Good question and we will have more on this point later. This apparent disconnect between the degrees of freedom in the theory (the quarks) and the degrees of freedom observed in the lab (the hadrons) has constituted one of the major intellectual challenges in particle physics. The previous experience has always been that you could “take things apart”, molecules \rightarrow atoms, atoms \rightarrow electrons and nuclei, nuclei \rightarrow neutrons and protons. When we try to take protons apart, we find more protons, plus pions and kaons, but no isolated quarks! On the spatial resolution scales larger than a fermi, we believe that quarks are *always* “confined” inside hadrons. We need to become comfortable with this new set of rules.

Interestingly, the “internal gear wheels” story currently ends here. No evidence for internal structure within quarks, or electrons, has yet been found. If quarks and leptons are discovered someday to be composite objects, bound states of some not-yet-known more fundamental constituents, then the length scale on which this binding occurs must be at least three orders of magnitude smaller than the femtometer (fm) scale of nucleons. This limit on the length scale is set by the corresponding energy scale (TeV) of the experimental measurements at both the Tevatron and the LHC, which, until now, have not exhibited any internal structure for quarks or leptons..

6.2 Stability of particles

Are protons, or electrons, or hydrogen atoms stable? Or can they spontaneously decay? In other words, if one of these particles (or atoms) is completely isolated, in a vacuum, can it eventually, spontaneously fall apart? It is important to recognize that this is a “bad” question. It is fundamentally unanswerable — because feasible experiments must necessarily last only a finite length of time. If there is no known evidence that a certain type of particle *can* decay, then the question one should ask is what *limits* can be placed on the stability of the particle.

For protons and electrons, we have no evidence whatsoever that these particles are unstable, and experimental bounds on the lifetimes of these particles, if they do decay, are very long,

$$\text{proton lifetime } \tau_p > 2.1 \times 10^{29} \text{ yr}, \quad (6.2.1)$$

$$\text{electron lifetime } \tau_e > 4.6 \times 10^{26} \text{ yr}. \quad (6.2.2)$$

We should be impressed with these limits, considering that they vastly exceed the age of the Earth (a mere 4.5 billion years) and the Universe (over 13 billion years). Suppose, hypothetically, that protons do decay with a lifetime of 10^{30} years. How could one ever know? The direct approach of watching one particle for 10^{30} years is obviously impractical. But if you can watch many identical particles simultaneously, and detect if (and when) a single one of them decays, then extremely long lifetimes can be measured.⁴ A cubic meter of water contains 2.7×10^{29} protons (and the same number of electrons). So if $\tau_p = 10^{30}$ yrs, then within a tank holding 100 cubic meters of water, 27 protons (on average) will decay every year. The challenge is in designing and operating an experiment which can detect the decay of individual protons within a large quantity of material. While the development

⁴The lifetime τ of an unstable particle is, by definition, the time interval (in its rest frame) for which the probability of the particle decaying is $1/e$. If you start with N_0 identical particles, then the mean number of particles which will *remain* after time t is given by $N(t) = N_0 e^{-t/\tau}$. If $N_0 \gg 1$ then, on average, one particle will have decayed by the time $t_1 = \tau/N_0$, since $N(t_1) \approx N_0 - 1$.

of such detectors (essentially instrumented large tanks of water) has not yet led to the observation of proton decay, it has resulted in detectors capable of detecting the neutrinos from our sun and from supernovas elsewhere in the galaxy. This is, in fact, a very nice story of the synergies that drive science. The initial push was to detect proton decay (yielding only a limit until now), but the technology developed contributed to the very exciting (and unexpected) discovery that neutrinos are not massless!

Next consider neutrons, the other basic constituents of nuclei besides protons. Unlike protons, an isolated neutron is known to be unstable, with a lifetime of about 15 minutes. The products of the decay are a proton, an electron, and a less familiar particle called an *electron antineutrino*, denoted $\bar{\nu}_e$. This decay is represented symbolically as

$$n \longrightarrow p + e^- + \bar{\nu}_e. \quad (6.2.3)$$

This decay process is referred to as a *beta decay*,⁵ and is a consequence of interactions known as *weak interactions*, which will be discussed more fully in a later chapter. Neutrinos are nearly massless, spin-1/2 particles which interact extremely weakly with ordinary matter and as a result are very difficult to detect. They come in several different types (distinguished by the charged lepton with which they are correlated by the weak interaction), and exhibit interesting quantum-mechanical phenomena which we will also examine later.

Although a single free neutron is unstable, when neutrons bind with protons to form nuclei the resulting bound states are, in many cases, effectively stable (meaning that their lifetimes, if finite, are in excess of billions of years). Such stable nuclei include deuterium (${}^2_1\text{H}$) which is a bound state of one proton with one neutron,⁶ helium-3 (${}^3_2\text{He}$) which contains two protons and one neutron, helium-4 (${}^4_2\text{He}$) consisting of two protons and two neutrons, and many progressively heavier nuclei (recall Figure 6.4) up to bismuth-209 (${}^{209}_{83}\text{Bi}$) which is the heaviest (known) nucleus that is essentially stable.⁷

6.3 Nuclear decays

In addition to (apparently) stable bound states, there are many more unstable nuclei with lifetimes that range from very long, billions of years, down to very short, less than femtoseconds. Stable nuclei have roughly the same number of protons and neutrons (or in heavier nuclei, slightly more neutrons than protons, recall Fig. 6.4). Many nuclei with an excess of neutrons, relative to the number of protons, undergo beta decay. This converts a neutron within the nucleus into a proton, while emitting an electron and an antineutrino. For example,

⁵This is a historical name which dates from the early 1900s, when three distinct types of radioactive decay, called α , β , and γ , had been identified. The different decay types were distinguished by the degree to which the particles emitted in the decay could penetrate ordinary matter. Alpha decays produce particles with very little penetrating power which were later identified as helium-4 nuclei. Gamma decays produce extremely penetrating particles, later identified to be high energy photons (“gamma rays”). Beta decays produce particles which penetrate farther than alphas, but less than gammas. These were subsequently identified to be electrons.

⁶Recall the notation for nuclei: the symbol ${}^A_Z\text{Sy}$ denotes the nucleus of the element “Sy” with A nucleons, of which Z are protons.

⁷In fact, bismuth-209 has recently been found to alpha decay with a lifetime of 2×10^{19} yr.

β decay		lifetime
${}^3_1\text{H}$	$\longrightarrow {}^3_2\text{He} + e^- + \bar{\nu}_e$	17.8 yr ,
${}^6_2\text{He}$	$\longrightarrow {}^6_3\text{Li} + e^- + \bar{\nu}_e$	1.16 s ,
${}^{10}_4\text{Be}$	$\longrightarrow {}^{10}_5\text{B} + e^- + \bar{\nu}_e$	2.18 Myr ,
${}^{14}_5\text{B}$	$\longrightarrow {}^{14}_6\text{C} + e^- + \bar{\nu}_e$	18.0 ms .

Some nuclei with an excess of protons, relative to the number of neutrons, can convert a proton into a neutron by capturing an electron from the cloud of electrons surrounding the nucleus, and then emitting a neutrino which carries off the excess energy,

electron capture		lifetime
${}^7_4\text{Be} + e^-$	$\longrightarrow {}^7_3\text{Li} + \nu_e$	76.9 day ,
${}^{41}_{20}\text{Ca} + e^-$	$\longrightarrow {}^{41}_{19}\text{K} + \nu_e$	1.50 Myr .

This mode of decay is only possible if the atom is not fully ionized, so that one or more electrons are bound to the nucleus. If that is not the case, neutron-poor nuclei can convert a proton into a neutron via *positron emission*. A *positron*, denoted e^+ , is a particle with the same mass as an electron, but with charge $+e$ instead of $-e$. It is an example of an *antiparticle*, discussed below. (Actually we have already “snuck” in the concept of antiparticles by mentioning both neutrinos *and* antineutrinos above.) The carbon-11 nucleus preferentially decays via positron emission even when it has an orbital electron it could otherwise capture,

positron emission		lifetime
${}^{11}_6\text{C}$	$\longrightarrow {}^{11}_5\text{B} + e^+ + \nu_e$	29.4 min .

Since this process creates a positron rather than absorbing an electron as above, the energy released by the change in the nucleus must be larger (to satisfy overall energy conservation).

Certain nuclei have multiple modes of decay with measurable rates. For example, potassium-40 (${}^{40}_{19}\text{K}$) has a lifetime of 1.8 billion years. In 89% of its decays, potassium-40 undergoes beta-decay to calcium-40, but in the remaining 11% of its decays, potassium-40 decays to argon-40 via electron capture or positron emission.

In addition to the above types of nuclear decay, in which a neutron is converted into a proton or vice-versa, some nuclei which are very proton-rich decay by simply ejecting a proton, or in some cases, an alpha particle. And some very neutron-rich nuclei simply eject a neutron.

Many excited states of nuclei decay to their ground states by emitting photons (just like excited atomic states). But in the case of nuclei, excited state energies are typically in the range of several MeV, so the photons emitted in nuclear decays are in the *gamma ray* portion of the electromagnetic spectrum (to the far right in Figure 6.5).

6.4 Photons

One other elementary particle which plays a major role in innumerable aspects of everyday life is the *photon*. Photons are quantized excitations of the electromagnetic field, and are the “force carriers” for the electromagnetic interaction (we say that an electromagnetic interaction has occurred when

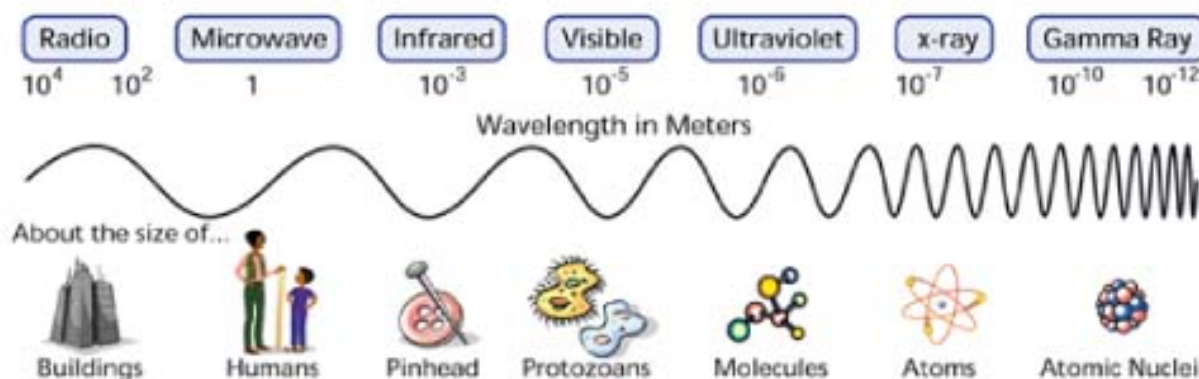


Figure 6.5: Illustration of the spectrum of electromagnetic waves.

a photon is exchanged). They have no rest mass (unlike the other particles we have discussed so far). The spectrum of electromagnetic radiation is illustrated in Figure 6.5 from long wave lengths (low energy photons) on the left to short wave lengths (large energy photons) on the right. Photons carry one unit of angular momentum, in units of \hbar . We typically express this point by the phrase “photons are spin 1 particles.”

In everyday life, quantum aspects of the electromagnetic field are not readily apparent. For a great many applications, a classical treatment of electromagnetism suffices, *i.e.*, the number of photons present is enormous and we cannot readily detect the emission or absorption of a single photon. But the quantized nature of light is revealed in phenomena such as the photoelectric effect, the presence of stimulated emission in lasers and masers, and the operation of sensitive photo-diodes which can detect *single* photons.⁸

6.5 Antiparticles

Early studies of cosmic rays revealed the existence of *positrons*, particles with the same mass as electrons but opposite charge. When a positron collides with an ordinary electron, they can interact and scatter like any two electrically charged particles. However, they can also under go a very special process, allowed only for the case of particle and antiparticle. They can annihilate with each other and produce (only) photons,

$$e^+ + e^- \longrightarrow \gamma + \gamma.$$

Accelerator-based scattering experiments have also revealed the existence of *antiprotons* and *antineutrons*, denoted \bar{p} and \bar{n} , respectively. They can similarly annihilate with their ordinary partners to produce photons,

$$\begin{aligned} p + \bar{p} &\longrightarrow \gamma + \gamma, \\ n + \bar{n} &\longrightarrow \gamma + \gamma. \end{aligned}$$

⁸Human vision, when fully dark-adapted, can nearly detect single photons of visible light. See, for example, the classic paper *Energy, Quanta, and Vision* by Hecht, Shlaer and Pirenne.

When one combines quantum mechanics and special relativity (leading to relativistic quantum field theory), a remarkable theoretical prediction is that antiparticles *must* exist. Charged particles must have distinct antiparticles with exactly the same mass and spin, but opposite electric charge. For certain neutral particles, such as the photon, there is no distinction between particle and antiparticle — one can say that the photon is its own antiparticle. At the moment it is unclear whether the other really neutral particle, the neutrino, is its own antiparticle or not. Our friends at CENPA (the UW physics laboratory on the other side of campus) are involved in the MAJORANA experiment that will test this idea by looking for neutrino-less double beta decays. Such decays can occur only if neutrinos are their own antiparticle (such self-conjugate fermions are called Majorana fermions). (Note that, while electrically neutral, the neutron does carry another “charge” called baryon number and is unambiguously distinct from the antineutron.) Although antimatter is not present in everyday life (this is another mystery - why is the universe so antisymmetric between matter and antimatter, which was presumably not the case at the time of the big bang?), antiparticles do exist, and the laws of nature are almost, but not quite, symmetric under the interchange of ordinary matter and antimatter. We will discuss these issues further in a later chapter.

6.6 Leptons

Electrons (e^-) and electron neutrinos (ν_e) are members of a class of particles known as *leptons*. Their antiparticles, the positron (e^+) and electron antineutrino ($\bar{\nu}_e$), are *antileptons*. Leptons (and antileptons) are spin 1/2 particles. All leptons participate in the weak interactions (*leptos* is Greek for weak) and the (electrically) charged leptons also participate in the electromagnetic interactions. However, leptons do not participate in the strong or nuclear interactions (*i.e.*, they are *not* bound states of quarks). In addition to the electron, two other charged leptons are known, the muon (μ^-) and the tau (τ^-). As the superscripts indicate, these particles are negatively charged; their charge is (apparently) *identical* to that of the electron. Their antiparticles are the antimuon (μ^+) and antitau (τ^+). There are distinct neutrinos associated, through the weak interactions, with each charged lepton. In addition to the electron neutrino, there is a muon neutrino (ν_μ) and a tau neutrino (ν_τ), as well as the corresponding antineutrinos ($\bar{\nu}_\mu$, $\bar{\nu}_\tau$). So an important question is - why 3 kinds of leptons?

The basic properties of the leptons are summarized in Table 6.1. The electric charge listed is in units of $|e|$. Neutrinos have much smaller rest masses than the charged leptons, so much smaller that it is extraordinarily difficult to measure neutrino masses (and we have not yet succeeded). On the other hand, the observation of neutrino oscillations, which will be discussed in a later chapter, implies that neutrinos must have non-zero masses. But at the moment only an upper bound on the actual values of the neutrino masses is known. (Note that research groups within the UW Department of Physics have played important roles in the experiments leading to our current understanding of the properties of neutrinos, and continue to do so.)

As indicated in Table 6.1, the “heavy” leptons decay into the light ones. The muon decays into an electron plus an electron antineutrino and a muon neutrino. The heavier tau has more options, decaying to both electrons and muons, *and* into a final state with hadrons (2 pions) and just the single lepton (the tau neutrino). In all of these processes “lepton number” is conserved. *Lepton number*, denoted L , is defined as the total number of leptons minus antileptons,

$$L \equiv (\# \text{ leptons}) - (\# \text{ antileptons}). \quad (6.6.1)$$

All known interactions conserve lepton number.⁹ In fact, the dominant interactions (but not, for example, neutrino oscillations) conserve lepton number separately for each lepton “flavor”, electron, muon and tau. So e^- and ν_e have $L_e = +1$, while e^+ and $\bar{\nu}_e$ have $L_e = -1$. Similar definitions hold for L_μ (+1 for μ^- and ν_μ , -1 for μ^+ and $\bar{\nu}_\mu$) and L_τ (+1 for τ^- and ν_τ , -1 for τ^+ and $\bar{\nu}_\tau$). In the decays noted in the table below we see, for example, $\mu^- \rightarrow e^- \bar{\nu}_e \nu_\mu$ with $L_e = 0$, $L_\mu = 1$ in both the initial and final states (and similarly for the decays of the τ^-). We will see later that this structure is built into the (perturbative) definition of the weak interactions.

particle	rest energy	lifetime	dominant decay	charge	L
ν_e	< 2 eV	\approx stable	—	0	1
ν_μ	< 0.19 MeV	\approx stable	—	0	1
ν_τ	< 18.2 MeV	\approx stable	—	0	1
e^-	0.511 MeV	stable	—	-1	1
μ^-	105.7 MeV	2.2 μ s	$e^- \bar{\nu}_e \nu_\mu$	-1	1
τ^-	1777 MeV	0.29 ps	$\pi^- \pi^0 \nu_\tau, e^- \bar{\nu}_e \nu_\tau, \mu^- \bar{\nu}_\mu \nu_\tau$	-1	1
$\bar{\nu}_e$	< 2 eV	\approx stable	—	0	-1
$\bar{\nu}_\mu$	< 0.19 MeV	\approx stable	—	0	-1
$\bar{\nu}_\tau$	< 18.2 MeV	\approx stable	—	0	-1
e^+	0.511 MeV	stable	—	+1	-1
μ^+	105.7 MeV	2.2 μ s	$e^+ \nu_e \bar{\nu}_\mu$	+1	-1
τ^+	1777 MeV	0.29 ps	$\pi^+ \pi^0 \bar{\nu}_\tau, e^+ \nu_e \bar{\nu}_\tau, \mu^+ \nu_\mu \bar{\nu}_\tau$	+1	-1

Table 6.1: Leptons and antileptons.

⁹Actually, this is not quite true. The current theory of weak interactions predicts that there are processes which can change lepton number. However, the rate of these processes is so small that lepton number violation is completely unobservable.

Chapter 7

Quarks and hadrons

Every atom has its ground state — the lowest energy state of its electrons in the presence of the atomic nucleus — as well as many excited states, which can decay to the ground state via emission of photons. Nuclei composed of multiple protons and neutrons also have their ground state plus various excited nuclear energy levels, which typically also decay via emission of photons (plus α and β radiation). But what about individual protons or neutrons?

It was asserted earlier that individual nucleons are also composite objects, and may be viewed as bound states of quarks. And just as atoms and nuclei have excited states, so do individual nucleons.

The force which binds quarks together into bound states is known as the *strong interaction*, and the theory which describes strong interactions (on distance scales small compared to a fermi) is called *quantum chromodynamics*, often abbreviated as QCD. The quarks carry a corresponding charge, the analogue of electric charge, which is labeled “color” charge, leading to the “chromo” in the name. Unlike electric charge, for which there is only a single variety - either plus or minus (with the underlying symmetry group $U(1)$), there are three possible “colors” for the color charge of a quark, along with the corresponding “anti-colors”. The group describing the underlying symmetry is $SU(3)$.¹ We will have more to say about QCD as we progress. But the justification for the validity of the following qualitative description of quarks and their bound states lies in the success of QCD as a description of what is observed in the laboratory. Using this theory, one can do detailed quantitative calculations of the masses and other properties of bound states of quarks and compare with experimental results. The theory works. (In fact, the story of how this theory has been verified in experiment, even though the theory has quarks as degrees of freedom, while experiment *never* detects individual quarks, is an interesting one indeed. We will have only a limited opportunity to discuss it this quarter, but I will note that an essential feature of this story is the emergence of “jets” of hadrons in the final states of high energy collisions, and I have spent much of my scientific career clarifying both the theoretical and experimental properties of these jets.)

In this Chapter we will introduce a variety of new concepts, which we will return to in more detail in subsequent discussion. In particular, we will be using several techniques from group theory. Now would be a good time to read both Chapters 10 and 11.

¹In the group theory language of Chapter 10 $U(1)$ is an Abelian group and the particle corresponding to the interaction, the photon, does *not* interact directly with itself, *i.e.*, the photon has zero electric charge. $SU(3)$, on the other hand, is a non-Abelian group and the gluons, the analogs of the photon, come in 8 varieties and *do* interact directly with each other, *i.e.*, have nonzero color charge.

7.1 Quark flavors

Quarks are spin-1/2 particles (fermions), which come in various species, (playfully) referred to as *flavors*. Different quark flavors have been given somewhat whimsical names, as shown in Table 7.1 (values from the PDG). Note that the table includes values for the quark “masses”, but care must be taken when interpreting these values as individual, *isolated* quarks are *never* observed experimentally. On the other hand, it should be clear that the masses for different flavors vary substantially.

flavor	symbol	mass	charge
up	u	$\approx 2.3^{+0.7}_{-0.5} \text{ MeV}/c^2$	$\frac{2}{3} e $
down	d	$\approx 4.8^{+0.5}_{-0.3} \text{ MeV}/c^2$	$-\frac{1}{3} e $
strange	s	$\approx 95 \pm 5 \text{ MeV}/c^2$	$-\frac{1}{3} e $
charm	c	$1.275 \pm 0.025 \text{ GeV}/c^2$	$\frac{2}{3} e $
bottom	b	$4.18 \pm 0.03 \text{ GeV}/c^2$	$-\frac{1}{3} e $
top	t	$173.07 \pm 0.52 \pm 0.72 \text{ GeV}/c^2$	$\frac{2}{3} e $

Table 7.1: Known quark flavors

Along with quarks, there are, of course, also *antiquarks*, denoted \bar{u} , \bar{d} , \bar{s} , etc., with the *same* masses but opposite electric charge as their partner. (So, for example, the \bar{u} antiquark has charge $-2/3$ and the \bar{d} has charge $+1/3$ - note the *non-integer* values.) As suggested above, quarks are distinguished from leptons by an additional quantum number that is called *color*, which takes three possible values: say red, blue, or green (and anti-red, anti-blue and anti-green for the antiquarks). These names are simply labels for different quantum states of the quark.² Since quarks have spin 1/2, they can also be labeled by their spin projection, \uparrow or \downarrow , along any chosen spin quantization axis. Hence, for each quark flavor, there are really six different types of quark, distinguished by the color (red, blue, green) and spin projection (up, down).

In addition to the curious names, two other things in Table 7.1 should strike you as odd: the enormous disparity of masses of different quarks, spanning five orders of magnitude, and the fact that quarks have fractional charge (in units of $|e|$). Both issues are at the core of ongoing research, including that at the LHC, seeking evidence of the dynamics of mass generation (now apparently at least partially clarified by the discovery of the Higgs boson) and the connection between quarks and leptons.

7.2 Hadrons

No (reproducible) experiments have detected any evidence for free, *i.e.*, isolated quarks. Moreover, there is no evidence for the existence of any isolated charged particle whose electric charge is not an integer multiple of the electron charge. This is referred to as *charge quantization*. Consistent with these observational facts, the theory of strong interactions predicts that quarks will always be

²These names are purely conventional — one could just as well label the different “color” states as 1, 2, and 3. But the historical choice of names explains why the theory of strong interactions is called quantum chromodynamics: a quantum theory of the dynamics of “color” — although this color has nothing to do with human vision!

trapped inside bound states with other quarks and antiquarks, never separated from their brethren by distances larger than about a fermi.³ The bound states produced by the strong interactions are called *hadrons* (*hadros* is Greek for strong).

Quantum chromodynamics, in fact, predicts that only certain types of bound states of quarks can exist, namely those which are “colorless”. (This can be phrased in a mathematically precise fashion in terms of the symmetries of the theory. More on this later. For now consider this situation as being similar to that in atomic physics where the low energy states, the atoms, are electrically neutral.) Recall that to make white light, one mixes together red, blue, and green light. Similarly, to make a colorless bound state of quarks one sticks together three quarks, one red, one blue, and one green. But this is not the only way. Just as antiquarks have electric charges which are opposite to their partner quarks, they also have “opposite” color: anti-red, anti-blue, or anti-green. So another way to make a colorless bound state is to combine three antiquarks, one anti-red, one anti-blue, and one anti-green. A final way to make a colorless bound state is to combine a quark and an antiquark, say red and anti-red, or better the truly colorless combination $r\bar{r} + b\bar{b} + g\bar{g}$. Bound states of three quarks are called *baryons*, bound states of three antiquarks are called *antibaryons*, both of which are fermions, while quark-antiquark bound states are called *mesons* and are bosons.

How these rules emerge from QCD will be described in a bit more detail later. For now, let’s just look at some of the consequences. The prescription that hadrons must be colorless bound states says nothing about the flavors of the constituent quarks and antiquarks. In the language of quantum mechanics we say that color dependent operators and flavor dependent operators commute,

$$[\text{color}, \text{flavor}] = 0. \quad (7.2.1)$$

Since quarks come in the multiple flavors of Table 7.1, we can (and will) enumerate the various possibilities for the hadrons. Similar comments apply to the spatial (orbital angular momentum) and spin parts of the wave function (*i.e.*, the dynamics of these various parts commute to a good approximation), and we can think of the wave functions describing the hadrons, again to a good approximation, as *products* of a color wave function, a flavor wave function, a spatial wave function and a spin wave function. The most important violation of this assumption of the factorization of the wave function arises due to the role of the Pauli Exclusion Principle for baryons, as we will shortly see.

For the lowest mass hadrons we may assume that the quarks are essentially at rest (except for the constraints of quantum mechanics) and that the orbital angular momentum vanishes, $\vec{L} = 0$. Thus the rest energy of such a hadron (like any bound state, although it is admittedly a bit more subtle in this case) may be regarded as the sum of the rest energies of its constituents plus the energy associated with the binding interaction. For the lowest mass hadrons, including nucleons, we will see that most of their total rest energy (mass) comes from the binding energy, *i.e.*, the “gluon cloud”, the analog of the electromagnetic potential for an atom. But the masses of quarks also play a part. Looking at the quark masses listed in Table 7.1, it is apparent that *u*, *d* and *s* quarks are quite light compared to the mass ($\approx 1 \text{ GeV}/c^2$) of a nucleon, while the other quark flavors are considerably

³Except at sufficiently high temperatures. Above a temperature of $T_c \approx 2 \times 10^{12} \text{ K}$ (or $kT \approx 170 \text{ MeV}$), hadrons “melt” or “vaporize” and quarks are liberated. This is important in the physics of the early universe, since temperatures are believed to have exceeded this value in the earliest moments after the big bang. Temperatures above T_c can also be produced, briefly, in heavy ion collisions. A nice overview of heavy ion collisions and quark gluon plasma may be found at www.bnl.gov/rhic/physics.asp. There is also an ongoing heavy ion program at the LHC, *i.e.*, some running time is dedicated to accelerating and colliding heavy nuclei rather than protons.

heavier. So it should not be surprising that the lightest hadrons will be those which are bound states of u and d quarks. Further, since a proton is composed of uud while a neutron is udd (*i.e.*, the difference is only $1u \rightarrow 1d$), we are gratified to see that m_n is only slightly larger than m_p (see Eqs. (4.1.3) and (4.1.4)). Substituting a strange quark for a u or d quark should be expected to raise the mass of the resulting bound by roughly 100 MeV (due to the larger s quark mass). And hadrons containing the other quark flavors (c , b , or t) should be substantially heavier.

As noted above, when enumerating possible combinations of quarks that could form hadrons, we must also consider the role of spin and flavor (electric charge, strangeness, *etc.*). For this purpose it is helpful to recall what you have learned in Quantum Mechanics (recall Chapter 5) and then go a step further to discuss how to combine spins (consistently with the principles of QM, see Chapter 11). In particular, combining the angular momentum of two spin 1/2 particles, in the case that there is no orbital angular momentum (*i.e.*, for $\vec{L} = 0$ in the ground state) can yield either spin 1 or 0. Importantly, the corresponding states have definite symmetry under the interchange of the two spins (symmetric for the spin 1 state and antisymmetric for the spin 0). Three spin 1/2 particles can combine to form either spin 3/2 or 1/2 (but *not* spin 1 or 0).

Let us review a bit of what you (may have) learned in your Quantum Mechanics course, some of which was discussed in Chapter 5. The new issue here is the question of *combining* the spin and flavor of the quarks to find the spin and flavor of the hadrons (see Chapters 10 and 11). (Your QM course may have included a discussion of combining spin and orbital angular to yield the total angular momentum.) The arithmetic of the construction of the combined spin states (*i.e.*, the wave functions) is encoded in the so-called Clebsch-Gordan coefficients ($\langle s_1 m_1, s_2 m_2 | SM \rangle$, where s_1 and s_2 are the two spins, m_1 and m_2 are the spin components along say the \hat{x}^3 direction and S and M are the combined (total) spin and \hat{x}^3 component). A table of these coefficients is included at the end of this chapter. The action of the raising/lowering or “ladder” operators, \hat{S}_{\pm} , allows us to change the value of m for a fixed value of s . As derived in Chapter 5 these operators obey the relation (the “ket” $|s, m\rangle$ represents a state of definite s and m)

$$\hat{S}_{\pm}|s, m\rangle = \sqrt{(s \mp m)(s \pm m + 1)}\hbar|s, m \pm 1\rangle. \quad (7.2.2)$$

Thus the (normalized) combined states of two spin 1/2 particles can be represented (in a hopefully familiar and obvious notation) as

$$|S, M\rangle = |1, 1\rangle = |\uparrow\uparrow\rangle, \quad |1, 0\rangle = (|\uparrow\downarrow\rangle + |\downarrow\uparrow\rangle)/\sqrt{2}, \quad |1, -1\rangle = |\downarrow\downarrow\rangle, \quad (7.2.3)$$

and

$$|0, 0\rangle = (|\uparrow\downarrow\rangle - |\downarrow\uparrow\rangle)/\sqrt{2}. \quad (7.2.4)$$

This notation makes explicit the fact that the spin 1 state is *symmetric* under the interchange of the two individual spins, while the spin 0 state is *antisymmetric*. (Recall our theme that symmetries are almost always the answer!). In the language of Group Theory and Representations (see Chapters 10 and 11, here we are thinking of representations of either $SO(3)$ or $SU(2)$, which are identical for our current purposes) we call spin 1/2 a doublet representation (two states, spin up and spin down) and label it as $\underline{2}$. Thus the combination and subsequent reduction (into irreducible representations) of two spin 1/2 states can be written

$$\underline{2} \otimes \underline{2} = \underline{3} \oplus \underline{1}, \quad (7.2.5)$$

where the triplet is the spin 1 state (corresponding to the 3 values of M) and the singlet is the spin 0 state (with only a single M value). The symbol \otimes tells us that we are *combining* the two doublets

on the left-hand-side, while \oplus means we are decomposing the result into irreducible representations on the right-hand-side. Note that our naive algebraic expectations are met by the counting of states in Eq. (7.2.5). The left-hand-side corresponds to $2 \times 2 = 4$ states while the right-hand-side again has $3 + 1 = 4$ states. Since the singlet representation has only 1 element, it is, as expected, invariant under rotations as a rotation acts to transform the elements of a representation into each other (and no transformation can occur for a single element). On the other hand, the 3 elements of the spin 1 (vector) representation can be transformed into one another by a rotation (as we discussed in some detail in Chapter 5).

The corresponding expression describing the combination of 3 spin 1/2 states is

$$\underline{2} \otimes \underline{2} \otimes \underline{2} = \underline{4} \oplus \underline{2} \oplus \underline{2}. \quad (7.2.6)$$

Thus we obtain the expected spin 3/2 state (the quartet) and *two* spin 1/2 states, corresponding to differing internal symmetry; think of combining two of the spins to yield either spin 1 or spin 0 as above and then combine the third spin to yield spin 3/2 and two forms of spin 1/2 (two doublets - either $\underline{3} \otimes \underline{2} = \underline{4} \oplus \underline{2}$ or $\underline{1} \otimes \underline{2} = \underline{2}$). (See the discussion in Chapter 11 for more about the “technology” of combining multiplets of the $SU(n)$ groups.) Again the total number of states is unchanged as $2 \times 2 \times 2 = 8$ matches $4 + 2 + 2 = 8$. Similarly to the structure in Eq. (7.2.5), the quartet state, $\underline{4}$, is symmetric under interchange of any pair of the spins.

Combining electric charge is easy: the charge of a hadron is just the *algebraic* sum of the charges of its constituent quarks. Combining the rest of the quarks’ flavor is more complicated but we can use the same technology as we used to combine spin above (recall the words of the famous theoretical physicist Richard Feynman, “The same equations have the same solutions.”). In particular, we can treat the nearly degenerate (nearly equal mass) u and d quarks as being members of a doublet representation of some underlying approximate $SU(2)$ symmetry and combine as above to find the corresponding flavor states. Since this structure parallels that for ordinary spin, the corresponding quantum number was historically labeled *isospin*, and we will use it to understand the flavor structure of the hadrons.

7.3 Mesons

Let us start with mesons and (for the moment) consider those “constructed” from just the lightest two flavors of quarks and antiquarks, u, \bar{u} and d, \bar{d} . We know that the color wave function is the trivial color singlet $r\bar{r} + b\bar{b} + g\bar{g}$ (*i.e.*, the singlet, $\underline{1}$, in the *color* $SU(3)$ representation decomposition $\underline{3} \otimes \bar{\underline{3}} = \underline{8} \oplus \underline{1}$). For the *lowest* mass states again we expect that there is no orbital angular momentum ($\vec{L} = 0$) and the spatial wave function is trivial. From Eq. (7.2.5) we know that the spin wave function can be reduced to either a spin singlet (spin 0 or scalar particle) or spin triplet (spin 1 or vector particle). Taking the u and d quarks to form an isospin doublet (as outlined above) we combine to find 4 flavor states as outlined in the first line of Table 7.2, where the lines are labeled by the number of strange plus antistrange quarks. Using the isospin language (in just the same way as we use the spin language - again “The same equations have the same solutions”) we can define these states in terms of “total isospin I ” and a single component I_3 . Historically the phases chosen for the antiquark doublet are $\bar{q}^T = (-\bar{d}, \bar{u})$ (note the minus sign which, although unmotivated here, does lead to some attractive features and you will see it in the literature, although sometimes with the

$(\#s) + (\#\bar{s})$	$Q = 1$	$Q = 0$	$Q = -1$
0	$u\bar{d}$	$u\bar{u}, d\bar{d}$	$d\bar{u}$
1	$u\bar{s}$	$s\bar{d}, d\bar{s}$	$s\bar{u}$
2		$s\bar{s}$	

Table 7.2: Possible light quark-antiquark combinations

minus sign attached to the \bar{u} instead of the \bar{d}). Thus the corresponding mesons (of definite isospin) are defined as

$$|I, I_3\rangle = |1, 1\rangle = -|u\bar{d}\rangle, \quad |1, 0\rangle = (|u\bar{u}\rangle - |d\bar{d}\rangle)/\sqrt{2}, \quad |1, -1\rangle = |d\bar{u}\rangle, \quad (7.3.1a)$$

$$|0, 0\rangle = (|u\bar{u}\rangle + |d\bar{d}\rangle)/\sqrt{2}. \quad (7.3.1b)$$

With little extra effort we can expand this discussion to include the *three* lightest quark flavors, u , d , and s , where we treat the s quark as an isospin singlet but carrying the “new” quantum number strangeness. Since it was mesons with the \bar{s} antiquark that were observed first (and labeled as having strangeness +1), the s quark is actually defined to have strangeness -1! By simple counting we see that we have added 4 extra possibilities to Table 7.2 in the second and third lines. We can interpret the states in Table 7.2 more precisely if we take the 3 lightest quarks to be members of a triplet of an approximate $SU(3)$ flavor symmetry, which is apparently even more badly “broken” than the $SU(2)$ of isospin, *i.e.*, the s quark mass differs by about 100 MeV/ c^2 . Without further discussion here (see Chapters 10 and 11) we will simply assert that, when we combine a triplet and anti-triplet of $SU(3)$, we obtain the expected 9 flavor states in the form

$$\mathbf{3} \otimes \bar{\mathbf{3}} = \mathbf{8} \oplus \mathbf{1}. \quad (7.3.2)$$

(Note this is the same structure we asserted above for color $SU(3)$ - the same equations have the same solutions, only the labels change). In particular, one of the states is a singlet under this flavor $SU(3)$ and we see a new representation, the octet (8 different individual states characterized by differing values of I, I_3 and strangeness).

ASIDE As noted earlier, in “color space” $q\bar{q}$ includes both an color octet and a color singlet. The latter is the singlet color state that we have claimed is the physical meson, while the former is precisely the description of the coupling between a quark pair and a gluon, where gluons are members of the octet, $\mathbf{8}$, representation (the so-called “adjoint” representation) of color $SU(3)$.

As suggested by Table 7.2, we want to be able represent the $\mathbf{8}$ of meson flavor in a 2-D form using I_3 (to the right) and strangeness (up) as the axes. So instead of the 1-D structure for the representations of $SU(2)$ in Eq. (7.2.2), described by the single quantum number m , for $SU(3)$ the representations are 2-D and labeled by the *two* quantum numbers I_3 and S (strangeness). We illustrate this 2-D idea first by (graphically) representing the quark and antiquark, $\mathbf{3}$ and $\bar{\mathbf{3}}$, including the minus sign in the antiquark in Table 7.3 (on the next page).

$$\begin{array}{ccccc} \underline{3} & d & u & & \bar{\underline{3}} & \bar{s} & \uparrow S \\ & & & \text{and} & & & \\ & s & & & \bar{u} & -\bar{d} & \rightarrow I_3 \end{array}.$$

Table 7.3: Flavor Triplet quarks and antiquarks

Next the flavor $\underline{8}$ can be represented as

$$\begin{array}{ccccccc} \underline{8} & d\bar{s} & & u\bar{s} & \uparrow S & & \\ & d\bar{u} & (u\bar{u} - d\bar{d})/\sqrt{2} & & -u\bar{d} & & \rightarrow I_3 \\ & & (u\bar{u} + d\bar{d} - 2s\bar{s})/\sqrt{6} & & & & \\ & s\bar{u} & & -s\bar{d} & & & \end{array}.$$

Table 7.4: Quark - antiquark octet

Finally we can combine the flavor (isospin and strangeness) of the quark and antiquark in a meson to find the flavor $\underline{1}$ in this notation as,

$$\underline{1} \quad (u\bar{u} + d\bar{d} + s\bar{s})/\sqrt{3}, \quad (7.3.3)$$

which we recognize as having the basic structure as the $SU(3)$ *color* singlet we mentioned earlier (*i.e.*, again “the same equations have the same solutions”).

Note that there are three *orthogonal* strangeness zero, $I_3 = 0$ states. One is the $SU(3)$ singlet (Eq. (7.3.3)), one is the isoscalar ($I = 0$) state in the $SU(3)$ octet (the lower line in the middle of Table 7.4), and the last is the $I_3 = 0, I = 1$ member of the $SU(3)$ octet (the upper line in the middle of Table 7.4).

ASIDE To make the concept of *orthogonal* more explicit think of a 3-D vector space defined by orthogonal “unit vectors” $|u\bar{u}\rangle$, $|d\bar{d}\rangle$ and $|s\bar{s}\rangle$, where

$$\langle u\bar{u}|u\bar{u}\rangle = 1, \text{ etc.}, \langle d\bar{d}|u\bar{u}\rangle = 0, \text{ etc.} \quad (7.3.4)$$

Then it follows that

$$\langle u\bar{u} - d\bar{d}|u\bar{u} + d\bar{d} + s\bar{s}\rangle = 1 - 1 = 0, \text{ etc.} \quad (7.3.5)$$

Since we expect that both the $SU(3)$ of flavor and the $SU(2)$ of isospin are “broken” symmetries in nature, we should anticipate that the quark content of the *physical* mesons with zero strangeness and zero electric charge may be mixtures of the combinations indicated above(as we will discuss shortly).

Summarizing the above discussion there are $3 \times 3 = 9$ different flavor possibilities for *both* spin 0 and spin 1. The actual observed lowest-mass mesons do indeed fall into just this pattern. In the same tabular form the names of the observed scalar mesons are indicated in Table 7.5. The corresponding “nonet” of vector mesons are labeled as in Table 7.6.

The “small print” associated with these tables includes the following. First, note that we have labeled these multiplets in terms of the spin but with superscript “-”, as in 0^- and 1^- . This label serves

$0^- \text{ } \underline{8}$	K^0	K^+	$\uparrow S$
	π^-	$\pi^0 (I = 1)$	$\pi^+ \rightarrow I_3$
		$\eta (I = 0)$	
	K^-	\bar{K}^0	
$0^- \text{ } \underline{1}$	$\eta' (I = 0, S = 0) .$		

Table 7.5: Scalar mesons

$1^- \text{ } \underline{8}$	K^{*0}	K^{*+}	$\uparrow S$
	ρ^-	$\rho^0 (I = 1)$	$\rho^+ \rightarrow I_3$
		$\omega (I = 0)$	
	K^{*-}	\bar{K}^{*0}	
$1^- \text{ } \underline{1}$	$\phi (I = 0, S = 0) .$		

Table 7.6: Vector mesons

to remind us, as we will discuss in more detail shortly, that these particles have negative intrinsic parity. We will come to understand this as arising from the fact that fermions and antifermions necessarily have *opposite* intrinsic parity. Thus a quark-antiquark pair with only trivial spatial wave function ($L = 0$) must have negative parity. Next, as suggested above, mixing is observed between the strangeless, chargeless states compared to the expectations expressed in Table 7.4. In particular, while the scalar mesons seem to match Table 7.4, the physical ϕ state seems to be essentially pure $s\bar{s}$, while the ω is $(u\bar{u} + d\bar{d})/\sqrt{2}$. As we will see, this last point is suggested by the fact that the ϕ decays dominantly into $K\bar{K}$ states. Finally the structure of decays of the neutral kaons is a special story unto itself and illustrates the “near” conservation of the combined symmetry CP (parity and charge conjugation) by the weak interactions.

These states (particles) are listed again in Table 7.7 for the scalars, along with their dominant decay modes and quark content, while Table 7.8 provides the same information for light spin one mesons.⁴ These are the lightest mesons.

In general, and as expected, mesons containing strange quarks are heavier than those with no strange quarks. But among the neutral, $S = 0$ mesons, it is noteworthy (and we have already noted it) that, while the η and η' in Table 7.7 have the naively expected quark content, the ω and ϕ in Table 7.8 do

⁴As noted earlier Tables 7.7 and 7.8 list parity odd mesons, as this is the parity of quark-antiquark bound states with no orbital excitation. We will discuss parity assignments in the next chapter. Note that the mass values listed in these and subsequent tables should really have units of MeV/c^2 (or the column should be labeled “rest energy” instead of “mass”). We will become increasingly sloppy about this distinction, since one can always insert a factor of c^2 , as needed, to convert mass to energy or vice-versa, and in the end we want to set $c = 1$.

meson	mass	lifetime	dominant decays	quark content
π^0	135.0 MeV	8.5×10^{-17} s	$\gamma\gamma$	$\frac{1}{\sqrt{2}}(u\bar{u} - d\bar{d})$
π^+	139.6 MeV	2.6×10^{-8} s	$\mu^+\nu_\mu$	$u\bar{d}$
π^-	139.6 MeV	2.6×10^{-8} s	$\mu^-\bar{\nu}_\mu$	$d\bar{u}$
K^+	493.7 MeV	1.2×10^{-8} s	$\mu^+\nu_\mu, \pi^+\pi^0$	$u\bar{s}$
K^-	493.7 MeV	1.2×10^{-8} s	$\mu^-\bar{\nu}_\mu, \pi^-\pi^0$	$s\bar{u}$
K^0, \bar{K}^0	497.6 MeV			$d\bar{s}, -s\bar{d}$
K_S^0	497.6 MeV	8.95×10^{-11} s	$\pi^+\pi^-, \pi^0\pi^0$	$\frac{1}{\sqrt{2}}(d\bar{s} - s\bar{d})$
K_L^0	497.6 MeV	5.1×10^{-8} s	$3\pi^0, \pi^+\pi^-\pi^0, \pi^\pm e^\mp \nu_e, \pi^\pm \mu^\mp \nu_\mu$	$\frac{1}{\sqrt{2}}(d\bar{s} + s\bar{d})$
η	547.9 MeV	5.1×10^{-19} s	$\gamma\gamma, \pi^+\pi^-\pi^0, \pi^0\pi^0\pi^0$	$\approx \frac{1}{\sqrt{6}}(u\bar{u} + d\bar{d} - 2s\bar{s})$
η'	957.8 MeV	3.3×10^{-21} s	$\pi^+\pi^-\eta, \rho^0\gamma, \pi^0\pi^0\eta$	$\approx \frac{1}{\sqrt{3}}(u\bar{u} + d\bar{d} + s\bar{s})$

Table 7.7: Light spin zero, parity odd mesons.

meson	mass	lifetime	dominant decays	quark content
ρ^+, ρ^0, ρ^-	775.5 MeV	4.4×10^{-24} s	$\pi\pi$	$u\bar{d}, \frac{1}{\sqrt{2}}(u\bar{u} - d\bar{d}), d\bar{u}$
ω	782.7 MeV	7.8×10^{-23} s	$\pi^+\pi^-\pi^0$	$\approx \frac{1}{\sqrt{2}}(u\bar{u} + d\bar{d})$
K^{*+}, K^{*-}	891.7 MeV	1.3×10^{-23} s	$K\pi$	$u\bar{s}, s\bar{u}$
K^{*0}, \bar{K}^{*0}	895.9 MeV	1.4×10^{-23} s	$K\pi$	$d\bar{s}, s\bar{d}$
ϕ	1019.5 MeV	1.5×10^{-22} s	$K^+K^-, K_L^0K_S^0$	$\approx s\bar{s}$

Table 7.8: Light spin one, parity odd mesons.

not. This reflects the possibility of quantum mechanical mixing among states with identical quantum numbers. In other words, eigenstates of the Hamiltonian can be linear combinations of basis states which have similar quark content. The form of this mixing will be discussed in more detail later, but the important conclusion here is that the basic description of mesons as bound states of quarks works! The introduction of the flavor symmetry of the quark picture brought simplicity out the previous chaos of all of the experimentally observed meson states

Finally It is also worthwhile noting the realization in Table 7.7 of the previous comment about the decays of the neutral kaons (K^0, \bar{K}^0). As we will discuss in more detail shortly the weak interactions (and these are weak decays, since we must change the flavor of the strange quark) approximately respect CP . Hence the neutral kaons dominantly decay through states of definite CP . The “even” CP state (CP eigenvalue +1) is allowed to decay into 2 pions and decays more quickly. Hence the label $K_{S(\text{hort})}^0$. The CP odd state (CP eigenvalue -1) labeled $K_{L(\text{ong})}^0$ decays more slowly into 3 pions or the more familiar states with leptons.

7.4 Baryons

One can go through a similar exercise for baryons. The primary differences are the further algebraic complexities of dealing with 3 instead of 2 quarks, and the fact that we are now dealing with 3 *identical* fermions. The exclusion principle now plays a role. In particular, the overall wavefunction describing the 3 quarks must be *anti*-symmetric with respect to the interchange of (all of) the quantum numbers of *any* pair of quarks. So, by way of introduction, we will quickly mention the expected behavior in color, space, spin and flavor and then we will go through the construction of baryons in more detail.

In the same notation we used for mesons (see Chapter 11), we find that combining 3 color triplets yields

$$\underline{3} \otimes \underline{3} \otimes \underline{3} = \underline{10} \oplus \underline{8} \oplus \underline{8} \oplus \underline{1}. \quad (7.4.1)$$

Thus the expected $3^3 = 27$ color states break-up into a “decuplet” ($\underline{10}$), two octets ($\underline{8}$), and a singlet ($\underline{1}$). The most important feature of these representations is that the decuplet is *symmetric* under the interchange of the color quantum number of *any* pair of quarks, while the singlet is purely *antisymmetric*.⁵ The two octets have (different) “mixed” symmetry for different quark pairs (see the discussion below). Most importantly, by our rule that the only viable physical states are color singlets, we choose to put the 3 quarks in the *antisymmetric* color $\underline{1}$. Since Pauli requires the complete wave function to be antisymmetric, the remaining bits of the wavefunction must be overall *symmetric*.

Since we are interested (first) in the least massive baryons, we take the spatial wavefunction to be symmetric (and trivial) with $L = 0$. Thus, due to the statistics of fermions, the combined spin and flavor wavefunctions of the lowest mass baryons must be *symmetric* under the interchange of any pair of quarks.

Combining three spin 1/2 objects can yield either spin 1/2 or 3/2, or in our new notation

$$\underline{2} \otimes \underline{2} \otimes \underline{2} = \underline{4} \oplus \underline{2} \oplus \underline{2}, \quad (7.4.2)$$

where the spin 3/2 is the quartet that is again the symmetric state. The two spin 1/2 doublets, like the color octets above, are of mixed symmetry. To illustrate that point more explicitly we can assemble the 3 quark state by first putting together 2 quarks. Now the spin arithmetic is the same as for the quark-antiquark above, $\underline{2} \otimes \underline{2} = \underline{3} \oplus \underline{1}$, where the triplet is symmetric under interchange of the 2 quarks and the singlet is antisymmetric. To be explicit, label the first two quarks as 1 and 2, $\underline{2}_1 \otimes \underline{2}_2 = \underline{3}_{S12} \oplus \underline{1}_{A12}$ with S for symmetric and A for antisymmetric. Now include quark 3 to yield $\underline{3}_{S12} \otimes \underline{2}_3 = \underline{4}_{S123} \oplus \underline{2}_{S12,3}$ and $\underline{1}_{A12} \otimes \underline{2}_3 = \underline{2}_{A12,3}$. The differing *mixed* qualities of the resulting symmetries for the doublets should be clear from the following more explicit version of Eq. (7.4.2)

$$\underline{2}_1 \otimes \underline{2}_2 \otimes \underline{2}_3 = \underline{4}_{S123} \oplus \underline{2}_{S12,3} \oplus \underline{2}_{A12,3}. \quad (7.4.3)$$

Similar mixed underlying symmetry is what distinguishes the $SU(3)$ octets above.

⁵The structure of this antisymmetric 3-quark color wave function is the (hopefully) familiar antisymmetric form in terms of the $SU(3)$ structure constant, $\epsilon_{jkl} q_j q_k q_l$ (ϵ_{jkl} is antisymmetric under the exchange of any pair of indices). We can tell this is a singlet because there are no “left-over” indices and this state cannot be transformed.

baryon	mass	lifetime	dominant decays	quark content
p	938.3 MeV	stable	—	uud
n	939.6 MeV	8.8×10^2 s	$pe^- \bar{\nu}_e$	udd
Λ	1115.7 MeV	2.6×10^{-10} s	$p\pi^-, n\pi^0$	uds
Σ^+	1189.4 MeV	0.80×10^{-10} s	$p\pi^0, n\pi^+$	uus
Σ^0	1192.6 MeV	7.4×10^{-20} s	$\Lambda\gamma$	uds
Σ^-	1197.4 MeV	1.5×10^{-10} s	$n\pi^-$	dds
Ξ^0	1314.9 MeV	2.9×10^{-10} s	$\Lambda\pi^0$	uss
Ξ^-	1321.7 MeV	1.6×10^{-10} s	$\Lambda\pi^-$	dss

Table 7.9: Light spin 1/2, parity even baryons.

baryon	mass	lifetime	dominant decays	quark content
$\Delta^-, \Delta^0, \Delta^+, \Delta^{++}$	1232 MeV	6×10^{-24} s	$p\pi, n\pi$	ddd, udd, uud, uuu
$\Sigma^{*-}, \Sigma^{*0}, \Sigma^{*+}$	1385 MeV	2×10^{-23} s	$\Lambda\pi$	dds, uds, uus
Ξ^{*-}, Ξ^{*0}	1530 MeV	7×10^{-23} s	$\Xi\pi$	dss, uss
Ω^-	1672 MeV	0.82×10^{-10} s	$\Lambda K^-, \Xi\pi$	sss

Table 7.10: Light spin 3/2, parity even baryons.

As expected the lightest observed baryons are, in fact, either spin 1/2 or spin 3/2. Tables 7.9 and 7.10 list the lightest spin 1/2 and spin 3/2 baryons, respectively. The intrinsic parity of these states is the same as the intrinsic parity as a quark, which, by *convention*, is chosen to be positive. Finally, we must consider the $SU(3)$ flavor structure. Again we note that “the same equations have the same solutions” and we use the same representation structure as for the $SU(3)$ of color above in Eq. (7.4.1), $\mathfrak{3} \otimes \mathfrak{3} \otimes \mathfrak{3} = \mathfrak{10} \oplus \mathfrak{8} \oplus \mathfrak{8} \oplus \mathfrak{1}$. The only step remaining is to pull the various factors together while ensuring overall antisymmetry under interchange of any 2 quarks.

We now discuss that final step in some detail (reviewing some of our previous discussions). Our goal is to understand the experimental observations summarized in Tables 7.9 and 7.10. In particular, Table 7.10 shows that the ten lightest $J = 3/2$ baryons form a decuplet representation of the $SU(3)$ of flavor (the triangular structure in the left-hand column in Table 7.10). This simply corresponds to matching the *symmetric* spin state (the $\mathfrak{4}$) with the *symmetric* flavor state (the $\mathfrak{10}$) (with the *antisymmetric* color $\mathfrak{1}$ state and the trivial but *symmetric* $L = 0$ spatial state). From the masses listed in Table 7.10 one sees that the Σ^* baryons, which contain one strange quark, are heavier than the Δ baryons, which contain only u and d quarks, by about 150 MeV. The Ξ^* baryons, which contain two strange quarks are heavier than the Σ^* by an additional ≈ 150 MeV, and the Ω^- baryon, containing three strange quarks, is yet heavier by about the same increment. This is consistent with our expectations that substituting heavier quarks for lighter quarks should increase the mass of bound states (by approximately the mass change of the substituted quark), since the binding dynamics due

to the color interactions remains the *same*, independent of the quark flavors involved.

As Table 7.9 shows, there are only eight light $J = 1/2$ baryons and they form a flavor $SU(3)$, spin $1/2$ baryon *octet*. There is only *one* combination of flavor octet and spin doublet with the correct overall symmetric behavior, which we then combine (as above) with the *antisymmetric* color $\underline{1}$ state and the trivial but *symmetric* $L = 0$ spatial state. As noted earlier, this difference between $J = 3/2$ and $J = 1/2$ bound states can be understood as a consequence of the Pauli exclusion principle.

7.5 Baryon wavefunctions

To understand how the fermionic nature of quarks produces the observed pattern of spin and flavor for the baryons, we first review in more detail the structure of the color wave function. Our goal is to characterize the form of the colorless bound state. What does this really mean? Just as you can think of a spin $1/2$ particle as having a wavefunction which is a two-component complex vector,

$$\begin{pmatrix} \psi_{\uparrow} \\ \psi_{\downarrow} \end{pmatrix} = \begin{pmatrix} \langle \uparrow | \psi \rangle \\ \langle \downarrow | \psi \rangle \end{pmatrix}, \quad (7.5.1)$$

the wavefunction of a quark (of definite flavor and spin) is a three-component vector in “color-space”,

$$\vec{\Psi} \equiv \begin{pmatrix} \psi_r \\ \psi_g \\ \psi_b \end{pmatrix} = \begin{pmatrix} \langle \text{red} | \psi \rangle \\ \langle \text{green} | \psi \rangle \\ \langle \text{blue} | \psi \rangle \end{pmatrix}. \quad (7.5.2)$$

The component ψ_r gives the amplitude for the quark color to be red, ψ_g is the amplitude to be green, *etc.* The assertion that hadrons must be “colorless” really means that the multi-quark wavefunction must not depend on the choice of basis in three-dimensional “color” space, *i.e.*, is unchanged by rotations in color space. Since the quark (color) wavefunction is a three-component vector, to build a colorless state from three quarks, q_1 , q_2 and q_3 , one must combine the three color vectors describing the individual quarks, $\vec{\psi}_{q_1}$, $\vec{\psi}_{q_2}$, and $\vec{\psi}_{q_3}$, in such a way that the result is basis *independent*. In practice this means that the three three-component color vectors must be combined to yield an expression with *no* left-over color indices, *i.e.*, no index on which a color transformation could act.

This may sound peculiar, but the mathematical problem is the same as asking how to build a rotationally-invariant scalar from three (ordinary) 3-vectors \vec{A} , \vec{B} and \vec{C} , in such a way that the result is a linear function of each of the vectors. You already know the (essentially unique) answer, namely the scalar triple-product of the three vectors, $\vec{A} \cdot (\vec{B} \times \vec{C})$. This triple product may be expressed in a variety of ways, including as the determinant of the components,

$$\vec{A} \cdot (\vec{B} \times \vec{C}) = \det \begin{vmatrix} A_1 & B_1 & C_1 \\ A_2 & B_2 & C_2 \\ A_3 & B_3 & C_3 \end{vmatrix} = \epsilon_{ijk} A_i B_j C_k. \quad (7.5.3)$$

In the last expression ϵ_{ijl} is the (by now familiar) totally antisymmetric (3x3x3) tensor which equals +1 when (ijk) is any cyclic permutation of (123), -1 when (ijk) is any cyclic permutation of (321), and zero otherwise.⁶ Recall that a determinant changes sign if any two columns (or rows) are

⁶The geometric definition of a cross product only makes sense for real vectors, but the expressions involving components of the vectors are equally well-defined for complex vectors.

interchanged. Consequently, the triple product is antisymmetric under any interchange of two of the vectors, which is encoded in the antisymmetry of the ϵ_{ijk} symbol. This is exactly the antisymmetric character that we claimed above for the color singlet state that arises when we combine 3 color triplets, $\mathbf{3} \otimes \mathbf{3} \otimes \mathbf{3} = \mathbf{1} \oplus \mathbf{8} \oplus \mathbf{8} \oplus \mathbf{1}$ (see footnote 5). As the *unique* 3x3x3 antisymmetric tensor it must also be the *structure constant* for both of the groups $SU(2)$ and $SO(3)$, as discussed in Chapters 1, 5 and 10.3. Thus the color singlet state of 3 quarks has the form

$$\Psi_{qqq,\text{singlet}} = \epsilon_{ijk} \Psi_{q_1,i} \Psi_{q_2,j} \Psi_{q_3,k}, \quad (7.5.4)$$

where the Ψ_q 's are the 3-component quark color (triplet) vectors of Eq. (7.5.2).

The complete wavefunction describing three quarks in a bound state must characterize not only the color of the quarks, but also their flavor, spin, and spatial location. To a good approximation, the wavefunction describing the lowest mass baryons will be a *product* of a color wavefunction (depending only on the color vectors as in Eq. (7.5.4)), a spatial wavefunction (depending only on the quark positions, or orbital angular momentum), and a flavor + spin wavefunction,

$$\Psi = \Psi_{\text{color}} \times \Psi_{\text{space}} \times \Psi_{\text{spin+flavor}}. \quad (7.5.5)$$

The essential point of the above discussion about triple products is that the color wavefunction for three quarks is antisymmetric under any interchange of the color vectors of any two quarks. The lightest hadrons which can be built from a given set of quark flavors will have a spatial wavefunction which is symmetric under interchange of quark positions, *i.e.*, it will have no orbital angular momentum. If this is not true (*i.e.*, if the orbital angular momentum is nonzero) then the wavefunction will have spatial nodes across which it changes sign, and this behavior invariably increases the kinetic energy of the state. The Pauli principle (or the fact that quarks are fermions) requires that the *total* wavefunction must be antisymmetric under interchange of any two quarks — which means simultaneous interchange of the positions, spins, flavors, and colors of the two quarks. Since the color part of the wavefunction must be antisymmetric, while the spatial part should be symmetric, this means that the flavor plus spin part must also be *symmetric* under permutations.⁷ Thus the symmetry properties of the spin and flavor wave functions must be *matched* to each other in order to provide the overall symmetry. This is why the two quantities are coupled in Eq. (7.5.5) (*i.e.*, we were planning ahead).

For a spin 3/2 baryon, such as the Δ^{++} , the flavor structure of the wavefunction is trivial, and totally symmetric, since all three quarks are the same type, namely uuu . For the $S_3 = 3/2$ state, the spin structure is also trivial, and totally symmetric, since all three quarks must individually have $S_3 = 1/2$ if the total spin projection is 3/2.⁸ Therefore the combined spin+flavor part of the wavefunction,

$$\Psi_{\text{spin+flavor}}^{\Delta^{++}} \sim (uuu) \times (\uparrow\uparrow\uparrow), \quad (7.5.6)$$

satisfies the above condition of symmetry under permutation of quark spins and flavors. Analogous spin+flavor wavefunctions may be constructed for all baryons (with any spin projection) in the spin

⁷This was actually an issue in the early days (~ 1970) when quarks had been postulated as the underlying degrees of freedom but the color quantum number had not yet appeared. The spatial, flavor and spin wavefunctions that matched the observed states are clearly symmetric, but the quarks are fermions. Where was Pauli? Then color appeared to save statistics *and* provide the needed interactions!

⁸This ignores the possibility of further constituents in the baryon in addition to the three up quarks, *e.g.*, gluons. Using an improved description of the structure of baryons does not change the essential conclusions of the following discussion.

3/2 decuplet. Just like the example in Eq. (7.5.6), these wavefunctions are independently symmetric under permutations of quark spins, or permutations of quark flavors. For completeness the quark content (in the same language as for the mesons) of the decuplet is presented in Table 7.11 and the particle names in Table 7.12 (which agree with the results already included in Table 7.10).

$\underline{10}$	ddd	ddu	duu	uuu	$\uparrow S$
	dds	dus	uus		$\rightarrow I_3$
	dss	uss			
	sss				

Table 7.11: 3 quark decuplet

$\frac{3}{2}^+ \underline{10}$	Δ^-	Δ^0	Δ^+	Δ^{++}	$\uparrow S$
	Σ^{*-}	Σ^{*0}	Σ^{*+}		$\rightarrow I_3$
	Ξ^{*-}	Ξ^{*0}			
	Ω^-				

Table 7.12: Baryon decuplet

Note that, since both the $\underline{10}$ of flavor $SU(3)$ and the $\underline{4}$ of spin are individually symmetric wavefunctions, there is no real coupling of spin and flavor in this case (the overall wave function is just a simple product). But note also that the need to match the symmetry properties of the spin and flavor wavefunctions means that, at least for the lowest mass baryons (with $L = 0$), there can be no spin 1/2 flavor decuplets ($\underline{10}$) and no spin 3/2 flavor octets ($\underline{8}$).

For $J = 1/2$ baryons, the situation is more complicated. As noted earlier the two possible spin doublets constructed from 3 spin 1/2 quarks have mixed symmetry with respect to interchanging pairs of the quark spins. Luckily the two flavor octets that can be constructed from 3 flavor triplets (quarks) have similar mixed symmetry and there is a combined spin 1/2, flavor octet 3 quark state that is overall symmetric under the simultaneous interchange of both the spin and flavor quantum numbers of any pair of quarks. To construct this state we recognize first that the case where all three quarks have the same flavor is *not* possible in the mixed symmetry octet. A flavor wavefunction such as uuu is totally symmetric. This explains why there are *no* light spin 1/2 baryons composed of three up (or three down, or three strange) quarks, in contrast to the case for spin 3/2 baryons. But if there are at least two distinct quark flavors involved, then it is possible to build a flavor wavefunction with the required symmetry. As an example, let us build a spin+flavor wavefunction for the proton. We need two u quarks and one d quark. A spin wavefunction of the form $(\uparrow\downarrow - \downarrow\uparrow)\uparrow$ describes a state in which the first two quarks have their spins combined to form an (antisymmetric) $S = 0$ state, so that adding the third spin yields a total spin of 1/2, as desired. Since this spin wavefunction is antisymmetric under interchange of the first two spins, we need a flavor wavefunction which is also antisymmetric under interchange of the first two quark flavors, namely $(ud - du)u$. If we multiply

these, we have a spin+flavor wavefunction,

$$[(ud - du)u] \times [(\uparrow\downarrow - \downarrow\uparrow)\uparrow] = (udu - duu)(\uparrow\downarrow\uparrow - \downarrow\uparrow\uparrow), \quad (7.5.7)$$

which is symmetric under combined spin and flavor exchange of the first two quarks. But we need a wavefunction which is symmetric under interchange of *any* pair of quark spins and flavors. This can be accomplished by simply adding terms which are related to the above by cyclic permutations (or in another words by repeating the above construction when it is the second and third, or first and third quarks which are combined to form spin zero). The result, which is unique up to an overall normalization factor, has the form

$$\Psi_{\text{spin+flavor}}^{\text{proton}} = (udu - duu)(\uparrow\downarrow\uparrow - \downarrow\uparrow\uparrow) + (uud - udu)(\uparrow\uparrow\downarrow - \uparrow\downarrow\uparrow) + (uud - duu)(\uparrow\uparrow\downarrow - \downarrow\uparrow\uparrow). \quad (7.5.8)$$

To be explicit we can also write this wavefunction with the terms multiplied out and normalized as

$$\begin{aligned} \Psi_{\text{spin+flavor}}^{\text{proton}} = \frac{1}{\sqrt{18}} [& 2u\uparrow u\uparrow d\downarrow + 2u\uparrow d\downarrow u\uparrow + 2d\downarrow u\uparrow u\uparrow \\ & - u\downarrow u\uparrow d\uparrow - u\uparrow u\downarrow d\uparrow - u\downarrow d\uparrow u\uparrow \\ & - u\uparrow d\uparrow u\downarrow - d\uparrow u\uparrow u\downarrow - d\uparrow u\downarrow u\uparrow]. \end{aligned} \quad (7.5.9)$$

You can, and should, check that these expressions satisfy the required condition of symmetry under interchange of spins and flavors of *any* pair of quarks. It should be clear that in this case the spin and flavor structures are truly intertwined. The wavefunction in Eq. (7.5.9) represents, for the case of 2 *u* quarks and 1 *d* quark, the unique member of an $SU(3)$ flavor octet and spin doublet (with $S_3 = +1/2$) that is symmetric under the interchange of any pair of quarks.

For fun (*e.g.*, in the HW) try to generate the following wavefunction for the neutron.

$$\begin{aligned} \Psi_{\text{spin+flavor}}^{\text{neutron}} = \frac{1}{\sqrt{18}} [& 2d\uparrow d\uparrow u\downarrow + 2d\uparrow u\downarrow d\uparrow + 2u\downarrow d\uparrow d\uparrow \\ & - d\downarrow d\uparrow u\uparrow - d\uparrow d\downarrow u\uparrow - d\downarrow u\uparrow d\uparrow \\ & - d\uparrow u\uparrow d\downarrow - u\uparrow d\uparrow d\downarrow - u\uparrow d\downarrow d\uparrow]. \end{aligned} \quad (7.5.10)$$

Similar constructions can be performed for all the other members of the spin 1/2 baryon octet. The expected quark content of the baryon octet is given in Table 7.13.

8	<i>udd</i>	<i>uud</i>	$\uparrow S$	
<i>sdd</i>	<i>sud</i> ($I = 1$)	<i>suu</i>	$\rightarrow I_3$	
	<i>sud</i> ($I = 0$)			
<i>ssd</i>	<i>ssu</i>			

Table 7.13: Baryon octet

The corresponding identification with the lowest mass baryons is provided in Table 7.14.

One notable feature of the set of octet baryons, shown in Tables 7.13 and 7.14, is the presence of two different baryons whose quark content is *sud*, specifically the Λ and the Σ^0 . This is not

$$\begin{array}{ccccccc}
\frac{1}{2}^+ \mathfrak{g} & & n & & p & & \uparrow S \\
\Sigma^- & & \Sigma^0 (I=1) & & \Sigma^+ & & \rightarrow I_3 \\
& & \Lambda^0 (I=0) & & & & \\
\Xi^- & & & & \Xi^0 & &
\end{array}$$

Table 7.14: Barvon octet

inconsistent. When three distinct flavors are involved, instead of just two, there are more possibilities for constructing a spin+flavor wavefunction with the required symmetry. A careful examination (left as a problem) shows that there are precisely two independent possibilities, completely consistent with the observed list of spin 1/2 baryons.

The mass values in Table 7.9 show that for spin 1/2 baryons, just as for spin 3/2 baryons, baryons with strange quarks are heavier than those with only up and down quarks; each substitution of a strange quark for an up or down raises the energy of the baryon by roughly 130–250 MeV.

7.6 Baryon number

Baryon number, denoted B , is defined as the total number of baryons minus the number of antibaryons, similarly to how we defined lepton number L in Eq. (6.6.1). Since baryons are bound states of three quarks, and antibaryons are bound states of three antiquarks, baryon number is the same as the number of quarks minus antiquarks, up to a factor of three,

$$B = (\# \text{ baryons}) - (\# \text{ antibaryons}) = \frac{1}{3} [(\# \text{ quarks}) - (\# \text{ antiquarks})]. \quad (7.6.1)$$

All known interactions conserve baryon number.⁹ High energy scattering processes can change the number of baryons, and the number of antibaryons, but not the net baryon number. For example, in proton-proton scattering, the reaction $p + p \rightarrow p + p + n + \bar{n}$ can occur, but not $p + p \rightarrow p + p + n + n$.

7.7 Hadronic decays

Turning to the decays of the various hadrons listed in Tables 7.7–7.10, it is remarkable how much can be explained using a basic understanding of the quark content of the different hadrons together with considerations of energy and momentum conservation. This is essentially the statement that understanding the basic symmetry properties, *i.e.*, the conserved quantum numbers, will get you a long way in the world of particle physics. This discussion will also help to illustrate the basic structure of the Standard Model (SM).

⁹This is not quite true. As with lepton number, the current theory of weak interactions predicts that there are processes which can change baryon number (while conserving $B-L$). The rate of these processes is so small that baryon number violation is (so far) completely unobservable in the laboratory. Of course, the observation that the majority of the observable universe (the stars) seem to be baryonic and not anti-baryonic means that baryon number violation must have been important at some point in the evolution of the universe (assuming a symmetric starting point).

As an example, consider the baryons in the spin $3/2$ decuplet. The rest energy of the Δ baryons is larger than that of a nucleon by nearly 300 MeV. This is more than the ≈ 140 MeV rest energy of a pion, which is the lightest meson. Consequently, a Δ baryon can decay to a nucleon plus a pion via strong interactions, which is the dominant way to produce pions as long as the process does not change the number of quarks minus antiquarks of each quark flavor (*i.e.*, the strong interactions preserve the net flavor). Specifically, a Δ^{++} can decay to $p\pi^+$, a Δ^+ can decay to either $p\pi^0$ or $n\pi^+$, a Δ^0 can decay to $p\pi^-$ or $n\pi^0$, and a Δ^- can decay to $n\pi^-$. These are the (overwhelmingly) dominant decay modes observed. The short lifetime of Δ baryons, $\tau \simeq 6 \times 10^{-24}$ s or $c\tau \simeq 1.8$ fm, is also indicative of a decay via strong interactions. (After all, “strong” should mean rapid interactions!) To set the scale, note that the time for light to travel a fermi is of order 3×10^{-24} s. The Δ baryon barely has time to “figure out” that it exists before it decays.

$1 \times 1/2$		$\begin{array}{ c } \hline 3/2 \\ \hline \end{array}$	$\begin{array}{ c } \hline +3/2 \\ \hline \end{array}$	$\begin{array}{ c } \hline 3/2 \\ \hline \end{array}$	$\begin{array}{ c } \hline 1/2 \\ \hline \end{array}$
$\begin{array}{ c } \hline +1 \\ \hline \end{array}$	$\begin{array}{ c } \hline +1/2 \\ \hline \end{array}$	$\begin{array}{ c } \hline 1 \\ \hline \end{array}$	$\begin{array}{ c } \hline +1/2 \\ \hline \end{array}$	$\begin{array}{ c } \hline 1/2 \\ \hline \end{array}$	$\begin{array}{ c } \hline 1/2 \\ \hline \end{array}$
	$\begin{array}{ c } \hline +1 \\ \hline \end{array}$	$\begin{array}{ c } \hline -1/2 \\ \hline \end{array}$	$\begin{array}{ c } \hline 1/3 \\ \hline \end{array}$	$\begin{array}{ c } \hline 2/3 \\ \hline \end{array}$	$\begin{array}{ c } \hline 3/2 \\ \hline \end{array}$
	$\begin{array}{ c } \hline 0 \\ \hline \end{array}$	$\begin{array}{ c } \hline +1/2 \\ \hline \end{array}$	$\begin{array}{ c } \hline 2/3 \\ \hline \end{array}$	$\begin{array}{ c } \hline -1/3 \\ \hline \end{array}$	$\begin{array}{ c } \hline 1/2 \\ \hline \end{array}$
		$\begin{array}{ c } \hline 0 \\ \hline \end{array}$	$\begin{array}{ c } \hline -1/2 \\ \hline \end{array}$	$\begin{array}{ c } \hline 2/3 \\ \hline \end{array}$	$\begin{array}{ c } \hline 1/3 \\ \hline \end{array}$
			$\begin{array}{ c } \hline -1 \\ \hline \end{array}$	$\begin{array}{ c } \hline -1/2 \\ \hline \end{array}$	$\begin{array}{ c } \hline 3/2 \\ \hline \end{array}$
2×1	$\begin{array}{ c } \hline 3 \\ \hline \end{array}$	$\begin{array}{ c } \hline +3 \\ \hline \end{array}$	$\begin{array}{ c } \hline 3 \\ \hline \end{array}$	$\begin{array}{ c } \hline 2 \\ \hline \end{array}$	$\begin{array}{ c } \hline 1 \\ \hline \end{array}$
				$\begin{array}{ c } \hline -1 \\ \hline \end{array}$	$\begin{array}{ c } \hline -1/2 \\ \hline \end{array}$

ASIDE: Let us take this opportunity to both study the isospin structure of the Δ decays in detail and utilize the Table of Clebsch-Gordan coefficients appended at the end of this Chapter. This table encodes in detail the way irreducible representations combine to form other irreducible representations. As we will illustrate below, this information can be obtained by application of the “ladder” operators of Chapter 5. For the current purposes

the appropriate section of the Table is the upper left section, which is shown to the left. The label on this section, “ $1 \times 1/2$ ”, tells us this section describes in detail how to combine isospin (or ordinary spin) 1 (a triplet representation) with isospin $1/2$ (a doublet representation), $\underline{3} \otimes \underline{2}$. We know that the resulting combined representations are an isospin $3/2$ quartet and an isospin $1/2$ doublet, $\underline{3} \otimes \underline{2} = \underline{4} \oplus \underline{2}$. This is precisely what we see along the uppermost line in each (funny shaped) box, $3/2$ or $1/2$. The next row of labels down tells us the specific value of I_3 (or J_3) for the joint representation. The labels to the left in the (funny shaped) boxes are the I_3 (or J_3) values labeling the specific states we are combining. Finally the numbers in the central, shaded regions of each of the (funny shaped) boxes are the actual Clebsch-Gordan coefficients. Strictly speaking, as explained near the top of the Table, the actual coefficient is the *square root* of the number in the Table, not including the sign. More specifically we interpret the upper left (funny shaped) box to mean $|I, I_3\rangle = |\frac{3}{2}, \frac{3}{2}\rangle = |1, 1\rangle|\frac{1}{2}, \frac{1}{2}\rangle$ or, in our particular application, $|\Delta^{++}\rangle = |\pi^+\rangle|p\rangle$. There is only *one* possible final state that conserves isospin and so the coefficient is necessarily one, *i.e.*, we can obtain this entry in the Table by “pure thought”, no calculation required.

Now consider the more interesting case in the second (funny shaped) box down. We read the Table to mean $|\frac{3}{2}, \frac{1}{2}\rangle = \sqrt{\frac{1}{3}}|1, 1\rangle|\frac{1}{2}, -\frac{1}{2}\rangle + \sqrt{\frac{2}{3}}|1, 0\rangle|\frac{1}{2}, \frac{1}{2}\rangle$ (no minus signs) or, in our case,

$$|\Delta^+\rangle = \sqrt{\frac{1}{3}}|\pi^+\rangle|n\rangle + \sqrt{\frac{2}{3}}|\pi^0\rangle|p\rangle. \quad (7.7.1)$$

We can obtain this result by appropriate application of the lowering ladder operator from Chapter 5. The essential “new” observation is that, since the isospin of the Δ is the *sum* of the isospin of the nucleon and the pion, the ladder operator for the Δ is the sum of the ladder operators for the nucleon and the pion,

$$I_{\pm}^{\Delta} = I_{\pm}^{\text{nucleon}} + I_{\pm}^{\pi}. \quad (7.7.2)$$

So, using the coefficients in Eq. (5.1.34), we have

$$\begin{aligned}
 I_-^\Delta |\Delta^{++}\rangle &= I_-^\Delta \left| \frac{3}{2}, \frac{3}{2} \right\rangle = \hbar \sqrt{\frac{3}{2} \frac{5}{2} - \frac{3}{2} \frac{1}{2}} \left| \frac{3}{2}, \frac{1}{2} \right\rangle = \hbar \sqrt{3} |\Delta^+\rangle, \\
 I_-^{\text{nucleon}} |\pi^+\rangle |p\rangle &= I_-^{\text{nucleon}} |1, 1\rangle \left| \frac{1}{2}, \frac{1}{2} \right\rangle = \hbar \sqrt{\frac{1}{2} \frac{3}{2} - \frac{1}{2} \left(-\frac{1}{2}\right)} |1, 1\rangle \left| \frac{1}{2}, -\frac{1}{2} \right\rangle = \hbar |\pi^+\rangle |n\rangle, \\
 I_-^\pi |\pi^+\rangle |p\rangle &= I_-^\pi |1, 1\rangle \left| \frac{1}{2}, \frac{1}{2} \right\rangle = \hbar \sqrt{1(2) - 1(0)} |1, 0\rangle \left| \frac{1}{2}, \frac{1}{2} \right\rangle = \hbar \sqrt{2} |\pi^0\rangle |p\rangle, \\
 \Rightarrow |\Delta^+\rangle &= \left| \frac{3}{2}, \frac{1}{2} \right\rangle = \sqrt{\frac{1}{3}} |1, 1\rangle \left| \frac{1}{2}, -\frac{1}{2} \right\rangle + \sqrt{\frac{2}{3}} |1, 0\rangle \left| \frac{1}{2}, \frac{1}{2} \right\rangle = \sqrt{\frac{1}{3}} |\pi^+\rangle |n\rangle + \sqrt{\frac{2}{3}} |\pi^0\rangle |p\rangle \quad (7.7.3)
 \end{aligned}$$

The corresponding isospin 1/2 state, the next column in the Table, can be determined (modulo an overall sign ambiguity - the choice in the Table is a convention) by the constraint of defining an “orthogonal” state. We have

$$\left| \frac{1}{2}, \frac{1}{2} \right\rangle = \sqrt{\frac{2}{3}} |1, 1\rangle \left| \frac{1}{2}, -\frac{1}{2} \right\rangle - \sqrt{\frac{1}{3}} |1, 0\rangle \left| \frac{1}{2}, \frac{1}{2} \right\rangle. \quad (7.7.4)$$

In principle, using ladder operators and orthogonality we could fill out the Table of coefficients. Being lazy but smart, we will simply use the Table.

Returning to the original discussion of Δ^+ decay, we see that it decays (strongly) 1/3 of the time to $\pi^+ n$,

$$|\langle \pi^+ | \langle n | \Delta^+ \rangle|^2 = |\sqrt{1/3}|^2 = \frac{1}{3}, \quad (7.7.5)$$

and 2/3 of the time to $\pi^0 p$,

$$|\langle \pi^0 | \langle p | \Delta^+ \rangle|^2 = |\sqrt{2/3}|^2 = \frac{2}{3}. \quad (7.7.6)$$

Note in particular that, in going from the *amplitude* with the Clebsch-Gordan coefficient (with the square root) to the *probability*, which is the amplitude *squared*, we *square* the Clebsch-Gordan coefficient.

The second column in the second (funny shaped) box describes the case with total isospin 1/2, which plays no role here as we are considering the strong decay of an isospin 3/2 particle and isospin is conserved by the strong interactions. Next we move down to the third (funny shaped) box to learn that $\left| \frac{3}{2}, -\frac{1}{2} \right\rangle = \sqrt{\frac{2}{3}} |1, 0\rangle \left| \frac{1}{2}, -\frac{1}{2} \right\rangle + \sqrt{\frac{1}{3}} |1, -1\rangle \left| \frac{1}{2}, \frac{1}{2} \right\rangle$ or, in our case,

$$|\Delta^0\rangle = \sqrt{\frac{2}{3}} |\pi^0\rangle |n\rangle + \sqrt{\frac{1}{3}} |\pi^-\rangle |p\rangle. \quad (7.7.7)$$

So a Δ^0 decays (strongly) 2/3 of the time to $\pi^0 n$, while 1/3 of the time it decays to $\pi^- p$. Finally the last (funny shaped) box tells us that the Δ^- decays *uniquely* to $\pi^- n$, as we have noted above. Being able to read the Table of Clebsch-Gordan coefficients is an extremely useful skill and we should all practice it.

Returning to the general discussion, we note that the lifetime of the Δ corresponds to a decay (or resonance) width $\Gamma_\Delta = \hbar/\tau \approx 120$ MeV, which is 10% of the rest energy of this baryon. We can think of the $\pi - p$ scattering amplitude as having a pole in the complex energy plane at the mass of the Δ , where the pole is off the real axis by the amount Γ_Δ . (Those readers not familiar with the

language of resonances looked at as poles in the complex energy plane, *i.e.*, if this was not covered in their Physics 227 class, are encouraged to read Lecture 5 from my version of 227 available here.)

While there is a potential issue concerning how long a particle must live in order to be called a particle, we should rather focus on the point that almost all the particles (states) we know about decay eventually (*i.e.*, nearly all of the poles are somewhat off the real axis). The real issue is the practical, experimental one. Does the state live long enough to be detected in the detector before it decays? Or is the lifetime so short that we see evidence of the state's existence *only* from a bump in an interaction rate (*i.e.*, we literally detect the pole in the complex energy plane). This is essentially the distinction in the tables provided by the PDG. In the former case, lifetimes are reported, while in the latter case the width (of the peak) is reported. Thus the PDG reports what is actually measured.

The Σ^* and Ξ^* baryons also have very short lifetimes, on the order of a few times 10^{-23} s. The Σ^* contains one strange quark. The Σ^* 's mass of $1285 \text{ MeV}/c^2$ is larger than the $1116 \text{ MeV}/c^2$ mass of the Λ , the lightest baryon containing a strange quark, by more than the mass of a pion. So strong interactions can cause a Σ^* to decay to a Λ plus a pion, which is the dominant observed decay. Similarly, Ξ^* baryons, containing two strange quarks, can decay via strong interactions to a Ξ (the lightest doubly strange baryon) plus a pion.

Looking ahead to Chapter 8.9 on parity, it is worthwhile noting a characteristic feature of these strong decays of members of the flavor decuplet (the flavor 10) into a member of the nucleon octet (the flavor 8) plus a pion. In J^P notation (spin and intrinsic parity) we have a $\frac{3}{2}^+$ going into a $\frac{1}{2}^+$ plus a 0^- . Thus naively neither the spin nor parity match, $\frac{3}{2} \neq \frac{1}{2} + 0$ and $+$ \neq $+$ \times $-$. But total angular momentum is conserved by all interactions and so we conclude that there must be one unit of *orbital* angular momentum in the final state, $L = 1\hbar$ (the proton and pion do not have zero impact parameter). This now allows the final state to have total angular momentum $\frac{3}{2} (= \frac{1}{2} + 1)$ and adds an extra parity factor of $(-1)^L = -1$. Thus everything works for a strong decay, conserving both angular momentum *and* parity.

The final member of the $J = 3/2$ decuplet, the Ω^- baryon, cannot decay via strong interactions to a lighter baryon plus a pion, because there are no lighter baryons containing three strange quarks (and strong interactions preserve the net number of strange quarks). It could, in principle, decay via strong interactions to a Ξ baryon (containing two strange quarks) and a K meson (containing one strange quark) — but it doesn't have enough energy. Its mass of $1672 \text{ MeV}/c^2$ is less than the sum of Ξ plus K masses. In fact, the Ω^- baryon cannot decay via *any* strong interaction process. Nor can it decay via electromagnetic processes, which also preserve the net flavor. But weak interactions are distinguished by the fact that they can *change* quarks of one flavor into a different flavor. Consequently, the Ω^- baryon can decay via weak interactions to a lighter baryon plus a meson with only 2 strange quarks remaining. The dominant decays involve the conversion of one strange quark into an up quark, leading to final states consisting of a Λ baryon plus a K^- meson, a Ξ^0 baryon plus a π^- , or a Ξ^- plus a π^0 . The Ω^- was, in some sense, easy to detect due to its characteristic “cascade” decay (first seen in a bubble chamber photograph)

$$\begin{array}{l} \Omega^- \rightarrow \Xi^0 \pi^- \\ \quad \searrow \Lambda^0 \pi^0 \\ \quad \quad \searrow p \pi^- . \end{array} \quad (7.7.8)$$

So overall the process in Eq.(7.7.8) is $\Omega^- \rightarrow p \pi^- \pi^- \pi^0$, but occurs in 3 distinct steps. Note that these final states conserve baryon number and electric charge, and are allowed by energy conservation.

The 10^{-10} s lifetime of the Ω^- is much *longer* than a typical strong interaction decay time, and is indicative of a weak interaction process.

Similar reasoning can be applied to the $J = 1/2$ baryons. The proton is (apparently) stable, while all the other members of the octet decay via weak interactions — except for the Σ^0 which can decay to a Λ plus a photon via electromagnetic interactions. Note that the 7×10^{-20} s lifetime of the Σ^0 is much shorter than a weak interaction lifetime, but is longer than typical strong interaction lifetimes. (In a very real sense the electromagnetic interactions are stronger than the weak interactions but weaker than the strong interactions.) The lifetimes of the Λ , Ξ , and Σ^\pm baryons are all around 10^{-10} s, typical of weak interaction decays. The 900 second lifetime of the neutron is vastly longer than a normal weak interaction lifetime. This reflects the fact that neutron decay is just barely allowed by energy conservation. The mass of the final proton plus electron (and antineutrino) is so close to the mass of the neutron that only about 8 MeV, or less than 0.1% of the rest energy of neutron, is available to be converted into the kinetic energy of the decay products.

Before ending this discussion, we can test our newly acquired understanding of conserved quantum numbers and decays by applying it to the case of mesons. Just like the spin 3/2 baryons tend to have strong interaction (*i.e.*, fast) decays into the spin 1/2 baryons plus a pion, the vector mesons of Table 7.6 have strong decays into the corresponding scalar mesons (*i.e.*, the states with the same number of strange quarks) plus a pion. Note, in particular, that all of the lifetimes are of order 10^{-22} to 10^{-24} seconds, typical strong (short) lifetimes. Again we should consider the correlation of spin and parity for the decay of a 1^- meson into two 0^- mesons. As with the strong baryon decays above we need 1 unit of orbital angular momentum to allow the conservation of total angular momentum, $1 = 0 + 0 + 1$, which is again just what we need to conserve parity, $-1 = -1 \times -1 \times -1$.

Also worthy of note is that the zero-strangeness, isoscalar state ω decays into 3 pions, while the similar zero-strangeness neutral member of the isovectors, the ρ^0 , decays into 2 pions. There is a combined transformation of charge conjugation (C) plus an isospin rotation (historically called “ G -parity” and described in more detail in the next Chapter) under which pions and the ω are odd (-1), while the ρ ’s are even. Since the strong interactions respect this symmetry, the decay of the ρ^0 must be into an even number of pions, while the ω decays into an odd number of pions. More generally a $q\bar{q}$ state made of u and d quarks has the following eigenvalue under G (depending on the orbital angular momentum, L , spin, S , and total isospin, I),

$$G|q\bar{q}(u, d), L, S, I\rangle = (-1)^{L+S+I}|q\bar{q}(u, d), L, S, I\rangle, \quad (7.7.9)$$

which yields, as noted above, (-1) for pions ($L = S = 0, I = 1$), (+1) for the ρ ($L = 0, S = I = 1$), and (-1) for the ω ($L = I = 0, S = 1$).

Figure 1 illustrates the construction of the matrix $Y_\ell^{-m} = (-1)^m Y_\ell^{m*}$ for $m=1$. The diagram shows a sequence of operations represented by boxes containing numbers and fractions. The boxes are arranged in a grid-like structure, with some boxes containing multiple rows of numbers. The values are:

- Box 1 (top left): 1×1 , 2 , $+1 +1$, $+2$, $2 \ 1$, $+1 +1$.
- Box 2 (top right): $+2 -1$, $1/15 \ 1/3 \ 3/5$, $+1 \ 0$, $8/15 \ 1/6 -3/10$, $0 +1$, $2/5 -1/2 \ 1/10$.
- Box 3 (middle left): $+1 \ 0$, $0 +1$, $1/2 \ 1/2$, $1/2 -1/2$.
- Box 4 (middle right): $2 \ 1 \ 0$, $0 \ 0 \ 0$, $+1 -1$, $1/6 \ 1/2 \ 1/3$, $0 \ 0$, $2/3 \ 0 -1/3$, $-1 +1$, $1/6 -1/2 \ 1/3$.
- Box 5 (bottom right): $+1 -1$, $0 \ 0$, $2 \ 1$, $-1 -1$, $0 -1$, $1/2 \ 1/2$, $-1 \ 0$, $1/2 -1/2$, -2 , $-1 -1$, 1 .

The final expression is $Y_\ell^{-m} = (-1)^m Y_\ell^{m*}$.

A further detail (not shown explicitly in Table 7.8) is that the two pion decay of the neutral ρ is $\rho^0 \rightarrow \pi^+ + \pi^-$ and *not* $\rho^0 \rightarrow \pi^0 + \pi^0$. This is an interesting realization of the how two isovectors (the pions) combine to form another isovector (the ρ). It supplies another opportunity to look in detail at the table of Clebsch-Gordan coefficients that appears at the end of this Chapter. The appropriate subsection (middle left) of the table is shown to the left here. For the case of two vectors combining to form another vector ($1 \times 1 \rightarrow 1$) where we want

to consider the neutral state of the final vector ($M = 0$), we focus on the middle column of the middle (funny shaped) box. We see from the 0 at the center of this box that there is no coupling to

the neutral states of the original vectors ($m_1 = m_2 = 0$), *i.e.*, $\rho^0 \rightarrow \pi^0 + \pi^0$. Instead there are equal couplings, up to the sign, to the 2 charged states,

$$|\rho^0\rangle = \sqrt{\frac{1}{2}}|\pi^+\rangle|\pi^-\rangle - \sqrt{\frac{1}{2}}|\pi^-\rangle|\pi^+\rangle. \quad (7.7.10)$$

ASIDE: Another way to look at the absence of this decay, $\rho^0 \rightarrow \pi^0 + \pi^0$, which illustrates again the “beautiful” self-consistency of physics, is to consider the symmetry required for identical bosons (see problem 11.1 in Das and Ferbel). We already noted that the two pion final state in ρ decay must correspond to $L = 1$ in order to conserve both total angular momentum and parity. Further, a two particle state with relative orbital angular momentum 1 is *antisymmetric* under the interchange of the two particles. This is why it adds a factor of (-1) to the overall parity of the state. On the other hand a two π^0 state must be *symmetric* under the interchange of the 2 π^0 's, since they are *identical* bosons. Hence there can be no $L = 1$ two π^0 state and the decay cannot occur. For the state composed of two oppositely charged pions *both* the isospin wave function of Eq. (7.7.10) and the $L = 1$ spatial wave function are antisymmetric under the interchange of the pions yielding the required overall symmetric state.

For the scalar mesons of Table 7.5 energy conservation rules out any strong decays. The (neutral) π^0 , η and η' all exhibit electromagnetic decays with photons in the final state and lifetimes of order 10^{-20} seconds (like the Σ^0). The kaons need to convert a strange quark into an up or down quark and so decay weakly with a 10^{-8} second lifetime (except the K_S , which is still a weak decay but slightly faster). Finally the charged pions decay weakly into leptons *only* with a similar 10^{-8} second lifetime.

To complete this discussion of decays we summarize in Table 7.15 the various additive and multiplicative quantum numbers and which interactions conserve them. We will discuss the related symmetries in more detail in the next chapter.

You are encouraged to look at the much more extensive listing of information about known mesons and baryons at the Particle Data Group website. Pick a few particles which have not been discussed above, and see if you can predict the dominant decay modes using the ideas we have discussed in this chapter.

It is worthwhile emphasizing again what “conserving” means in this context. Consider the decay of one particle into 2 particles, $A \rightarrow B + C$. For the scalar quantities like electric charge, baryon number and lepton number (or 4-vectors like energy-momentum) conserving means that the value of this quantity for particle A is equal to the simple sum of these quantities for particles B and C . For more complex quantities like angular momentum, isospin and quark flavor (with nontrivial group theory structure), we need to think a bit harder. Now conserving means that the “total” quantum number like J^2 (*i.e.*, defining the representation) and the simple additive quantities like J_3 (*i.e.*, defining the individual element of the representation) must match before and after the decay. Since the total angular momentum before the decay is just the spin of A , this representation must match one of the possible angular momentum representations that arise when we combine the spin of B with the spin of C and any possible *orbital* angular momentum of the BC pair. We have introduced the techniques of Young diagrams (see the sample problems in the next sub-chapter and Chapter 11) precisely so that we can calculate which angular momentum representations can be present in the final state when we combine the spins of B and C and the orbital angular momentum ($\mathcal{S}^B \otimes \mathcal{S}^C \otimes \mathcal{L}^{BC}$). Likewise, if we specify the spin polarization (S_3^A) in the initial state, it must be matched by that in the final state. Here the arithmetic is simple addition, $S_3^A = S_3^B + S_3^C + L_3^{BC}$.

Conserved quantity	Strong	EM	Weak
<i>Additive</i>			
Energy-momentum	Yes	Yes	Yes
Angular Momentum	Yes	Yes	Yes
Electric Charge	Yes	Yes	Yes
Baryon Number	Yes	Yes	Yes
Lepton Number	Yes	Yes	Yes
Quark Flavor	Yes	Yes	No
Isospin	Yes	No	No
<i>Multiplicative</i>			
Parity - P	Yes	Yes	No
Charge Conjugation - C	Yes	Yes	No
Time Reversal - T (or CP)	Yes	Yes	$\sim 10^{-3}$ viol
CPT	Yes	Yes	Yes
G - parity	Yes	No	No

Table 7.15: Conserved quantum numbers.

For isospin the calculation is very similar to angular momentum, since it again involves only $SU(2)$ representations (*i.e.*, the same equations have the same solutions). In fact, it is even simpler because there is *no* analogue of *orbital* angular momentum in isospin space! The specific element of the isospin representation is specified by I_3 , which satisfies $I_3^A = I_3^B + I_3^C$. For the $SU(3)$ of quark flavor (for the u , d and s quarks) we must match representations of $SU(3)$ before and after the decay. The additive conservation is now of 2 quantum numbers, typically I_3 and strangeness (the number of strange quarks).

7.8 Sample calculations

At several points in this Chapter we used results for the combinations of multiplets of both $SU(2)$ and $SU(3)$. The “slickest” technique for calculating these results is the method of Young diagrams described in Chapter 11. Here we present a brief summary of those results and the reader is strongly encouraged to read Chapter 11 now! This will allow you to understand the (somewhat peculiar) notation used below.

Consider first combining 2 *fundamental* multiplets (representations) as in Eq. (11.2.7). In the Young diagram language we have (“multiplying” the boxes in the obvious 2 ways)

$$\square \otimes \square = \square\square \oplus \begin{array}{c} \square \\ \square \end{array}. \quad (7.8.1)$$

To proceed we put the appropriate integers in the boxes, evaluate and divide by the “hooks”, and

calculate. For the $SU(2)$ case of adding 2 doublets (*e.g.*, 2 spin 1/2 fermions) we have

$$\underline{2} \otimes \underline{2} = \underline{2} \otimes \underline{2} = \begin{array}{|c|c|} \hline & \\ \hline \end{array} \oplus \begin{array}{|c|} \hline \\ \hline \end{array} = \begin{array}{|c|c|} \hline 2 & 3 \\ \hline \end{array} / \begin{array}{|c|c|} \hline 2 & 1 \\ \hline \end{array} \oplus \begin{array}{|c|} \hline 2 \\ \hline 1 \\ \hline \end{array} = \frac{2 \cdot 3}{2 \cdot 1} \oplus \frac{2 \cdot 1}{2 \cdot 1} = \underline{3} \oplus \underline{1}. \quad (7.8.2)$$

In words, combining 2 spin 1/2 particles yields one spin 1 state (the $\underline{3}$) and one spin 0 state (the $\underline{1}$).

For combining 2 triplets of $SU(3)$ (of color or flavor) we have

$$\underline{3} \otimes \underline{3} = \underline{3} \otimes \underline{3} = \begin{array}{|c|c|} \hline & \\ \hline \end{array} \oplus \begin{array}{|c|} \hline \\ \hline \end{array} = \begin{array}{|c|c|} \hline 3 & 4 \\ \hline \end{array} / \begin{array}{|c|c|} \hline 2 & 1 \\ \hline \end{array} \oplus \begin{array}{|c|} \hline 3 \\ \hline 2 \\ \hline 1 \\ \hline \end{array} = \frac{3 \cdot 4}{2 \cdot 1} \oplus \frac{3 \cdot 2}{2 \cdot 1} = \underline{6} \oplus \underline{\bar{3}}. \quad (7.8.3)$$

Note the (symmetry) distinction here between $\underline{\bar{3}}$ and $\underline{3}$ ($\begin{array}{|c|} \hline \\ \hline \end{array}$ versus $\begin{array}{|c|} \hline \\ \hline \end{array}$ and the corresponding distinction illustrated in Table 7.3, ∇ versus \triangle) that does not exist for the $SU(2)$ $\underline{\bar{2}}$ and $\underline{2}$, which are identical ($\begin{array}{|c|} \hline \\ \hline \end{array}$ versus $\begin{array}{|c|} \hline \\ \hline \end{array}$). This means that in the $SU(3)$ case we can obtain something new and different from combining a $\underline{3}$ and $\underline{\bar{3}}$ (*i.e.*, as distinct from combining two $\underline{3}$'s). We have

$$\underline{3} \otimes \underline{\bar{3}} = \begin{array}{|c|} \hline \\ \hline \end{array} \otimes \begin{array}{|c|} \hline \\ \hline \end{array} = \begin{array}{|c|c|} \hline & \\ \hline \end{array} \oplus \begin{array}{|c|} \hline \\ \hline \end{array} = \begin{array}{|c|c|} \hline 3 & 4 \\ \hline 2 \\ \hline \end{array} / \begin{array}{|c|c|} \hline 3 & 1 \\ \hline 1 \\ \hline \end{array} \oplus \begin{array}{|c|} \hline 3 \\ \hline 2 \\ \hline 1 \\ \hline \end{array} = \frac{3 \cdot 4 \cdot 2}{3 \cdot 1 \cdot 1} \oplus \frac{3 \cdot 2 \cdot 1}{3 \cdot 2 \cdot 1} = \underline{8} \oplus \underline{1}. \quad (7.8.4)$$

Next consider combining three fundamental representations, where we must immediately distinguish $SU(2)$ and $SU(3)$ (we cannot antisymmetrize 3 objects if there are only 2 different kinds). For $SU(2)$ (3 fermions) we have, as in Eq. (7.4.2),

$$\begin{aligned} \underline{2} \otimes \underline{2} \otimes \underline{2} &= \left(\begin{array}{|c|c|} \hline & \\ \hline \end{array} \oplus \begin{array}{|c|} \hline \\ \hline \end{array} \right) \otimes \begin{array}{|c|} \hline \\ \hline \end{array} = \begin{array}{|c|c|c|} \hline & & \\ \hline \end{array} \oplus \begin{array}{|c|c|} \hline & \\ \hline \end{array} \oplus \begin{array}{|c|c|} \hline & \\ \hline \end{array} \\ &= \begin{array}{|c|c|c|} \hline 2 & 3 & 4 \\ \hline \end{array} / \begin{array}{|c|c|c|} \hline 3 & 2 & 1 \\ \hline \end{array} \oplus \begin{array}{|c|c|} \hline 2 & 3 \\ \hline 1 \\ \hline \end{array} / \begin{array}{|c|c|} \hline 3 & 1 \\ \hline 1 \\ \hline \end{array} \oplus \begin{array}{|c|c|} \hline 2 & 3 \\ \hline 1 \\ \hline \end{array} / \begin{array}{|c|c|} \hline 3 & 1 \\ \hline 1 \\ \hline \end{array} = \frac{2 \cdot 3 \cdot 4}{3 \cdot 2 \cdot 1} \oplus \frac{2 \cdot 3 \cdot 1}{3 \cdot 1 \cdot 1} \oplus \frac{2 \cdot 3 \cdot 1}{3 \cdot 1 \cdot 1} \\ &= \underline{4} \oplus \underline{2} \oplus \underline{2}. \end{aligned} \quad (7.8.5)$$

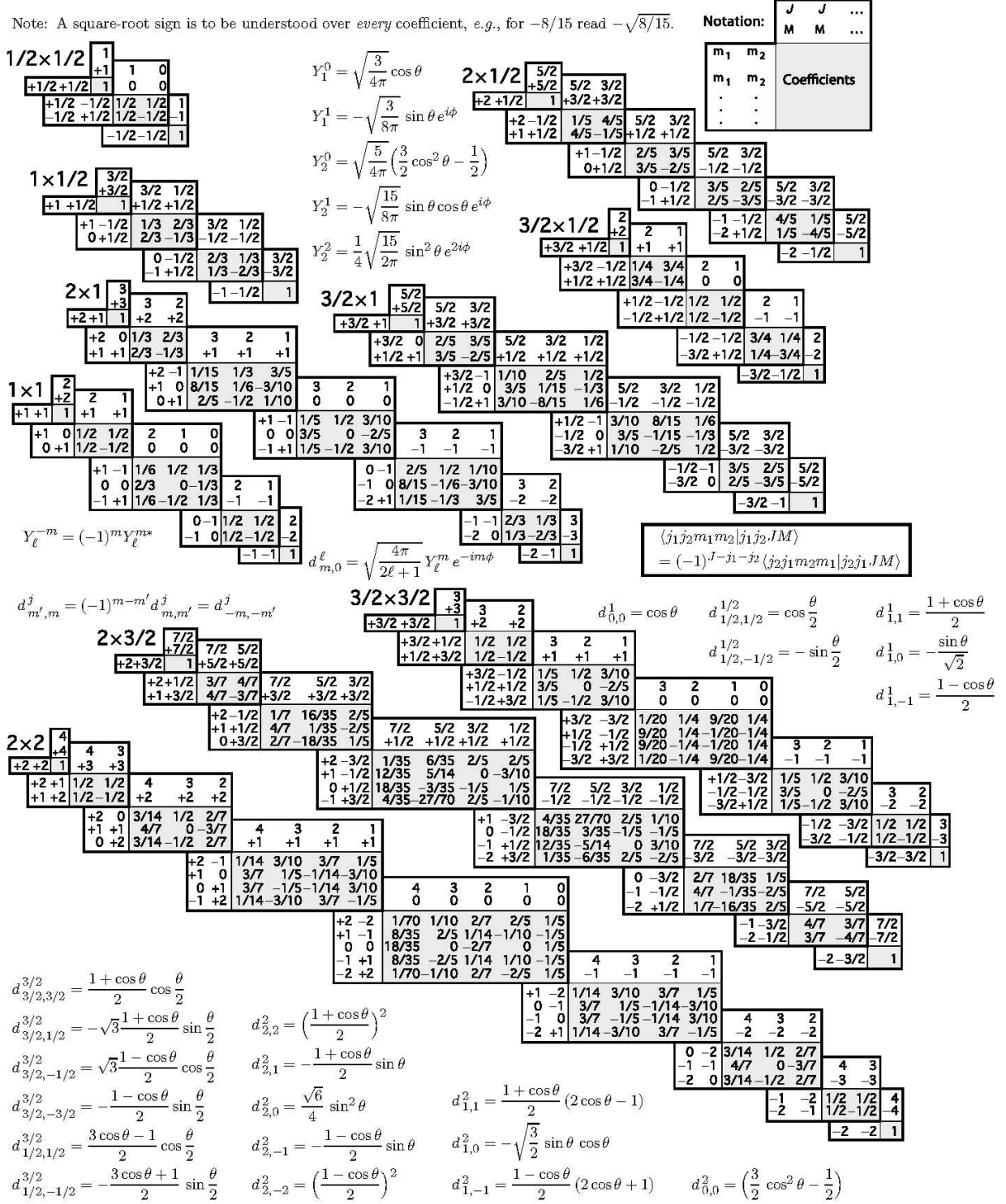
For the distinct case of three $SU(3)$ triplets we have instead, as in Eq. (7.4.1), (note the differences

from the $SU(2)$ case above)

$$\begin{aligned}
 \boxed{3} \otimes \boxed{3} \otimes \boxed{3} &= \left(\boxed{} \oplus \boxed{} \right) \otimes \boxed{} = \boxed{} \oplus \boxed{} \oplus \boxed{} \oplus \boxed{} \\
 &= \boxed{3 \ 4 \ 5} / \boxed{3 \ 2 \ 1} \oplus \boxed{3 \ 4} / \boxed{2} \oplus \boxed{3 \ 4} / \boxed{2} \oplus \boxed{3} / \boxed{2 \ 1} \\
 &= \frac{3 \cdot 4 \cdot 5}{3 \cdot 2 \cdot 1} \oplus \frac{3 \cdot 4 \cdot 2}{3 \cdot 1 \cdot 1} \oplus \frac{3 \cdot 4 \cdot 2}{3 \cdot 1 \cdot 1} \oplus \frac{3 \cdot 2 \cdot 1}{3 \cdot 2 \cdot 1} \\
 &= 10 \oplus 8 \oplus 8 \oplus 1.
 \end{aligned} \tag{7.8.6}$$

40. CLEBSCH-GORDAN COEFFICIENTS, SPHERICAL HARMONICS, AND d FUNCTIONS

Note: A square-root sign is to be understood over every coefficient, e.g., for $-8/15$ read $-\sqrt{8/15}$.



Chapter 8

Symmetries

8.1 Quantum dynamics

Let us begin this topic by reviewing some “quantum mechanical” notation (recall Chapter 5). The state, or *ket*, vector $|\psi\rangle$ of a physical system completely characterizes the system at a given instant. The corresponding *bra* vector $\langle\psi|$ is the Hermitian conjugate of $|\psi\rangle$. Properly normalized states satisfy the relation that the “bra-ket” is unity, $\langle\psi|\psi\rangle = 1$. We are imagining that $|\psi\rangle$ is defined in a “vector space” of states that is spanned by some *complete* set of appropriately chosen (see below) *linearly independent* (typically orthogonal) basis states (or vectors), $|\psi_n\rangle$, such that *any* state vector can be expressed as a sum of these basis vectors,

$$|\psi\rangle = \sum_n c_n |\psi_n\rangle \quad (8.1.1)$$

(exactly analogously to the way we think of expressing an ordinary 3-vector as the sum of coefficients times the usual unit vectors $\hat{x}^1, \hat{x}^2, \hat{x}^3$).

Let $|\psi(t)\rangle$ denote the state of a system at time t . Given an initial state $|\psi(0)\rangle$, the goal of quantum dynamics is to predict $|\psi(t)\rangle$ for $t \neq 0$. The superposition principle of quantum mechanics implies that there is a *linear* operator $U(t)$, called the *time-evolution operator*, which maps any state at time zero into the corresponding state at time t ,

$$|\psi(t)\rangle = U(t) |\psi(0)\rangle. \quad (8.1.2)$$

Time evolution must map any properly normalized state at one time into a *normalized* state at another time, *i.e.*, probability is conserved. This implies that the time evolution operator is *unitary*,

$$1 = \langle\psi(t)|\psi(t)\rangle = \langle\psi(0)|U(t)^\dagger U(t)|\psi(0)\rangle = \langle\psi(0)|\psi(0)\rangle \Rightarrow U(t)^\dagger = U(t)^{-1}. \quad (8.1.3)$$

Here we are assuming that the system under study is “isolated” in the sense that no probability can be added or leak away, *i.e.*, $\langle\psi(t)|\psi(t)\rangle$ is a constant.

It is often convenient to consider a differential form of time evolution. The time derivative of any state must again (by the superposition principle) be given by some linear operator acting on the state. That linear operator, times $i\hbar$, is called the Hamiltonian, denoted H . In other words,

$$i\hbar \frac{d}{dt} |\psi(t)\rangle = H |\psi(t)\rangle. \quad (8.1.4)$$

This is the (time dependent) Schrödinger equation. It is a linear first order differential equation, whose solution can be (formally) written immediately (through the wonders of the functions of mathematical physics in Chapter 1) in terms of an exponential (defined as an operator by its power series expansion),

$$|\psi(t)\rangle = e^{-iHt/\hbar} |\psi(0)\rangle. \quad (8.1.5)$$

ASIDE To make this discussion more concrete we recall the explicit forms of some Hamiltonians of interest. For a *free*, non-relativistic particle of mass m moving in one dimension the classical Hamiltonian is $H = p^2/2m = m\dot{x}^2/2$. Quantum mechanically ($p \rightarrow -i\hbar\partial/\partial x$) we have $H = -(\hbar^2/2m)\partial^2/\partial x^2$. For a (non-relativistic) harmonic oscillator ($\omega^2 = k/m$) the classical Hamiltonian is $H = p^2/2m + kx^2/2$, while the quantum mechanical version is $H = -(\hbar^2/2m)\partial^2/\partial x^2 + m\omega^2 x^2/2$. Comparing (8.1.5) with the definition (8.1.2), one sees that this exponential of the Hamiltonian (times $-it/\hbar$) is precisely the desired time evolution *operator*,

$$U(t) = e^{-iHt/\hbar}. \quad (8.1.6)$$

The Hamiltonian must be Hermitian, $H^\dagger = H$, in order for $U(t)$ to be unitary,

$$U(t)^\dagger = e^{+iH^\dagger t/\hbar} \Rightarrow U(t)^\dagger U(t) = 1 = e^{i0} = e^{+i(H^\dagger - H)t/\hbar}. \quad (8.1.7)$$

Typically we choose our basis states to be the *eigenstates* of the Hamiltonian,

$$H|\psi_n\rangle = E_n|\psi_n\rangle, \quad (8.1.8)$$

as they have particularly trivial time dependence

$$|\psi_n(t)\rangle = e^{-iE_n t/\hbar} |\psi_n(0)\rangle. \quad (8.1.9)$$

Since H is Hermitian (for our isolated, probability conserving system), its eigenvalues, the E_n , are *real* and the time dependence is simply a *phase* that changes linearly with time. This is what ensures that the normalization measured by $\langle\psi(t)|\psi(t)\rangle$ does not change with time. Such states are often referred to as *stationary* states (since the time dependence is trivial).

In principle, we have just solved all quantum dynamics! Of course, actually evaluating this exponential of the Hamiltonian can be (and usually is) a challenge. A quantum system whose space of states is N -dimensional (*i.e.*, n above runs from 1 to N) will have a Hamiltonian which is an $N \times N$ matrix. Most systems of interest will have a very large, or infinite, dimensional space of states.

8.2 Symmetries

We have spent much of this quarter discussing the analysis “power” associated with symmetries. Now we want to develop some of the formalism associated with symmetries. (See also Chapter 10 on Group Theory.) A general linear transformation T (not to be confused with the time reversal operator - see Section 8.12), which maps an arbitrary state $|\psi\rangle$ into some different state $|\tilde{\psi}\rangle = T|\psi\rangle$, is called a *symmetry* if T is unitary, $T^\dagger = T^{-1}$, so that probability is conserved, *and* T commutes with the time evolution operator,¹

$$TU(t) = U(t)T \text{ or } [T, U(t)] = 0. \quad (8.2.1)$$

¹This definition applies to time-independent symmetry transformations. A more general formulation is required for Lorentz boosts and time-reversal transformations, which have the effect of changing the meaning of time.

To understand this, consider some arbitrary initial state $|\psi(0)\rangle$, and imagine that you have worked out how this state evolves in time so that you know $|\psi(t)\rangle = U(t)|\psi(0)\rangle$. Applying the transformation T to the initial state $|\psi(0)\rangle$ will, in general, produce a different state $|\tilde{\psi}(0)\rangle = T|\psi(0)\rangle$. This transformed initial state will evolve in time into $|\tilde{\psi}(t)\rangle = U(t)|\tilde{\psi}(0)\rangle = U(t)T|\psi(0)\rangle$. But, if condition (8.2.1) is satisfied, then one can interchange $U(t)$ and T and write this result as $|\tilde{\psi}(t)\rangle = TU(t)|\psi(0)\rangle = T|\psi(t)\rangle$. In other words, if T is a symmetry transformation, transforming and then time-evolving any state is the same as first time-evolving, and then applying the symmetry transformation. This is summarized by the diagram²

$$\begin{array}{ccc} |\psi(t)\rangle & \xrightarrow{T} & |\tilde{\psi}(t)\rangle \\ U(t) \uparrow & & \uparrow U(t) \\ |\psi(0)\rangle & \xrightarrow{T} & |\tilde{\psi}(0)\rangle \end{array} \quad (8.2.2)$$

showing that $|\tilde{\psi}(t)\rangle$ can be constructed from $|\psi(0)\rangle$ by following *either* path.

The condition (8.2.1) that the transformation T commute with the time evolution operator is equivalent to the condition that T commute with the Hamiltonian,

$$[T, H] \equiv TH - HT = 0. \quad (8.2.3)$$

Symmetries have many useful consequences. One class of applications follows directly from the basic definition embodied in the diagram (8.2.2) — if you understand how some state $|\psi\rangle$ evolves in time, you can immediately predict how the transformed state $|\tilde{\psi}\rangle$ will evolve. For example, we will be discussing a transformation known as charge conjugation which interchanges particles and antiparticles, *e.g.*, turning a proton into an antiproton, a π^+ into a π^- , *etc.* Charge conjugation is a symmetry of the strong and electromagnetic interactions (recall the table at the end of Chapter 7). This symmetry directly implies that the rate at which a Δ^{++} baryon decays to a proton and a π^+ , which is a strong interaction decay, is the same as the rate at which the $\bar{\Delta}^{--}$ antibaryon (the antiparticle of the Δ^{++}) decays to an antiproton and a π^- . And it implies that the cross section for π^+ scattering on protons must be the same as the cross section for π^- mesons to scatter on antiprotons. So we have understood a lot about the strong interactions, knowing only one of its symmetry properties, but no other details about the dynamics.

A second category of applications follows from the commutativity (8.2.3) of a symmetry transformation with the Hamiltonian. Recall, from linear algebra, that two matrices (or linear operators) are simultaneously diagonalizable if and only if they commute. Consequently, if T is a symmetry then there exist states $\{|\psi_n\rangle\}$ which are *simultaneous* eigenstates of the Hamiltonian *and* of the transformation T ,

$$H|\psi_n\rangle = E_n|\psi_n\rangle, \quad (8.2.4a)$$

$$T|\psi_n\rangle = t_n|\psi_n\rangle. \quad (8.2.4b)$$

The eigenvalue E_n of the Hamiltonian is the energy of the state $|\psi_n\rangle$ — as noted above the Hamiltonian eigenstates are called energy eigenstates or *stationary states*. As already noted the time-dependent Schrödinger equation (8.1.4) implies that the time evolution of an energy eigenstate is just $|\psi_n(t)\rangle = e^{-iE_nt/\hbar}|\psi_n(0)\rangle$. Hence, an eigenstate of the Hamiltonian is also an eigenstate of the

²Mathematicians call this a *commutative diagram*.

time evolution operator $U(t)$, with eigenvalue $e^{-iE_nt/\hbar}$, and the state is stationary in time except for the overall *phase*. Finally we have also noted that, since the Hamiltonian is a Hermitian operator, its eigenvalues E_n must be *real* (no decays included yet).

Because the symmetry transformation T is also a unitary operator, its eigenvalues t_n must also be (simply) phase factors, $t_n = e^{i\phi_n}$ for some *real* phase ϕ_n .³ The simultaneous diagonalizability of H and T implies that energy eigenstates can also be labeled by an additional (quantum) number, t_n (or equivalently ϕ_n), which characterizes the effect of the symmetry transformation T on the state. Phrased differently, the eigenvalues of a symmetry transformation T define a quantum number which distinguishes different classes of eigenstates, and we can make use of this more detailed labeling of the states. Note also that the eigenvalues of T are *constants* of the motion, *i.e.*, are unchanged under the time dependence specified by $U(t)$ and correspond to conserved quantities.

There are many examples of this. A particle moving in a (one dimensional) square well potential, $V(x) = \begin{cases} 0, & |x| < L/2; \\ \infty, & \text{otherwise,} \end{cases}$ is an example of a system in which a parity transformation, $x \rightarrow -x$, is a symmetry, since the potential is unchanged by this transformation. Consequently, energy eigenstates in this potential can be labeled by their parity; their wavefunctions must either be even, $\psi_n(x) = \psi_n(-x)$, or odd, $\psi_n(x) = -\psi_n(-x)$, under $x \rightarrow -x$.

More generally, real particles (in infinite, empty space) can be labeled by their momentum and energy, as well as their angular momentum (spin), electric charge, baryon number, lepton number and intrinsic parity. As we have discussed several times before (and will again), these are all examples of quantum numbers which are associated with specific symmetries.

A third category of applications of symmetries involves time evolution of states which are eigenstates of some symmetry T but which are not (simple) eigenstates of the Hamiltonian (*i.e.*, can exhibit nontrivial time dependence). Let $|\psi_{\text{in}}\rangle$ be some initial state which is an eigenstate of the symmetry T with eigenvalue t_{in} . Let $|\psi_{\text{out}}\rangle$ be some final state which is an eigenstate of the symmetry T with eigenvalue t_{out} . For example, think of $|\psi_{\text{in}}\rangle$ as the initial state of some scattering experiment involving two incoming particles of types a and b , while $|\psi_{\text{out}}\rangle$ is a final state describing outgoing particles of types c and d . Can the scattering process $a + b \rightarrow c + d$ occur? In other words, can the matrix element $\langle\psi_{\text{out}}|U(t)|\psi_{\text{in}}\rangle$, giving the amplitude for the initial state to evolve into the chosen final state, be non-zero? The answer is no — *unless* the symmetry eigenvalues (*i.e.*, the conserved quantum numbers) of the initial and final states coincide. Written mathematically we have

$$\langle\psi_{\text{out}}|U(t)|\psi_{\text{in}}\rangle = 0 \quad \text{if} \quad t_{\text{in}} \neq t_{\text{out}}. \quad (8.2.5)$$

To see that Eq. (8.2.5) must be true we use the fact that T is unitary to write $1 = T^\dagger T$. Inserting the identity operator changes nothing, so

$$\langle\psi_{\text{out}}|U(t)|\psi_{\text{in}}\rangle = \langle\psi_{\text{out}}|T^\dagger T U(t)|\psi_{\text{in}}\rangle = \langle\psi_{\text{out}}|T^\dagger U(t)T|\psi_{\text{in}}\rangle. \quad (8.2.6)$$

The last step used the condition that T is a symmetry to interchange T and $U(t)$. By assumption, $|\psi_{\text{in}}\rangle$ is an eigenstate of T , $T|\psi_{\text{in}}\rangle = t_{\text{in}}|\psi_{\text{in}}\rangle$, and similarly $T|\psi_{\text{out}}\rangle = t_{\text{out}}|\psi_{\text{out}}\rangle$. Taking the Hermitian conjugate of this last relation gives $\langle\psi_{\text{in}}|T^\dagger = \langle\psi_{\text{in}}|t_{\text{in}}^*$. Use these eigenvalue relations for $|\psi_{\text{in}}\rangle$ and

³To show this, multiply each side of the eigenvalue condition (8.2.4b) by its Hermitian conjugate to obtain $\langle\psi_n|T^\dagger T|\psi_n\rangle = t_n^* t_n \langle\psi_n|\psi_n\rangle$. The left hand side is just $\langle\psi_n|\psi_n\rangle$ since T is unitary, so this condition can only be satisfied if $|t_n| = 1$.

$\langle \psi_{\text{out}} |$ to simplify $\langle \psi_{\text{out}} | T^\dagger U(t) T | \psi_{\text{in}} \rangle$. The result is

$$\langle \psi_{\text{out}} | U(t) | \psi_{\text{in}} \rangle = t_{\text{out}}^* t_{\text{in}} \langle \psi_{\text{out}} | U(t) | \psi_{\text{in}} \rangle. \quad (8.2.7)$$

Note that exactly the same matrix element appears on both sides. To satisfy this equation either $t_{\text{out}}^* t_{\text{in}}$ must equal 1, *or else* the matrix element $\langle \psi_{\text{out}} | U(t) | \psi_{\text{in}} \rangle$ must vanish. Because the eigenvalues of unitary T are pure phases, it follows that $t_{\text{out}}^* = 1/t_{\text{out}}$. Hence the condition that $t_{\text{out}}^* t_{\text{in}} = 1$ is the same as the statement that $t_{\text{in}} = t_{\text{out}}$. The interested reader may be more familiar with the corresponding statement about the eigenstates of the corresponding Hermitian *generator* Q of the transformation T , $T = e^{iQa}$. Such eigenvectors constitute an orthogonal set of basis states and the eigenvalues of the Hermitian operator Q are *real*. That is precisely what we have just proved. If the eigenvalue of Q is q_n , then $t_n = e^{iq_n a}$ is the required pure phase, $t_n^* = e^{-iq_n a}$. The symmetry respecting physics in $U(t)$ will not change the eigenvalue of either T or Q in going from the “in” state to the “out” state. States corresponding to *different* eigenvalues of the symmetry operators are orthogonal, $\langle \psi_m | \psi_n \rangle = 0$ for $q_m \neq q_n$, and remain so as they evolve in time!

The key point here is that symmetries can be used to understand what types of final states can, or cannot, occur in many scattering experiments and decays, even without detailed knowledge of the dynamics. Conservation laws for energy, momentum, angular momentum, electric charge, and baryon and lepton number (and more) can all be viewed as particular cases of this general result.

A final type of application (related to category 2 above) concerns sets of multiple symmetry transformations. Suppose transformations T_1 and T_2 are both symmetries, and hence both commute with the Hamiltonian. But suppose that T_1 and T_2 do *not* commute with each other. (As an example, discussed in detail below, let T_1 be J_3 , angular momentum along the $\hat{3}$ -axis, and T_2 be J_1 or J_2 , or one of the ladder operators J_\pm - recall Chapter 5.) Then one *cannot* simultaneously diagonalize the Hamiltonian and *both* T_1 and T_2 , although one can find a basis in which H and, say, T_1 are diagonal. Let $|\psi_n\rangle$ be one of these basis states, so that $H|\psi_n\rangle = E_n|\psi_n\rangle$ and $T_1|\psi_n\rangle = t_{1,n}|\psi_n\rangle$. Applying the symmetry transformation T_2 to the state $|\psi_n\rangle$ will produce some state $|\tilde{\psi}_n\rangle$, which must also be an eigenstate of the Hamiltonian with exactly the same energy E_n .⁴ It may be a linearly independent state, *i.e.*, $|\tilde{\psi}_n\rangle$ need not be proportional to $|\psi_n\rangle$. Consequently, the existence of symmetry transformations that do not mutually commute can lead to *degenerate* energy levels, *i.e.*, multiple linearly independent states with exactly the same energy. Further these energy degenerate states will comprise *complete* representations of the underlying symmetry group. This is necessarily true because there are (symmetry) transformations that move us around within the representation while simultaneously commuting with the Hamiltonian. Angular momentum eigenstates provide a familiar example of this. In any theory which is rotationally invariant, every energy eigenstate with non-zero angular momentum must be part of a degenerate multiplet. If the (total) angular momentum is $j\hbar$, then the multiplet will contain $(2j+1)$ states, since the projection of the angular momentum along a(n arbitrarily) chosen quantization axis can take any of $2j+1$ values, $\{-j, -j+1, \dots, j-1, j\}$, but the energy cannot depend on the value of this projection else it would not be rotationally invariant. Thus, knowing something about the symmetries of a given physical system, already tells us a good deal about the possible states of the system. The mathematics of symmetries and the associated representations is called Group Theory and we will present a brief introduction in this class. The representation with $2j+1$ elements, which we just described, is a representation of the symmetry group $SO(3)$, and the nearly identical group $SU(2)$. Hopefully you have learned something about

⁴This follows from the given assumption that T_2 is a symmetry, so that it commutes with H . Consequently, $H|\tilde{\psi}_n\rangle = H(T_2|\psi_n\rangle) = T_2(H|\psi_n\rangle) = T_2(E_n|\psi_n\rangle) = E_n|\tilde{\psi}_n\rangle$.

these groups either in your quantum or classical mechanics classes or in math methods (or in Chapter 5). Symmetries, and the corresponding conserved quantities, are important for understanding both classical and quantum physics, but probably more useful in the latter. We will apply the techniques from Chapter 5 at the end of this chapter to the 3-quark spin wave functions that we discussed in Chapter 7.

We can apply similar ideas to the $SU(3)$ (approximate) flavor symmetry. The statement that QCD is flavor symmetric means that the eigenstates of QCD, the hadrons, should appear in *complete* representations of $SU(3)$, the (now) familiar $\underline{8}$'s and $\underline{10}$'s. If the masses of the u, d, s quarks were identical and we ignored the weak and electromagnetic interactions, all members of the same representation would have the *same* rest mass. However, as we discussed in Chapter 7, the s quark mass is substantially larger than the u and d quark masses (although still much smaller than a typical hadron mass). Thus the full flavor $SU(3)$ symmetry is broken and the masses of hadrons within the same flavor $SU(3)$ multiplet (*i.e.*, representation) are not quite degenerate, but exhibit a small breaking that depends linearly on the number of s quarks. On an even smaller scale, the d quark mass is slightly larger than the u quark mass leading to a tiny breaking of isospin symmetry. Note that in the case of $SU(3)$ the representations are 2-D and the individual states are labeled by the “total $SU(3)$ quantum number”, *e.g.*, $\underline{8}$ or $\underline{10}$, along with isospin I , I_3 , and the strangeness content S . Also there are 2 linearly independent sets of ladder operators typically chosen as the ones for isospin, I_{\pm} , and one of the sets of operators that move us along the diagonals, typically labeled U_{\pm} and V_{\pm} , *i.e.*, $SU(3)$ contains 3 $SU(2)$ subgroups with ladder operators but only 2 sets are linearly independent. (The existence of the 3 subgroups is related to the 3-fold symmetry of the graphical forms of the representations in Chapter 7.)

8.3 Continuous symmetries

Continuous symmetries are symmetries which depend (continuously!) on some parameter that controls the magnitude of the transformation. Examples include translations and rotations. Let $T(a)$ denote a continuous symmetry transformation depending on the real parameter a . Assume (without loss of generality) that $a = 0$ corresponds to doing nothing, so that $T(0)$ equals the identity (no change) operator. One can always choose to define the parameterization so that $(T(a/2))^2 = T(a)$, or more generally that $(T(a/N))^N = T(a)$ for any N . Here we are appealing to our intuitive expectation that rotating twice through an angle $\theta/2$ $[(T(\theta/2))^2]$ about some axis is equivalent to rotating once through angle θ $[T(\theta)]$ about the same axis. This implies that $T(a)$ depends exponentially on the parameter a (as we have already suggested), so that one can write

$$T(a) = e^{iQa}, \quad (8.3.1)$$

for some operator Q , which is called the *generator* of the symmetry $T(a)$. (See also the discussion in Chapters 1, 5 and 10. We are here discussing the structure of the Lie Groups, of which we have discussed $U(1)$, $SU(2)$ and $SU(3)$.) In order for $T(a)$ to be unitary (as required), the generator Q must be Hermitian. Note that the relation between $T(a)$ and Q is completely analogous to the relation between the time evolution operator and the Hamiltonian; the Hamiltonian (divided by $-\hbar$) is the generator of time evolution.

The condition (8.2.3) that $T(a)$ commute with H implies that the generator Q of any continuous symmetry must also commute with the Hamiltonian,

$$[Q, H] = 0. \quad (8.3.2)$$

Once again, this implies that Q and H are simultaneously diagonalizable.

Note that, given some continuous symmetry transformation $T(a)$, one can extract the associated generator Q by performing a Taylor series expansion of $T(a)$ about $a = 0$. Keeping just the first non-trivial term gives $T(a) = 1 + iQa + \dots$, so that

$$Q = -i \left. \frac{d}{da} T(a) \right|_{a=0}. \quad (8.3.3)$$

Alternatively, given any Hermitian operator Q which commutes with the Hamiltonian, one can construct a unitary symmetry transformation by exponentiating iQ (times an arbitrary real number), as in (8.3.1). So one can regard either the generator Q , or the finite transformation $T(a)$, as defining a continuous symmetry. In the language of Group Theory, as outlined in Chapter 10, $T(a)$ is an element of the underlying symmetry group (a Lie Group) while Q is an element of the *algebra* corresponding to the group.

8.4 Spacetime symmetries

Spacetime symmetries are symmetries which characterize the underlying geometry of Minkowski space. Translations in both space and time, spatial rotations, and Lorentz boosts are all continuous spacetime symmetries. These are symmetries of the laws of physics, as currently understood. As we have already suggested, associated with those continuous symmetries that commute with the Hamiltonian, there are *additive* conserved quantities or quantum numbers (and corresponding conserved Noether currents - see Chapter 10 and the NYT article on our webpage, <http://courses.washington.edu/partsym/14Spr/Noether.pdf>), such as linear momentum, energy and angular momentum.

ASIDE In addition to continuous transformations certain *discrete* transformations are also useful as we have already mentioned, *e.g.*, parity ($P: \vec{x} \rightarrow -\vec{x}$), charge conjugation ($C: \text{particle} \rightarrow \text{anti-particle}$) and time reversal ($T: t \rightarrow -t$). These symmetries are associated with *multiplicative* quantum numbers. While these symmetries are *not* exact symmetries (*i.e.*, do *not* commute with *all* of the interactions), they do provide useful approximate symmetries (*i.e.*, commute with “most” of the interactions), as we will discuss later.

Returning to continuous spacetime symmetries, the total momentum operator \vec{P} (divided by \hbar) is the generator of spatial translations ($-i\nabla$). Hence, the unitary operator $T_{\text{trans}}(\Delta\vec{x})$, which has the effect of performing a spatial translation through a displacement $\Delta\vec{x}$, is an exponential of momentum,⁵

$$T_{\text{trans}}(\Delta\vec{x}) = e^{i\vec{P} \cdot \Delta\vec{x}/\hbar}. \quad (8.4.1)$$

In any translationally invariant theory, the total momentum \vec{P} commutes with the Hamiltonian (and hence with the time evolution operator). Therefore, conservation of momentum is a direct consequence of spatial translation invariance.

⁵Recall that in single particle quantum mechanics in, for simplicity, one dimension, the coordinate representation of the momentum operator is $\frac{\hbar}{i} \frac{\partial}{\partial x}$. So acting with $\exp(iP\Delta x/\hbar) = 1 + iP\Delta x/\hbar - \frac{1}{2}(iP\Delta x/\hbar)^2 + \dots$ on an arbitrary state $|\Psi\rangle$ is the same as acting with $\exp(\Delta x \frac{\partial}{\partial x}) = 1 + \Delta x \frac{\partial}{\partial x} + \frac{1}{2}(\Delta x \frac{\partial}{\partial x})^2 + \dots$ on the wavefunction $\Psi(x)$. This produces $\Psi(x) + \Delta x \Psi'(x) + \frac{1}{2}(\Delta x)^2 \Psi''(x) + \dots$ which is a Taylor series expansion of the translated wavefunction $\Psi(x + \Delta x)$.

The Hamiltonian H (divided by $-\hbar$) is the generator of time translations, and the associated unitary operator which has the effect of performing a time translation through an interval Δt is precisely the time evolution operator

$$U(\Delta t) = e^{-iH\Delta t/\hbar}. \quad (8.4.2)$$

The Hamiltonian commutes with itself, and therefore it satisfies the conditions defining the generator of a symmetry. Since the Hamiltonian is the operator which measures energy, this shows that conservation of energy is a direct consequence of time translation invariance.

A general *spacetime* translation with displacement $\Delta x = (\Delta x^0, \Delta \vec{x})$ is just a combination of a spatial translation through $\Delta \vec{x}$ and a time translation through $\Delta t = \Delta x^0/c$. The unitary operator which implements this spacetime translation is the product of $T_{\text{trans}}(\Delta \vec{x})$ and $U(\Delta t)$.⁶ Defining $P^0 \equiv H/c$ allows one to write this as a single exponential of a Minkowski space dot product,

$$T_{\text{trans}}(\Delta x) = T_{\text{trans}}(\Delta \vec{x}) \times U(\Delta x^0/c) = e^{-iP^\mu \Delta x_\mu/\hbar}. \quad (8.4.3)$$

The total angular momentum \vec{J} (divided by \hbar) is the generator of rotations. The unitary operator which implements a rotation through an angle θ about an axis defined by a unit vector \hat{n} is an exponential of the component of angular momentum along \hat{n} ,

$$T_{\text{rot}}(\theta, \hat{n}) = e^{i\theta \hat{n} \cdot \vec{J}/\hbar}. \quad (8.4.4)$$

The total angular momentum \vec{J} commutes with the Hamiltonian in any rotationally invariant theory. Hence, conservation of angular momentum is a direct consequence of spatial rotation invariance.

One can also define operators \vec{G} which are the generators of Lorentz boosts, so that the unitary operator which implements a boost along some direction \hat{n} can be written as an exponential,

$$T_{\text{boost}}(y, \hat{n}) = e^{iy \hat{n} \cdot \vec{G}/\hbar}. \quad (8.4.5)$$

The parameter y , which, as we have noted earlier, is called *rapidity*, is not the velocity of the boost, but it determines the boost velocity via $v/c = \tanh y$. If we consider a particle whose motion (energy E , momentum p) is characterized by this boost, we have $\tanh y = pc/E$. In contrast to the situation with rotations and translations, the boost generators \vec{G} do not commute with H because Lorentz boosts change the meaning of time.⁷ Because of this, invariance under Lorentz boosts does not lead to any *additional* conserved quantities analogous to momentum or angular momentum.

ASIDE: Since in this class we often discuss physics as it is being studied at particle colliders, *e.g.*, the LHC, we should note that, in the context of such experiments, the term rapidity is typically used for boosts *only* along the direction of the beams, say the \hat{x}^3 direction. Thus one will often see (as

⁶Because $T_{\text{trans}}(\Delta \vec{x})$ commutes with $U(\Delta t)$ (or equivalently, because \vec{P} commutes with H), the order in which one performs this product doesn't matter.

⁷The boost generators \vec{G} depend explicitly on time, and the required condition that they must satisfy turns out to be $\frac{\partial}{\partial t} \vec{G} + i[H, \vec{G}] = 0$.

we did in the HW) rapidity defined via

$$\begin{aligned}
 p^\mu &= (E/c, p_1, p_2, p_3) \\
 p_T &= \sqrt{p_1^2 + p_2^2} \\
 E &= \sqrt{m^2 c^4 + p_T^2 c^2} \cosh y \\
 p_3 &= \sqrt{m^2 c^2 + p_T^2} \sinh y \\
 \tanh y &= p_3 c / E, \quad y = 0.5 \ln [(E + p_3 c) / (E - p_3 c)].
 \end{aligned} \tag{8.4.6}$$

Thus the 3-momenta of particles are characterized by the “transverse” momentum (transverse to the beam) p_T and the rapidity y , where the 2-vector direction of \vec{p}_T is described by the azimuthal angle ϕ . Note that making the \hat{x}^3 direction “special” in this sense is reasonable as the initial state (beams colliding along the \hat{x}^3 direction) clearly breaks the overall rotational symmetry and the final states clearly exhibit this lack of symmetry. The many particles produced at the LHC are observed to be distributed approximately uniformly in the rapidity defined in Eq. (8.4.6) but are restricted to small values of the transverse momentum, p_T (of order a few GeV/c). In those rare events which produce particles at large p_T , the large p_T particles are observed to cluster into localized (in rapidity and azimuthal angle ϕ) “jets” of particles and are interpreted to arise from the large angle scattering of the underlying quarks and gluons. (These jets as a characteristic feature of high energy collisions is a topic I have studied for much of my scientific career.)

8.5 Charge, lepton, and baryon number

The electric charge Q is an operator which, when acting on any state containing particles with individual charges $\{q_i\}$ ($i = 1, \dots, N$), measures the sum of all these charges,

$$Q|\Psi\rangle = q_{\text{tot}}|\Psi\rangle \tag{8.5.1}$$

with $q_{\text{tot}} \equiv \sum_{i=1}^N q_i$. So, as its name suggests, Q measures the total electric charge of any state. More precisely, each q_i should be understood as the charge of a particle in units of $|e|$. The electric charges of all known particles which can be produced in isolation at distances large compared to a fermi (*i.e.*, not including quarks) are integer multiples of $|e|$; this is known as *charge quantization*. Hence the operator Q will always have integer eigenvalues.

Maxwell’s equations are inconsistent if electric charge is not conserved. Therefore, Q must commute with the Hamiltonian (or with the time evolution operator), and hence Q is the generator of a continuous symmetry,

$$T_Q(\alpha) \equiv e^{i\alpha Q}. \tag{8.5.2}$$

Applying this transformation to any state (of definite charge) multiplies the state by a *phase* proportional to its electric charge, $T_Q(\alpha)|\Psi\rangle = e^{i\alpha q_{\text{tot}}}|\Psi\rangle$. Note that this is also a continuous symmetry (characterized by the continuous parameter α) that operates not in spacetime like translations, but simply in the space of complex numbers. In the group theory language of Chapter 10, the invariance of electromagnetism is described by the group $U(1)$. The corresponding conserved additive quantum number is just the charge q_{tot} .

In precisely the same fashion, one may regard baryon number B and lepton number L ⁸ as quantum operators which measure the total baryon number or lepton number, respectively, with underlying $U(1)$ symmetries. And one may exponentiate either of these operators to form continuous symmetry transformations generated by B and L ,

$$T_B(\alpha) \equiv e^{i\alpha B}, \quad T_L(\alpha) \equiv e^{i\alpha L}. \quad (8.5.3)$$

8.6 Approximate symmetries

There are many circumstances where it is useful to consider transformations which are almost, but not quite, symmetries of a theory. Consider, for example, a hydrogen atom in a weak background magnetic field. If the magnetic field were zero, then the Hamiltonian describing the dynamics of the atom would be rotationally invariant. As noted above, this implies that energy eigenstates with non-zero angular momentum must form *degenerate* energy levels. Turning on a magnetic field breaks three dimensional rotation invariance, since the Hamiltonian will now contain terms which depend on the direction of the background magnetic field. (More precisely, turning on a magnetic field reduces the symmetry from three dimensional rotation invariance down to one dimensional rotation invariance with respect to rotations about the direction of the magnetic field.) The presence of the magnetic field will perturb the energy levels of the atom, and lift the degeneracy of energy eigenstates with differing angular momentum projections along the direction of the field. But, if the magnetic field is sufficiently weak, the energy splitting induced by the field will be small (compared to the spacings between the non-degenerate energy levels in the absence of the field). In this circumstance, it makes sense to regard the Hamiltonian of the system as the sum of a “large” rotationally invariant piece H_0 , which describes the atom in the absence of a magnetic field, plus a “small” perturbation ΔH , which describes the interaction with the weak magnetic field,

$$H = H_0 + \Delta H. \quad (8.6.1)$$

One can systematically calculate properties of the atom as a series expansion in the size of ΔH , or more correctly in the size of ΔH divided by the appropriate eigenvalue of H_0 , since the expansion parameter must be dimensionless. The starting point involves ignoring ΔH altogether and understanding the properties of H_0 , including the properties of its eigenstates. And when studying the physics of H_0 alone, one can use full three dimensional rotation symmetry to characterize the corresponding energy eigenstates. These eigenstates, in turn, provide a complete, orthogonal set of basis states useful for analyzing this system *even* when the perturbation, ΔH , is included.

Exactly the same approach can be applied to particle physics to separate the effects of weak and electromagnetic interactions (small) from those of strong interactions (large). Similarly we can study the effects of the (small) symmetry breaking due to the quark masses on the flavor $SU(3)$ symmetric limit. In fact, this approach of using “perturbation theory” to analyze a system in terms of perturbations around a simple, completely analyzable symmetric limit, is an extremely important tool to have in your physics toolbox!

⁸Note that we are (perhaps unfortunately) using the same symbol L for both lepton number and orbital angular momentum. The meaning in any specific situation will (hopefully) be obvious from the context.

8.7 Flavor symmetries

Strong interactions, as described by quantum chromodynamics, preserve the net number of quarks of each flavor. Strong interactions can cause the creation or annihilation of quark-antiquark pairs of any given flavor, but this does not change the number of quarks minus antiquarks of each flavor. This is also true of electromagnetic interactions, but not weak interactions. Consequently, in a hypothetical world in which weak interactions are turned off, operators which measure the number of quarks minus antiquarks of each flavor,

$$N_u = (\# u \text{ quarks}) - (\# \bar{u} \text{ quarks}), \quad (8.7.1a)$$

$$N_d = (\# d \text{ quarks}) - (\# \bar{d} \text{ quarks}), \quad (8.7.1b)$$

$$N_s = (\# s \text{ quarks}) - (\# \bar{s} \text{ quarks}), \quad (8.7.1c)$$

$$N_c = (\# c \text{ quarks}) - (\# \bar{c} \text{ quarks}), \quad (8.7.1d)$$

$$N_b = (\# b \text{ quarks}) - (\# \bar{b} \text{ quarks}), \quad (8.7.1e)$$

$$N_t = (\# t \text{ quarks}) - (\# \bar{t} \text{ quarks}), \quad (8.7.1f)$$

all commute with the QCD (and EM) Hamiltonian. Therefore, all these operators may be regarded as generators of continuous symmetries. Note that baryon number equals the total number of quarks minus antiquarks, divided by three,

$$B = \frac{1}{3} [(\# \text{ quarks}) - (\# \text{ antiquarks})] = \frac{1}{3} \sum_{f=u,d,s,c,b,t} N_f, \quad (8.7.2)$$

since baryons contain three quarks, while antibaryons contain three antiquarks. For historical reasons, it is conventional to refer to *strangeness* as the number of strange *anti*-quarks minus quarks,

$$S \equiv -N_s = (\# \bar{s} \text{ quarks}) - (\# s \text{ quarks}). \quad (8.7.3)$$

This definition assigns strangeness +1 to the (originally observed) K^+ meson (with a \bar{s} antiquark), and this convention predates the development of QCD and the quark model of hadrons.

In a world without weak interactions, π mesons would be absolutely stable because there are no lighter strongly interacting hadrons into which pions could decay. Kaons (K mesons) would also be stable, even though they are over three times heavier than pions, because K mesons are the lightest hadron with nonzero strangeness (and strangeness is conserved by strong interactions). Similarly, the Ω^- baryon, containing three strange quarks, would be stable because there is no other combination of hadrons it might decay into that has lower energy, together with baryon number one and strangeness minus three.

Completely analogous arguments apply to hadrons containing the heavier charm and bottom quarks. In the absence of weak interactions, there would be many additional stable hadrons containing nonzero net “charmness”, or net “bottomness” (or “topness”).

8.8 Isospin

Figure 8.1 graphically displays the mass spectrum of light mesons and baryons. Looking at this figure, or the tables containing information about hadrons in the previous chapter, many degeneracies or near-degeneracies are immediately apparent. For example, the masses of the π^+ and π^- mesons are

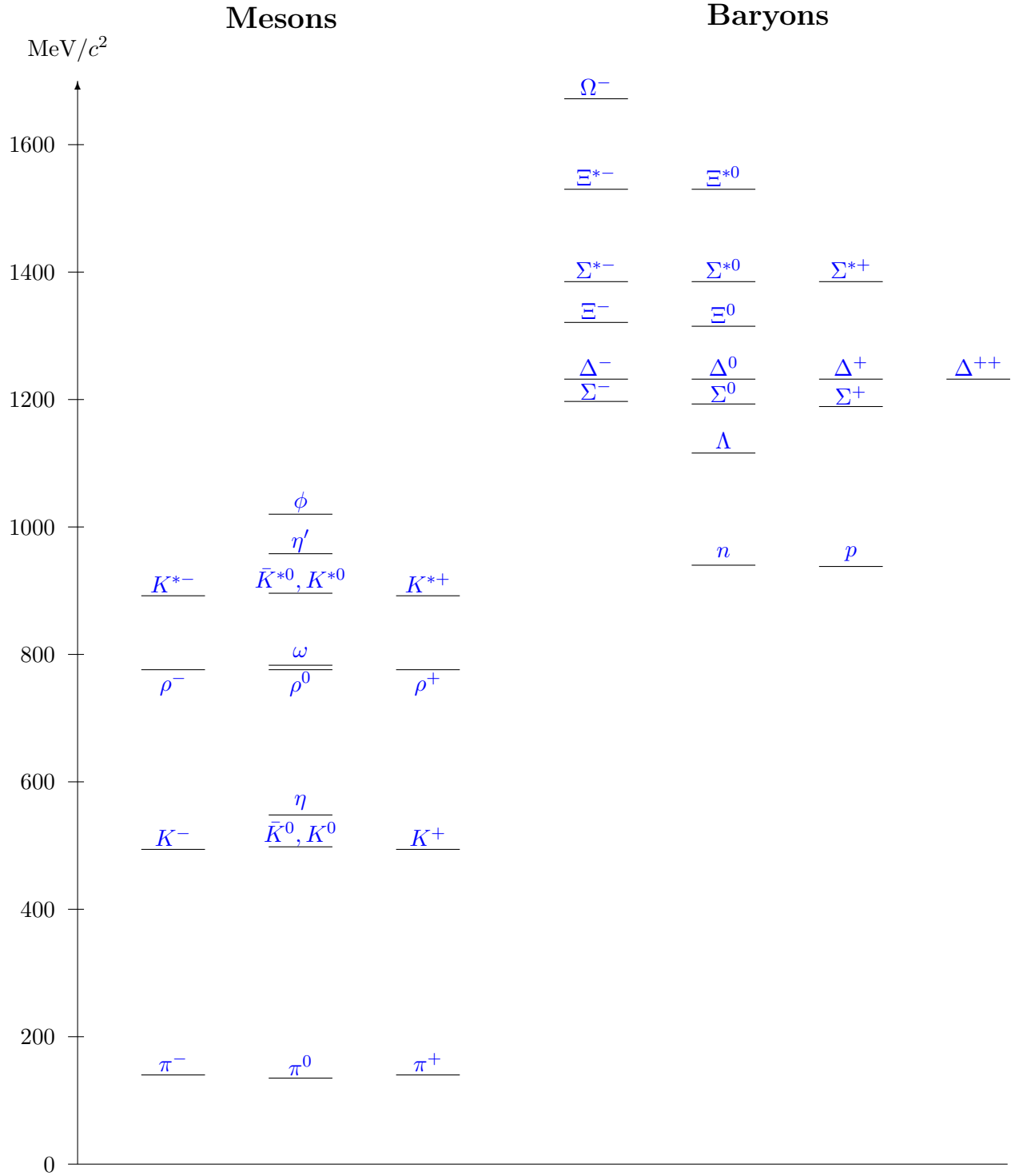


Figure 8.1: The mass spectrum of light mesons and baryons. Each column show mesons or baryons with a particular charge. (Only the lightest mesons and baryons with given quark content and spin are shown. Many additional excited states with higher masses are known, but are not displayed in this plot.)

the same, while the mass of the π^0 meson is only slightly different. The mass of the neutron is quite close to that of the proton. The masses of the Σ^+ , Σ^0 and Σ^- baryons are all different from the other hadrons, but very close to each other. And likewise for many other “multiplets” of mesons and baryons. As we have noted several times earlier, the most obvious “near degeneracy” is that due to isospin symmetry, a topic which we discuss again and more completely in the following.

Comparing the quark content of various hadrons (and referring to Tables 7.7 - 7.10 as needed), one sees that the near-degeneracies are all associated with substitutions of u for d quarks, or vice-versa. For example, the Σ^+ baryon has two up and one strange quark. Replacing one up quark by a down converts the Σ^+ into a Σ^0 , whose mass is larger than that of the Σ^+ by $3.3 \text{ MeV}/c^2$, which is less than 0.3% of the Σ^+ mass (a true “near degeneracy”). Replacing the remaining up quark by a down converts the Σ^0 into a Σ^- , whose mass is an additional $4.8 \text{ MeV}/c^2$ larger.

The mass differences among the three Σ baryons, the three π mesons, between the neutron and proton, or within any of the other nearly degenerate multiples, must arise from some combination of two effects. First, the masses of up and down quarks are not quite the same. The mass of a down quark (*c.f.* Table 7.1) is a few MeV/c^2 larger than that of an up quark. This mass difference is tiny compared to the masses of hadrons, but it is comparable to the few MeV/c^2 mass splittings within the various near-degenerate multiplets.

Second, while the color interactions of up and down quarks are identical, the other interactions are different. They have differing electric charges ($2/3$ for u , and $-1/3$ for d), which means that their electromagnetic interactions are not the same. Their weak interactions also differ. But, as far as hadronic masses are concerned, the effects of weak and electromagnetic interactions are numerically small perturbations on top of the dominant effects due to strong interactions, and strong interactions are flavor-blind. In a hypothetical world in which weak and electromagnetic interactions are absent, and in which up and down quarks have the same mass, these near-degeneracies would all become *exact* degeneracies.

As we have discussed many times, you are encouraged to interpret this situation as being similar to angular momentum multiplets (because it is very similar - the *same* equations have the *same* solutions). In any rotationally invariant theory, every state with angular momentum $J\hbar$ is part of a multiplet containing $2J + 1$ degenerate states. A rotation transforms the different states in the multiplet into linear combinations of each other. The simplest non-trivial case is $J = \frac{1}{2}\hbar$, the so-called *fundamental* representation, whose multiplet contains two (linearly independent) states conventionally chosen to have angular momentum up or down along some given axis, and represented by $|\uparrow\rangle$ and $|\downarrow\rangle$. The action of a rotation corresponds to a linear transformation,

$$\begin{pmatrix} |\uparrow'\rangle \\ |\downarrow'\rangle \end{pmatrix} = M \begin{pmatrix} |\uparrow\rangle \\ |\downarrow\rangle \end{pmatrix}. \quad (8.8.1)$$

For a rotation about an axis \hat{n} through an angle θ , the matrix M has the form (note especially the factors of $1/2$ in the angle and the representation of the operator $\vec{I} = \vec{\sigma}/2$)

$$M = e^{i\theta\hat{n}\cdot\vec{I}} = e^{i(\theta/2)\hat{n}\cdot\vec{\sigma}} = \left(\cos \frac{\theta}{2} + i\hat{n} \cdot \vec{\sigma} \sin \frac{\theta}{2} \right), \quad (8.8.2)$$

with $\vec{\sigma}$ denoting the Pauli matrices, $\sigma_1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$, $\sigma_2 = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}$, and $\sigma_3 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$.⁹ You can easily (and should) check that M is a unitary matrix with determinant equal to one. As noted earlier, the

⁹Of special interest are the combinations $\sigma_{\pm} = (\sigma_1 \pm i\sigma_2)/2$ which provide a realization of the ladder operators of Eq. (5.1.11): $\sigma_+ = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}$, $\sigma_- = \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}$. The former turns $|\downarrow\rangle$ into $|\uparrow\rangle$ and the latter does the opposite.

space (or group) of such 2×2 matrices is called $SU(2)$, the Special Unitary Group in 2 dimensions.¹⁰

In the limit that the u and d quarks are degenerate in mass (and weak and electromagnetic interactions are turned off), there is an analogous symmetry which transforms up and down quark states into new linear combinations of the two flavors,

$$\begin{pmatrix} |u'\rangle \\ |d'\rangle \end{pmatrix} = M \begin{pmatrix} |u\rangle \\ |d\rangle \end{pmatrix}, \quad (8.8.3)$$

where M is any 2×2 unitary matrix with determinant one. This is the formal definition of the symmetry called *isospin* (or isotopic spin) with M an isospin “rotation”.

The mathematical structure of isospin rotations is completely analogous to spatial rotations (although isospin has nothing to do with ordinary spatial rotations, but, as usual, “the same equations have the same solutions”). There are three generators of isospin rotations, I_1 , I_2 and I_3 . Their commutation relations have the same form as the commutation relations of angular momentum operators (which are the generators of spatial rotations),¹¹

$$[I_a, I_b] = i\epsilon_{abc} I_c. \quad (8.8.4)$$

Total isospin is denoted by I , and can have either integer or half-integer values. An up quark has $I_3 = +1/2$, while a down quark has $I_3 = -1/2$. Hence an up quark behaves (with respect to isospin rotations) just like an up spin does (with respect to spatial rotations). This parallel is the origin of the names ‘up’ and ‘down’ for the two lightest quarks. For antiquarks, the assignments are reversed, a \bar{u} quark has $I_3 = -1/2$ while a \bar{d} has $I_3 = +1/2$.¹²

When we build states containing multiple up and down quarks (or antiquarks), the addition of isospin works just like adding angular momentum. For example, combining two isospin one-half objects can yield either isospin 0 or isospin 1, *i.e.*, $\underline{2} \otimes \underline{2} = \underline{3} \oplus \underline{1}$. An antisymmetric combination of u and d quarks,

$$(ud - du), \quad (8.8.5)$$

gives $I = 0$ isospin singlet, while a symmetric combination gives the isospin one triplet state(s). Hence, the three $I = 1$ flavor states of two quarks are

$$(uu), (ud + du), (dd), \quad (8.8.6)$$

with I_3 for these states equaling $+1$, 0 , and -1 , respectively. Similarly, when three u or d quarks are combined (as in a baryon), the resulting isospin can be either $1/2$ or $3/2$, *i.e.*, $\underline{2} \otimes \underline{2} \otimes \underline{2} = \underline{4} \oplus \underline{2} \oplus \underline{2}$.

Looking back at the nearly degenerate set of particles shown in Figure 8.1, the π^+ , π^0 and π^- mesons form an $I = 1$ multiplet, whose masses would be exactly equal were it not for the perturbing effects

¹⁰ As discussed in the Group Theory Lecture in Chapter 10, the two groups $SO(3)$ and $SU(2)$ are intimately related with exactly the same “algebra”. Thus it is no surprise that the current discussion looks like the (hopefully) more familiar case of angular momentum.

¹¹ This is just the statement that the two groups have the same algebra as already noted in the previous footnote.

¹² It is (note the now familiar minus sign) $\begin{pmatrix} -|\bar{d}\rangle \\ |\bar{u}\rangle \end{pmatrix}$ which transforms in the same manner as $\begin{pmatrix} |u\rangle \\ |d\rangle \end{pmatrix}$, namely $\begin{pmatrix} -|\bar{d}'\rangle \\ |\bar{u}'\rangle \end{pmatrix} = M \begin{pmatrix} -|\bar{d}\rangle \\ |\bar{u}\rangle \end{pmatrix}$.

of weak and electromagnetic interactions and the up and down quark mass difference. Similarly, the K^+ and K^0 mesons (whose quark contents are $u\bar{s}$ and $d\bar{s}$, respectively) form an isospin $1/2$ multiplet with strangeness one, while the K^- and \bar{K}^0 mesons (with quark content $s\bar{u}$ and $s\bar{d}$) form another $I = 1/2$ multiplet with $S = -1$. The three rho mesons form another $I = 1$ multiplet. Turning to the baryons, the two nucleons (*i.e.*, the proton and the neutron) form an $I = 1/2$ multiplet (as do the Ξ baryons), while the three Σ baryons have $I = 1$ and the four Δ baryons have $I = 3/2$.

Conservation of isospin (by the strong interactions) can also be used to explain a variety of more detailed hadronic properties, including the fraction of Δ^+ decays which yield $p\pi^0$ versus $n\pi^+$, or the fraction of different pion pairs produced by ρ decays. While the total rate of Δ^+ or ρ^0 decays will depend on the details of the QFT version of the strong interactions, the *relative* branching ratios into differing particle-pair states, *e.g.*, $p\pi^0$ versus $n\pi^+$ or $\pi^0\pi^0$ versus $\pi^+\pi^-$, will (in the limit of conserved isospin) be determined by the corresponding Clebsch-Gordan coefficient. Recall that a table of C-G coefficients was included at the end of Chapter 7. We will work through some detailed examples in the HW.

Isospin conservation can also be used to explain the absence of many unseen decay modes. For example, the $\Lambda(1690)$ is an excited state of the $\Lambda(1116)$ baryon (still $I = 0$, but $J^P = \frac{3}{2}^-$, *i.e.*, 1 unit of orbital angular momentum), with 1690 MeV rest energy and quark content uds . Roughly 25% of the time, a $\Lambda(1690)$ decays to a $\Lambda(1116)$ plus two pions. But it *never* decays to a $\Lambda(1116)$ plus a single pion, despite that fact that more energy would be available for conversion into kinetic energy if only a single pion were produced. To understand why decays to a $\Lambda(1116)$ plus two pions are favored, note that the $\Lambda(1116)$ baryon, and its excited states like the $\Lambda(1690)$ have $I = 0$, while pions have $I = 1$. So the decay $\Lambda(1690) \rightarrow \Lambda(1116) + \pi$ would have $\Delta I = 1$ — an initial state with isospin zero and a final state of isospin one. Thus this strong interaction decay is not allowed, since the strong interactions conserve isospin. But in the final state of the observed decay $\Lambda(1690) \rightarrow \Lambda(1116) + \pi + \pi$, the total isospin is the combination of two $I = 1$ pions plus the $I = 0$ $\Lambda(1116)$ baryon. Adding two isospin one objects can yield isospin two, one, or zero. So, if the final pions combine to form zero isospin, then isospin will be conserved in this decay, and a strong interaction decay is allowed. The two pion decay can conserve isospin, while the single pion case cannot. Similarly we can understand why the $\Lambda(1690)$ can decay into a Σ ($I = 1$) and one or two pions.

Because isospin is only an approximate symmetry, predictions one can make using isospin invariance are not exact results in the real world. However, because the up and down quark mass difference is so small, and weak and electromagnetic interactions are *much* weaker than strong interactions, predictions which follow from isospin invariance are quite accurate — violations are typically at or below the 1% level.

8.9 Parity

The most familiar of the relevant *discrete* symmetries with *multiplicative* eigenvalues is parity. A parity transformation, denoted P , has the effect of reversing all spatial coordinate axes. Therefore, a parity transformation acting on a state of a single particle located at some spatial position \vec{x} produces a state in which the particle is located at $-\vec{x}$. Fourier transforming to the momentum representation, one can equally well say that a parity transformation acting on a single particle state with momentum \vec{p} will produce a state with momentum $-\vec{p}$. Written symbolically, this suggests that if $|\vec{p}\rangle$ represents a state of some particle with momentum \vec{p} , then the parity transformed state should be $P|\vec{p}\rangle = |-\vec{p}\rangle$. This is not quite right, however, as the unitary transformation P can also produce

a change in the overall phase of the state. Therefore, in general one must write

$$P |\vec{p}\rangle = \eta_P |-\vec{p}\rangle, \quad (8.9.1)$$

where η_P is some phase factor which can depend on the type of particle under consideration. A parity transformation does *not* change the spin or angular momentum of a particle.¹³

Applying two parity transformations amounts to reversing the directions of all spatial coordinate axes, and then reversing them all over again. This overall transformation has no net effect. Hence, as an operator, parity must square to the identity, $P^2 = 1$. This implies that the phase η_P appearing in Eq. (8.9.1) must square to one, $\eta_P^2 = 1$, so either $\eta_P = +1$ or $\eta_P = -1$. This sign is called the *intrinsic parity* of a particle. Some particles (such as protons and neutrons) have positive intrinsic parity, while others (such as pions and photons) have negative parity. One can show (from relativistic quantum mechanics) that for particles which are bosons, the intrinsic parities of antiparticles are the same as the corresponding particles, while for fermions, antiparticles have intrinsic parities which are *opposite* to the corresponding particle.

For multiparticle states, the form of the wavefunction describing the relative motion of the particles also affects the behavior of the state under a parity transformation. If two particles A and B (viewed in their mutual center-of-momentum frame) have orbital angular momentum L , then an additional factor of $(-1)^L$ appears in the result of a parity transformation,¹⁴

$$P |\Psi_L^{A+B}\rangle = \eta_P^A \eta_P^B (-1)^L |\Psi_L^{A+B}\rangle, \quad (8.9.2)$$

where η_P^A and η_P^B are the intrinsic parities of the individual particles. Note, as claimed earlier, that the resulting total response to the parity transformation is the *product* of the individual bits.

Intrinsic parities can be assigned to particles in such a way that parity is a symmetry of the strong and electromagnetic interactions. In particular, the light mesons in Tables 5.7 and 5.8 are all parity-odd (*i.e.*, they have negative intrinsic parity).¹⁵ The photon is also parity-odd. The baryons listed in Tables 5.9 and 5.10 are all parity-even corresponding to the *conventional* choice that the lowest mass baryons and the quarks have positive intrinsic parity, while the lowest mass antibaryons and the antiquarks have negative intrinsic parity. The excited baryons, with nonzero internal orbital angular momentum, can have negative parity (*e.g.*, the $J^P = \frac{3}{2}^-$, isospin $\frac{1}{2}$ baryon with mass 1520 MeV/ c^2 or the $\Lambda(1690)$ mentioned above).

Physics became much more interesting when it was realized that parity is *not* respected by all interactions. In particular, parity is *not* a symmetry of the weak interactions, and the discovery that all interactions do not respect parity was a BIG deal historically (the 1957 Nobel Prize in Physics). This will be discussed further in the next chapter. Thus parity is an approximate symmetry, useful for understanding strong or electromagnetic processes, but is not a true symmetry of all nature.

¹³Recall that $\vec{L} = \vec{r} \times \vec{p}$. Since a parity transformation reverses both \vec{r} and \vec{p} , the (orbital) angular momentum \vec{L} does not change. The intrinsic spin transforms in the same fashion as \vec{L} . We label vectors like \vec{L} that do *not* change sign under parity as axial or pseudo-vectors (in contrast to the more familiar polar vectors that do change sign).

¹⁴This factor comes from the behavior of spherical harmonics, which describe states of definite orbital angular momentum, under the transformation $\vec{x} \rightarrow -\vec{x}$, namely $Y^{LM}(-\hat{x}) = (-1)^L Y^{LM}(\hat{x})$.

¹⁵All these mesons are *s*-wave quark-antiquark bound states, so they have no orbital angular momentum. Their negative parity reflects the *opposite* intrinsic parities of fermions and antifermions, here applied to quarks and antiquarks. Higher energy even-parity mesons do exist; these may be understood as bound states with non-zero orbital angular momentum.

8.10 Charge conjugation

A charge conjugation transformation, denoted C , has the effect of interchanging particles and antiparticles, $C|A\rangle = |\bar{A}\rangle$. So, for example, charge conjugation turns a proton into an antiproton, an electron into a positron, and a π^+ into a π^- . For particles which are their own antiparticles (“self-conjugate” particles), such as the photon and π^0 , there can also be an overall (eigenvalue) phase factor,

$$C|A\rangle = \eta_C |A\rangle \quad (\text{self-conjugate particles}). \quad (8.10.1)$$

Similarly to the parity transformation, applying charge conjugation twice takes us back to where we started so that $\eta_C^2 = 1$, $\eta_C = \pm 1$. These phases, which again depend on particle type, can be defined in such a way that charge conjugation is a symmetry of the strong and electromagnetic interactions (and match conventional choices for the phases - recall the minus sign in our definition of the antiquark doublet). However, charge conjugation is not an invariance of weak interactions. So charge conjugation is only an approximate symmetry, like parity, but is very useful when considering strong or electromagnetic processes.

Charge conjugation has no effect on momenta or spins of particles, but the electric charge and other (additive) flavor quantum numbers (B , L , S , I_3) all have their signs changed by the charge conjugation transformation. Hence, only particles which are neutral (and whose strangeness, I_3 , baryon, and lepton numbers all vanish) can be self-conjugate.

The photon is charge-conjugation odd (*i.e.*, its phase $\eta_C = -1$). To understand why, consider a classical electromagnetic field produced by some charge or current density. A charge conjugation transformation would change the electrically charged particles which are the source of the electromagnetic field into their oppositely charged antiparticles. In other words, the charge and current densities appearing in Maxwell’s equations would change sign. Since Maxwell’s equations are linear, this implies that the electromagnetic field itself would change sign. The photon is a quantized excitation in the electromagnetic field. Its behavior under charge conjugation reflects the behavior of a classical EM field: it changes sign.

Since a single photon is charge-conjugation odd, a multi-photon state containing N photons is charge-conjugation even if N is even, and charge-conjugation odd if N is odd. The neutral pion (dominantly) decays to two photons, while π^0 decay to three photons has not been observed. Neutral pion decay is an electromagnetic process, for which charge conjugation is a symmetry. Hence, the neutral pion is charge-conjugation *even*.

As an example of the utility of charge conjugation symmetry, consider positronium (as we did in the HW, but here using slightly different tools). This is the name given to bound states of an electron and a positron. Since an electron and positron have opposite electric charges, they have an attractive Coulomb interaction, and consequently form Coulombic bound states — just like the electron and proton in a hydrogen atom. Relative to hydrogen, there are two noteworthy differences. First, because the positron mass equals the electron mass (instead of being much much heavier like a proton), spacings between energy levels in positronium are half the corresponding spacings in hydrogen (the reduced mass is smaller by a factor of 2). More importantly, positronium is not stable. Unlike a hydrogen atom, the electron and positron in positronium can (and eventually will) annihilate into photons.

Consider positronium in its $1s$ ground state. How many photons will be produced when it decays? Answering this requires a consideration of symmetries, not hard calculations. Energy and momentum conservation forbid decay into a single photon since the photon has zero rest mass and positronium

does not. To understand whether decay into two photons is possible, we need to specify the initial state more carefully. Since the electron and positron each have spin $1/2$, the total spin of positronium can be either $S = 0$ or $S = 1$. Since a $1s$ state has no orbital angular momentum, the total angular momentum is the same as the spin. The singlet ($S = 0$) state of positronium is known as *para-positronium*, while the triplet ($S = 1$) state is called *ortho-positronium*. Recall that when two spins are combined to form $S = 1$, the spin wavefunction [either $\uparrow\uparrow$, $(\uparrow\downarrow + \downarrow\uparrow)$, or $\downarrow\downarrow$, depending on the value of S_3] is *symmetric* under interchange of the two spins. But the singlet spin wavefunction, $(\uparrow\downarrow - \downarrow\uparrow)$, is *antisymmetric* under interchange of spins.

Charge conjugation operating on positronium interchanges the types of the two fermions, without affecting their spins or positions. Interchanging the positions of the electron and positron flips their relative separation, $\vec{r} \rightarrow -\vec{r}$. Since a $1s$ state has a rotationally invariant spatial wavefunction, swapping the positions of the electron and positron does not change the value of the spatial wavefunction. In the $S = 1$ spin triplet (ortho-positronium), swapping the two spins also does not change the value of the spin wavefunction, since the spin wavefunction is symmetric. Hence, the action of charge conjugation on ortho-positronium is the same as completely interchanging the two particles (because the state is symmetric under interchange of positions and spins). Finally, similarly to our discussion of the intrinsic parities of fermions and antifermions, the anticommutation properties of the corresponding creation (and annihilation) *operators* introduces an extra factor of (-1) . Consequently, ortho-positronium must be charge-conjugation *odd*. In contrast (but by the same reasoning), in the $S = 0$ spin singlet (para-positronium), the action of charge conjugation differs from that of a complete interchange of the two fermions by an extra minus sign coming from the antisymmetry of the spin wavefunction. Therefore, para-positronium is charge-conjugation *even*.

More generally a fermion-antifermion pair with definite values of orbital angular momentum (L) and total spin (S) has the following property under C ,

$$C|f\bar{f}\text{pair}, L, S\rangle = (-1)^L(-1)^{S+1}(-1)|f\bar{f}\text{pair}, L, S\rangle = (-1)^{L+S}|f\bar{f}\text{pair}, L, S\rangle, \quad (8.10.2)$$

where the middle expression explicitly displays the corresponding behavior of the spatial wavefunction, the spin wavefunction and the intrinsic phase. Thus, as already noted, the $1s$ ortho-positronium state ($L = 0, S = 1$) is C odd, while para-positronium ($L = 0, S = 0$) is C even.

We noted above that a multi-photon state is charge conjugation even or odd depending on whether the number of photons is even or odd. Hence, charge conjugation invariance (of electromagnetic interactions) implies that para-positronium must decay to an *even* number of photons, while ortho-positronium must decay to an *odd* number of photons. Every additional photon in the final state decreases the rate of decay (by at least one factor of the fine structure constant α). Therefore, singlet positronium should decay to two photons, while triplet positronium should decay, more slowly, to three photons. This is precisely what is observed. The lifetime of spin singlet para-positronium is $125 \text{ ps} = 125 \times 10^{-12} \text{ s}$, while the lifetime of spin triplet ortho-positronium has the much larger value $142 \text{ ns} = 142 \times 10^{-9} \text{ s}$. Recall that we found this same distinction between the decays of the positronium states in our analysis in the HW, although there we made use of detailed considerations of angular momentum conservation.

Similarly applying Eq. 8.10.2 to the neutral (quark-antiquark) π^0 ($L = S = 0$) tells us that the π^0 is an eigenstate of C with eigenvalue $+1$, consistent with the fact that it decays into 2 photons (via the C conserving electromagnetic interactions).

Finally we return briefly to the topic of G -parity introduced in the previous chapter. This transfor-

mation is the product of both C and an isospin rotation of π radians about the 2-axis,

$$G = Ce^{i\pi I_2}. \quad (8.10.3)$$

Since the strong interactions respect both C and I , G is a symmetry of the strong interactions and is helpful in organizing multi-pion states. The analog of Eq. 8.10.2 is

$$G|f\bar{f}\text{pair}, L, S, I\rangle = (-1)^{L+S+I}|f\bar{f}\text{pair}, L, S, I\rangle, \quad (8.10.4)$$

i.e., the isospin rotation (by π) generates an extra factor $(-1)^I$ as the analog of the more familiar $(-1)^L$ from an ordinary rotation. For example, the pion, $L = 0$, $S = 0$, $I = 1$, has *odd* $((-1)^1)$ G-parity, as does the ω , $L = 0$, $S = 1$, $I = 0$. Hence the strong decays of the ω must involve odd numbers of pions (3 in this case). On the other hand, the ρ , $L = 0$, $S = 1$, $I = 1$, has *even* $((-1)^2)$ G-parity and decays into 2 pions.

8.11 CP

As we have discussed the charged current part of the weak interactions (*i.e.*, the exchange of the charged W 's) explicitly and separately violate C and P , coupling only to left-handed particles and right-handed antiparticles. On the other hand this very structure is connected by the (simultaneous) operation of CP , which turns a left-handed particle into a right-handed antiparticle. Until the 1960's it was believed that the weak interactions conserve the eigenvalue of the combined operator CP . As already mentioned the interesting structure seen in the decays of the $|K^0\rangle - |\bar{K}^0\rangle$ system explicitly confirms the importance of CP conservation. Combining the results of the previous two Sections tells us that for a fermion - anti-fermion system, like the π^0 , the CP eigenvalue is defined by

$$CP|f\bar{f}\text{pair}, L, S\rangle = (-1)^L(-1)^{S+1}(-1) * (-1)(-1)^L|f\bar{f}\text{pair}, L, S\rangle = (-1)^{S+1}|f\bar{f}\text{pair}, L, S\rangle. \quad (8.11.1)$$

Thus the π^0 is CP odd and the decay of a state of definite CP into π^0 's will be determined by the initial CP , even $CP \rightarrow$ an even number of π^0 's (typically 2) while odd $CP \rightarrow$ an odd number π^0 's (typically 3). Further, in this argument, we can replace a pair of π^0 's by a $\pi^+\pi^-$ pair with $L = 0$, which dominates at low energies. As we consider in some detail in the HW, this means that the CP even combination of $|K^0\rangle$ and $|\bar{K}^0\rangle$ can decay (rather quickly) to 2 pions, while the CP odd state decays (more slowly) to 3 pions. So these definite CP states are labeled K_{short} , or K_S , and K_{long} , or K_L . Note also that the mixing of $|K^0\rangle$ and $|\bar{K}^0\rangle$ can happen via the weak interactions because the weak interactions do not conserve the strange quark number. Finally note that with our conventions $C|K^0\rangle = (+1)|\bar{K}^0\rangle$ (although this is not uniformly true in the literature) and thus, similarly to the neutral pion, $CP|K^0\rangle = (-1)|\bar{K}^0\rangle$. This means the CP eigenstates are proportional to $|K^0\rangle \mp |\bar{K}^0\rangle$ with the minus sign corresponding to *even* CP .

ASIDE: A relevant way to think about the K^0 system is in terms of 2 sets of basis vectors. The kaons are typically *produced* via the strong interactions where the strangeness conserving basis is the appropriate one, while the decay process via the weak interaction is more simply viewed in the CP eigenstate basis,

$$\text{Strangeness: } \begin{pmatrix} |K^0\rangle \\ |\bar{K}^0\rangle \end{pmatrix} \quad \text{vs} \quad CP: \begin{pmatrix} |K_S\rangle \\ |K_L\rangle \end{pmatrix} = \frac{1}{\sqrt{2}} \begin{pmatrix} |K^0\rangle - |\bar{K}^0\rangle \\ |K^0\rangle + |\bar{K}^0\rangle \end{pmatrix}. \quad (8.11.2)$$

These two choices of basis are clearly related by a simple rotation.

Finally we must note that in the 1960's it was learned that the weak interactions, in fact, do *not* precisely conserve CP , but rather exhibit a tiny (1 part in a thousand) violation. The experimental observation of CP violation in the neutral kaon system led to the 1980 Nobel Prize in Physics. This is an extremely interesting story, now being repeated at the LHC in the bottom quark sector, but we do not have time to discuss it more fully here.

8.12 Time reversal and CPT

Time reversal, denoted T , is a transformation which has the effect of flipping the sign of time, $t \rightarrow -t$.¹⁶ So time reversal interchanges the past and the future. If some state $|\Psi_1\rangle$ evolves into state $|\Psi_2\rangle$ after a time interval Δt , then the time-reversed final state $T|\Psi_2\rangle$ will evolve into the time-reversed initial state $T|\Psi_1\rangle$ (after the same time interval Δt) — if time reversal is a symmetry of the dynamics.

As with C and P , time reversal is a symmetry of strong and electromagnetic interactions, but not of weak interactions. However, the product of charge conjugation, parity, and time reversal, or CPT , is a symmetry of all known interactions. In fact, one can prove that any Lorentz invariant theory (which satisfies causality) must be CPT invariant. This is one of the deepest results which follows from combining special relativity and Lorentz invariance, and essentially follows from analytic continuation applied to Lorentz transformations.¹⁷ Note finally that CPT invariance implies that the violation of time reversal invariance is intimately related to CP violation. Considerable effort, including locally at the UW, is now going into directly detecting T violating observables.¹⁸

8.13 Sample calculations

Let us illustrate the use the technology of the ladder operators to determine the various 3-quark spin wavefunctions for the baryons we discussed in Chapter 7. We start by combining the notation of the raising/lowering operators of Chapter 5 with the arrow notation (for the individual quarks) of Chapter 7 and applying them to the *symmetric* spin $\frac{3}{2}$ state,

$$|\frac{3}{2}, \frac{3}{2}\rangle = |\uparrow\uparrow\uparrow\rangle. \quad (8.13.1)$$

For a single quark state operated on by the lowering ladder operator we have

$$J_{-,1}|\frac{1}{2}, \frac{1}{2}\rangle = \hbar\sqrt{\left(\frac{1}{2} + \frac{1}{2}\right)\left(\frac{1}{2} - \frac{1}{2} + 1\right)}|\frac{1}{2}, -\frac{1}{2}\rangle = \hbar|\frac{1}{2}, -\frac{1}{2}\rangle, \quad (8.13.2)$$

¹⁶Because this transformation changes the meaning of time, it is not represented by a unitary operator which commutes with the Hamiltonian. In fact, unlike all other symmetries discussed so far, time reversal, in quantum mechanics, is not represented by a linear operator, but rather by an “anti-linear” operator. Such operators do not satisfy the defining relation of linear operators, $\mathcal{O}(c_1|\Psi_1\rangle + c_2|\Psi_2\rangle) = c_1(\mathcal{O}|\Psi_1\rangle) + c_2(\mathcal{O}|\Psi_2\rangle)$. Instead, for anti-linear operators, $\mathcal{O}(c_1|\Psi_1\rangle + c_2|\Psi_2\rangle) = c_1^*(\mathcal{O}|\Psi_1\rangle) + c_2^*(\mathcal{O}|\Psi_2\rangle)$.

¹⁷The Wikipedia entry on CPT symmetry has a nice sketch of the proof of the CPT theorem, together with a summary of its history.

¹⁸Results from the *BaBar* experiment reported in autumn 2012, *Phys. Rev. Lett.* **109**, 211801 (2012), and describing the $B^0\bar{B}^0$ system exhibit the expected correlation between CP and T violation with very good precision.

or

$$J_{-,1}|\uparrow\rangle = \hbar|\downarrow\rangle. \quad (8.13.3)$$

To apply to the 3-quark state we just need to note that, although for the total angular momentum we need to understand how to add *representations*, for the components it is just the ordinary (linear, algebraic) sum,

$$J_{3,\text{tot}} = J_{3,1} + J_{3,2} + J_{3,3} \quad \text{and} \quad J_{\pm,\text{tot}} = J_{\pm,1} + J_{\pm,2} + J_{\pm,3}. \quad (8.13.4)$$

The overall normalization is given by Eq. (5.1.33) via

$$J_{-, \text{tot}}|\frac{3}{2}, \frac{3}{2}\rangle = \hbar\sqrt{\left(\frac{3}{2} + \frac{3}{2}\right)\left(\frac{3}{2} - \frac{3}{2} + 1\right)}|\frac{3}{2}, \frac{1}{2}\rangle = \hbar\sqrt{3}|\frac{3}{2}, \frac{1}{2}\rangle. \quad (8.13.5)$$

Applying the total lowering ladder to the explicit spin state we have

$$J_{-, \text{tot}}|\uparrow\uparrow\uparrow\rangle = J_{-,1}|\uparrow\uparrow\uparrow\rangle + J_{-,2}|\uparrow\uparrow\uparrow\rangle + J_{-,3}|\uparrow\uparrow\uparrow\rangle = \hbar(|\downarrow\uparrow\uparrow\rangle + |\uparrow\downarrow\uparrow\rangle + |\uparrow\uparrow\downarrow\rangle). \quad (8.13.6)$$

Combining we have the desired result for the normalized spin wavefunction (note that the \hbar factors cancel)

$$|\frac{3}{2}, \frac{1}{2}\rangle = \frac{1}{\sqrt{3}}(|\downarrow\uparrow\uparrow\rangle + |\uparrow\downarrow\uparrow\rangle + |\uparrow\uparrow\downarrow\rangle). \quad (8.13.7)$$

Repeating this process while being explicit about the terms generated and the normalization, we have

$$J_{-, \text{tot}}|\frac{3}{2}, \frac{1}{2}\rangle = \hbar\sqrt{\left(\frac{3}{2} + \frac{1}{2}\right)\left(\frac{3}{2} - \frac{1}{2} + 1\right)}|\frac{3}{2}, -\frac{1}{2}\rangle = 2\hbar|\frac{3}{2}, -\frac{1}{2}\rangle \quad (8.13.8)$$

and (recall that the lowering operator yields zero when applied to a spin-down state)

$$\begin{aligned} J_{-, \text{tot}}\frac{1}{\sqrt{3}}(|\downarrow\uparrow\uparrow\rangle + |\uparrow\downarrow\uparrow\rangle + |\uparrow\uparrow\downarrow\rangle) \\ = \frac{\hbar}{\sqrt{3}}(0 + |\downarrow\downarrow\uparrow\rangle + |\downarrow\uparrow\downarrow\rangle + |\downarrow\downarrow\uparrow\rangle + 0 + |\uparrow\downarrow\downarrow\rangle + |\downarrow\uparrow\downarrow\rangle + |\uparrow\downarrow\downarrow\rangle + 0) \\ = \frac{2\hbar}{\sqrt{3}}(0 + |\downarrow\downarrow\uparrow\rangle + |\downarrow\uparrow\downarrow\rangle + |\uparrow\downarrow\downarrow\rangle), \end{aligned} \quad (8.13.9)$$

which combine to give the expected result (*i.e.*, the factors of \hbar and 2 cancel),

$$|\frac{3}{2}, -\frac{1}{2}\rangle = \frac{1}{\sqrt{3}}(|\downarrow\downarrow\uparrow\rangle + |\downarrow\uparrow\downarrow\rangle + |\uparrow\downarrow\downarrow\rangle). \quad (8.13.10)$$

Finally the last step is given by

$$J_{-, \text{tot}}|\frac{3}{2}, -\frac{1}{2}\rangle = \hbar\sqrt{\left(\frac{3}{2} - \frac{1}{2}\right)\left(\frac{3}{2} + \frac{1}{2} + 1\right)}|\frac{3}{2}, -\frac{3}{2}\rangle = \sqrt{3}\hbar|\frac{3}{2}, -\frac{3}{2}\rangle \quad (8.13.11)$$

and

$$\begin{aligned} J_{-, \text{tot}}\frac{1}{\sqrt{3}}(|\downarrow\downarrow\uparrow\rangle + |\downarrow\uparrow\downarrow\rangle + |\uparrow\downarrow\downarrow\rangle) \\ = \frac{\hbar}{\sqrt{3}}(0 + 0 + |\downarrow\downarrow\downarrow\rangle + 0 + |\downarrow\downarrow\downarrow\rangle + 0 + |\downarrow\downarrow\downarrow\rangle + 0 + 0) \\ = \sqrt{3}\hbar|\downarrow\downarrow\downarrow\rangle, \end{aligned} \quad (8.13.12)$$

which combine to give the expected result (*i.e.*, now the factors of \hbar and $\sqrt{3}$ cancel),

$$|\frac{3}{2}, -\frac{3}{2}\rangle = |\downarrow\downarrow\downarrow\rangle. \quad (8.13.13)$$

Note that each of these states is symmetric under the interchange of any pair of quarks.

Recall from Eq. (7.4.2) that when we combine 3 spin $\frac{1}{2}$ quarks there are two resulting total spin $\frac{1}{2}$ representations, which we distinguished in Eq. (5.4.3) by whether the states were antisymmetric in quarks 1 and 2 (*i.e.*, quarks 1 and 2 have pair spin 0) or symmetric in quarks 1 and 2 (*i.e.*, quarks 1 and 2 have pair spin 1). (Note that the choice here to focus on quarks 1 and 2 is another *arbitrary* labeling choice, and eventually in Chapter 7 there was a sum over all choices in the overall wavefunction.) Further we know that both possibilities must be orthogonal to the corresponding spin $\frac{3}{2}$ wave function in Eq. (8.13.7). In the last Chapter we choose the former case (antisymmetric in 1 and 2, recall we made this *choice* to match the mixed symmetry of the flavor wave functions), which has the normalized wave function

$$|\frac{1}{2}, \frac{1}{2}\rangle_{A12,3} = \frac{1}{\sqrt{2}} (|\uparrow\downarrow\uparrow\rangle - |\downarrow\uparrow\uparrow\rangle), \quad (8.13.14)$$

which is clearly orthogonal to the wavefunction in Eq. (8.13.7). The spin down version, which can be obtained easily with the lowering operator (and noting the cancelations), is

$$|\frac{1}{2}, -\frac{1}{2}\rangle_{A12,3} = \frac{1}{\sqrt{2}} (|\uparrow\downarrow\downarrow\rangle - |\downarrow\uparrow\downarrow\rangle), \quad (8.13.15)$$

which is orthogonal to the wavefunction in Eq. (8.13.10). Finding the other total spin $\frac{1}{2}$ wavefunction takes a bit more thought, but it must be orthogonal to the two we have already defined (and be symmetric in quarks 1 and 2). Except for the question of the overall sign it must look like

$$|\frac{1}{2}, \frac{1}{2}\rangle_{S12,3} = \frac{1}{\sqrt{6}} (|\uparrow\downarrow\uparrow\rangle + |\downarrow\uparrow\uparrow\rangle - 2|\uparrow\uparrow\downarrow\rangle), \quad (8.13.16)$$

which has the correct symmetry properties and is orthogonal to the previous versions. Applying the lowering operator we quickly obtain

$$|\frac{1}{2}, -\frac{1}{2}\rangle_{S12,3} = \frac{1}{\sqrt{6}} (2|\downarrow\downarrow\uparrow\rangle - |\downarrow\uparrow\downarrow\rangle - |\uparrow\downarrow\downarrow\rangle). \quad (8.13.17)$$

Chapter 9

Weak interactions

As already discussed, weak interactions are responsible for many processes which involve the transformation of particles from one type to another. Weak interactions cause nuclear beta decay, as well as the decays of muons, charged pions, kaons, and many other hadrons. All processes which involve production or scattering of neutrinos, the conversion of quarks from one flavor to another, or the conversion of leptons from one type to another, involve weak interactions.

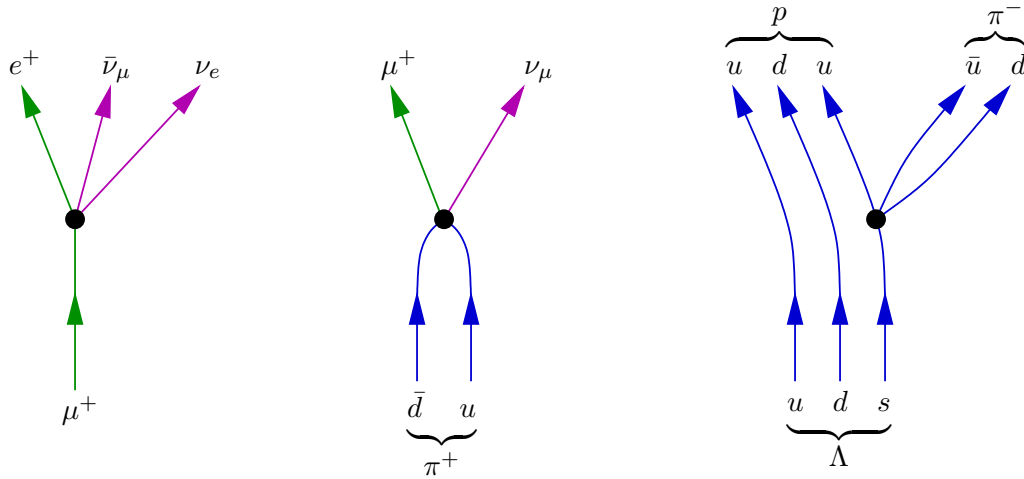


Figure 9.1: Depictions, at the level of quarks and leptons, of the weak decays $\mu^+ \rightarrow e^+ + \bar{\nu}_\mu + \nu_e$, $\pi^+ \rightarrow \mu^+ + \nu_\mu$, and $\Lambda \rightarrow p + \pi^-$.

Figures 9.1 and 9.2 depict, at the level of quarks and leptons, some of these weak interaction processes. As these figures illustrate, every weak interaction vertex (the black dot in the figures) involves four fermions (where here we are counting fermions and antifermions equally), either one fermion turning into three (as in muon decay) or two incoming fermions scattering and producing two outgoing fermions (as in neutrino scattering in Fig. 9.2). As the Λ baryon decay in Fig. 9.1 illustrates, there can also be *spectator* quarks which are constituents of the hadrons involved but not direct participants in the weak interaction process.

The complete Hamiltonian which describes particle interactions can be written as a sum of contributions from strong, electromagnetic, and weak interactions,

$$H = H_{\text{strong}} + H_{\text{EM}} + H_{\text{weak}} . \quad (9.0.1)$$

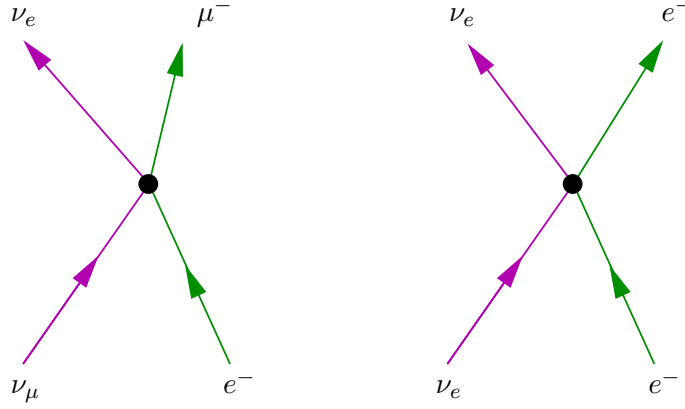


Figure 9.2: Left: inelastic neutrino scattering, $\nu_\mu + e^- \rightarrow \nu_e + \mu^-$. Right: elastic neutrino scattering, $\nu_e + e^- \rightarrow \nu_e + e^-$.

Because weak interactions are truly weaker than strong or electromagnetic interactions, it is useful to think of H_{weak} as a small perturbation to the dynamics generated by strong and electromagnetic interactions.

9.1 Muon decay

Consider (anti)muon decay, $\mu^+ \rightarrow e^+ + \bar{\nu}_\mu + \nu_e$, in Fig. 9.1. Let the ket $|\mu^+(\vec{p}=0)\rangle$ denote an initial state containing a single μ^+ at rest. Let the bra $\langle e^+(\vec{p}_e) \bar{\nu}_\mu(\vec{p}_{\bar{\nu}}) \nu_e(\vec{p}_\nu) |$ denote a final state describing a positron with spatial momentum \vec{p}_e , a muon antineutrino with momentum $\vec{p}_{\bar{\nu}}$, and an electron neutrino with momentum \vec{p}_ν . The existence of muon decay means that the time evolution of the initial state $|\mu^+(\vec{p}=0)\rangle$ will have a non-zero projection onto the final state $\langle e^+(\vec{p}_e) \bar{\nu}_\mu(\vec{p}_{\bar{\nu}}) \nu_e(\vec{p}_\nu) |$. This can only happen if the Hamiltonian, which generates time evolution, has a non-zero matrix element connecting these states. And this can only be due to the weak interaction part of the Hamiltonian. In other words, the existence of muon decay implies that the amplitude

$$M \equiv \langle e^+(\vec{p}_e) \bar{\nu}_\mu(\vec{p}_{\bar{\nu}}) \nu_e(\vec{p}_\nu) | H_{\text{weak}} | \mu^+(\vec{p}=0) \rangle, \quad (9.1.1)$$

is non-zero. The rate of decay must be proportional to the *square* of this amplitude. Because there are many different final states corresponding to different values of the final momenta p_e , $p_{\bar{\nu}}$ and p_ν , the complete decay rate Γ will involve a sum over all possible final states. Schematically,

$$\Gamma \sim \sum_{\text{final states}} |M|^2. \quad (9.1.2)$$

The amplitude M , when properly defined (see footnote below), will include momentum conservation, *i.e.*, the constraint that $\vec{p}_e + \vec{p}_{\bar{\nu}} + \vec{p}_\nu = 0$. When momentum is conserved, $\vec{p}_{\bar{\nu}}$ will equal $-(\vec{p}_\nu + \vec{p}_e)$, so M may be regarded as function of two independent momenta, p_e and p_ν . This amplitude can, in principle, depend in some complicated fashion on these two final momenta. But the simplest possibility is for the amplitude to have minimal dependence on the outgoing momenta. Physically, this corresponds to a point-like interaction, for which the spatial variation of wavefunctions (due to their momentum) plays no role.

This approximation (that the weak interaction occurs essentially at a point) turns out to work remarkably well, and we will shortly discuss how we can understand this perhaps surprising result. If the amplitude M is momentum independent then, with just a little calculation, one can perform the sum over final states in Eq. (9.1.2) (strictly speaking, the 3-momenta are *continuous* variables and the sum is really an integral) and predict the muon decay spectrum as a function of either the positron energy or momentum. The spectrum is the fraction of decays in which the positron has energy between E and $E + dE$, or momentum between p and $p + dp$ with $E = \sqrt{p^2 + m_e^2}$. Figure 9.3 shows the comparison between experimental data for the decay spectrum versus momentum and the result of this calculation. The agreement is excellent.

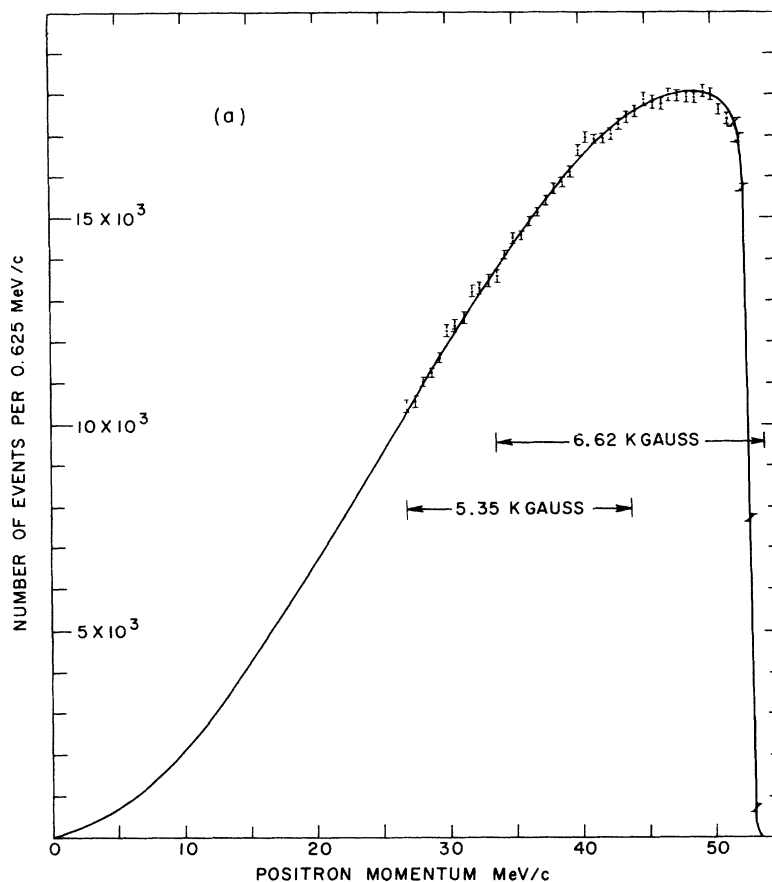


Figure 9.3: Energy spectrum of positrons emitted from decays of positively charged muons. The solid curve is the theoretical prediction; data points are shown with error bars. [From M. Bardon *et al.*, Phys. Rev. Lett. 14, 449 (1965)].

To characterize the value of the amplitude M , it will be useful to begin with some dimensional analysis. To make this as easy as possible, it will be convenient to use “natural units” in which $\hbar = c = 1$ (recall the discussion in section 3.7). Since c has ordinary dimensions of [length/time], setting $c = 1$ means that we are regarding length and time as having the same dimensions. Since \hbar has dimensions of [energy \times time], setting $\hbar = 1$ means that we are regarding energy and frequency (or inverse time) as having the same dimensions. Setting both \hbar and c to unity means that we are treating length and inverse energy as dimensionally equivalent. After using natural units in any

calculation, one can always reinsert factors of \hbar and c as needed to restore conventional dimensions. In particular, the value $\hbar c \simeq 197 \text{ MeV fm}$ may be regarded as a conversion factor which allows one to convert lengths measured in femtometers into lengths measured in MeV^{-1} , $1 \text{ fm} = \frac{1}{197} \text{ MeV}^{-1}$.

The Hamiltonian is the operator which measures energy. Its eigenvalues are the energies of stationary states. Therefore, the Hamiltonian must have dimensions of energy. If $|\Psi\rangle$ is any physical, normalized state, then the matrix element $\langle\Psi|H|\Psi\rangle$ is the expectation value of the energy in state $|\Psi\rangle$. Hence, matrix elements of the Hamiltonian, such as the muon decay amplitude M in Eq. (9.1.1), also have dimensions of energy, provided the states appearing in the matrix element are appropriately normalized as we will now discuss.

The wavefunction describing a particle with definite momentum \vec{p} is proportional to the plane wave $e^{i\vec{p}\cdot\vec{x}/\hbar}$. To normalize such a state, it is convenient to imagine that space is not infinite, but rather is limited to some finite, but arbitrarily large region \mathcal{V} . The condition that a state is normalized then becomes

$$1 = \int_{\mathcal{V}} d^3x |\Psi(\vec{x})|^2, \quad (9.1.3)$$

where the integral only includes the interior of the region \mathcal{V} . For simplicity, suppose that this region is a cube whose edges have length L (and hence volume L^3). A normalized state describing a particle with momentum \vec{p} will thus have a wavefunction

$$\Psi(\vec{x}) = e^{i\vec{p}\cdot\vec{x}/\hbar} / L^{3/2}. \quad (9.1.4)$$

The *absolute square* of this wavefunction gives a constant probability density of $1/L^3$ whose volume integral over the region \mathcal{V} equals one, as desired.

Now consider the muon decay amplitude M . The initial muon, with zero spatial momentum, will have a *constant* wavefunction, $\psi_\mu(\vec{x}) = 1/L^{3/2}$. The final positron, with momentum \vec{p}_e , will have a plane-wave wavefunction $\psi_e(\vec{x}) = e^{i\vec{p}_e\cdot\vec{x}/\hbar} / L^{3/2}$, and similarly the final neutrino and antineutrino will have wavefunctions $\psi_{\nu_e}(\vec{x}) = e^{i\vec{p}_\nu\cdot\vec{x}/\hbar} / L^{3/2}$ and $\psi_{\bar{\nu}_e}(\vec{x}) = e^{i\vec{p}_{\bar{\nu}}\cdot\vec{x}/\hbar} / L^{3/2}$, respectively.

Since the point-like weak interaction event can occur at any point in space, the complete amplitude will involve an integral over space with an integrand that is the product of the amplitude $\psi_\mu(\vec{x})$ to find the muon at some point \vec{x} , times the product of conjugate wavefunctions $\psi_e(\vec{x})^* \psi_{\nu_e}(\vec{x})^* \psi_{\bar{\nu}_e}(\vec{x})^*$, giving the amplitudes for the created positron, neutrino, and antineutrino all to be at (the same) point \vec{x} , all multiplied by some overall constant which will control the rate of this process,

$$M = \left[\int_{\mathcal{V}} d^3x \psi_e(\vec{x})^* \psi_{\nu_e}(\vec{x})^* \psi_{\bar{\nu}_e}(\vec{x})^* \psi_\mu(\vec{x}) \right] \times (\text{const.}). \quad (9.1.5)$$

For the weak interaction the overall constant is known as the *Fermi constant*, G_F , divided by $\sqrt{2}$. (Including this factor of $\sqrt{2}$ is merely a convention, but is required so that G_F matches its historical definition.) The integrand appearing in this matrix element is just a constant,

$$\psi_e(\vec{x})^* \psi_{\nu_e}(\vec{x})^* \psi_{\bar{\nu}_e}(\vec{x})^* \psi_\mu(\vec{x}) = \frac{e^{-i(\vec{p}_e + \vec{p}_\nu + \vec{p}_{\bar{\nu}})\cdot\vec{x}/\hbar}}{(L^{3/2})^4} = L^{-6}, \quad (9.1.6)$$

provided the momenta satisfy conservation of 3-momentum, $\vec{p}_e + \vec{p}_\nu + \vec{p}_{\bar{\nu}} = 0$.¹ Integrating over the

¹The mathematically astute will recognize that the integral over all (3 dimensional) space of an expression like the one in Eq. (9.1.6) produces precisely the Dirac delta function that ensures 3-momentum conservation, *i.e.*, translational invariance means integrating over all space which yields the momentum conserving delta function.

region \mathcal{V} thus simply yields a factor of the volume, L^3 , of this region. Hence, we find

$$M = \frac{G_F/\sqrt{2}}{L^3}. \quad (9.1.7)$$

We noted above that the decay amplitude M must have dimensions of energy. Since $1/L^3$ has dimensions of energy cubed (having set $\hbar = c = 1$), we learn that the Fermi constant G_F must have dimensions of $1/(\text{energy})^2$.

The value of the Fermi constant G_F may be fixed by demanding that the muon decay rate Γ calculated from Eq. (9.1.2) agrees with the experimentally determined value. The decay rate is just the inverse of the lifetime, so $\Gamma = 1/\tau_\mu = 1/(2 \mu\text{s})$. Performing the sum over final states in Eq. (9.1.2) involves integrating over the final momenta subject to the constraints of energy and momentum conservation. Details of this calculation, which is straightforward, will be omitted. One finds that $\Gamma = G_F^2 m_\mu^5/(192\pi^3)$. Note that the resulting natural units from $G_F^2 m_\mu^5 = \text{energy}^{-4} \text{energy}^5 = \text{energy} = \text{time}^{-1}$, just what we expect for a *rate* of decay. Equating this with the inverse of the experimentally measured decay rate and solving for G_F yields

$$G_F = 1.16637(1) \times 10^{-5} \text{ GeV}^{-2} \simeq 1.2 \times 10^{-5} \text{ GeV}^{-2} = 12 \text{ TeV}^{-2}. \quad (9.1.8)$$

9.2 Neutrino scattering

The significance of the determination of the Fermi constant described above comes from the fact that a factor of G_F will appear in *every* weak interaction amplitude. Consider, for example, the *inelastic* neutrino *scattering* process,

$$\nu_\mu + e^- \rightarrow \nu_e + \mu^-, \quad (9.2.1)$$

depicted on the left in Fig. 9.2 in which the “flavor” of the charged lepton changes. With sufficient experimental skill and resources, this is a measurable process. The cross section for this scattering process equals the rate of scattering events divided by the incident flux of neutrinos and the density of target electrons. For a neutrino beam with constant flux, the scattering rate is just the probability of scattering in time Δt , divided by Δt . And the probability, as always in quantum mechanics, is the absolute square of a probability amplitude which involves a matrix element of the weak interaction Hamiltonian between the relevant incoming and outgoing states, $M = \langle \text{out} | H_{\text{weak}} | \text{in} \rangle$. This weak interaction amplitude must also be proportional to G_F , so that²

$$\sigma \propto |M|^2 \propto G_F^2. \quad (9.2.2)$$

Now do some more dimensional analysis. A cross section is an area, with dimensions of length squared or (in natural units) $[\text{energy}]^{-2}$. The Fermi constant G_F also has dimensions of $[\text{energy}]^{-2}$, but G_F appears *squared* in the cross section. Therefore the cross section must equal G_F^2 times “something else” with dimensions of $[\text{energy}]^2$. What can this “something else” depend on? One possibility, which is surely relevant, is the neutrino energy. But the energy of a particle is frame-dependent. Since the cross section is an invariant concept (effectively the “size” *transverse* to the direction of the

²In fact, analytic continuation in the four-momenta relates the amplitude for inelastic neutrino scattering, $\nu_\mu + e^- \rightarrow \nu_e + \mu^-$, to the amplitude for μ^+ decay. This relation, which involves replacing particles in the initial state by their antiparticles in the final state (or vice-versa) is known as *crossing symmetry*.

beam, which is not changed by boosts along the beam), we must be able to express the cross section in terms of Lorentz invariant quantities. A Lorentz invariant measure of the scattering energy is $s \equiv (p_{\nu_\mu} + p_e)^2 = (p_{\nu_e} + p_\mu)^2 = E_{\text{CM}}^2$. At low energies, the value of the cross section will also depend on the electron and muon masses. After all, if $E_{\text{CM}} < m_\mu c^2$, then the reaction $\nu_\mu + e^- \rightarrow \nu_e + \mu^-$ cannot possibly occur. So by “pure thought” we conclude that it must be possible to express the cross section in the (dimensionally consistent and invariant!) form

$$\sigma = G_F^2 s \times f\left(\frac{m_e}{\sqrt{s}}, \frac{m_\mu}{\sqrt{s}}\right), \quad (9.2.3)$$

where f is some *dimensionless* function of the *dimensionless* ratios m_e/E_{CM} and m_μ/E_{CM} . Further we expect (by energy-momentum conservation) that this function will be non-vanishing only when both arguments are less than one.

The simplest regime to consider is large energy relative to the muon mass, $E_{\text{CM}} \gg m_\mu c^2$. In this domain, the ratios m_e/E_{CM} and m_μ/E_{CM} are both tiny. Since the cross section can be expressed in the form (9.2.3), understanding the behavior of the cross section when the energy is large is the same problem as understanding the behavior of the cross section in a hypothetical world where the value of the electron and muon masses are arbitrarily small (compared to some reference mass scale).

A crucial observation is that there is no reason to expect anything dramatic, or singular, to happen in the limit of vanishingly small electron and muon mass (at fixed energy E_{CM}). In the relativistic relation between (total) energy and momentum, the zero mass limit is perfectly smooth, and just leads to the energy-momentum relation of a massless particle,³

$$E(\vec{p}) = \sqrt{\vec{p}^2 + m^2} = |\vec{p}| + \frac{1}{2} \frac{m^2}{|\vec{p}|} + \cdots \xrightarrow{m/E \rightarrow 0} |\vec{p}|. \quad (9.2.4)$$

Similarly, the massless limit of the function $f(\frac{m_e}{\sqrt{s}}, \frac{m_\mu}{\sqrt{s}})$ appearing in the cross section (9.2.3) should be expected to be finite and non-zero, so that $A \equiv f(0, 0)$ is just some pure number like 2 or π . A detailed calculation shows that, for the process (9.2.1), the number A is $1/\pi$. Therefore, the inelastic neutrino cross section is given by

$$\sigma_{\nu_\mu e^- \rightarrow \nu_e \mu^-} = \frac{G_F^2 E_{\text{CM}}^2}{\pi}, \quad (9.2.5)$$

when $E_{\text{CM}} \gg m_\mu c^2 (\gg m_e c^2)$. This quadratic rise of the cross section with center-of-mass energy (for energies above the relevant particle masses) also applies to other weak interaction scattering processes, including neutrino scattering with nucleons and elastic neutrino-electron scattering. In the latter example, the cross section is

$$\sigma_{\nu_e e^- \rightarrow \nu_e e^-} = 0.551 \frac{G_F^2 E_{\text{CM}}^2}{\pi}. \quad (9.2.6)$$

Recall, as we have already mentioned and will discuss in more detail in the next section, the weak interactions involve not only the exchange of the electrically charged W bosons (the so-called “charged current” weak interaction), as happens in the inelastic process of Eq. (9.2.5), but also the exchange of

³In contrast, the non-relativistic (kinetic) energy $E_{\text{NR}}(\vec{p}) = \vec{p}^2/(2m)$ is *not* well-behaved if $m \rightarrow 0$ for fixed momentum \vec{p} .

the electrically neutral Z boson (the so-called “neutral current” weak interaction), which contributes also to the elastic process in Eq. (9.2.6) (see the Feynman diagrams in Fig. 9.6). The differences in the two processes (one being only W exchange and the other both W and Z) leads to the different coefficient in Eq. (9.2.6) versus Eq. (9.2.5). The coefficient *decreases* by a factor of about 2 due to *destructive* interference between the W and Z exchanges (recall that quantum physics is all about interference).

For completeness we should also note the elastic cross section for the related (by crossing) process with antineutrinos,

$$\sigma_{\bar{\nu}_e e^- \rightarrow \bar{\nu}_e e^-} = 0.231 \frac{G_F^2 E_{\text{CM}}^2}{\pi}. \quad (9.2.7)$$

This cross section is smaller by approximately a factor of 3, which arises from the further fact (not present in our original simple point model) that the W (but not the Z) couples only to fermions of a definite spin projection. The W couples only to “left-handed” fermions (spin projection opposite the direction of motion) and “right-handed” antifermions (spin projection along the direction of motion). (Recall this is why the weak interactions violate both P , which transforms left-handed to right-handed, and C , which transforms particle to antiparticle.) In particular, this means that for the elastic neutrino-electron scattering process of Eq. (9.2.6) we have a left-handed neutrino scattering from a left-handed electron. Thus, viewed in the CM frame, the two spins are oppositely oriented and we have (total) $J_3 = 0$ for both the initial and final states. Thus angular momentum conservation imposes no constraints. In contrast, the anti-neutrino elastic scattering process of Eq. (9.2.7) involves a right-handed antineutrino scattering from a left-handed electron. Hence we have (total) $J_3 = 1$ along the direction of motion of the antineutrino in both the initial and final states. Hence “back”-scattering of the antineutrino (and the electron) is not allowed by angular momentum conservation, since it would require turning $J_3 = +1$ into $J_3 = -1$. In fact, compared to the neutrino process, the scattering cross section of the antineutrino process is reduced at all scattering angles except exactly forward ($\theta = 0$). Thus the “handedness” in the weak dynamics explains both why the weak interactions do not respect parity and the extra factor of about 1/3 in Eq. (9.2.7).

These predictions of neutrino cross sections increasing with increasing energy, which arise directly from our simple picture of the weak interactions as a point interaction, have been confirmed experimentally for energies in the multi-MeV to multi-GeV range.⁴ But the prediction of quadratically rising cross sections raises an immediate puzzle: can cross sections really grow with increasing energy *forever*? Or is there some point at which the behavior must change?

In fact, cross sections *cannot* become arbitrarily large. The number of scattering events in any scattering experiment is proportional to the cross section. But ultimately, the number of scatterings cannot be larger than the total number of projectiles! A quantum mechanical analysis shows that for point-like (or so-called s -wave) scattering, the cross-section must satisfy the bound

$$\sigma < \frac{\lambda^2}{4\pi} = \frac{\pi}{\vec{p}^2}, \quad (9.2.8)$$

where $\lambda = 2\pi\hbar/|\vec{p}|$ is the de Broglie wavelength of the projectile in the center-of-mass frame. This is referred to as a *unitarity* bound.

⁴See, for example, the plots of the (anti)neutrino-nucleon total cross section at the particle data group website (see Figure 48.1). Note that for neutrino scattering on a nucleus, the lab frame energy is proportional to the square of the center-of-mass energy, $E_{\text{lab}} \propto E_{\text{CM}}^2$, when E_{lab} is large compared to the target mass. So the quadratic rise of the cross section with E_{CM} is equivalent to linear growth as a function of E_{lab} , and a *constant* behavior for σ/E_{lab} , as plotted.

For an ultra-relativistic scattering, viewed in the center-of-mass frame, the energy of each particle is almost the same as the magnitude of its momentum (times c), and hence $E_{\text{CM}} \simeq 2|\vec{p}|$. Equating expression (9.2.5) for the neutrino cross section with the unitarity bound (9.2.8), one finds that the cross section (9.2.5) violates the unitarity bound when the center-of-mass energy exceeds

$$E^* \equiv \sqrt{\frac{2\pi}{G_F}} \approx 700 \text{ GeV}. \quad (9.2.9)$$

Therefore, at some energy below 700 GeV, something must dramatically change the behavior of weak interaction cross sections to stop their quadratic rise with increasing energy. In fact, before we reach this energy the weak interactions begin to exhibit the fact that they do not really correspond to a *point* interaction, but rather to the exchange of the aforementioned W 's and Z 's.

9.3 Weak gauge bosons

At energies somewhat below E^* , weak interaction cross sections become comparable to electromagnetic cross sections. At this point, one might anticipate significant changes in the behavior of both electromagnetic and weak interactions. This turns out to be true. Figure 9.4 shows the cross section for electron-positron annihilation into hadrons as a function of $\sqrt{s} = E_{\text{CM}}$. At energies below about 50 GeV, one sees that the cross section generally decreases with increasing energy (note the logarithmic scale). Since the electromagnetic coupling is dimensionless (unlike the weak interactions, there is no symmetric breaking for EM and no mass scale), the same dimensional analysis we applied to the weak interactions yields electromagnetic cross sections that behave like α^2/s (instead of $G_F^2 s$) and this is the general fall-off we see in the upper plot in Figure 9.4. The lower plot in Figure 9.4 shows the *ratio* between $e^+e^- \rightarrow \text{hadrons}$ and $e^+e^- \rightarrow \mu^+\mu^-$, which cancels out the $1/s$ behavior. In fact, the levels of the various “flat” sections in this plot are easily understood in terms of *sum* over the squares of the electric charges of the quarks to which the photon couples, with the scale of the charge e^2 canceled out by the charge (squared) of the muon in the denominator. Thus at low energies, where only u, d, s quarks are produced, the ratio is

$$\begin{aligned} R_{\text{low}} &= 3 \left[\left(\frac{2}{3} \right)^2 + \left(\frac{-1}{3} \right)^2 + \left(\frac{-1}{3} \right)^2 \right] \\ &= 3 \left[\frac{4}{9} + \frac{1}{9} + \frac{1}{9} \right] = 2. \end{aligned} \quad (9.3.1)$$

In this expression the preceding factor of 3 accounts for the fact that quarks come in 3 colors, which contribute equally to the coupling to photons but are distinct and do not interfere (*i.e.*, you square the amplitude for each color *first* and *then* sum over colors). As the energy increases, quarks with larger masses contribute and the value of R takes a step up at each threshold (to produce a new flavor pair). At the c quark threshold (where we also see the $c\bar{c}$ bound state resonances J/ψ and $\psi(2S)$) we add an additional $3 \times 4/9$ and R increases to $10/3$. At the b quark threshold (marked by the $b\bar{b}$ Υ resonance) R is increased by $3 \times 1/9$ to $11/3$. The fact that this simple picture of quarks with the specified (if peculiar) fractional electric charges and in 3 colors is in such good agreement with the data was an essential step in the general acceptance of the current Standard Model of particle physics.

The other dramatic feature of Figure 9.4 is the appearance of the various spin one, parity odd hadronic resonances — the broad ρ and ρ' , the narrower ω and ϕ , and the very narrow “spikes”

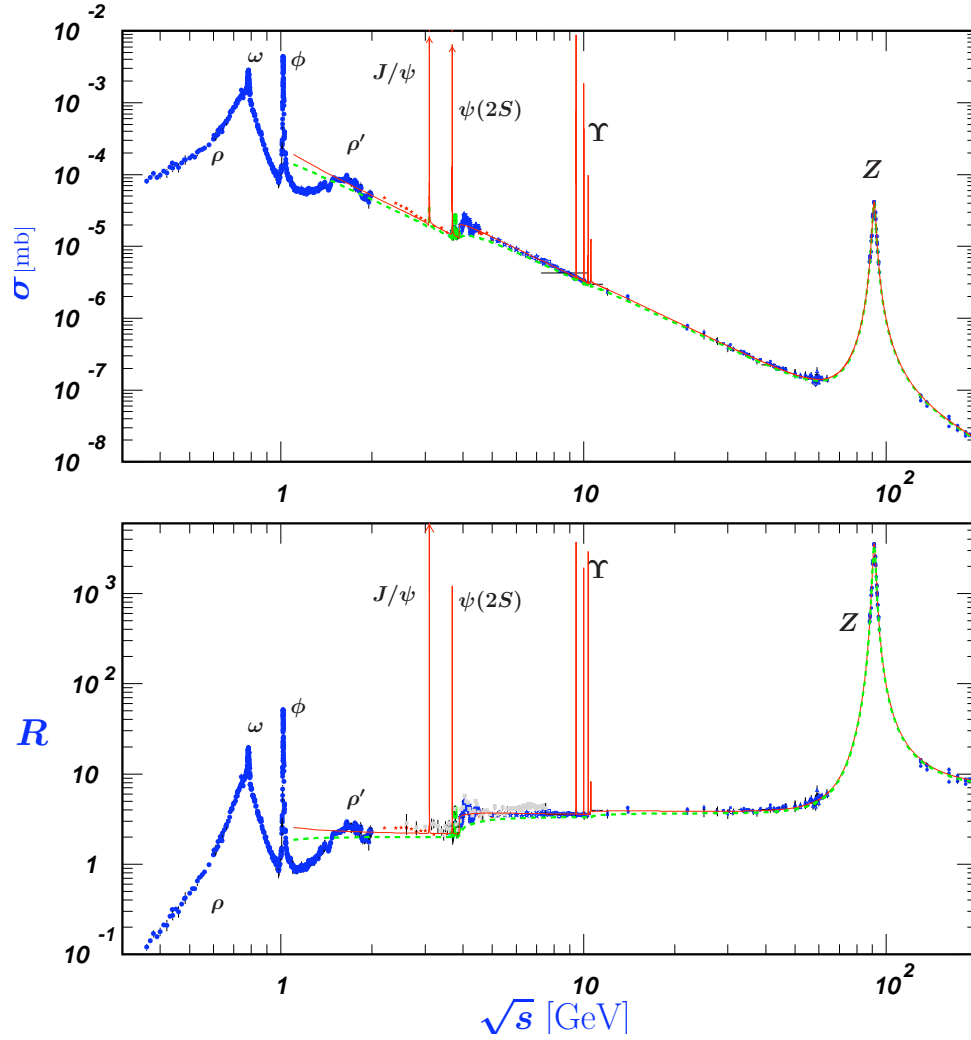


Figure 9.4: Top: Cross section for e^+e^- annihilation to hadrons as a function of $\sqrt{s} = E_{\text{c.m.}}$. Bottom: Ratio of cross sections for e^+e^- annihilation to hadrons versus annihilation to muon pairs, $R = \sigma_{e^+e^- \rightarrow \text{hadrons}} / \sigma_{e^+e^- \rightarrow \mu^+\mu^-}$. particle data group website (see page 5).]

associated with $c\bar{c}$ and $b\bar{b}$ heavy quark states (already mentioned above). The J/ψ and $\psi(2s)$ are $c\bar{c}$ bound states with energies close to twice the charm quark mass, while the upsilon (Υ) states near $2m_b$ are $b\bar{b}$ states. But then, at a much higher energy near 90 GeV, there is a very large resonance which is something new and relevant to our current discussion. This is not a quark-antiquark bound state, but rather a new type of particle which is called the Z boson. The same resonance appears in neutrino scattering. There is also a closely related pair of charged particles known as the W^+ and W^- . These are not seen in Figure 9.4 because a single W^+ or W^- cannot result from e^+e^- annihilation — this would violate charge conservation! However, the W bosons are present in those interactions where a quark or lepton changes type (*i.e.*, flavor) and they can be pair ($W^+ + W^-$) produced in $e^+ + e^-$ annihilation at energies above $2M_W c^2$, *i.e.*, off the above plot to the right..

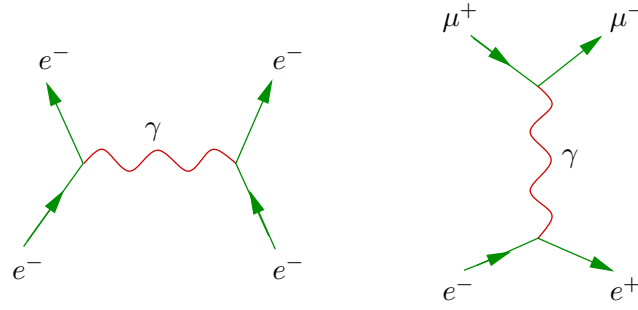


Figure 9.5: Feynman diagrams for Coulomb scattering: $e^-e^- \rightarrow e^-e^-$ (left), and electron-positron annihilation to muons: $e^+e^- \rightarrow \mu^+\mu^-$ (right).

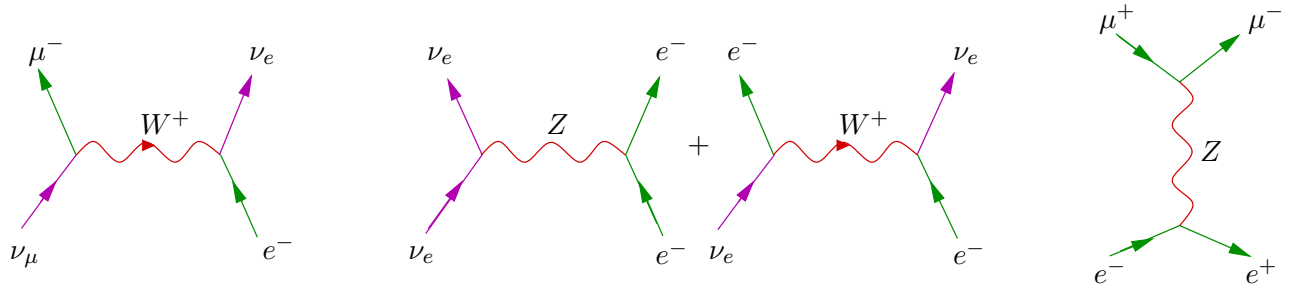


Figure 9.6: Feynman diagrams for inelastic neutrino scattering: $\nu_\mu + e^- \rightarrow \nu_e + \mu^-$ (left), elastic neutrino scattering: $\nu_e + e^- \rightarrow \nu_e + e^-$ (middle), and the weak interaction contribution to $e^+e^- \rightarrow \mu^+\mu^-$ (right).

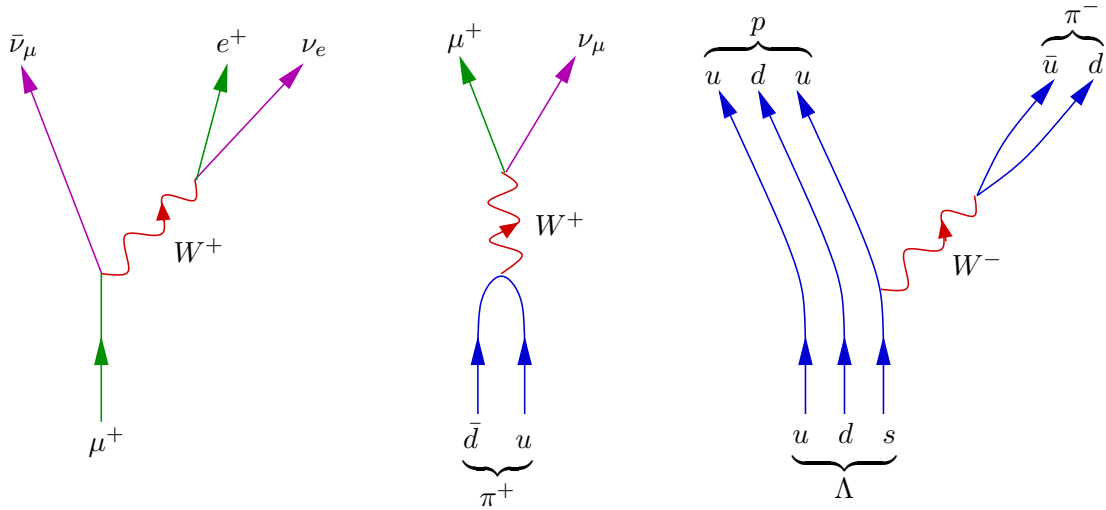


Figure 9.7: Depictions of the weak decays $\mu^+ \rightarrow e^+ + \bar{\nu}_\mu + \nu_e$ (left), $\pi^+ \rightarrow \mu^+ + \nu_\mu$ (middle), and $\Lambda \rightarrow p + \pi^-$ (right), showing the exchange of weak gauge bosons.

Together, the W^\pm and Z are known as *weak gauge bosons*.⁵ They are spin one particles with masses

$$m_W = 80.385 \pm 0.015 \text{ GeV}, \quad m_Z = 91.1876 \pm 0.0021 \text{ GeV}. \quad (9.3.2)$$

The simplest current picture is that these masses for the weak gauge bosons arise from the interaction between the weak gauge bosons and the (apparently now found) Higgs boson, which itself is assumed to have a nonzero “vacuum expectation value” or $\langle 0|h|0 \rangle \neq 0$. (The data from the LHC suggest pretty clearly that a Higgs boson with the expected properties and a mass of about $125 \text{ GeV}/c^2$ has been detected.) The weak gauge bosons mediate the weak interactions, in the same sense that the photon is responsible for mediating electromagnetic interactions. Coulomb interactions may be viewed as resulting from the exchange of photons between charged particles, and a process like $e^+e^- \rightarrow \mu^+\mu^-$ may be regarded as occurring via the annihilation of the electron and positron into a (virtual) photon, which lives only a very short time before converting into the final μ^+ and μ^- . The diagrams of Figure 9.5 depict these electromagnetic processes.

In the same fashion, weak interactions may be regarded as arising from the exchange of W and Z bosons. Figure 9.6 depicts the same weak interaction scattering processes illustrated in Figure 9.2, plus the weak interaction contribution to $e^+e^- \rightarrow \mu^+\mu^-$, explicitly showing the exchange of weak gauge bosons. Figure 9.7 does the same for the weak decays of Figure 9.1. These diagrams illustrate the fact that the (lowest order) weak interactions are not really “point” interactions, but rather localized to a small, but nonzero scale of order $1/M_W \sim 2 \times 10^{-3}$ fermi. For particles with de Broglie wavelengths short enough to probe this sort of distance (*i.e.*, $E \gg 100 \text{ GeV}$), the character of the weak interactions is moderated and the cross section stops increasing quadratically with the (CM) energy (and eventually decreases as $1/E_{\text{CM}}^2$). Note also that it is the difference between the left-hand diagram and the middle diagram in Figure 9.6 that explains the difference between the cross sections in Eqs. (9.2.6) and (9.2.5).

The diagrams of Figures 9.5–9.6 are examples of *Feynman diagrams* (named after the theorist Richard P. Feynman who introduced them). They actually do more than merely depict some process — these diagrams encode precise rules for how to calculate the quantum mechanical amplitude associated with each process. But developing this in detail (*e.g.*, in QFT) will have to be left for a subsequent class.

With this brief sketch of the current understanding of weak interactions, we must conclude our introduction to particles and symmetries. Hopefully it has whetted your appetite to learn more about this subject.

⁵The word “gauge” appears here because the underlying (broken) $SU(2)$ symmetry is of a type called a “gauge symmetry”.

Chapter 10

Introduction to Group Theory

Since symmetries described by groups play such an important role in modern physics, we will take a little time to introduce the basic structure (as seen by a physicist) of this most interesting field of study. We will first define groups in the abstract and then proceed to think about their representations, typically in the form of matrices. In physics we are typically interested in "real" operators that act on "real" states, which provide concrete representations of the more abstract concept of groups. The underlying arithmetic will be familiar from your previous experience with matrices.

10.1 Definitions

Group (G): A set of (perhaps abstract) elements (*i.e.*, things) - g_1, \dots, g_n , plus a definition of the multiplication operation, *i.e.*, a definition of the product or combination of two of the elements, such that

1. $g_i \bullet g_j = g_k \in G$ - products of elements are also elements of the group (a property of groups called closure),
2. the multiplication operation is associative - $(g_i \bullet g_j) \bullet g_k = g_i \bullet (g_j \bullet g_k)$,
3. the identity element exists as an element of the group, $\mathbf{1} \in G$, where $\mathbf{1} \bullet g_j = g_j \bullet \mathbf{1} = g_j$ (sometimes the left and right identities are distinct, but not generally in the context of physics),
4. the group includes a *unique* inverse for each element, $g_i \in G \Rightarrow g_i^{-1} \in G$ such that $g_i \bullet g_i^{-1} = g_i^{-1} \bullet g_i = \mathbf{1}$ (as with matrices the inverse is sometimes defined separately for left and right multiplication but this situation will not arise in this discussion).

Note: it is *not* necessary that the multiplication be commutative (and the interesting cases are those that are *not* commutative):

- if $g_i \bullet g_j = g_j \bullet g_i$ (commutative), the group is an Abelian group, Abelian group
- if $g_i \bullet g_j \neq g_j \bullet g_i$ (non-commutative), the group is a non-Abelian group. non-Abelian group

10.2 Finite Groups

If the number of elements n is finite ($n < \infty$), then the group is called a finite or discrete group of order n . Finite groups are commonly used in the study of solid state physics where discrete symmetries arise regularly. There is a trivial group corresponding to $n = 1$ with $g = 1$ only. Clearly the group properties are all satisfied but in a trivial way.

How about $n = 2$? Call the elements of the group $\mathbf{1}$ and \mathbf{P} . Evidently $\mathbf{P}^{-1} = \mathbf{P}$, $\mathbf{P} \bullet \mathbf{P} = \mathbf{1}$ in order to satisfy the condition of being a group. There are, in fact, two related and physically interesting realizations of this group. One case is the reflection group where \mathbf{P} is a reflection in a plane (*i.e.*, one of the 3 possible planes in 3-D space). For example, reflection in the xy plane means $\mathbf{P}f(x, y, z) = f(x, y, -z)$ so that $\mathbf{P} \bullet \mathbf{P}f(x, y, z) = \mathbf{P}f(x, y, -z) = f(x, y, z)$ as required. Another $n = 2$ group corresponds to reflection through the origin (in 3-D space), $\mathbf{P}(x, y, z) \rightarrow (-x, -y, -z)$. Again $\mathbf{P} \bullet \mathbf{P} = \mathbf{1}$, $\mathbf{P} \bullet \mathbf{P}(x, y, z) = \mathbf{P}(-x, -y, -z) = (x, y, z)$. This is the parity operation that (as we will see) plays an important role in the context of atomic, nuclear, and particle physics. Another (simpler) representation of the group of order 2 are the numbers $(1, -1)$ coupled with ordinary multiplication. The multiplication table for all of these order 2 groups is given in Table 10.1, where the notation is that a given entry is the result of multiplying the column label on the left by the row label (and the fact that \mathbf{P} acts like (-1) is made explicit).

	$\mathbf{1}$	$\mathbf{P}(-1)$
$\mathbf{1}$	$\mathbf{1}$	$\mathbf{P}(-1)$
$\mathbf{P}(-1)$	$\mathbf{P}(-1)$	$\mathbf{1}$

Table 10.1: Order 2 Group multiplication table

Since the structure of all of the order 2 groups (the multiplication table) is identical, we say that the groups are isomorphic (the formal mathematical term for identical). Note that the order 2 group is (trivially) Abelian, as must be the case if it can be (faithfully) represented by numbers (without needing matrices).

Next consider the representation afforded by the 3 complex numbers $(1, e^{2\pi i/3}, e^{4\pi i/3}) = (1, A, A^2)$, which serve to define a order 3 group. Again the multiplication table, as in Table 10.2, is unique and Abelian (the reader is encouraged to prove this).

	$\mathbf{1}$	\mathbf{A}	\mathbf{A}^2
$\mathbf{1}$	$\mathbf{1}$	\mathbf{A}	\mathbf{A}^2
\mathbf{A}	\mathbf{A}	\mathbf{A}^2	$\mathbf{1}$
\mathbf{A}^2	\mathbf{A}^2	$\mathbf{1}$	\mathbf{A}

Table 10.2: Order 3 Group multiplication table

Groups of order n of the form $(1, A, A^2, \dots, A^{n-1} = 1)$ are called cyclic groups. So Table 10.2 indicates that the $n = 3$ group is cyclic. For $n = 4$ the cyclic group has one representation provided by the set of numbers $(1, i, -1, -i)$, where $(i)^{-1} = -i$ and $(-1)^{-1} = -1$. The multiplication table for the

$n = 4$ cyclic group is indicated in Table 10.3, including the just mentioned explicit representation.

	1	$\mathbf{A}(i)$	$\mathbf{A}^2(-1)$	$\mathbf{A}^3(-i)$
1	1	$\mathbf{A}(i)$	$\mathbf{A}^2(-1)$	$\mathbf{A}^3(-i)$
$\mathbf{A}(i)$	$\mathbf{A}(i)$	$\mathbf{A}^2(-1)$	$\mathbf{A}^3(-i)$	1
$\mathbf{A}^2(-1)$	$\mathbf{A}^2(-1)$	$\mathbf{A}^3(-i)$	1	$\mathbf{A}(i)$
$\mathbf{A}^3(-i)$	$\mathbf{A}^3(-i)$	1	$\mathbf{A}(i)$	$\mathbf{A}^2(-1)$

Table 10.3: Order 4 Cyclic Group multiplication table

However, for the case of $n = 4$ there is a second possible multiplication table as indicated in Table 10.4. This group, often called the 4's Group, is still Abelian, but is *not* cyclic. While it is still true that we have $\mathbf{AB} = \mathbf{BA} = \mathbf{C}$ (or $\mathbf{AA}^2 = \mathbf{A}^2\mathbf{A} = \mathbf{A}^3$ in the cyclic case), we now have $\mathbf{A}^2 = \mathbf{B}^2 = \mathbf{C}^2 = \mathbf{1}$ (instead of $\mathbf{A}^2 = \mathbf{A}^2$, $(\mathbf{A}^2)^2 = \mathbf{1}$ and $(\mathbf{A}^3)^2 = \mathbf{A}^2$).

	1	A	B	C
1	1	A	B	C
A	A	1	C	B
B	B	C	1	A
C	C	B	A	1

Table 10.4: 4's Group multiplication table

Note that the above groups are all Abelian (the elements commute and the multiplication tables are symmetric about the diagonal) and they can be represented by ordinary (complex) numbers. Faithful representations (*i.e.*, representations of the group that are faithful to its properties) of non-Abelian groups will require the use of matrices in order to exhibit nonzero commutators.

In general a group will contain subgroups, *i.e.*, subsets of the elements which themselves form groups. The full (original) group and the unit element are called the trivial subgroups, while other subgroups are called proper subgroups. Clearly the elements $(1, -1) = (1, \mathbf{A}^2)$ constitute a proper subgroup of order 2 of the order 4 cyclic group. In the study of finite groups the concepts of conjugate elements, classes and characters play an important role. Two elements, \mathbf{A} and \mathbf{B} , of a group are conjugate if $\mathbf{B} = \mathbf{C}^{-1}\mathbf{A}\mathbf{C}$ with \mathbf{C} another element of the group. The set of elements conjugate to \mathbf{A} when the element \mathbf{C} is allowed to vary over all members of the group form a class, *i.e.*, they are all related by a similarity transformation. The elements of a class can be thought as representing the same transformation but for differ choices of the basis vectors (recall that a similarity transformation takes us to a different basis set). Since the trace of a matrix is unchanged by a similarity transformation, the traces of all matrices in a single class must be equal. This common trace is called the character of the class. Thus the characters help us to classify the structure of finite groups.

It is also useful to ask whether, through a judicious choice of basis (corresponding to a special similarity transformation), we can make all the matrices representing a group block diagonal. These sub-matrices will also constitute representations of the group, and we say that the original repre-

sensation is reducible. If this diagonalization process is not possible, the original representation is said to be irreducible. Clearly life is easiest if we can work with the lowest dimension (faithful) irreducible representation. This is called the *fundamental* representation. The various fundamental representations play a special role in the process of labeling the multiplets of particles that appear in particle physics.

10.3 Lie Groups

In many (most?) circumstances the groups of interest in physics have an infinite number of elements, but the individual elements are specified by (*i.e.*, are functions of) a finite number (N) of parameters, $g = G(x_1, \dots, x_N)$. Of particular interest are those groups where the parameters vary continuously over some range. Thus the number of parameters is finite but the number of group elements is infinite. If the range of all of the parameters is bounded, the group is said to be compact, *e.g.*, the parameter space of the compact group $SO(3)$ (rotations in 3-D as discussed in Chapter 1) is (the interior of) a sphere of radius π . On the other hand, the group of linear translations in 3-D is non-compact, *i.e.*, the magnitude of the translation can be arbitrarily large. Further, the groups we employ in physics often have the added feature that the derivatives ($\partial g / \partial x_i$) with respect to all parameters exist. Groups with this property are called *Lie Groups*. Lie Groups play an essential role in our understanding of particle physics and we will pursue this discussion of Lie Groups a bit further.

First we focus on the behavior near the origin of the parameter space. By definition the group element at the origin in parameter space,

$$g(0, \dots, 0) \equiv \mathbf{1}, \quad (10.3.1)$$

is the identity element. Near the origin of the parameter space the group elements correspond to infinitesimal transformations (arbitrarily close to the identity operator) and the derivatives are especially important. As a result they have a special name - the *generators* X_k (as already mentioned in Chapters 1 and 5),

$$\left. \frac{\partial g}{\partial x_k} \right|_{x_j=0, \text{all } j} \equiv iX_k. \quad (10.3.2)$$

The factor i in the previous equation (and subsequent equations) arises from the conventional choice to deal with Hermitian generators represented by Hermitian matrices. Since the generators are finite in number ($k = 1$ to N), it is easier to discuss the generators than the infinite number of group elements (we are lazy and smart). The generators serve to define a N -dimensional Lie algebra (a vector space) where both addition (of elements of the algebra) and multiplication by constants are defined. The general element of this Lie algebra can be expressed as a linear combination of the generators,

$$\vec{X} = \sum_{k=1}^N c_k X_k. \quad (10.3.3)$$

This is analogous to the familiar 3-dimensional vector space except that here the generators are the basis vectors (instead of $\hat{x}, \hat{y}, \hat{z}$). Further, we can think of the generators as allowing a "Taylor series" expansion of the group elements near the origin. General group elements can be obtained from the elements of the algebra via exponentiation.

The algebra also supports the definition of an outer (or vector) product (like the familiar cross product) that produces another element of the algebra, *i.e.*, the algebra is closed under this operation. This “product” is just the familiar commutator

$$[X_k, X_l] \equiv X_k X_l - X_l X_k = i C_{klm} X_m. \quad (10.3.4)$$

The tensor C_{klm} is called the structure constant(s) of the algebra. It fully specifies the structure of the algebra and therefore of the structure of the group itself near the origin of the parameter space. For example, the Pauli matrices provide a representation of the unitary group operating on 2-D vectors, or $SU(2)$ (as discussed in Chapter 5). Thus the algebra of $SU(2)$ is given by

$$\left[\frac{\sigma_j}{2}, \frac{\sigma_k}{2} \right] = i \epsilon_{jkl} \frac{\sigma_l}{2}, \sigma_1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \sigma_2 = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \sigma_3 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}, \quad (10.3.5)$$

with the structure constant equal to the (now familiar) fully antisymmetric 3x3x3 tensor ϵ_{jkl} (the Levi-Civita symbol). In fact, this is the *unique* such 3x3x3 antisymmetric tensor and all Lie groups with 3 generators must have the same structure constant and thus the same Lie algebra. In particular, the group of rotations in 3D, $SO(3)$ (as discussed in Chapter 1), has the same algebra,

$$[J_j, J_k] = i \epsilon_{jkl} J_l, \quad (10.3.6)$$

with the J_l standing for the generator of rotations (we are taking $\hbar = 1$ in this discussion). In quantum mechanics this operator will become familiar as the angular momentum operator. An explicit form for these matrices (appropriate for the rotations of ordinary location 3-vectors as in Chapter 1) is given by the following (and the reader is encouraged to check the commutation relations in Eq. 10.3.6),

$$J_1 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & -i \\ 0 & i & 0 \end{pmatrix}, J_2 = \begin{pmatrix} 0 & 0 & i \\ 0 & 0 & 0 \\ -i & 0 & 0 \end{pmatrix}, J_3 = \begin{pmatrix} 0 & -i & 0 \\ i & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}. \quad (10.3.7)$$

While the precise form of these matrices may not be intuitively obvious, the general form should be clear from our understanding of how ordinary rotations work. For example, a rotation about the 1-axis is a rotation in the 2-3 plane. It should serve to mix the 2 and 3 components of an ordinary 3-vector. Modulo the issue of phases, this is precisely what the form of the J_1 matrix in Eq. (10.3.7) does. It transforms a 3-component into a 2-component and a 2-component into a 3-component.

Since the algebras of $SU(2)$ and $SO(3)$ are identical, it is amusing to ask how the groups themselves differ. This issue is addressed below.

While it will not be demonstrated here, a very important concept is the connection between the symmetries of a physical system and the conserved quantities which characterize its dynamics. For example, if a system is rotationally symmetric, *i.e.*, invariant under the operation of the group elements of $SO(3)$, its motion will exhibit a constant, *i.e.*, conserved, angular momentum. Translational invariance in space implies conserved linear momentum, while translation invariance in time implies energy conservation. These connections are realizations of Noether’s Theorem (*i.e.*, symmetries mean conserved quantities, see <http://courses.washington.edu/partsym/14Spr/Noether.pdf>).

To complete this introduction to group theory we will look in general terms at the two Lie groups that seem to arise most often in physics - the Orthogonal group $SO(n)$ and the Unitary group $SU(n)$. The former appears in the study of real n -D vector spaces, *e.g.*, Euclidean space-time, and are defined by being transformations of the vector space that preserve the length of vectors or, more generally

preserving any scalar product (appropriately defined, if there is a nontrivial metric). Thus, if two vectors have zero scalar product in one reference frame, *i.e.*, they are orthogonal, they will remain orthogonal in the rotated frame - hence the name of the group. The Unitary group appears in the study of complex n -D vector spaces, *e.g.*, quantum mechanics, and are defined by again preserving the length of (state) vectors, *i.e.*, probability. [Note that in both cases scalar products are those products that “use” all indices - nothing is left to “operate on”. Hence scalar products are left unchanged by the transformations.]

To see what the properties of the groups these statements imply consider first a n -D real vector and its square, r_1 and $r_1 \bullet r_1 = r_1^T r_1$. Now consider the same vector in a transformed reference frame (or after transforming the vector) where the transformation (rotation) is represented by a real matrix Λ (not a boost here), $r'_1 = \Lambda r_1$. We demand for the Orthogonal group that $r_2 \bullet r_1$ be preserved by the transformation for any r_1, r_2 ,

$$r_2'^T r'_1 = (\Lambda r_2)^T \Lambda r_1 = r_2^T \Lambda^T \Lambda r_1 = r_2^T r_1, \quad (10.3.8)$$

which leads us to

$$\Lambda^T \Lambda = \mathbf{1}, \quad \Lambda^{-1} = \Lambda^T. \quad (10.3.9)$$

So the characteristic feature of the Orthogonal Group is that it is represented by real orthogonal matrices, *i.e.*, matrices whose inverses are their transposes. If the scalar product is defined with a non-trivial metric g , as with the Lorentz transformations of the group $SO(3,1)$ (where the argument (3,1) reminds us of the plus/minus signs in the metric) we have instead (where $\det[g]^2 = 1$)

$$r_2'^T g r'_1 = (\Lambda r_2)^T g \Lambda r_1 = r_2^T \Lambda^T g \Lambda r_1 = r_2^T g r_1, \quad (10.3.10)$$

or

$$\Lambda^T g \Lambda = g, \quad \Lambda^{-1} = g \Lambda^T g. \quad (10.3.11)$$

Note that it follows from these equations and the properties of determinants that

$$\det [\Lambda^T \Lambda] = \det [\Lambda^T] \det [\Lambda] = \det [\Lambda]^2 = 1, \quad \det [\Lambda] = \pm 1, \quad (10.3.12)$$

or

$$\det [g \Lambda^T g \Lambda] = \det [g]^2 \det [\Lambda^T] \det [\Lambda] = \det [\Lambda]^2 = 1. \quad (10.3.13)$$

Recall that typically we want only the “Special” (hence the “ S ” in the label of the group) or unimodular group (no reflections) and we require that the determinant of Λ be +1 (*i.e.*, a -1 means that a reflection is present in the transformation).

Using (complex) exponentiation (recall Eq. (10.3.2)) to go from the algebra to the group, we write $\Lambda = e^{i\alpha S}$, where α is a *real* parameter and, in order for Λ to be real, S is a purely imaginary $n \times n$ matrix (recall Eq. (10.3.7)). The orthogonal form means

$$(e^{i\alpha S})^T = e^{i\alpha S^T} = (e^{i\alpha S})^{-1} = e^{-i\alpha S} \implies S^T = -S = S^* \implies S^\dagger = (S^*)^T = S. \quad (10.3.14)$$

Thus (with our choice of i factors) the generator of a real, orthogonal transformation is represented by a Hermitian matrix (again recall the matrices in Eq. (10.3.7)). The constraint we imposed on the determinant of Λ translates into a constraint on the trace of S (the reader should convince herself of this result)

$$\det [e^{i\alpha S}] = +1 \implies \text{Tr} [S] = 0, \quad (10.3.15)$$

which is trivially satisfied by an anti-symmetric matrix (*e.g.*, any purely imaginary, Hermitian matrix; recall this property for the J_k matrices above). For the more general case of a scalar product defined with a metric g , S is still traceless and satisfies $gS^T g = S^*$, *i.e.*, S displays mixed symmetry as defined by g .

As an explicit example for $SO(3)$ (recall Chapter 1) consider a rotation by an angle θ about the 3 (or z -axis). It follows from the properties of the matrices in Eq. (10.3.7) that powers of these matrices (like the Pauli matrices) are simple,

$$J_3^{2n} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \quad J_3^{2n+1} = J_3. \quad (10.3.16)$$

Thus a rotation matrix defined by the exponential form $\Lambda = e^{i\alpha S}$ is given by its power series expansion as

$$\begin{aligned} g(\theta) &= e^{i\theta J_3} = \mathbf{1} + \sum_{n=1}^{\infty} \frac{(i\theta J_3)^n}{n!} \\ &= \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix} + \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix} \sum_{n=0}^{\infty} \frac{(i\theta)^{2n}}{(2n)!} + J_3 \sum_{n=0}^{\infty} \frac{(i\theta)^{2n+1}}{(2n+1)!} \\ &= \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix} + \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix} \cos \theta + iJ_3 \sin \theta \\ &= \begin{pmatrix} \cos \theta & \sin \theta & 0 \\ -\sin \theta & \cos \theta & 0 \\ 0 & 0 & 1 \end{pmatrix}. \end{aligned} \quad (10.3.17)$$

We recognize that, as expected, this transformation does not change the 3 (or z) component of a vector, while the 1-2 ($x - y$) components are mixed in the just the way we expect for a rotation of the basis vectors in the 1-2 plane. (Note that our choice of signs yields a “passive” rotation, we are rotating the basis vectors not the physics vectors.)

Next we can determine the number of independent components of S , *i.e.*, the dimensionality of the corresponding algebra (also called the order of the Lie group). A purely imaginary Hermitian matrix (i times a real, anti-symmetric matrix) has zeroes on the diagonal and all components below the diagonal are determined by (-1 times) those above. Thus we want 1/2 the number of off-diagonal elements in an $n \times n$ matrix. Hence the algebra of $SO(n)$ has dimension

$$\begin{aligned} N[SO(n)] &= \frac{n^2 - n}{2} = \frac{n(n-1)}{2}, \\ N \begin{bmatrix} n=2 \\ n=3 \\ n=4 \end{bmatrix} &= \begin{matrix} 1 \\ 3 \\ 6 \end{matrix}. \end{aligned} \quad (10.3.18)$$

The corresponding discussion for the Unitary group now involves complex numbers and complex conjugation in the scalar product. Thus, if the unitary transformation is described by a matrix U , we have

$$r_2'^{\dagger} r_1' = (Ur_2)^{\dagger} Ur_1 = r_2^{\dagger} U^{\dagger} U r_1 = r_2^{\dagger} r_1, \quad (10.3.19)$$

i.e., U is a unitary matrix. In (complex) exponential notation (where T is not necessarily purely imaginary as U is not real, but β is real)

$$\begin{aligned} U &= e^{i\beta T}, \quad U^\dagger = e^{-i\beta T^\dagger} = U^{-1} = e^{-i\beta T} \\ &\Rightarrow T^\dagger = T. \end{aligned} \quad (10.3.20)$$

The generator is again represented by a Hermitian matrix, as we expect from our earlier discussion. We also have

$$\det [U^\dagger U] = \det [U]^* \det [U] = 1, \quad \det [U] = \pm 1 \quad (10.3.21)$$

and we focus on the Special version of the group, $SU(n)$,

$$\det [U] \equiv 1 \Rightarrow \text{Tr} [T] = 0. \quad (10.3.22)$$

As an explicit example consider the analogue of the rotation in Eq. (10.3.17), but now replacing the generator J_3 with the corresponding Pauli matrix $\sigma_3/2$ (recall Eq. (10.3.5)). These definitions yield the following expressions

$$\begin{aligned} e^{i\sigma_3\theta/2} &= \mathbf{1} \sum_{n=0}^{\infty} \frac{(i\theta/2)^{2n}}{(2n)!} + \sigma_3 \sum_{n=0}^{\infty} \frac{(i\theta/2)^{2n+1}}{(2n+1)!} \\ &= \mathbf{1} \cos \theta/2 + i\sigma_3 \sin \theta/2 \\ &= \begin{pmatrix} \cos \theta/2 + i \sin \theta/2 & 0 \\ 0 & \cos \theta/2 - i \sin \theta/2 \end{pmatrix} \\ &= \begin{pmatrix} e^{i\theta/2} & 0 \\ 0 & e^{-i\theta/2} \end{pmatrix}. \end{aligned} \quad (10.3.23)$$

So the 2-components of the 2-spinor (for a spinor quantized in the z -direction, recall Chapter 5) are just multiplied by a phase (proportional to the component of the spin, $\pm 1/2$). Note that the transformation is just a factor of -1 for $\theta = 2\pi$. Unlike the rotation of “real” vectors, where a rotation angle of 2π always brings you back to where you started, for spin $1/2$ we end up at -1 times where we started! We need to rotate through 4π to get back to where we started (see below).

So we conclude that the algebra of $SU(n)$ is defined by traceless, Hermitian (but not necessarily imaginary) matrices in the appropriate number of dimensions. In n -D a $n \times n$ complex matrix has 2 times n^2 components. Being Hermitian reduces this by a factor of 2 and the constraint of zero trace removes another degree of freedom. Thus the order of the special unitary algebra in n -D is

$$\begin{aligned} N[SU(n)] &= \frac{2n^2}{2} - 1 = n^2 - 1, \\ N \begin{bmatrix} n=1(\text{really } U(1)) \\ n=2 \\ n=3 \end{bmatrix} &= \begin{bmatrix} 1 \\ 3 \\ 8 \end{bmatrix}. \end{aligned} \quad (10.3.24)$$

Note that the algebras of $U(1)$ and $SO(2)$ have the same (trivial) dimension (1). You might expect that they are related and they are! They are identical or isomorphic as groups, $U(1) \cong SU(2)$. This becomes obvious when we recall that rotations in a (single) plane, $SO(2)$, can be performed in any order, *i.e.*, $SO(2)$ is an Abelian group like $U(1)$. You might have thought that $SO(2)$ had

2-D representations, unlike $U(1)$, but, in fact, these representations can, by an appropriate choice of basis vectors, be *reduced* to the canonical 1-D representations $e^{in\theta}$, which are the irreducible representations of $U(1)$. Another way to see this is to note that 2-D problems, *i.e.*, $SO(2)$, can always be mapped onto the complex plane, *i.e.*, $U(1)$.

The groups $SO(3)$ and $SU(2)$ are also related. As mentioned earlier the algebras are identical since there is a unique choice for the anti-symmetric tensor C_{jkl} . It must be equal to ϵ_{jkl} since this is the *only* 3x3x3 fully antisymmetric tensor - an application of the “what else can it be?” theorem! This implies that these groups must be identical near the origin of the 3-D parameter space. On the other hand, the groups are not isomorphic (identical) when we consider the full parameter space. Instead $SU(2)$ is, in some sense, a larger group. For every element of $SO(3)$ there are two elements in $SU(2)$. This relationship is called a homomorphism with a 2 to 1, $SU(2)$ to $SO(3)$, mapping. Another way to think about this is that the parameter space of $SO(3)$ is like the interior of a sphere of radius π (think about this!). Next consider how the parameter space maps onto the group space. Each point inside the sphere specifies a direction from the origin, which is the axis of the rotation, and a distance from the origin, which is the magnitude of the rotation. When we get to the surface at π , we must identify antipodal points, a rotation through π in one direction is equivalent to a rotation of π about the exactly opposite direction. This means we can define a path in both the parameter space and the group space by starting at the origin, going out to π in one direction, hopping to (exactly) the other side of the sphere, and coming back to the origin. This is a *closed* path in the group space that *cannot* be shrunk to zero in the parameter space! Thus the space is not simply connected. On the other hand for $SU(2)$ we define a similar picture but the sphere extends to 2π and now, no matter what direction we left the origin along, we reach the transformation -1 at 2π (recall that, when we rotate a spin 1/2 state by 2π , we don't get back to the original state but to minus the original state). Thus the entire surface at 2π is identified as a single point in group space (the group element -1 with no issues about only antipodal points in this case). All closed paths can be shrunk to zero and the group space is *simply* connected. $SU(2)$ is called the *covering* group for $SO(3)$.

Before we finish this discussion, let's think just a bit more about the representations of groups. Just as in our initial discussion of finite groups, we need the concept of reducible and irreducible. If a representation (*i.e.*, the matrices) of the group elements can be reduced to block diagonal form by some choice of basis vectors,

$$\left[\begin{array}{cc|cc|cc} [&] & 0 & & 0 & \\ 0 & [& &] & 0 & \\ 0 & & 0 & & [&] \end{array} \right], \quad (10.3.25)$$

then that representation is reducible. If it cannot be written in this form, it is irreducible. In fact, the irreducible representations of Abelian groups are all 1-D. The smallest dimension representation that faithfully represents the group, *i.e.*, displays all of its structure (the commutation relations), is called the defining or fundamental representation. All groups have 1-D (scalar) representations but they are not faithful for non-Abelian groups. [Recall that 1-D representations are just numbers, which must commute unlike matrices.] For $SO(3)$ the fundamental representation is the vector representation, **3**. For $SU(2)$ the fundamental representation is the spinor representation, **2**. The half-integer spin representations ($J = 1/2, 3/2, \dots$), are often referred to as the spinor representations of $SO(3)$ but they are strictly the representations of $SU(2)$. The fundamental representation of the algebra of

$SU(2)$ and $SO(3)$ are provided by the matrices in Eqs. (10.3.5) and (10.3.7), respectively.

Interestingly we can also interpret the algebra itself as providing a representation of the group, called the *adjoint* representation. The generators themselves are the basis vectors. The transformation of a basis vector by a generator is defined as the commutator of the generator with the “basis vector” (*i.e.*, the other generator). Hence the structure constants, the C_{jkl} , define a matrix representation of the algebra and the group (by exponentiation). For $SO(3)$ the adjoint representation and the fundamental representation are identical, *i.e.*, the matrices in Eq. (10.3.7) can be written in the form $[J_j]_{kl} = i\epsilon_{jlk}$, where it is important to note the order of the last two indices. Thus for $SO(3)$ the adjoint representation provided by the structure constants (properly defined) is identical to the fundamental representation. Similarly the generators of $SU(2)$ provide not only the adjoint representation of $SU(2)$ but also form a fundamental (and adjoint) representation of $SO(3)$, *i.e.*, the Pauli matrices transform like a 3-vector under rotation.

Chapter 11

Multiplets and Young Diagrams

11.1 Basic Definitions

For the question of decomposing products of $SU(N)$ representations into irreducible representations (e.g., $N = 2$ for spin or $N = 3$ for color or flavor), the most efficient notation is that of Young diagrams. These are just left justified arrays of boxes with a specific set of (seemingly ad hoc) rules for their manipulation and interpretation. Without derivation the rules include the following.

1. Each horizontal row of boxes is at least as long as the horizontal row below it.
2. We can think of the horizontal direction as symmetrization (with respect to some internal index) and the vertical direction as anti-symmetrization. There are at most N rows for the case of $SU(N)$.
3. For the $SU(3)$ representation (p,q) the first row has p more boxes than the second row and the second row has q more boxes than the third row. Thus we have

$$\begin{aligned} (1,0) = \underline{3} = \begin{array}{|c|} \hline \square \\ \hline \end{array}, \quad (0,1) = \bar{\underline{3}} = \begin{array}{|c|} \hline \square \\ \hline \square \\ \hline \end{array}, \quad (0,0) = \underline{1} = \begin{array}{|c|} \hline \square \\ \hline \square \\ \hline \square \\ \hline \end{array}, \\ (1,1) = \underline{8} = \begin{array}{|c|c|} \hline \square & \square \\ \hline \square & \square \\ \hline \end{array}, \quad (2,0) = \underline{6} = \begin{array}{|c|c|} \hline \square & \square \\ \hline \end{array}, \quad (3,0) = \underline{10} = \begin{array}{|c|c|c|} \hline \square & \square & \square \\ \hline \end{array}. \end{aligned} \quad (11.1.1)$$

4. The counting of states within a given representation (Young diagram) involves three steps. First you “fill in” the boxes starting with the upper left hand corner based on the symmetry group. For $SU(N)$ you put N in the upper left corner box and then increase the number when moving to the right and decrease the number when moving down (see the examples below). For example, for $SU(3)$ the diagram $\begin{array}{|c|c|} \hline \square & \square \\ \hline \square & \square \\ \hline \end{array}$ becomes

$$\begin{array}{|c|c|} \hline \square & \square \\ \hline \square & \square \\ \hline \end{array} \Rightarrow \begin{array}{|c|c|} \hline 3 & 4 \\ \hline 2 & \end{array}. \quad (11.1.2)$$

The next step involves again putting numbers in the boxes but this time the number correspond to the length the “hook” that has that box as the “elbow” of the “hook” (this is the most confusing part of this game). For each individual box in the Young diagram we count the number of boxes to

the right of the starting box in the *same* row and the number of boxes below the given box in the *same* column. The sum of these two integers, plus 1 for the original box, is the length of the (right) “hook” with the original box at the “elbow”. We put this length (*i.e.*, this integer) in the box and proceed until all boxes contain the lengths of the associated hooks.¹ As an example, consider again the diagram $\begin{array}{|c|c|} \hline \square & \square \\ \hline \square & \\ \hline \end{array}$. Applying the hook rules just defined populates the Young diagram in the following way,

$$\begin{array}{|c|c|} \hline \square & \square \\ \hline \square & \\ \hline \end{array} \Rightarrow \begin{array}{|c|c|} \hline 3 & 1 \\ \hline 1 & \\ \hline \end{array}. \quad (11.1.3)$$

Without proof, we state that the number of states in the representation corresponding to a Young diagram is given by the product of all of the numbers in the boxes based on the N of the symmetry divided by the product of all the numbers in the boxes corresponding to the lengths of the hooks. For the example above for $SU(3)$, we have

$$N \left(\begin{array}{|c|c|} \hline \square & \square \\ \hline \square & \\ \hline \end{array} \right) \equiv \frac{\begin{array}{|c|c|} \hline 3 & 4 \\ \hline 2 & \\ \hline \end{array}}{\begin{array}{|c|c|} \hline 3 & 1 \\ \hline 1 & \\ \hline \end{array}} \equiv \frac{3 \cdot 4 \cdot 2}{3 \cdot 1 \cdot 1} = 8, \quad (11.1.4)$$

as expected. Two other examples to test your understanding are

$$\begin{aligned} N \left(\begin{array}{|c|} \hline \square \\ \hline \square \\ \hline \square \\ \hline \end{array} \right) &= \frac{\begin{array}{|c|} \hline 3 \\ \hline 2 \\ \hline 1 \\ \hline \end{array}}{\begin{array}{|c|} \hline 3 \\ \hline 2 \\ \hline 1 \\ \hline \end{array}} = \frac{3 \cdot 2 \cdot 1}{3 \cdot 2 \cdot 1} = 1, \\ N \left(\begin{array}{|c|c|c|} \hline \square & \square & \square \\ \hline \square & \square & \square \\ \hline \square & \square & \square \\ \hline \end{array} \right) &= \frac{\begin{array}{|c|c|c|} \hline 3 & 4 & 5 \\ \hline 3 & 2 & 1 \\ \hline \end{array}}{\begin{array}{|c|c|c|} \hline 3 & 2 & 1 \\ \hline \end{array}} = \frac{3 \cdot 4 \cdot 5}{3 \cdot 2 \cdot 1} = 10. \end{aligned} \quad (11.1.5)$$

11.2 Combine Multiplets

To actually combine multiplets, *i.e.*, define a product of representations, we need to carefully label things. Here we use the notation of the PDG.

(See <http://pdg.lbl.gov/2013/reviews/rpp2013-rev-young-diagrams.pdf>.)

Consider the product of 2 octets defined with the following notation

$$\mathfrak{8} \otimes \mathfrak{8} = \begin{array}{|c|c|} \hline \square & \square \\ \hline \square & \\ \hline \end{array} \otimes \begin{array}{|c|c|} \hline a & a \\ \hline b & \\ \hline \end{array}, \quad (11.2.1)$$

where we use boxes to represent the first octet and lettered boxes for the second (with “a” for the first row, “b” for the second row, etc.). Now we proceed to “add the boxes” with the following rules.

1. Start with the left-hand Young diagram (the empty boxes) $\begin{array}{|c|c|} \hline \square & \square \\ \hline \square & \\ \hline \end{array}$.

¹We thank auditor Hanan Bell, Spring 2014, for suggesting this description of the hooks. Several other descriptions are possible.

2. Now add the “a”’s in all ways that produce a valid Young diagram, but with no more than a single “a” in each column (initially symmetric labels cannot be antisymmetrized)

$$\begin{array}{|c|c|} \hline & \\ \hline & \\ \hline \end{array} \Rightarrow \begin{array}{|c|c|c|c|} \hline & & a & a \\ \hline & & & \\ \hline & & a & \\ \hline \end{array} \oplus \begin{array}{|c|c|c|} \hline & & a \\ \hline & & \\ \hline & a & \\ \hline \end{array} \oplus \begin{array}{|c|c|c|} \hline & & a \\ \hline & & \\ \hline & a & \\ \hline \end{array} \oplus \begin{array}{|c|c|c|} \hline & & a \\ \hline & & \\ \hline & a & \\ \hline \end{array} \oplus \begin{array}{|c|c|c|} \hline & & a \\ \hline & & \\ \hline & a & \\ \hline \end{array}. \quad (11.2.2)$$

3. Starting in the *second* row (where the “b”’s were initially) add the “b”’s subject to the constraint that, reading from right to left starting at the right end of the first row and then moving on to the second row, the number of “a”’s must be \geq the number of “b”’s (\geq the number of “c”’s, *etc.*) at each point in the reading process. Thus, when reading through the boxes in the prescribed fashion, we can come to the first “b” only after we have passed *at least* one “a”. We come to the second “b” only after passing the second “a”, *etc.* The allowed Young diagrams for our current example are

$$\begin{array}{|c|c|c|c|} \hline & & a & a \\ \hline & & & \\ \hline & b & & \\ \hline \end{array} \oplus \begin{array}{|c|c|c|c|} \hline & & a & a \\ \hline & & & \\ \hline & b & & \\ \hline \end{array} \oplus \begin{array}{|c|c|c|} \hline & & a \\ \hline & & \\ \hline & a & b \\ \hline \end{array} \oplus \begin{array}{|c|c|c|} \hline & & a \\ \hline & & \\ \hline & b & a \\ \hline \end{array} \oplus \begin{array}{|c|c|c|} \hline & & a \\ \hline & & \\ \hline & a & b \\ \hline \end{array} \oplus \begin{array}{|c|c|c|} \hline & & a \\ \hline & & \\ \hline & a & b \\ \hline \end{array} \oplus \begin{array}{|c|c|c|} \hline & & a \\ \hline & & \\ \hline & a & b \\ \hline \end{array} \\ = \begin{array}{|c|c|c|c|} \hline & & & \\ \hline & & & \\ \hline & & & \\ \hline & & & \\ \hline \end{array} \oplus \begin{array}{|c|c|c|c|} \hline & & & \\ \hline & & & \\ \hline & & & \\ \hline & & & \\ \hline \end{array} \oplus \begin{array}{|c|c|c|c|} \hline & & & \\ \hline & & & \\ \hline & & & \\ \hline & & & \\ \hline \end{array} \oplus \begin{array}{|c|c|c|c|} \hline & & & \\ \hline & & & \\ \hline & & & \\ \hline & & & \\ \hline \end{array} \oplus \begin{array}{|c|c|c|c|} \hline & & & \\ \hline & & & \\ \hline & & & \\ \hline & & & \\ \hline \end{array} \oplus \begin{array}{|c|c|c|c|} \hline & & & \\ \hline & & & \\ \hline & & & \\ \hline & & & \\ \hline \end{array}. \quad (11.2.3)$$

4. Using the rules noted earlier we can work out the multiplicity of each of these irreducible representations (using the notation introduced above).

$$\begin{array}{|c|c|c|c|} \hline & & & \\ \hline & & & \\ \hline & & & \\ \hline & & & \\ \hline \end{array} \oplus \begin{array}{|c|c|c|c|} \hline & & & \\ \hline & & & \\ \hline & & & \\ \hline & & & \\ \hline \end{array} \oplus \begin{array}{|c|c|c|c|} \hline & & & \\ \hline & & & \\ \hline & & & \\ \hline & & & \\ \hline \end{array} \oplus \begin{array}{|c|c|c|c|} \hline & & & \\ \hline & & & \\ \hline & & & \\ \hline & & & \\ \hline \end{array} \oplus \begin{array}{|c|c|c|c|} \hline & & & \\ \hline & & & \\ \hline & & & \\ \hline & & & \\ \hline \end{array} \oplus \begin{array}{|c|c|c|c|} \hline & & & \\ \hline & & & \\ \hline & & & \\ \hline & & & \\ \hline \end{array} \\ = \frac{\begin{array}{|c|c|c|c|} \hline 3 & 4 & 5 & 6 \\ \hline 2 & 3 & & \\ \hline \end{array}}{\begin{array}{|c|c|c|c|} \hline 5 & 4 & 2 & 1 \\ \hline 2 & 1 & & \\ \hline \end{array}} \oplus \frac{\begin{array}{|c|c|c|c|} \hline 3 & 4 & 5 & 6 \\ \hline 2 & & & \\ \hline 1 & & & \\ \hline \end{array}}{\begin{array}{|c|c|c|c|} \hline 6 & 3 & 2 & 1 \\ \hline 2 & & & \\ \hline 1 & & & \\ \hline \end{array}} \oplus \frac{\begin{array}{|c|c|c|} \hline 3 & 4 & 5 \\ \hline 2 & 3 & \\ \hline \end{array}}{\begin{array}{|c|c|c|} \hline 4 & 3 & 2 \\ \hline 3 & 2 & 1 \\ \hline \end{array}} \\ \oplus \frac{\begin{array}{|c|c|c|} \hline 3 & 4 & 5 \\ \hline 2 & 3 & \\ \hline 1 & & \\ \hline \end{array}}{\begin{array}{|c|c|c|} \hline 5 & 3 & 1 \\ \hline 3 & 1 & \\ \hline 1 & & \\ \hline \end{array}} \oplus \frac{\begin{array}{|c|c|c|} \hline 3 & 4 & 5 \\ \hline 2 & 3 & \\ \hline 1 & & \\ \hline \end{array}}{\begin{array}{|c|c|c|} \hline 5 & 3 & 1 \\ \hline 3 & 1 & \\ \hline 1 & & \\ \hline \end{array}} \oplus \frac{\begin{array}{|c|c|} \hline 3 & 4 \\ \hline 2 & 3 \\ \hline 1 & 2 \\ \hline \end{array}}{\begin{array}{|c|c|} \hline 4 & 3 \\ \hline 3 & 2 \\ \hline 2 & 1 \\ \hline \end{array}} \\ = \frac{3 \cdot 4 \cdot 5 \cdot 6 \cdot 2 \cdot 3}{5 \cdot 4 \cdot 2 \cdot 2 \cdot 1 \cdot 1} \oplus \frac{3 \cdot 4 \cdot 5 \cdot 6 \cdot 2 \cdot 1}{6 \cdot 3 \cdot 2 \cdot 2 \cdot 1 \cdot 1} \oplus \frac{3 \cdot 4 \cdot 5 \cdot 2 \cdot 3 \cdot 4}{4 \cdot 3 \cdot 3 \cdot 2 \cdot 2 \cdot 1} \\ \oplus \frac{3 \cdot 4 \cdot 5 \cdot 2 \cdot 3 \cdot 1}{5 \cdot 3 \cdot 3 \cdot 1 \cdot 1 \cdot 1} \oplus \frac{3 \cdot 4 \cdot 5 \cdot 2 \cdot 3 \cdot 1}{5 \cdot 3 \cdot 3 \cdot 1 \cdot 1 \cdot 1} \oplus \frac{3 \cdot 4 \cdot 2 \cdot 3 \cdot 1 \cdot 2}{4 \cdot 3 \cdot 3 \cdot 2 \cdot 2 \cdot 1} \\ = 27 \oplus 10 \oplus \bar{10} \oplus 8 \oplus \bar{8} \oplus 1 \\ = (2, 2) \oplus (3, 0) \oplus (0, 3) \oplus (1, 1) \oplus (1, 1) \oplus (0, 0). \quad (11.2.4)$$

Note that, as we have seen before, the “a-b” notation makes clear that the internal symmetry structure of the 2 octets is different. So finally we have the result that

$$8 \otimes 8 = 27 \oplus 10 \oplus \bar{10} \oplus 8 \oplus \bar{8} \oplus 1 \quad (11.2.5)$$

Looking ahead to the application to the $SU(3)$ of color we can reproduce some other results that we have already used. Consider the product of a quark and antiquark,

$$\begin{aligned} \mathfrak{3} \otimes \bar{\mathfrak{3}} &= \square \otimes \begin{array}{|c|} \hline a \\ \hline b \\ \hline \end{array} = \begin{array}{|c|c|} \hline & a \\ \hline b & \\ \hline \end{array} \oplus \begin{array}{|c|} \hline a \\ \hline b \\ \hline \end{array} \\ &= \mathfrak{8} \oplus \mathfrak{1}. \end{aligned} \quad (11.2.6)$$

Next consider the product of 3 quarks, but begin by first looking at 2 quarks,

$$\mathfrak{3} \otimes \mathfrak{3} = \square \otimes \square = \begin{array}{|c|c|} \hline & \\ \hline & \\ \hline \end{array} \oplus \begin{array}{|c|} \hline \\ \hline \\ \hline \end{array} = \mathfrak{6} \oplus \bar{\mathfrak{3}}. \quad (11.2.7)$$

With the third quark we have

$$\begin{aligned} \mathfrak{3} \otimes \mathfrak{3} \otimes \mathfrak{3} &= (\mathfrak{6} \oplus \bar{\mathfrak{3}}) \otimes \mathfrak{3} = \left(\begin{array}{|c|c|} \hline & \\ \hline & \\ \hline \end{array} \oplus \begin{array}{|c|} \hline \\ \hline \\ \hline \end{array} \right) \otimes \square \\ &= \begin{array}{|c|c|c|} \hline & & \\ \hline & & \\ \hline \end{array} \oplus \begin{array}{|c|c|} \hline & \\ \hline & \\ \hline \end{array} \oplus \begin{array}{|c|c|} \hline & \\ \hline & \\ \hline \end{array} \oplus \begin{array}{|c|} \hline \\ \hline \\ \hline \end{array} \\ &= \mathfrak{10} \oplus \mathfrak{8} \oplus \mathfrak{8} \oplus \mathfrak{1}. \end{aligned} \quad (11.2.8)$$

In the context of color we are interested only in the color singlets for the mesons and baryons respectively. Note, as already discussed, that the singlet is the completely antisymmetric state. Applied to the $SU(3)$ of flavor, we see again that the mesons should appear in octets and singlets while the baryons should form decuplets, octets (of differing internal permutation symmetry) and singlets of flavor. However, not all of these states can be combined (with space, color and spin wave functions) to yield states with the required overall asymmetry under permutations. For example, the antisymmetric color wave function requires net symmetry in the other quantum numbers. For the ground state we expect the space wave function to be symmetric. The spin wave function is either symmetric ($S = 3/2$) or mixed ($S = 1/2$). Thus only the flavor symmetric $\mathfrak{10}$, with spin $3/2$, and the appropriately mixed symmetry $\mathfrak{8}$, with spin $1/2$, can appear in the baryon ground state.

For further discussion (including how to connect the integers (p,q) to the “shapes” of the multiplets in isospin-strangeness plane) see the PDG report at

<http://pdg.lbl.gov/2013/reviews/rpp2013-rev-young-diagrams.pdf>

The brief summary is that p counts the number of “spaces” between occupied states at the top (largest strangeness) of the multiplet and q counts the number of spaces between occupied states at the bottom (most negative strangeness).

Index

- G_F , *see* Fermi constant
- 3-velocity, 62
- 4-acceleration, 62
- 4-force, 62
- 4-momentum, 60
 - conservation, 68
- 4-velocity, 58
- , 154
- antimatter, 93
 - antibaryon, 98
 - antilepton, 94
 - antineutrino, 94
- baryon, 98
 - decuplet, 105
 - number, 111
 - octet, 107
 - wavefunction, 107
- beta decay, 91
- boost, *see* Lorentz transformation
- bra vector, 121
- causality, 49
- charge conjugation, 137
- charge quantization, 129
- Clebsch-Gordan, 121
- CM frame, 69
- color, 97
- cosmic ray, 27
- CP, 139
- CPT, 140
- electromagnetism, 67
- electron, 85
- energy, 61
- event, 32
- Exponential Function, 10
- Fermi constant, 146, 147
- Feynman diagrams, 153
- field strength tensor, 68
- finite groups, 155
- flavor, *see* quark, flavor
- G-parity, 138
- Galilean relativity, 22
- gamma factor, 25
- gamma rays, 92
- GPS, 27
- group theory, 154
- hadron, 96
 - decay, 111
 - spectrum, 132
- Hamiltonian, 121
- Hyperbolic Functions, 10
- ideal clock, 24
- inertial frame, 32
- intrinsic parity, 136
- invariant interval, 44
- isospin, 131
- ket vector, 121
- lab frame, 69
- Ladder Operators, 73
- lepton, 94
 - number, 94
- Lie Groups, 157
- lightcone, 34
- lightlike, 45
- Lorentz
 - contraction, 26
 - force, 68
 - transformation, 36
- meson, 98
- Minkowski space, *see* spacetime

- momentum, 61
- muon, 94
 - decay spectrum, 145
 - shower, 27
- natural units, 42, 145
- neutrino, 91, 94
- neutron, 87
- Newton's laws, 22
- Noether's Theorem, 158
- nucleon, 89
- nucleus, 85
- nuclide, 88
- parity, 135
- Pauli principle, 89
- photon, 92
- plane wave, 42, 65
- positron, 92
- Power Series, 9
- proper time, 57
- proton, 87
- QCD, *see* quantum chromodynamics
- quantum chromodynamics, 96
- quantum dynamics, 121
- quark, 96
 - flavor, 97
 - masses, 97
- reference frame, 32
- relativistic addition of velocities, 59
- relativity postulates, 24
- rest energy, 61
- rest mass, 60
- scattering, 68
- Schrödinger equation, 122
- simultaneity, 34
- Sinusoidal Functions, 13
- spacelike, 45
- spacetime, 32
 - dot product, 41
 - Minkowski, 44
 - vector, 39
- speed of light, 24
- Spin and Statistics, 89
- strangeness, 131
- strong interaction, 96
- surface of simultaneity, 33
- symmetry, 121
 - approximate, 130
 - continuous, 126
 - crossing, 147
 - generator, 126
 - spacetime, 127
- tau lepton, 94
- time dilation, 25
- time reversal, 140
- timelike, 45
- unitarity bound, 149
- weak gauge boson, 153
- weak interaction, 91, 143
- worldline, 32
- worldvolume, 32
- Young diagrams, 164