

Preface

The preparation of these notes began in 2008 when I taught the first offering of a 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 (bras and kets), quantum time evolution, observables and expectation values, spin-1/2 and related two-state systems, and quantized angular momentum. Facility with calculus, linear algebra, and basic mathematical methods is also assumed. Brief appendices on basic mathematics and quantum mechanics summarize some of this needed background material. Prior exposure to special relativity, or particle physics, is not required.

This version of these notes incorporates or adapts a number of suggestions due to my colleague, Stephen D. Ellis, who has taught *Particles and Symmetries* multiple times starting in 2011. His contributions are gratefully acknowledged.

Some words regarding conventions: Arrows are used to indicate three-dimensional spatial vectors, such as \vec{x} . Components of spatial vectors are written as x^i , with a Latin index (such as i) which runs from 1 to 3. Four-dimensional spacetime vectors, which are introduced in chapter 2, are not 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 only sensible choice (in the opinion of this author), which makes the dot product of spacetime vectors having vanishing time component the same as the usual three-dimensional dot product, and allows plane waves in space and spacetime to have the same $e^{ik \cdot x}$ form. Pay no attention to anyone urging use of the other convention!

Laurence G. Yaffe
March 2015

Introduction

As we start this study of *Particles and Symmetries* it is appropriate to begin with a description of the overall goal of the course, which is to provide an introduction to an area of physics that has seen dramatic progress in the last 50 years — elementary particle physics. A central tool underlying this progress has been the exploitation of symmetries, reflected in the interactions of particles, hence the title of this course. The understanding which has emerged is encoded in the so-called Standard Model of particle physics, which identifies the fundamental particles and interactions among these particles relevant for describing nearly all of the physical universe. When one includes collective behavior (quarks bound in nuclei, electrons bound in atoms, atoms bound in solid matter) plus classical gravity, the result is a nearly complete explanation for the physics observed from the largest distance scales, *e.g.*, the evolution of the universe from very early times, down to the shortest distances probed at particle accelerators. Full command of this fundamental understanding requires some tools not at our disposal, namely quantum field theory. However, using only special relativity, basic quantum mechanics, and symmetries, one can understand a surprisingly large portion of particle physics in an accessible and relatively quantitative fashion.

From a pedagogical perspective, this endeavor provides an opportunity to discuss special relativity in detail, and practice using it to describe the kinematics of particle collisions at high energy. Key concepts will include 4-dimensional momentum conservation (which is itself associated with the invariance of physics under translations in space and time), and the universal speed limit set by the speed of light. Developing practical facility with 4-vector notation and the transformations (boosts) that relate quantities in different inertial reference frames will be emphasized. From quantum mechanics, heavy use will be made of the uncertainty principle, and the key role played by simultaneous eigenstates of mutually commuting operators. You should have seen some of this structure in the context of quantized angular momentum. We will use symmetries, and a little bit of the underlying mathematics of group theory, to make many testable predictions. Important applications will involve approximate symmetries, where there is not exact invariance under some transformation, but rather the transformation induces “small” perturbations. This will allow us to separate big effects from small effects, or more formally organize perturbative expansions in these small effects — another essential tool in physics. Throughout the course, one emphasis will be honing your skill for making order-of-magnitude estimates, *i.e.*, quickly estimating a rough value for some physical quantity even when fine details are not known. Do not be overly concerned if some of these concepts are unclear on first reading — clarity should improve as the course progresses.

Chapter 1

Special relativity

1.1 Galilean relativity

Newton's laws of motion,

$$\frac{d\vec{p}}{dt} = \vec{F}, \quad \frac{d\vec{x}}{dt} = \frac{\vec{p}}{m}, \quad (1.1.1)$$

retain the same form if one substitutes

$$\vec{x} \rightarrow \vec{x}' + \vec{u}t, \quad \vec{p} \rightarrow \vec{p}' + m\vec{u}, \quad (1.1.2)$$

for any velocity \vec{u} which is constant (independent of time). In other words, equations (1.1.1) and (1.1.2) imply that

$$\frac{d\vec{p}'}{dt} = \vec{F}, \quad \frac{d\vec{x}'}{dt} = \frac{\vec{p}'}{m}. \quad (1.1.3)$$

This shows that changing coordinates to those of a 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. It is an example of an *invariance*, a change in the description of a system (in this case, a change in the coordinate system) which preserves the form of the equations of motion. An intrinsic aspect of Galilean relativity is the assumption that time has the same meaning in all inertial frames, so t represents time as measured by any good (and synchronized) clock, regardless of whether that clock is moving.

Consider a particle, or wave, which moves with some velocity \vec{v} when viewed in the unprimed frame, so that 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 (1.1.2), 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

$$\vec{v}' = \vec{v} - \vec{u}. \quad (1.1.4)$$

This *shift* in velocities upon transformation to a moving frame is completely in accord with everyday experience. For example, as illustrated in Figure 1.1, if a person standing on the ground sees a

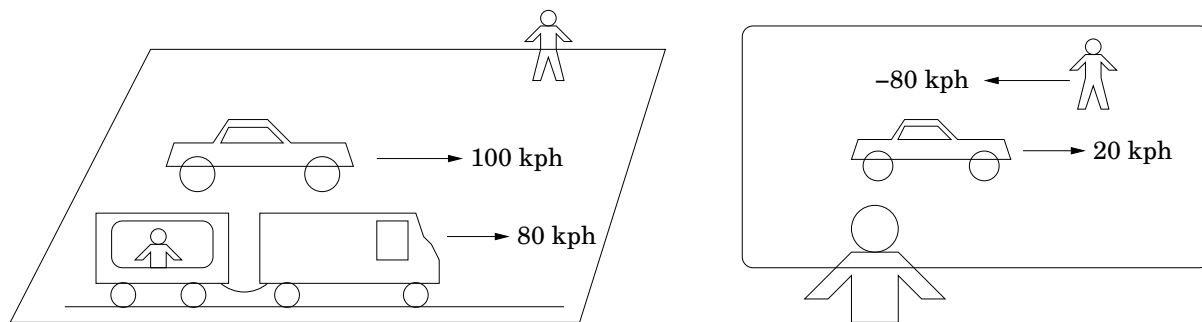


Figure 1.1: A moving train and car, as seen from the ground (left), and from the train (right).

car moving at 100 kph (kilometers per hour) parallel to a train moving at 80 kph, then a person sitting in the train will see that car moving with a relative velocity of 20 kph = (100 – 80) 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,

$$f' = \frac{v_s - u}{\lambda} = f \left(1 - \frac{u}{v_s} \right). \quad (1.1.5)$$

This is the familiar *Doppler shift* for the case of a moving observer and stationary source with respect to the medium. Recall that the medium plays an important role here. If, with respect to the medium, it is the observer who is stationary while the source moves away from the observer at speed u , then the result for the Doppler shift becomes

$$f' = f / \left(1 + \frac{u}{v_s} \right). \quad (1.1.6)$$

The two results coincide to first order in u/v_s (*i.e.*, (1.1.6) approaches (1.1.5) for $|u/v_s| \ll 1$), but as u approaches v_s (so that $u/v_s \rightarrow 1$) the two expressions are very different. This reflects the fact that for sound, there is a physically distinguished *special* reference frame, the rest frame of the medium through which the sound propagates.

1.2 Constancy of c

When applied to light (electromagnetic radiation), the Galilean relativity velocity transformation (1.1.4) 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 (or source). It has been unequivocally demonstrated that (1.1.4) does not apply to light. Moreover, unlike sound, light requires no medium in which to propagate.

Newton's laws, and the associated Galilean relativity relations (1.1.2) and (1.1.4), provide extremely accurate descriptions for the dynamics of particles and waves which move slowly compared to the speed of light,¹

$$c = 2.99\,792\,458 \times 10^8 \text{ m/s}. \quad (1.2.1)$$

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 small compared to c . Our goal is to find a formulation of 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.

1.3 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 sundials. 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 sundial changes with the season and, on much longer time scales, changes due to slowing of the Earth's rotation caused by tidal friction.

An idealized clock, which is particularly simple to analyze, is shown in Fig. 1.2. 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 $\Delta t = 2L/c$.

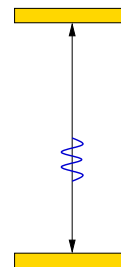


Figure 1.2: An idealized clock in which a pulse of light repeatedly bounces between two mirrors.

¹This value is exact — because the meter is defined by this value for c and the international standard for time.

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.

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. 1.3. 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

$$\Delta t' = \frac{\Delta t}{\sqrt{1 - (u/c)^2}}. \tag{1.3.1}$$

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 (called the *Lorentz factor*) of

$$\gamma \equiv \frac{1}{\sqrt{1 - (u/c)^2}}, \tag{1.3.2}$$

where u is the speed with which the clock is moving. Note that γ is greater than one for any non-zero speed u which is less (in magnitude) than c .

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. 1.4. Analyzing this case is also instructive.

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

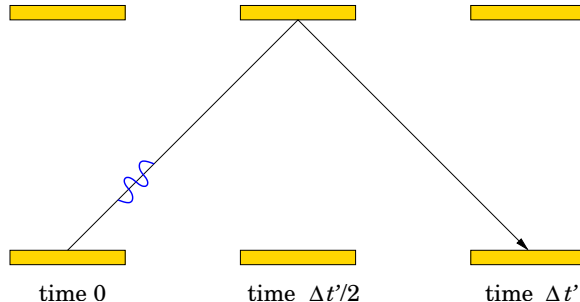


Figure 1.3: Three snapshots of the same clock viewed from a moving frame.

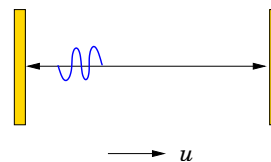


Figure 1.4: Our idealized clock, now rotated so that its axis is parallel to the direction of motion.

³After all, if some other good clock remains synchronized with our idealized clock when viewed in their common rest frame, then postulate 2 implies that the same synchronization between the two clocks must also be present when the two clocks are viewed in a moving frame.

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, by postulate 2 they must also remain synchronized when viewed from a moving frame whose velocity is parallel to one clock and perpendicular to the other.

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

$$\Delta t' = \frac{L'}{c + u} + \frac{L'}{c - u} = \frac{2cL'}{c^2 - u^2} = \gamma^2 (2L'/c). \quad (1.3.3)$$

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 ,

$$L' = \frac{L}{\gamma} = L \sqrt{1 - (u/c)^2}. \quad (1.3.4)$$

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).

1.4 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 observed directly.

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 (usually a proton or atomic nucleus) strikes

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.

an air molecule in the upper reaches of the atmosphere (typically above 20 km), 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 slow 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 is 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 flying 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, but not the fact that high energy muons *can* reach the ground from the upper atmosphere.

1.5 Superluminal motion?

The time dilation (1.3.1) and Lorentz contraction (1.3.4) equations make no sense (*i.e.*, are no longer real) if $u > c$. As we will discuss further 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 . Consequently, there was considerable excitement in the autumn of 2011 when the OPERA neutrino experiment at the Gran Sasso laboratory in Italy reported that neutrinos (which are thought to have a very small but nonzero mass) appeared 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 precise synchronization between clocks in Geneva and in Gran Sasso. A precision better than 50 nanoseconds is needed; this is achievable, with difficulty, using the GPS system.

The OPERA result fundamentally conflicts with special relativity. Either our postulates, or the experimental measurement, must be in error. All indications are that it is the measurement which

was in error. In early 2012, the OPERA team reported that their original measurement may have suffered from a synchronization error caused by a loose connection in the cable which relayed GPS signals to the experiment's clocks. Moreover, in April 2012 the independent ICARUS experiment, also at Gran Sasso, reported their measurement of neutrino speed, which was found to be consistent with the speed of light.

1.6 Example problems

1.6.1 Earthbound clocks

Two ideal clocks, initially synchronized, sit on the surface of the Earth, one at the south pole and the other on the equator. (Ignore any difference in elevation between the two sites.)

(a) Q: Will these two clocks remain synchronized?

A: No, because the Earth rotates. As viewed in the frame in which the Earth's center of mass is at rest, which is also the frame in which the south pole is (essentially) at rest, the clock on the equator is moving, and hence will exhibit time dilation.

(b) Q: For the equatorial clock, how much does γ differ from 1?

A: The Earth's circumference is about 25,000 miles or 40,000 km. The rotation period of 1 day = $24 \times 60 \times 60$ seconds = 86,400 s. So the rotation speed of a point on the equator is $v \approx 40/86.4$ km/s = 463 m/s. Dividing by c gives $v/c \approx 1.5 \times 10^{-6}$, or $(v/c)^2 \approx 2.4 \times 10^{-12}$.

To evaluate $\gamma - 1$, one must be careful not to lose all numerical accuracy. Blindly using (1.3.2), one must evaluate the denominator keeping more than 12 significant digits, or else the effect of the $(v/c)^2$ term will be completely lost. Since v/c is tiny, a better approach is to use the first two terms in the binomial expansion $1/\sqrt{1-x} = 1 + \frac{1}{2}x + \frac{3}{8}x^2 + \dots$. Hence, $\gamma - 1 \approx \frac{1}{2}(v/c)^2 \approx 1.2 \times 10^{-12}$.

(c) Q: After one year, what is the time difference between the two clocks? Is this measurable (using current technology)?

A: Relative to the polar clock, the equatorial clock will have lost $(\gamma - 1) \times 86,400$ seconds, which is just over 100 ns. This is measurable. The best atomic clocks have long term frequency stability of a few parts in 10^{16} , much better than the part in 10^{12} effect of time dilation due to Earth's rotation.

1.7 Further resources

GPS and Relativity, R. Pogge

Relativity in the Global Positioning System, N. Ashby

Michelson-Morley experiment, Wikipedia

Global Positioning System, Wikipedia

Cosmic ray, Wikipedia

OPERA experiment, Wikipedia

ICARUS experiment, Wikipedia

Chapter 2

Minkowski spacetime

2.1 Events

An *event* is some occurrence which takes place at some instant in time at some particular point in 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.

2.2 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 *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, *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. This is illustrated in Figure 2.1. In general, the *rest frame* of an inertial observer (or object) means the inertial frame in which the specified observer (or object) is at rest.

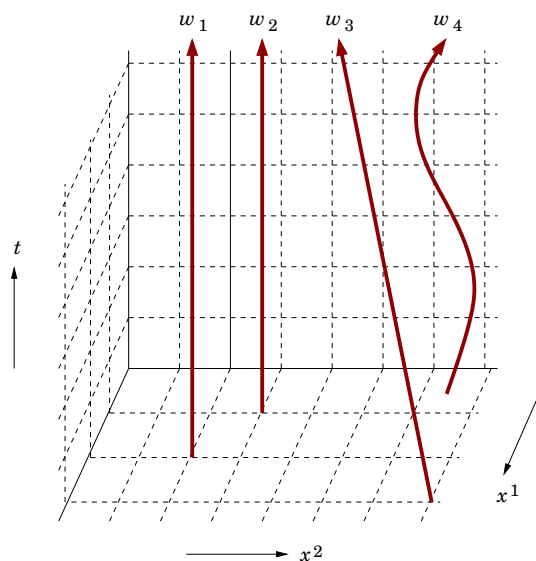


Figure 2.1: An inertial reference frame. Worldlines w_1 and w_2 represent observers at rest in this reference frame, w_3 is the spacetime trajectory of an inertial observer who is moving in this frame, and w_4 is the spacetime trajectory of a non-inertial object whose velocity and acceleration fluctuates.

As Figure 2.2 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.¹ You can imagine every observer carrying a notebook and recording the time, according to his or her clock, of events of interest.

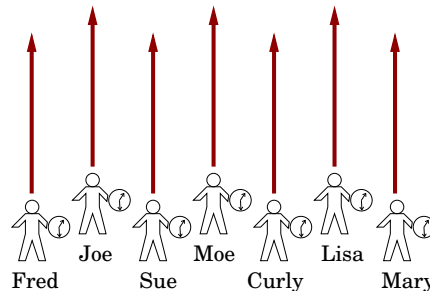


Figure 2.2: A family of inertial observers, all with synchronized watches and mutually at rest, defines an inertial reference frame.

As an example, consider the statement that 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 2.3.

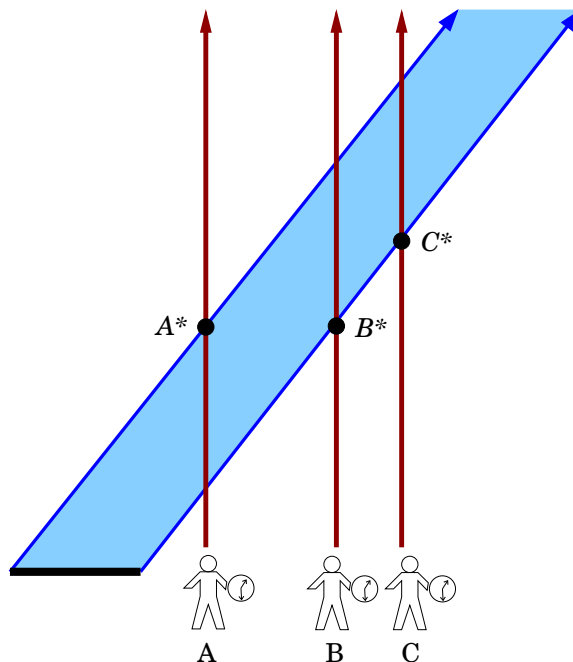


Figure 2.3: A moving rod which passes by three inertial observers who are at rest.

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^* (in Fig. 2.3, B^* is displaced precisely horizontally from A^*), while event C^* is not. The length of the rod, in this reference frame, is the spatial distance between events A^* and B^* , marking the endpoints of the rods on a surface of simultaneity. This is the same as the distance between observers A and B, who are mutually at rest. 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

$$d_{AB} = [(x_B^1 - x_A^1)^2 + (x_B^2 - x_A^2)^2 + (x_B^3 - x_A^3)^2]^{1/2} . \tag{2.2.1}$$

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

¹Achieving this synchronization can be a challenge in practice — as evident from the discussion of the OPERA experiment in section 1.5.

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 drift in the value of Δt indicates that the two observers are not mutually at rest (or that their clocks are failing to remain synchronized).

2.3 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 length

$$x^0 \equiv ct, \quad (2.3.1)$$

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.*, 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 $[(x^1)^2 + (x^2)^2 + (x^3)^2]^{1/2} = x^0$. This set of events form a cone, illustrated in Figure 2.4. The intersection of this cone with the x^0 - x^1 plane is 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 $\pm x^1$ direction.

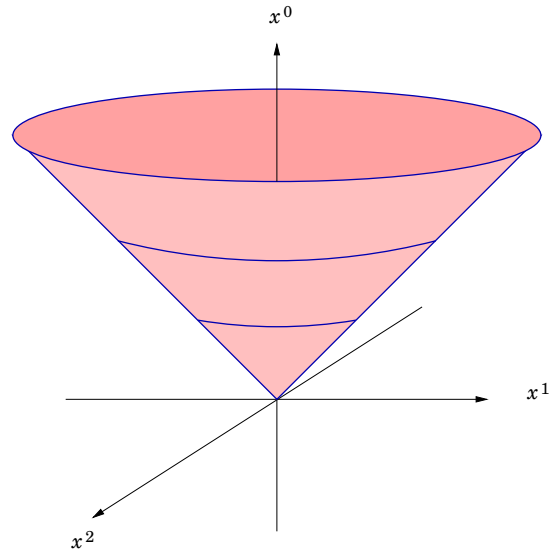


Figure 2.4: The “lightcone” of a flash of light emitted from the origin.

2.4 Simultaneity

Next, consider the reference frames of two different inertial 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, so that B's position satisfies

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

Figure 2.5 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, 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) is a tilted line with a slope of c/v , since this corresponds to all events with $x^0 = (c/v)x^1$ (and vanishing x^2 and x^3). Note that as $v \rightarrow 0$ the slope diverges and the line becomes vertical, coinciding with the worldline of A. As $v \rightarrow c$, the slope approaches one and the worldline of B approaches a 45° line lying on the lightcone.

Surfaces of simultaneity for observer A correspond to horizontal planes in this diagram, because such planes represent 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), and are shown in the figure as thin red lines labeled $x'^0 = -1, 0, \text{ or } 1$.

One way to see that this must be the case is to note that the 45° worldline of a light ray traveling from the origin in the $+x^1$ direction (the dashed line with unit slope in the figure) bisects the angle between observer A's worldline (the x^0 axis in the figure) and his surface of simultaneity defined by $x^0 = 0$. Exactly the same statement must be true for observer B — she will also describe the path of the light as bisecting the angle between her worldline and her surface of simultaneity (the red $x'^0 = 0$ line) which contains the origin. This is an application of our second postulate (physics looks the same in all inertial reference frames). Hence, when plotted in A's reference frame, 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 which one observer views as occurring simultaneously (but at different locations) will not be simultaneous when viewed by a different observer moving at a non-zero relative velocity.

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 2.6. Two flashes of light (shown as black lines in the figure) are emitted at events R and S and meet at event T . In observer B's frame, shown on the left, the emission events are *simultaneous* and displaced 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 frame, $w_{B''}$ is the worldline of an observer sitting at rest a distance L' away, and $w_{B'}$ is the worldline of an

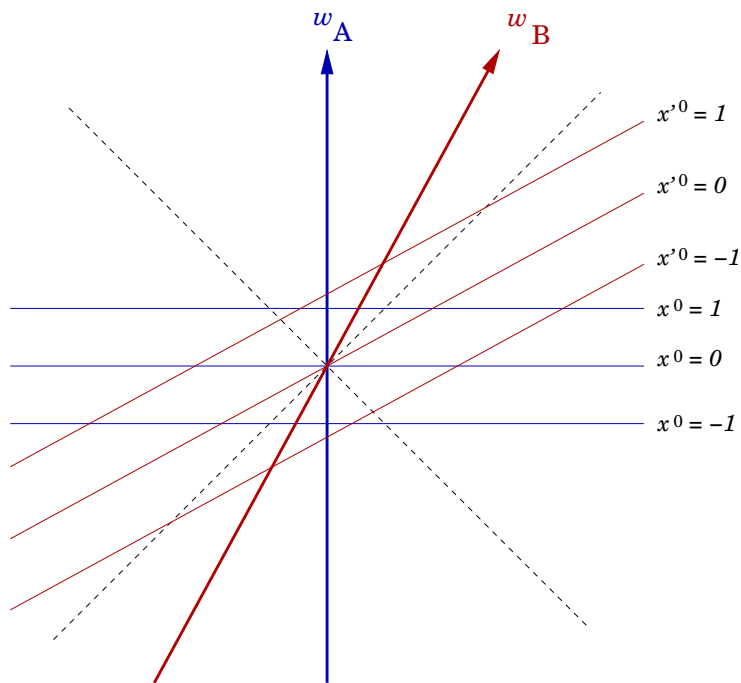


Figure 2.5: Worldline of two observers, and corresponding surfaces of simultaneity. The dashed lines show the lightcone of the origin.

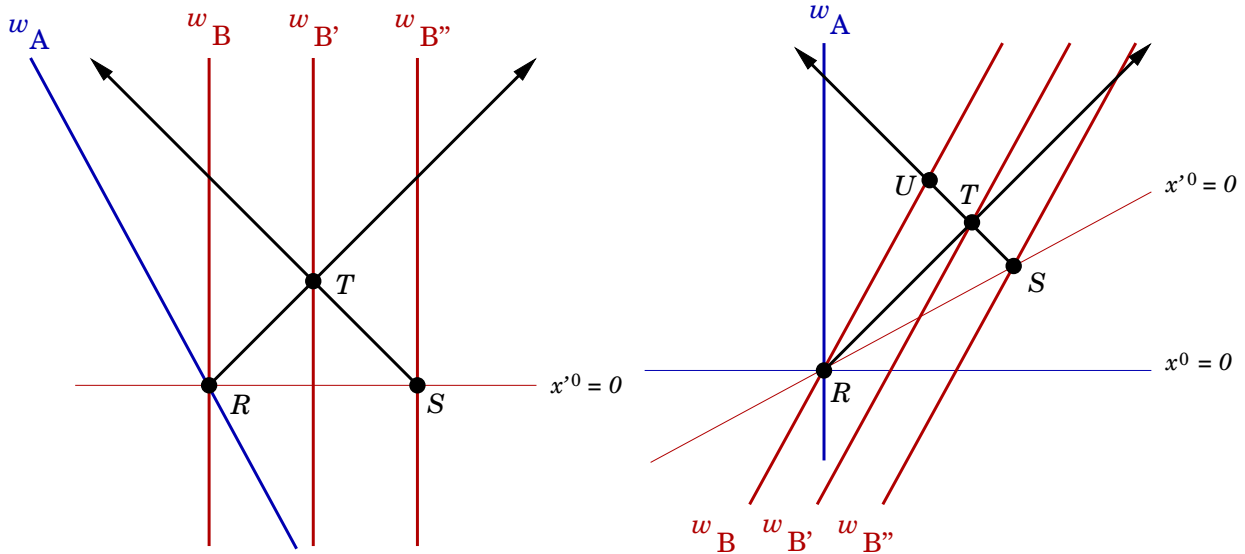


Figure 2.6: Two flashes of light emitted at events R and S which meet at event T , as described in two different frames.

observer at rest halfway in between B and B'' .)

In observer A 's frame, shown in the right panel of Figure 2.6, the worldlines of observers at rest in frame B are now tilted lines with slope c/v . But the paths of the light rays (propagating within the plane shown) lie at $\pm 45^\circ$, 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 event U lies in the left panel, and show that in this panel triangles RTU and RTS are again similar.)

2.5 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*.

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,

$$\begin{pmatrix} x^0 \\ x^1 \\ x^2 \\ x^3 \end{pmatrix} = \Lambda \begin{pmatrix} x'^0 \\ x'^1 \\ x'^2 \\ x'^3 \end{pmatrix}, \quad (2.5.1)$$

where Λ is some 4×4 matrix. 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 Lorentz contraction will only affect lengths in the 1-direction, leaving the 2 and 3 directions unaffected. Therefore, we should have

$$x^2 = x'^2, \quad x^3 = x'^3 \quad (\text{for boost along } x^1), \quad (2.5.2)$$

implying that the boost matrix Λ has the block diagonal form

$$\Lambda = \begin{pmatrix} M & N & 0 & 0 \\ P & Q & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}, \quad (2.5.3)$$

with an identity matrix in the lower 2×2 block, and some non-trivial 2×2 matrix in the upper 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 his coordinate system. Specializing to this worldline, the transformation (2.5.1) gives

$$x^0 = M x'^0, \quad x^1 = P x'^0, \quad (2.5.4)$$

implying that $x^1 = (P/M)x^0$. 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 $P/M = v/c$. 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,

$$\gamma = \frac{\Delta t_A}{\Delta t_B} = \frac{dx^0}{dx'^0} = M. \quad (2.5.5)$$

Combining this with the required value of P/M implies that $P = \gamma(v/c)$. This determines the first column of the Lorentz boost matrix (2.5.3).

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 (2.5.1) gives

$$x^0 = N x'^1, \quad x^1 = Q x'^1, \quad (2.5.6)$$

or $x^0 = (N/Q)x^1$. Therefore, we must have $N/Q = v/c$. 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 (2.5.1) gives $x^0 = (M + N)x'^0$, and $x^1 = (P + Q)x'^0$. Therefore, we must have $M + N = P + Q$. Inserting $M = \gamma$, $P = (v/c)\gamma$, $N = (v/c)Q$ and solving for Q yields $Q = \gamma$. Putting it all together, we have

$$\Lambda = \begin{pmatrix} \gamma & \gamma(v/c) & 0 & 0 \\ \gamma(v/c) & \gamma & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}, \quad (2.5.7)$$

for a boost along the 1-direction with velocity v . With this matrix, multiplying out the transformation (2.5.1) gives

$$x^0 = \gamma(x'^0 + \frac{v}{c}x'^1), \quad x^2 = x'^2, \quad (2.5.8a)$$

$$x^1 = \gamma(\frac{v}{c}x'^0 + x'^1), \quad x^3 = x'^3. \quad (2.5.8b)$$

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 $\hat{n} = (n_x, n_y, n_z)$ is given by

$$\Lambda = \begin{pmatrix} \gamma & \gamma(v/c)n_x & \gamma(v/c)n_y & \gamma(v/c)n_z \\ \gamma(v/c)n_x & 1 + (\gamma-1)n_x^2 & (\gamma-1)n_xn_y & (\gamma-1)n_xn_z \\ \gamma(v/c)n_y & (\gamma-1)n_xn_y & 1 + (\gamma-1)n_y^2 & (\gamma-1)n_yn_z \\ \gamma(v/c)n_z & (\gamma-1)n_xn_z & (\gamma-1)n_yn_z & 1 + (\gamma-1)n_z^2 \end{pmatrix}. \quad (2.5.9)$$

Finally, it is always possible for two inertial reference frames to differ by a spatial rotation, in addition to a boost. The coordinate transformation corresponding to a spatial rotation may also be written in the form (2.5.1), but with a transformation matrix which has the block-diagonal form

$$\Lambda = \begin{pmatrix} 1 & 0 \\ 0 & R \end{pmatrix} \quad (\text{spatial rotation}), \quad (2.5.10)$$

where R is some 3×3 rotation matrix (an orthogonal matrix with determinant one). 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 (2.5.10) and a boost of the form (2.5.9),

$$\Lambda = \Lambda_{\text{boost}} \times \Lambda_{\text{rotation}}. \quad (2.5.11)$$

2.6 Rapidity

The mixing of time and space components of a four-vector generated by the Lorentz transformation matrix (2.5.7) may seem reminiscent of the mixing of spatial components of a vector undergoing a rotation. A closer connection is apparent if one introduces the ‘‘rapidity’’ η , which is monotonically related to v/c via

$$\tanh \eta = v/c. \quad (2.6.1)$$

The rapidity η ranges from $-\infty$ to $+\infty$ as v/c varies between -1 and $+1$. A short exercise (using the hyperbolic identity $1 - \tanh^2 z = 1/\cosh^2 z$) shows that

$$\cosh \eta = \gamma, \quad \sinh \eta = \gamma (v/c), \quad (2.6.2)$$

so the non-trivial upper 2×2 block of the Lorentz transformation matrix (2.5.7) takes the form $\begin{pmatrix} \cosh \eta & \sinh \eta \\ \sinh \eta & \cosh \eta \end{pmatrix}$, with hyperbolic functions replacing the usual trigonometric functions appearing in a rotation. Characterizing a boost by its rapidity (instead of v/c) is often convenient; rapidity is commonly used when analyzing data from particle colliders such as the Large Hadron Collider (LHC) near Geneva.

2.7 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 coordinate axes), and then express arbitrary vectors as linear combinations of the chosen basis vectors,

$$\vec{v} = \sum_{i=1}^3 \hat{e}_i v^i. \quad (2.7.1)$$

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,

$$v = \sum_{\mu=0}^3 \hat{e}_\mu v^\mu. \quad (2.7.2)$$

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 sometimes we will use Latin letters i, j, k to represent spatial indices 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:

$$\hat{e}_\mu v^\mu \equiv \sum_{\mu=0}^3 \hat{e}_\mu v^\mu. \quad (2.7.3)$$

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,

$$x = \hat{e}_\mu x^\mu \equiv \hat{e}_0 x^0 + \hat{e}_1 x^1 + \hat{e}_2 x^2 + \hat{e}_3 x^3. \quad (2.7.4)$$

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 (2.5.1). 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 (2.5.1) may be compactly rewritten as

$$x^\mu = \Lambda^\mu{}_\nu x'^\nu. \quad (2.7.5)$$

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

$$x'^\mu = (\Lambda^{-1})^\mu{}_\nu x^\nu, \quad (2.7.6)$$

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 changes reference frames.

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

$$\hat{e}'_\nu = \hat{e}_\mu \Lambda^\mu{}_\nu. \quad (2.7.7)$$

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,

$$x = \hat{e}'_\mu x'^\mu = \hat{e}_\nu x^\nu. \quad (2.7.8)$$

Recall that the *dot product* of two spatial vectors, $\vec{a} \cdot \vec{b}$, may be 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 that the dot product of spatial vectors is a scalar.

What is the appropriate generalization of dot products 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 be *independent* of the choice of (inertial) reference frame one uses to specify the components of these vectors. And it should reduce to the usual spatial dot product if both a and b lie within a common surface of simultaneity. There is a unique solution to these requirements: given two spacetime vectors a and b whose components in some inertial frame are a^μ and b^μ , the dot product of these vectors is

$$a \cdot b \equiv -a^0 b^0 + a^1 b^1 + a^2 b^2 + a^3 b^3, \quad (2.7.9)$$

²For boost matrices of the form (2.5.7) or (2.5.9), changing the sign of v 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.

or with an implied summation on spatial indices, $a \cdot b = -a^0 b^0 + a^i b^i$. This differs from the normal definition of a dot product (in four dimensional Euclidean space) merely by the change in sign of the time component term. This definition satisfies the required linearity and reduces to the usual spatial dot product if the time components of both four vectors vanish.³

To see that the dot product definition (2.7.9) is frame-independent, and thus defines a scalar, it is sufficient to check the effect of a boost of the form (2.5.7), 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. (2.7.6), using the boost (2.5.7) gives

$$a'^0 = \gamma \left(a^0 - \frac{v}{c} a^1 \right), \quad a'^1 = \gamma \left(a^1 - \frac{v}{c} a^0 \right), \quad a'^2 = a^2, \quad a'^3 = a^3, \quad (2.7.10a)$$

$$b'^0 = \gamma \left(b^0 - \frac{v}{c} b^1 \right), \quad b'^1 = \gamma \left(b^1 - \frac{v}{c} b^0 \right), \quad b'^2 = b^2, \quad b'^3 = b^3. \quad (2.7.10b)$$

Hence

$$\begin{aligned} -a'^0 b'^0 + a'^1 b'^1 + a'^2 b'^2 + a'^3 b'^3 & \\ &= \gamma^2 \left[- \left(a^0 - \frac{v}{c} a^1 \right) \left(b^0 - \frac{v}{c} b^1 \right) + \left(a^1 - \frac{v}{c} a^0 \right) \left(b^1 - \frac{v}{c} b^0 \right) \right] + a^2 b^2 + a^3 b^3 \\ &= \gamma^2 \left[1 - (v/c)^2 \right] \left(-a^0 b^0 + a^1 b^1 \right) + a^2 b^2 + a^3 b^3 \\ &= -a^0 b^0 + a^1 b^1 + a^2 b^2 + a^3 b^3, \end{aligned} \quad (2.7.11)$$

where the last step used $\gamma^2 \equiv 1/[1 - (v/c)^2]$. Therefore, as claimed, the value of the dot product (2.7.9) is independent of the specific inertial frame one uses to define the vector coefficients.

The spacetime dot product (2.7.9) is a useful construct in many applications. As a preview of things to come, consider some plane wave (acoustic, electromagnetic, or any other type) propagating with angular frequency ω and wave-vector \vec{k} . One normally writes the complex amplitude for such a wave as some overall coefficient times the complex exponential $e^{-i\omega t + i\vec{k} \cdot \vec{x}}$. 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 compactly as a spacetime dot product,

$$e^{-i\omega t + i\vec{k} \cdot \vec{x}} = e^{ik \cdot x}. \quad (2.7.12)$$

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 spacetime momentum vector p with time component $p^0 \equiv E/c$, then this phase factor may also be written as a spacetime dot product,

$$e^{-iEt/\hbar + i\vec{p} \cdot \vec{x}/\hbar} = e^{ip \cdot x}. \quad (2.7.13)$$

2.8 Minkowski spacetime

In Euclidean space, the dot product of a vector with itself gives the square of the norm (or length) of the vector, $\vec{v} \cdot \vec{v} \equiv |\vec{v}|^2$. Proceeding by analogy, we will *define* the square of a spacetime vector

³Regrettably some physicists, and some textbooks, define the dot product of spacetime vectors with an overall minus sign relative to (2.7.9), so that $a \cdot b \stackrel{\text{bad}}{\equiv} a^0 b^0 - a^i b^i$. This makes spacetime dot products reduce to *minus* the usual three-dimensional dot product when time components vanish. As long as one uses a single convention consistently, no physical result can depend on the choice. However, this author strongly recommends using, exclusively, the “mostly plus” convention (2.7.9).

using the dot product (2.7.9), so that

$$(a)^2 \equiv a \cdot a = -(a^0)^2 + (a^1)^2 + (a^2)^2 + (a^3)^2. \quad (2.8.1)$$

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

$$s^2 \equiv -(\Delta x^0)^2 + (\Delta x^1)^2 + (\Delta x^2)^2 + (\Delta x^3)^2. \quad (2.8.2)$$

Spacetime in which the “distance” between events is defined by this expression is called *Minkowski spacetime*.^{4,5,6}

The definition of the invariant interval (2.8.2), or the square of a vector (2.8.1), differ from the usual Euclidean space relations merely due to the minus sign in front of the time component terms. But this is a fundamental change. Unlike Euclidean distance, the spacetime interval s^2 can be positive, negative, or zero. 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 ,

$$s^2 = 0 \implies (\Delta \vec{x})^2 = (\Delta x^0)^2 = (c \Delta t)^2 \quad (\text{lightlike separation}). \quad (2.8.3)$$

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 negative, then the spatial separation is less than the time separation (times c),

$$s^2 < 0 \implies (\Delta \vec{x})^2 < (\Delta x^0)^2 = (c \Delta t)^2 \quad (\text{timelike separation}). \quad (2.8.4)$$

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 positive, then the spatial separation is greater than the time separation (times c),

$$s^2 > 0 \implies (\Delta \vec{x})^2 > (\Delta x^0)^2 = (c \Delta t)^2 \quad (\text{spacelike separation}). \quad (2.8.5)$$

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 2.7.

⁴Those authors who choose to define spacetime dot products with an overall minus sign (“mostly minus” convention), mentioned in footnote 3, also define the spacetime interval with an overall minus sign relative to our definition (2.8.2). Our sign convention is more convenient — use it, but beware of differing conventions in the literature.

⁵Minkowski 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 (2.8.2). We will largely ignore gravity.

⁶A further word about index conventions may 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, differ merely by flipping the sign of the time component, so that $a_0 \equiv -a^0$ 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 Minkowski spacetime, the metric tensor is diagonal, $\|g_{\mu\nu}\| = \text{diag}(-1, +1, +1, +1)$.

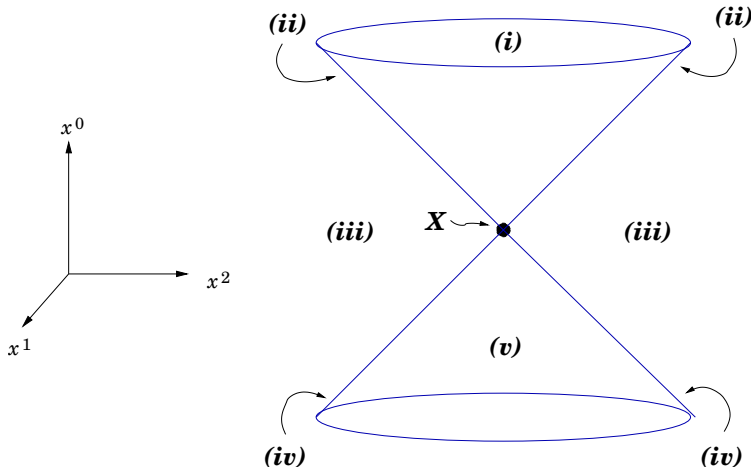


Figure 2.7: The past and future lightcones of an event X separate spacetime into those events which are: (i) timelike separated and in the future of X , (ii) lightlike separated and in the future of X , (iii) spacelike separated, (iv) lightlike separated and in the past of X , and (v) timelike separated and in the past of X .

2.9 The pole and the barn

A classic puzzle illustrating basic aspects of special relativity is the pole and the barn, sketched in Figure 2.8. 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 length of the barn 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 $\frac{\sqrt{3}}{2}c \simeq 0.866c$, and hence the pole (in your frame) is Lorentz contracted by a factor of $1/\gamma = \sqrt{1 - (v/c)^2} = 1/2$. Consequently, 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 entered the barn through the other door.

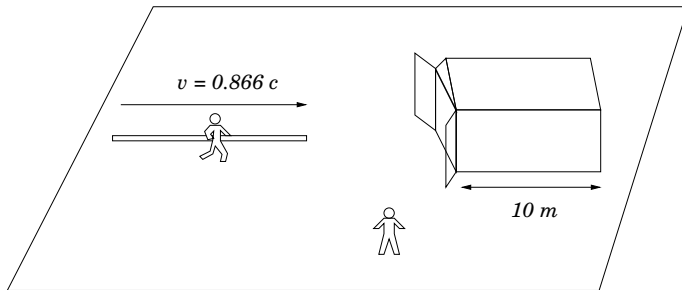


Figure 2.8: A relativistic runner, carrying a long pole, approaches and passes through a barn. Does the pole fit within the barn?

Figure 2.8: A relativistic runner, carrying a long pole, approaches and passes through a barn. Does the pole fit within the barn?

But now consider 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 most 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 also often 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.⁷

⁷Recall that the equation $x^2 - y^2 = s^2$ defines a hyperbola in the (x, y) plane whose asymptotes are the 45° lines

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 runner's frame.) Try doing this yourself before reading the following step-by-step description of Figure 2.9.

Orient coordinates so that the ends of the barn are at $x^1 = 0$ and $x^1 = 10$ 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 $x^1 = 10$ 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 2.4 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 2.9, 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.

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 intervals from event A using the runner's frame coordinates, instead of barn frame coordinates. Within the 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 is at 5 m, while event D must lie on the hyperbola whose intersection with the x^1 axis is at 20 m, as shown.

$y = \pm x$. 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.

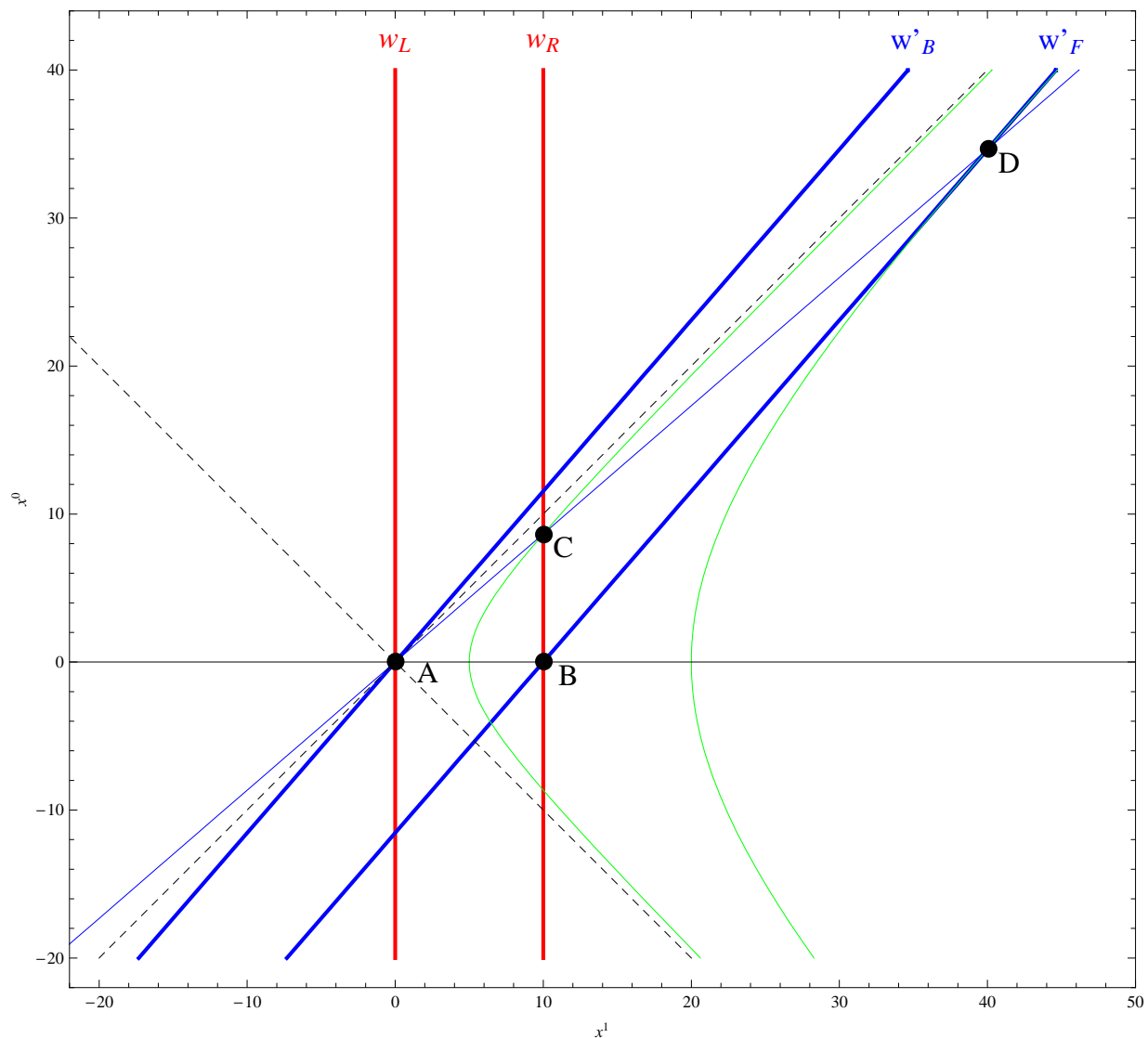


Figure 2.9: A spacetime diagram of the pole and the barn, showing events in the rest frame of the barn. The red vertical lines are the worldlines w_L and w_R of the left and right ends of the barn. The blue lines labeled w'_F and w'_B are the worldlines of the front and back of the pole, respectively. The thin blue line passing through events A, C, and D is a surface of simultaneity in the runner's reference frame. The green hyperbola passing through event C shows events at invariant interval $s^2 = (5 \text{ m})^2$ relative to event A. This hyperbola intercepts the x^1 axis at 5 m. The other green hyperbola passing through event D shows events at invariant interval $s^2 = (20 \text{ m})^2$ relative to event A. Note that this hyperbola intercepts the x^1 axis at 20 m.

2.10 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. 2.10, 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 others 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.

2.11 Example problems

2.11.1 Proper time intervals⁸

The time interval between two events is called a *proper time interval* in some given inertial frame if the two events occur at the same spatial location in that frame. Consider two frames of reference: the rest frame (frame S) of the Earth and the rest frame (frame S') of a spaceship moving with velocity $v = 0.6c$ with respect to Earth. The spaceship skims the surface of the Earth at some instant — call this event 1. Assume that coordinates and clocks are adjusted so that event 1 has coordinates

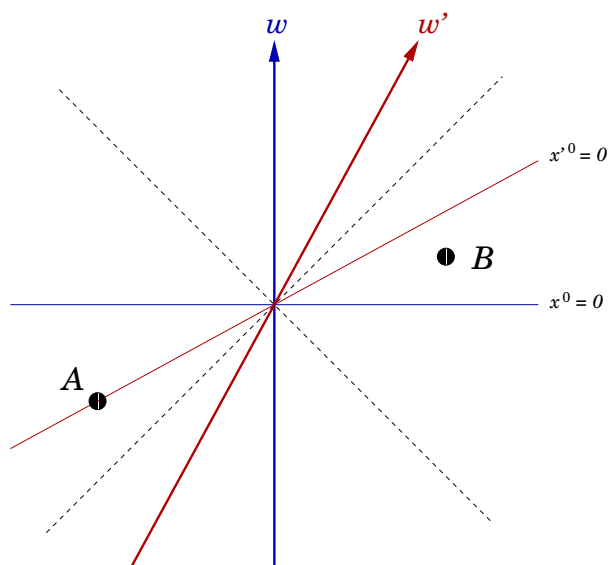


Figure 2.10: Two spacelike separated events A and B . In the unprimed frame, B is in the future of A , but in the primed frame B is in the past of A .

⁸Adapted from Kogut problem 2-1.

$t_1 = 0, x_1 = 0$ in frame S , and $t'_1 = 0, x'_1 = 0$ in frame S' . Event 2 marks the emission of a pulse of light from the Earth towards the spaceship at $t_2 = 10$ minutes. Event 3 marks the detection of the light pulse by observers in the spaceship. (Neglect the sizes of both the Earth and the spaceship.)

- (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 spatial location in frame S (*i.e.*, on the Earth), but not at the same location in frame S' on the spaceship. Hence the time interval between events 1 and 2 is a proper time interval in the Earth frame, but not in the spaceship frame.

- (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 locations 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 location on the spaceship (frame S'), but not at the same point on the Earth. Hence the time interval between events 1 and 3 is a proper time interval in the spaceship frame, but not in the Earth frame.

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

A: We want to determine the time t'_2 of the light emission in frame S' . This time interval (from event 1) is not a proper time interval and we must account for time dilation (with respect to the proper time interval in frame S). We have $\gamma = 1/\sqrt{1 - (v/c)^2} = 1/\sqrt{1 - (0.6)^2} = 1.25$, and hence $t'_2 = \gamma t_2 = 1.25 \times 10 \text{ min} = 12.5 \text{ min}$.

- (e) Q: In the spaceship frame, how far away is the Earth when the light pulse is emitted?

A: We need to determine the distance to the Earth from the spaceship at the moment (in frame S') when the light is emitted. This is just the distance traveled at velocity v during the time interval $\Delta t' = t'_2 - t'_1$ (as measured in frame S') between events 1 and 2. Using the value for t'_2 from part (d), and $t'_1 = 0$, we have:

$$l'_2 = v \Delta t' = 0.6 \times (3.0 \times 10^8 \text{ m/s}) \times (12.5 \text{ min}) \times (60 \text{ s/min}) = 1.35 \times 10^{11} \text{ m}.$$

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

A: We need the time of event 3 in frame S' . We already know both the time and distance to the Earth at the emission of the pulse, and we know that light travels at c in *all* frames. Thus, we merely need to add the light travel time to the emission time (all in frame S'),

$$t'_3 = t'_2 + l'_2/c = 12.5 \text{ min} + (1.35 \times 10^{11} \text{ m})/(3 \times 10^8 \text{ m/s}) \times (1 \text{ min}/60 \text{ s}) = 20 \text{ min}.$$

- (g) Q: Analyzing everything in the Earth frame, find the time of event 3 according to Earth's clock.

A: The light pulse is emitted (in frame S) at $t_2 = 10 \text{ min}$. At that moment the distance to the spaceship, in Earth's frame, is $l_2 = vt_2 = 0.6 \times (3 \times 10^8 \text{ m/s}) \times 10 \text{ min} \times 60 \text{ s/min} = 1.08 \times 10^{11} \text{ m}$.

Between events 2 and 3, the light pulse moves away from Earth at speed c while the spaceship continues to recede at speed v . The light reaches the spaceship when $c(t_3 - t_2) = l_2 + v(t_3 - t_2)$, so

$$t_3 - t_2 = \frac{l_2}{c - v} = \frac{1.08 \times 10^{11} \text{ m}}{1.2 \times 10^8 \text{ m/s}} = 900 \text{ s} = 15 \text{ min}.$$

Finally, $t_3 = t_2 + (t_3 - t_2) = 10 \text{ min} + 15 \text{ min} = 25 \text{ min}$.

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

A: We noted in (c) that the time interval between events 1 and 3 is a proper time interval in frame S' . In frame S , it will appear that the spaceship clock is running slow, due to time dilation, so that

$$t_3 = \gamma t'_3 = 1.25 \times 20 \text{ min} = 25 \text{ min}.$$

Reassuringly, this agrees with our result from (g).

2.11.2 Passing in the night⁹

Two rockets, A and B, pass each other while moving in opposite directions. The rockets have identical proper lengths (*i.e.*, lengths in their respective rest frames) of 100 m. Consider two events: Event 1 is when the front of B passes the front end of A. Event 2 is when the front of B passes the back end of A. In frame A (the rest frame of rocket A), the time interval Δt_A between the two events is 1.5×10^{-6} s.

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

A: We know the length of rocket A in its rest frame, 100 m, and the time for the front of rocket B to travel that distance (as measured in frame A). The distance/time ratio gives the velocity of B as measured in frame A, and this is the relative velocity of the two rockets. Hence,

$$v_{\text{rel}} = \frac{100 \text{ m}}{1.5 \times 10^{-6} \text{ s}} = 6.667 \times 10^7 \text{ m/s}.$$

(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 as viewed from A. Hence, the time Δt_B for the front of A to pass the entire length of rocket B, as measured in frame B, is again 1.5×10^{-6} s.

(c) Q: According to the clocks on rocket B, how much time passes between events 1 and 2 (*i.e.*, between the passage of the front of B by the front of A, and the passage of the front of B by the rear of A)? Does this agree with your answer to (b)? Should it?

A: In frame B, the length of rocket A is Lorentz contracted, $L_{A \text{ in B}} = (100 \text{ m})/\gamma$, with $\gamma = 1/\sqrt{1 - (v_{\text{rel}}/c)^2} = 1.0257$. So $L_{A \text{ in B}} = 97.50 \text{ m}$, and

$$\Delta t = \frac{L_{A \text{ in B}}}{v_{\text{rel}}} = \frac{\Delta t_B}{\gamma} = 1.46 \times 10^{-6} \text{ s}.$$

This results does not, and should not, agree with Δt_B .

⁹Adapted from Kogut problem 2-2.

2.11.3 Emission and absorption¹⁰

Q: The emission and absorption of a light ray define two distinct spacetime events, which are separated by a distance ℓ in the common rest frame of the emitter and the absorber. Find 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.

A: Three different methods for solving the problem, each of which are instructive, are presented:

Method #1: Thought-experiment

Choose the x^1 direction to coincide with the direction of the light ray. In the original frame, the light ray travels a distance $\Delta x^1 = \ell$ in a time $\Delta t = \ell/c$. Now consider the emission and 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) $\Delta t'$, so in this frame the light ray travels a distance $c \Delta t'$. Since the distance between the emission and absorption locations equals ℓ in the original frame, that separation is now ℓ/γ in the boosted frame due to Lorentz contraction. But it is *essential* to realize that while the emission and absorption locations are fixed in the original frame, they are moving in the boosted frame. In particular, the location of the absorption event moves a distance $-v \Delta t'$ while the light is traveling, which must be added to ℓ/γ to obtain the net distance traveled by the light in this frame. Therefore, $c \Delta t' = \ell/\gamma - v \Delta t'$. Solve for $c \Delta t'$:

$$c \Delta t' = \frac{\ell/\gamma}{1 + v/c} = \ell \sqrt{\frac{1 - v/c}{1 + v/c}}.$$

This is the distance between emission and absorption events in the boosted frame; the time between these events (in the boosted frame) is just $\Delta t' = \frac{\ell}{c} \sqrt{\frac{1-v/c}{1+v/c}}$, since the speed of light is frame-independent. Notice that this result is not just given by time dilation. For positive v , the time interval between emission and absorption in the boosted frame is less than in the original frame, while for negative v , the boosted frame time interval is greater.¹¹

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. Under a boost, the origin is mapped to the origin so the emission event also occurs at the origin of

¹⁰Adapted from Kogut problem 2-6.

¹¹This result allows us to make a nice connection with the discussion in Chapter 1. Suppose that the light ray, instead of being absorbed, is reflected back and detected at the emitter. The total time interval between emission and detection (in the original frame), $\Delta t_{\text{tot}} = 2\ell/c$, is just the time between ticks of the clock discussed in Chapter 1. Taking into account the different direction of motion of the light after reflection, the total time interval as observed in the moving frame (in the configuration of Figure 1.4) is

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

in agreement with the original time dilation result.

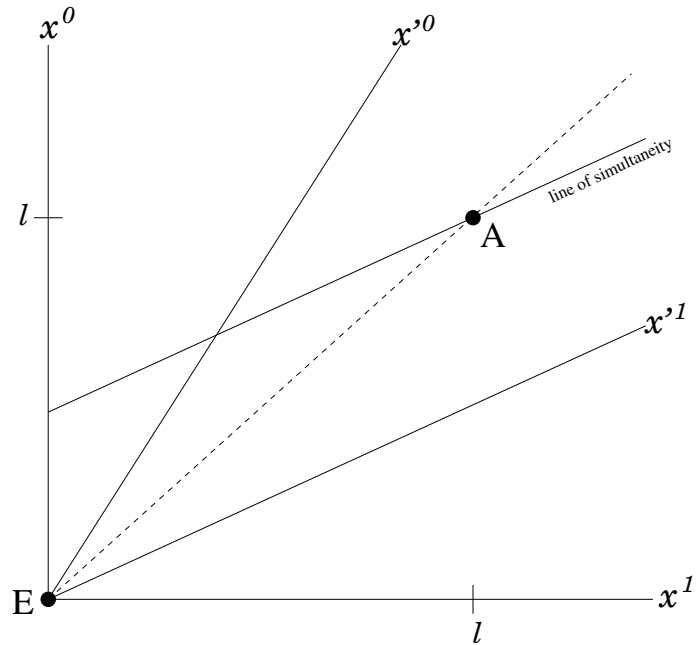
the boosted frame (since we assumed that this was the synchronizing event). The absorption event coordinates, in the boosted frame, are given by

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

The spatial separation is given by $x'^1 = \gamma\ell(1 - v/c)$, which simplifies to the same answer given above for $c\Delta t'$, namely $\ell\sqrt{\frac{1-v/c}{1+v/c}}$. Since the events lie on the lightcone, the time separation (times c) and spatial separation are identical.

Method #3: Spacetime diagram

In the diagram we have drawn the lines of simultaneity for the boosted observer that intersect the emission and absorption events, labeled E and A, respectively. The upper line of simultaneity is described by the equation $(x^0 - \ell)/(x^1 - \ell) = v/c$ which, written in more familiar slope-intercept form, is $x^0 = (v/c)x^1 + \ell(1 - v/c)$. The x^0 -intercept is $\ell(1 - v/c)$ and, as you can see from the diagram, it gives the time between emission and absorption events for the boosted observer (times c). 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 by clocks in the original frame, not those in the boosted frame. But we already know how to convert time intervals between 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 $x'^0 = \gamma x^0 = \gamma\ell(1 - v/c) = \ell\sqrt{(1 - v/c)/(1 + v/c)}$.



2.11.4 Changing frame (I)¹²

Q: An event has coordinates $(x')^\mu = (c \times 9 \times 10^{-8} \text{ s}, 100 \text{ m}, 0, 0)$ in frame S' . Frame S' moves with velocity $v/c = 4/5$ along the x^1 axis with respect to the S frame. Determine the location of the event in frame S .

A: Assume, for convenience, that the spacetime origins of the two frames coincide. The boost factor relating the frames is $\gamma = 1/\sqrt{1 - (v/c)^2} = 5/3$, and hence the relevant Lorentz transformation is:

$$x = \Lambda(v) x' = \begin{pmatrix} 5/3 & 4/3 & 0 & 0 \\ 4/3 & 5/3 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 27 \text{ m} \\ 100 \text{ m} \\ 0 \\ 0 \end{pmatrix} = \begin{pmatrix} 178.3 \text{ m} \\ 202.7 \text{ m} \\ 0 \\ 0 \end{pmatrix} = \begin{pmatrix} c \times 59.4 \times 10^{-8} \text{ s} \\ 202.7 \text{ m} \\ 0 \\ 0 \end{pmatrix}.$$

¹²Adapted from Kogut problem 4-3.

2.11.5 Changing frame (II)¹³

Q: Two events have coordinates $(x_1)^\mu = (L, L, 0, 0)$ and $(x_2)^\mu = (L/2, 2L, 0, 0)$ in frame S . The two events are simultaneous in frame S' . Find the velocity \vec{v} of frame S' as seen from frame S . Assume the spacetime origins of both frames coincide. When do these events occur in frame S' ?

A: We have $\Delta x^0 = L/2$ and $\Delta x^1 = -L$ (with $\Delta x \equiv x_1 - x_2$). Coordinates in frame S' will be related to those in frame S by some boost in the x^1 direction with velocity $-\vec{v} = -v \hat{e}_1$, $x' = \Lambda(-v)x$. (It is $\Lambda(-v) = \Lambda(v)^{-1}$ since we have interchanged x and x' relative to Eq. 2.5.1.) Using the explicit form (2.5.7), with $v \rightarrow -v$, we have $\Delta x'^0 = \gamma(\Delta x^0 - \frac{v}{c} \Delta x^1) = \gamma L(\frac{1}{2} + \frac{v}{c})$. For this to vanish, we must have $v/c = -1/2$, implying that frame S' moves with velocity $\vec{v} = -(c/2) \hat{e}_1$ as seen in frame S . The common time of the two events in the S' frame is $t' = \gamma(ct_1 - \frac{v}{c} x_1)/c = \frac{2}{\sqrt{3}}(L + \frac{1}{2}L)/c = \sqrt{3}L/c$. As a check, the same result is obtained using the second event's coordinates.

¹³Adapted from Kogut problem 4-4.

Chapter 3

Relativistic dynamics

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

$$\vec{f} = m \vec{a}, \quad (3.0.1)$$

and acceleration is the second time derivative of position,

$$\vec{a} = \frac{d\vec{v}}{dt} = \frac{d^2\vec{x}}{dt^2}. \quad (3.0.2)$$

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 this regime the relation (3.0.1) 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 (3.0.1) and (3.0.2) predict that a particle subject to a constant force (and initially at rest) will acquire a velocity which can become arbitrarily large,

$$\vec{v}(t) = \int_0^t \frac{d\vec{v}}{dt'} dt' = \frac{\vec{f}}{m} t \longrightarrow \infty \quad \text{as } t \rightarrow \infty. \quad (3.0.3)$$

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).

3.1 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 3.1. 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) 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.

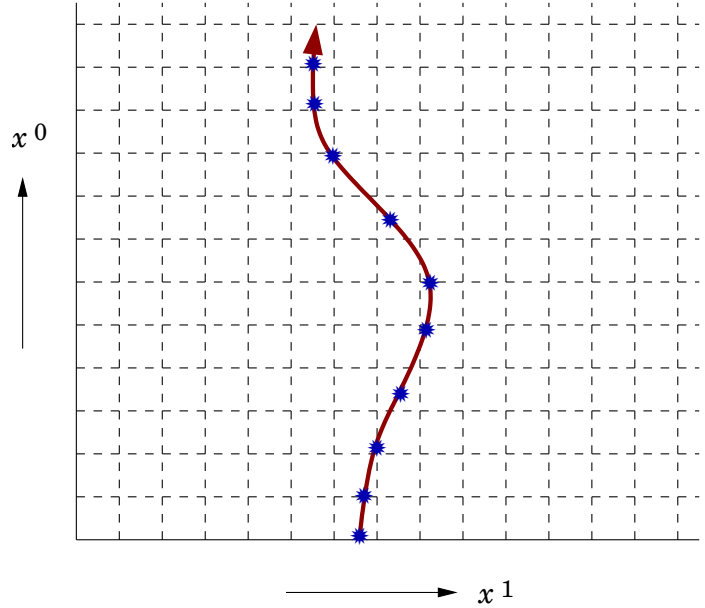


Figure 3.1: The worldline of a non-inertial particle, with tick marks at constant intervals of proper time.

3.2 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,

$$ct = x^0(\tau). \quad (3.2.1)$$

The *four-velocity* of a particle is the derivative of its spacetime position with respect to proper time,

$$u \equiv \frac{dx(\tau)}{d\tau}. \quad (3.2.2)$$

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

$$u^0 = c \frac{dt}{d\tau}. \quad (3.2.3)$$

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,

$$u^k = \frac{dx^k}{d\tau} = \frac{dt}{d\tau} \frac{dx^k}{dt} = \frac{u^0}{c} v^k. \quad (3.2.4)$$

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

$$\frac{u^0}{c} = \frac{dt}{d\tau} = \gamma = \left[1 - \frac{\vec{v}^2}{c^2}\right]^{-1/2}. \quad (3.2.5)$$

Combined with Eq. (3.2.4), this shows that the spatial components of the 4-velocity equal the three-velocity times a factor of γ ,

$$u^k = \gamma v^k = \frac{v^k}{\sqrt{1 - \vec{v}^2/c^2}}. \quad (3.2.6)$$

We can now use Eqs. (3.2.5) and (3.2.6) to evaluate the square of the 4-velocity,

$$u^2 = -(u^0)^2 + (u^k)^2 = -\gamma^2 (c^2 - \vec{v}^2) = -c^2. \quad (3.2.7)$$

So a four-velocity vector *always* squares to $-c^2$, regardless of the value of the 3-velocity.

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 vector which has unit "length" everywhere along the worldline. The fact that u^2 is negative shows that the 4-velocity is always a timelike vector.

Having picked a specific reference frame in which to evaluate the components of the four-velocity u , Eqs. (3.2.5) and (3.2.6) 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 four-velocity u is a more natural quantity to use — it has geometric meaning which is *independent* of any choice of reference frame. Moreover, the components u^μ of four-velocity transform linearly under a Lorentz boost in exactly the same fashion as any other 4-vector. [See Eq. (2.7.5)]. In contrast, under a Lorentz boost the components of 3-velocity transform in a somewhat messy fashion. (Example problem 3.10.1 below works out the precise form for the case of parallel velocities.)

3.3 4-momentum

The *rest mass* of any object, generally denoted m , is the mass of the object as measured in its rest frame. The *four-momentum* of a particle (or any other object) with rest mass m is defined to be m times the object's four-velocity,

$$p = m u. \quad (3.3.1)$$

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 momentum which you learned in introductory physics, $\vec{p} = m \vec{v}$, is wrong — this is a non-relativistic approximation which is only useful for slowly moving objects. This is important, so let us repeat,

$$\vec{p} \neq m \vec{v}. \quad (3.3.2)$$

Momentum is *not* mass times 3-velocity. Rather, momentum is mass times 4-velocity.¹

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 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!

If the spatial components of the four-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

$$p^0 = E/c. \quad (3.3.3)$$

Using the relation (3.3.1) between 4-momentum and 4-velocity, plus the result (3.2.5) for u^0 , allows one to express the the total energy E of an object in terms of its rest mass and its velocity,

$$E = cp^0 = mcu^0 = mc^2\gamma = \frac{mc^2}{\sqrt{1 - \vec{v}^2/c^2}} = mc^2 \cosh \eta, \quad (3.3.4)$$

where the last form uses the relation (2.6.2) between rapidity and γ . In other words, the relativistic gamma factor of any object is equal to the ratio of its total energy to its rest energy,

$$\gamma = \frac{E}{mc^2}. \quad (3.3.5)$$

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 (3.3.4) in powers of \vec{v}^2/c^2 . This gives

$$E = mc^2 + \frac{1}{2}m\vec{v}^2 + \dots, \quad (3.3.6)$$

and shows that 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 kinetic energy, are suppressed by additional powers of \vec{v}^2/c^2 . One can define a relativistic *kinetic energy* K , as simply the difference between the total energy and the rest energy, $K = E - mc^2$.

Combining the 4-momentum definition (3.3.1), and the relation (3.2.4) between three- and four-velocity components, yields the relation between the spatial components of the relativistic momentum and the 3-velocity,

$$\vec{p} = m\vec{v}\gamma = \frac{m\vec{v}}{\sqrt{1 - \vec{v}^2/c^2}} = m\hat{v} \sinh \eta, \quad (3.3.7)$$

where the last form uses rapidity and a unit spatial vector \hat{v} pointing in the direction of the 3-velocity. Expanding in powers of v/c shows that, for low velocities, the relativistic spatial momentum reduces to the non-relativistic form,

$$\vec{p} = m\vec{v} + \dots, \quad (3.3.8)$$

up to higher order corrections suppressed by powers of \vec{v}^2/c^2 .

We saw above that four-velocities square to $-c^2$. Because four-momentum is just mass times four-velocity, the four-momentum of any object with mass m satisfies

$$p^2 = -m^2c^2. \quad (3.3.9)$$

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

$$E^2 = c^2 \vec{p}^2 + (mc^2)^2, \quad (3.3.10a)$$

or

$$E = \sqrt{c^2 \vec{p}^2 + (mc^2)^2}. \quad (3.3.10b)$$

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.

But what if you want to find the ordinary 3-velocity? Return to the relation $u^k = \gamma v^k$ [Eq. (3.2.6)] 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 $E = m c^2 \gamma$, we have $p^k = (E/c^2) v^k$ or

$$v^k = \frac{p^k}{E/c^2}. \quad (3.3.11)$$

Three-velocity is *not* equal to momentum divided by mass — forget this falsehood! Rather, the ordinary 3-velocity equals the spatial momentum divided by the total energy (over c^2). And its magnitude never exceeds c , no matter how large the momentum (and energy) become.

3.4 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 four-momentum and proper time. The appropriate generalization of Newtonian dynamics which is consistent with our relativity postulates is

$$\frac{dx}{d\tau} = \frac{p}{m}, \quad (3.4.1a)$$

$$\frac{dp}{d\tau} = f. \quad (3.4.1b)$$

Eq. (3.4.1a) is just the definition (3.2.2) of 4-velocity rewritten in terms of 4-momentum, while Eq. (3.4.1b) 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 (3.4.1) 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 *four-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

³Eq. (3.4.1b) 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 $dp/d\tau = f$.

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 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 (3.4.1), 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. (3.3.9)]. 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 four-force must always be orthogonal to the four-momentum,

$$p \cdot f = 0. \quad (3.4.2)$$

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

$$f^0 = \frac{p^k f^k}{p^0} = \frac{\vec{v}}{c} \cdot \vec{f}, \quad (3.4.3)$$

showing that the time component of the 4-force is completely determined by the spatial force components (and the 3-velocity).

3.5 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 four-force must always be orthogonal to the momentum. So it is impossible for the 4-force $f(\tau)$ to be a fixed four-vector, independent of τ . However, it is possible for the force to have components which are 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 force of magnitude F , pointing in the x^1 direction, acts on the particle. So 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 four-velocity at later times may be written as some time-dependent Lorentz boost acting on the initial four-velocity,

$$u(\tau) = \Lambda_{\text{boost}}(\tau) u_0. \quad (3.5.1)$$

The condition that the force is constant in a co-moving frame amounts to the statement that the *same* Lorentz boost relates the four-force at any time τ to the initial force,

$$f(\tau) = \Lambda_{\text{boost}}(\tau) f_0. \quad (3.5.2)$$

At all times, $u^2 = -c^2$ (because u is a four-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

$$u^\mu(\tau) = (u^0(\tau), u^1(\tau), 0, 0), \quad f^\mu(\tau) = (f^0(\tau), f^1(\tau), 0, 0), \quad (3.5.3)$$

with $u^0(0) = c$, $u^1(0) = 0$ and $f^0(0) = 0$, $f^1(0) = F$. 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 force must be given by

$$f^\mu(\tau) = \frac{F}{c} (u^1(\tau), u^0(\tau), 0, 0). \quad (3.5.4)$$

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

$$\frac{du^0(\tau)}{d\tau} = \frac{F}{mc} u^1(\tau), \quad \frac{du^1(\tau)}{d\tau} = \frac{F}{mc} u^0(\tau). \quad (3.5.5)$$

This is easy to solve if you remember that $\frac{d}{dz} \sinh z = \cosh z$ and $\frac{d}{dz} \cosh z = \sinh z$. To satisfy Eq. (3.5.5), and our initial conditions, we need

$$u^0(\tau) = c \cosh \frac{F\tau}{mc}, \quad u^1(\tau) = c \sinh \frac{F\tau}{mc}. \quad (3.5.6)$$

The ordinary velocity $v^k = u^k (c/u^0)$ [Eq. (3.2.4)], so the speed of this particle subject to a constant force is

$$v(\tau) = c \tanh \frac{F\tau}{mc}. \quad (3.5.7)$$

Since $\tanh z \sim z$ for small values of the argument, 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. From the definition (2.6.1) of rapidity, $v/c = \tanh \eta$, one sees that the result (3.5.7) for the speed just corresponds to rapidity growing linearly with proper time,

$$\eta(\tau) = \frac{F\tau}{mc}. \quad (3.5.8)$$

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. The integrals are elementary,

$$x^0(\tau) = \int_0^\tau d\tau' u^0(\tau') = c \int_0^\tau d\tau' \cosh \frac{F\tau'}{mc} = \frac{mc^2}{F} \sinh \frac{F\tau}{mc}, \quad (3.5.9a)$$

$$x^1(\tau) = \int_0^\tau d\tau' u^1(\tau') = c \int_0^\tau d\tau' \sinh \frac{F\tau'}{mc} = \frac{mc^2}{F} \left[\cosh \frac{F\tau}{mc} - 1 \right]. \quad (3.5.9b)$$

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 $2e^{-F\tau/mc} \sim mc/(Ft)$.

3.6 Plane waves

Next, we want to discuss how waves (of any type) may be described using relativistic notation. Consider some plane wave with spatial wave-vector \vec{k} and (angular) frequency ω , as measured in some inertial frame. The amplitude of the wave may be 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 which move at speed $v = \omega/|\vec{k}|$ in the direction of \vec{k} .

As mentioned earlier (2.7.12), it is natural to combine ω and \vec{k} into a spacetime wave-vector k with components

$$k^\mu = (\omega/c, k^1, k^2, k^3), \quad (3.6.1)$$

so that $\omega = ck^0$ and the complex exponential $e^{i\vec{k}\cdot\vec{x}-i\omega t} = e^{ik\cdot x}$ only involves a spacetime dot product. 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 event 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.

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). 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,

$$\omega_{\text{obs}} = -u_{\text{obs}} \cdot k. \quad (3.6.2)$$

This expression is now written in a completely general fashion which is observer-dependent but frame-independent. That is, the expression (3.6.2) depends explicitly on the observer's 4-velocity u_{obs} , but is independent of the frame used to evaluate the dot product between u_{obs} and k . Therefore, the frequency which is measured by *any* observer will be given by (minus) the dot product of the observer's 4-velocity u and the wave-vector k . Once again, this dot product may be evaluated using whatever reference frame is most convenient.

For light waves (in a vacuum), the wave speed $v = c$ and $\omega = c|\vec{k}|$. The resulting spacetime wavevector (3.6.1) is automatically a lightlike 4-vector which squares to zero,

$$k_{\text{light}}^\mu = \frac{\omega}{c} (1, \hat{k}), \quad k_{\text{light}}^2 = 0. \quad (3.6.3)$$

A nice application of Eq. (3.6.2), demonstrating the value of writing physical quantities in frame independent form, is illustrated in Figure 3.2.⁴ 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 ?

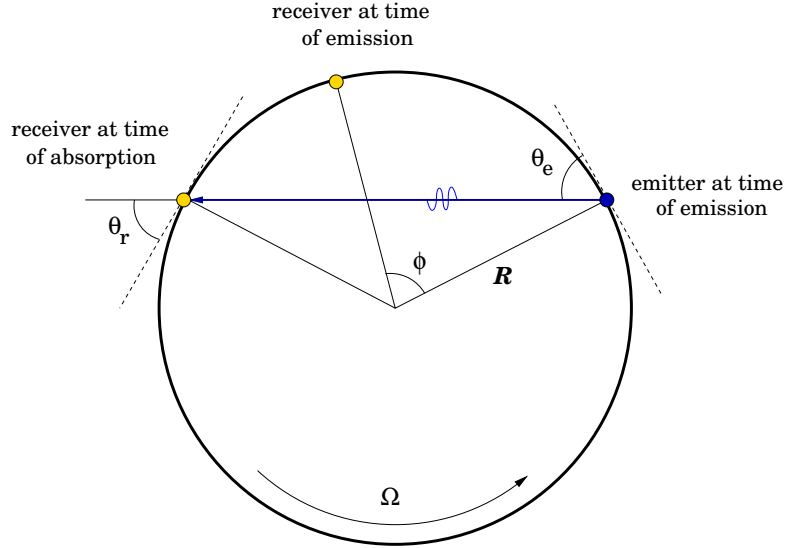


Figure 3.2: Inside a rotating centrifuge, light is emitted at one point and later received at another point. Is there a Doppler shift between the frequencies of emission and reception?

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 emitter and receiver. Given the three-dimensional geometry involved, this is rather involved.

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 (3.6.2) for the frequency. We need to compute

$$\frac{\nu_r}{\nu_e} = \frac{-u_r \cdot k}{-u_e \cdot k} = \frac{u_r^0 k^0 - \vec{u}_r \cdot \vec{k}}{u_e^0 k^0 - \vec{u}_e \cdot \vec{k}}. \quad (3.6.4)$$

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), then we can express the spatial dot products in terms of cosines of these angles,

$$\frac{\nu_r}{\nu_e} = \frac{u_r^0 k^0 - |\vec{u}_r| |\vec{k}| \cos \theta_r}{u_e^0 k^0 - |\vec{u}_e| |\vec{k}| \cos \theta_e}. \quad (3.6.5)$$

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

⁴This discussion is an adaptation of an example in *Gravitation* by Misner, Thorne and Wheeler.

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 (3.6.5) 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!

3.7 Electromagnetism

As already seen in the discussion of lightcones, plane waves, and Doppler shifts, the techniques we are developing are particularly useful for understanding the propagation of light. Unfortunately, we do not have time for extensive explorations of other relativistic aspects of electromagnetism, which will be left for later 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. 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 called the *field strength tensor*, whose components are⁵

$$\|F^\mu{}_\nu\| = \begin{pmatrix} 0 & E_x & E_y & E_z \\ E_x & 0 & cB_z & -cB_y \\ E_y & -cB_z & 0 & cB_x \\ E_z & cB_y & -cB_x & 0 \end{pmatrix}. \quad (3.7.1)$$

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

$$f_{\text{Lorentz}}^\mu = \frac{q}{c} F^\mu{}_\nu u^\nu. \quad (3.7.2)$$

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.

3.8 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. Interestingly, dramatic changes during collisions can also occur for elementary particles. Studying

⁵The explicit form of the field strength tensor depends on the choice of units one uses for electric and magnetic fields. Expression (3.7.1) is applicable with SI units, where $|\vec{E}|$ is measured in Newtons/Coulomb, and $|\vec{B}|$ in Tesla. One can (and should) check that (3.7.1) is dimensionally consistent — using SI quantities, multiplying by c converts the units for B into the units for E .

the collisions of elementary particles is a primary method used to investigate fundamental interactions (and is the reason for building large high-energy particle colliders such as the LHC).

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 3.3, the total energy E and spatial momentum \vec{p} of any object may be combined to form the four-momentum $p^\mu = (E/c, \vec{p})$. Consequently, conservation of energy plus conservation of (spatial) momentum may be compactly rephrased as the conservation of four-momentum: in the absence of any external forces, the total four-momentum p_{tot} of any system cannot change,

$$\frac{d}{dt} p_{\text{tot}}(t) = 0. \quad (3.8.1)$$

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 four-momentum is just the sum of the four-momentum of each object,

$$p_{\text{in}} = \sum_{a=1}^{N_{\text{in}}} p_a, \quad (3.8.2)$$

(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 four-momentum is the sum of the four-momenta of all outgoing objects,

$$p_{\text{out}} = \sum_{b=1}^{N_{\text{out}}} p_b. \quad (3.8.3)$$

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

$$p_{\text{in}} = p_{\text{out}}. \quad (3.8.4)$$

As with any four-vector equation, one may choose to write out the components of this equation in whatever reference frame is most convenient. 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*. Experiments of this variety are known as “fixed target” experiments; the rest-frame of the actual laboratory 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*,⁶ the components of the total four-momentum are

$$p_{\text{CM}}^\mu = (E_{\text{CM}}/c, 0, 0, 0), \quad (3.8.5)$$

where E_{CM} is the total energy of the system in the CM frame.

⁶CM 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.

As an application of these ideas, consider 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. (3.3.5), 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,

$$p + p \rightarrow X,$$

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 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 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 four-vector conservation equation (3.8.4). In the lab frame, if we orient coordinates so that the z -axis is the collision axis, then

$$p_{\text{in}} = p_{\text{projectile}} + p_{\text{target}} = \begin{pmatrix} E_{\text{in}}/c \\ 0 \\ 0 \\ p_{\text{in}} \end{pmatrix} + \begin{pmatrix} m_p c \\ 0 \\ 0 \\ 0 \end{pmatrix}. \quad (3.8.6)$$

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

$$p_{\text{out}} = p_X = \begin{pmatrix} E_X \\ p_X^1 \\ p_X^2 \\ p_X^3 \end{pmatrix}. \quad (3.8.7)$$

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. (3.3.10), applied to the projectile proton (with known mass), may be used to relate the incident spatial momentum and energy, $p_{\text{in}}^2 = (E_{\text{in}}/c)^2 + (m_p c)^2$. The same relation (3.3.10), 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. (3.3.9)] that the square of any four-momentum directly gives the rest mass of the object, $p^2 = -m^2 c^2$. Hence

$$\begin{aligned} -m_X^2 c^2 &= p_X^2 = p_{\text{out}}^2 = p_{\text{in}}^2 = (p_{\text{projectile}} + p_{\text{target}})^2 \\ &= p_{\text{projectile}}^2 + p_{\text{target}}^2 + 2 p_{\text{projectile}} \cdot p_{\text{target}} \\ &= -2 m_p^2 c^2 - 2 E_{\text{in}} m_p. \end{aligned} \quad (3.8.8)$$

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.

Most of the energy of the projectile is needed to provide the kinetic energy of the outgoing particle X , which is necessarily associated with the conserved spatial momentum. More generally, the maximum mass grows (only) like the square root of the lab frame energy, $m_X^{\max} \sim \sqrt{2E_{\text{in}}m_p/c^2}$, when $E_{\text{in}} \gg m_p c^2$.

This illustrates why “colliders” in which two beams of particles are aimed at each other, so that the lab and CM frames coincide, are more efficient when hunting for new heavy particles. If the colliding particles have equal mass, then they will also have equal energy (E_{in}) when their spatial momenta are equal and opposite. In this case, the total spatial momentum vanishes and the maximum mass of a produced particle is limited only by the total energy, $m_X^{\max} = 2E_{\text{in}}/c^2$, which grows *linearly* with the beam energy E_{in} .

3.9 Units and sizes

It may be helpful at this point to say a few words about units and the size of things. For “dimensionfull” quantities (*i.e.*, quantities which are not pure numbers and whose measurement requires some standard for comparison), the value of the quantity depends on one’s choice of units. It only makes sense to say that a dimensionfull quantity is “large” or “small” in comparison to some other quantity with the *same* units. For velocities, the universal value of the speed of light makes c the natural standard for comparison; an object is moving slowly (and non-relativistic dynamics can be a good approximation) if its speed is small *compared* to the speed of light, $|v|/c \ll 1$. Similarly, classical mechanics can provide a good approximation when quantum interference effects produced by a wave function such as (2.7.13) vary so rapidly that they become unresolvable. This is the case when $p \cdot x$ is large *compared* to Planck’s constant \hbar .

In the SI (or MKS) system, there are *three* independent fundamental units, length (m), mass (kg), and time (s). These units are convenient for describing many phenomena which occur on human scales. But they are *not* convenient for describing atomic, nuclear, or particle physics phenomena. For example, the mass of a proton is 1.67×10^{-27} kg, and the size of a proton is conveniently measured in fermi, not meters. A fermi (fm) is shorthand for one femtometer, $1 \text{ fm} = 1 \text{ femtometer} = 10^{-15} \text{ m}$. Likewise, the lifetime of a typical particle that decays via the strong interactions (discussed in chapter 4) is of order 10^{-23} s, roughly the time needed for light to travel across a distance of 1 fm.

As most of the physics we will discuss in this course is both relativistic and quantum mechanical, it will often be convenient to use units in which the speed of light and Planck’s constant \hbar have numerical values close to unity. In fact, one is free to choose “natural” units in which both c and \hbar are *exactly* equal to unity. By declaring that

$$c = 2.99\,792\,458 \times 10^8 \text{ m/s} = 1, \quad (3.9.1)$$

one is choosing to regard time and distance as having the *same* units; one second is the *same* as $2.99 \cdots \times 10^8$ meters. As a measure of distance, one second means one “lightsecond,” the distance light travels in a second. The speed of light, when expressed in m/s, is just a conversion factor between two different units for distance, meters and seconds, in the same way that $1 = 2.54 \text{ cm/in}$ or $1 = 6 \text{ ft/fathom}$ are conversion factors relating other measures of distance.

Similarly, by declaring that

$$\hbar = 1.05\,457\,148 \cdots \times 10^{-34} \text{ J s} = 1, \quad (3.9.2)$$

1 kg	=	5.61×10^{26} GeV	[GeV/ c^2]
1 m	=	5.07×10^{15} GeV $^{-1}$	[$\hbar c$ /GeV]
1 s	=	1.52×10^{24} GeV $^{-1}$	[$\hbar c$ /GeV]
1 fm $\equiv 10^{-15}$ m	=	5.07 GeV $^{-1}$	[$\hbar c$ /GeV]
$(1 \text{ fm})^2 = 10 \text{ mb} = 10^{-30} \text{ m}^2$	=	25.7 GeV $^{-2}$	[($\hbar c$ /GeV) 2]
$\hbar c$	=	197 MeV fm	
$(\hbar c)^2$	=	0.389 GeV 2 mb	

Table 3.1: Useful approximate conversion factors. The last column shows the appropriate units with \hbar and c included.

one is choosing to regard energy and frequency (inverse time) as having the same units. Since quantum states with energy E evolve in time with an amplitude $e^{-iEt/\hbar}$, one sees that their frequency of oscillation is always directly related to their energy by a factor of Planck’s constant, $\omega = E/\hbar$. This relation applies to photons, electrons, and any other particle. So it is natural to regard Planck’s constant \hbar , expressed in J/s (or any other traditional units), as just a conversion factor between two different measures for energy (or frequency).

With c set equal to unity, time has the same dimensions as distance. Moreover, mass, momentum and energy all have the same units (since factors of c can convert one to the other). With \hbar also set equal to unity, mass and energy have the same units as 1/distance or 1/time. The net result is that there is only one fundamental independent dimension, say energy, which requires a choice of unit. We could use Joules, ergs, or any other measure of energy, but it will be most convenient to choose a unit which is comparable to energy scales relevant for particle physics — such as the proton’s rest energy, $m_p c^2$. This is about 1.5×10^{-10} J, showing that Joules are not a very nice choice for our purposes. It is preferable, and conventional, to instead use SI-prefixed (*e.g.*, kilo-, mega-, giga-, ...) electron volts, namely keV = 10^3 eV, MeV = 10^6 eV, GeV = 10^9 eV, TeV = 10^{12} eV, *etc.* As the proton rest energy is very close to one GeV, $m_p c^2 = 0.938$ GeV, giga-electron volts (GeV) will be especially convenient.

As noted above, the fermi is a useful measure for lengths in particle physics applications. A convenient conversion factor is $1 = 197 \text{ MeV fm}$, (or $\hbar c = 197 \text{ MeV fm}$ with \hbar and c retained), so $1 \text{ fm} \approx 1/(0.2 \text{ GeV})$. For measuring areas (*e.g.*, cross sections for scattering), the “barn”, defined as 10^{-28} m^2 , is commonly used in nuclear physics. For particle physics applications, millibarn ($\text{mb} = 10^{-31} \text{ m}^2$), microbarn ($\mu\text{b} = 10^{-34} \text{ m}^2$), or nanobarn ($\text{nb} = 10^{-37} \text{ m}^2$) are generally more convenient. One square fermi is 10 millibarn. Table 3.1 lists a number of conversion factors relating traditional and natural particle physics units. In these notes, we will initially retain explicit factors of c and \hbar , but you should gradually become comfortable using natural units with $c = \hbar = 1$.

As a final illustration of the relation between different units, Table 3.2 compares the sizes, in both meters and GeV $^{-1}$, of a wide variety of objects. Note the huge range of sizes that characterize our universe. The last quantity listed, the Planck length, is the length scale, or inverse mass scale, where quantum fluctuations in the geometry of spacetime (*i.e.*, quantum gravity effects) are expected to become significant.

observable universe	$\sim 10^{26}$ m	$\approx 5 \times 10^{41}$ GeV ⁻¹	($\sim 10^{11}$ galaxies)
galaxy supercluster	$\sim 10^{24}$ m	$\approx 5 \times 10^{39}$ GeV ⁻¹	
galaxy	$\sim 10^{21}$ m	$\approx 5 \times 10^{36}$ GeV ⁻¹	($\sim 10^{11}$ stars)
star	$\sim 10^9$ m	$\approx 5 \times 10^{24}$ GeV ⁻¹	
Earth	$\sim 10^7$ m	$\approx 5 \times 10^{22}$ GeV ⁻¹	
human	$\sim 10^0$ m	$\approx 5 \times 10^{15}$ GeV ⁻¹	
atom	$\sim 10^{-10}$ m	$\approx 5 \times 10^5$ GeV ⁻¹	
nucleus	$\sim 10^{-14}$ m	$\approx 5 \times 10^1$ GeV ⁻¹	
proton	$\sim 10^{-15}$ m	$\approx 5 \times 10^0$ GeV ⁻¹	
present observational limit	$\sim 10^{-19}$ m	$\approx 5 \times 10^{-4}$ GeV ⁻¹	
Planck length	$\sim 10^{-35}$ m	$\approx 5 \times 10^{-20}$ GeV ⁻¹	

Table 3.2: Characteristic sizes of various objects (to within factors of 2–3).

3.10 Example problems

3.10.1 Relativistic velocity addition

Q: Frame S' moves in the x^1 direction with velocity v_0 relative to frame S . A point particle moves with velocity v' in the x^1 direction as seen in frame S' . Find the 3-velocity of the particle in frame S .

A: In frame S' , the components of the 4-velocity of the particle are $(u')^\mu = (\gamma_{v'} c, \gamma_{v'} v', 0, 0)$, with $\gamma_{v'} = (1 - v'^2/c^2)^{-1/2}$. Since the S' frame is moving, relative to frame S , by velocity v_0 in the x^1 direction, 4-vector components in frame S are related to those in frame S' by the Lorentz transformation matrix

$$\Lambda(v_0) = \begin{pmatrix} \gamma_0 & \gamma_0 (v_0/c) & 0 & 0 \\ \gamma_0 (v_0/c) & \gamma_0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix},$$

with $\gamma_0 = (1 - v_0^2/c^2)^{-1/2}$. Applying this matrix to the components $(u')^\mu$ yields the components u^μ of the particle's 4-velocity in frame S ,

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

The ordinary 3-velocity is related to the 4-velocity via the relation (3.2.4), or $v^k = u^k/(u^0/c)$. Inserting the S frame components u^μ yields a 3-velocity (pointing in the 1 direction),

$$v = v^1 = \frac{v_0 + v'}{1 + v_0 v'/c^2}, \quad [v^2 = v^3 = 0].$$

The numerator of this answer is the familiar Galilean result, but the denominator reflects relativistic corrections. If either v_0 or v' are small compared to c , then the denominator is close to the 1 and the Galilean result is approximately correct. But if either (or both) of the initial three-velocities approach c , then the final velocity v also approaches, but never exceeds, c .

The reader is encouraged to: (a) verify that the result for u satisfies $u^2 = -c^2$, (b) show that the results for u^0 and v satisfy the usual relation $u^0 = \gamma_v c$ with $\gamma_v = (1 - v^2/c^2)^{-1/2}$, and (c) show that

if velocities are described by equivalent rapidities, $v_0 = c \tanh \eta_0$, $v' = c \tanh \eta'$, and $v = c \tanh \eta$, then $\eta = \eta_0 + \eta'$. In other words, rapidities (of collinear boosts) add linearly, but 3-velocities do not.

3.10.2 Doppler shift

Q: Using Eqs. (3.6.2) and (3.6.3), derive the relativistic Doppler shift of light — find the frequency seen by an observer moving *away* from a source of light with (ordinary) frequency ν_0 , as measured in the rest frame of the source.

A: Take the observer to be moving in the x^1 direction with velocity v . The most convenient reference frame is the rest frame of the source (since this is the frame in which we have information about both the light and the observer). In the source frame, the observer's 4-velocity has components $u_{\text{obs}}^\mu = c\gamma(1, v/c, 0, 0)$. The angular frequency of the light is $\omega_0 = 2\pi\nu_0$, and the spacetime wavevector (for the light moving in the x^1 direction which reaches the observer) has components $k^\mu = (\omega_0/c)(1, 1, 0, 0)$. Using (3.6.2), we have

$$\frac{\nu_{\text{obs}}}{\nu_0} = \frac{\omega_{\text{obs}}}{\omega_0} = -\frac{u_{\text{obs}} \cdot k}{\omega_0} = \gamma(1 - v/c) = \sqrt{\frac{1 - v/c}{1 + v/c}}.$$

For $v > 0$, corresponding to the source and observer receding from each other, we have $\nu_{\text{obs}}/\nu_0 < 1$, so the light appears to be red-shifted. For an observer approaching the source, simply change the sign of v ; in this case $\nu_{\text{obs}}/\nu_0 > 1$ and the light appears blue-shifted to a higher frequency.

3.10.3 Kinetic energy, speed, and momentum⁷

Q: A relativistic particle has kinetic energy equal to twice its rest energy. Find the speed of the particle (relative to c) and its spatial momentum.

A: Total energy is kinetic energy plus rest energy, $E = K + mc^2 = 3mc^2$. Total energy is also γ times rest energy, so $\gamma \equiv (1 - v^2/c^2)^{-1/2} = E/(mc^2) = 3$. Solving for v/c gives $v/c = \sqrt{1 - \frac{1}{9}} = 0.943$. The (magnitude of the) particle's spatial momentum is $p = \gamma mv = 3mc(v/c) = 2.83mc$. This could also be evaluated directly using Eq. (3.3.10), which may be rearranged as $p = \sqrt{(E/c)^2 - (mc)^2} = \sqrt{8}mc$.

3.10.4 Light propulsion⁸

Q: The most fuel-efficient rocket exhaust is photons (*i.e.*, light), as this has the fastest exit velocity for any given energy. Suppose a rocket, emitting only light (in the backward direction), has initial mass M_i and final mass M_f . Find its final velocity (in the frame in which it starts from rest).

A: The hard way to do this problem is to integrate the relativistic version of Newton's equations (3.4.1) with a time-dependent mass. It is much easier to just use conservation of total energy and momentum. Working in the initial rest frame of the rocket, the total initial energy and spatial momentum are $E_{\text{tot}} = M_i c^2$ and $\vec{p}_{\text{tot}} = 0$, respectively, since the rocket is at rest. At the final time the rocket, now with mass M_f , is moving in some direction (call it $+\hat{x}$) with velocity \vec{v} , and all the emitted photons are moving in the opposite ($-\hat{x}$) direction. Hence, the total energy at the final

⁷Adapted from Kogut problem 6-11.

⁸Adapted from Kogut problem 6-16.

time is $E_{\text{tot}} = \gamma M_f c^2 + E_{\text{photons}}$, and the total final spatial momentum $\vec{p}_{\text{tot}} = \gamma M_f \vec{v} + \vec{p}_{\text{photons}} = (\gamma M_f v - E_{\text{photons}}/c) \hat{x}$. (Note that $|\vec{p}_{\text{photon}}| = E_{\text{photon}}/c$, since the 4-momentum of a photon is a light-like vector.)

Requiring that the final total spatial momentum agree with the initial value of 0 implies that $E_{\text{photons}}/c = \gamma M_f v$, while demanding that the total initial and final energies agree implies that $E_{\text{photons}}/c = M_i c - \gamma M_f c$. Equating these two results for E_{photons}/c gives

$$M_i = \gamma M_f \frac{c+v}{c} = M_f \sqrt{\frac{1+v/c}{1-v/c}}.$$

Solving for v/c yields

$$\frac{v}{c} = \frac{(M_i/M_f)^2 - 1}{(M_i/M_f)^2 + 1} = \frac{M_i^2 - M_f^2}{M_i^2 + M_f^2}.$$

So reaching a relativistic velocity, $v \approx c$, requires that the final mass (including the payload) be very much smaller than the initial mass ($M_f \ll M_i$), even with the most efficient idealized propulsion imaginable.

Chapter 4

Known particles

4.1 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?

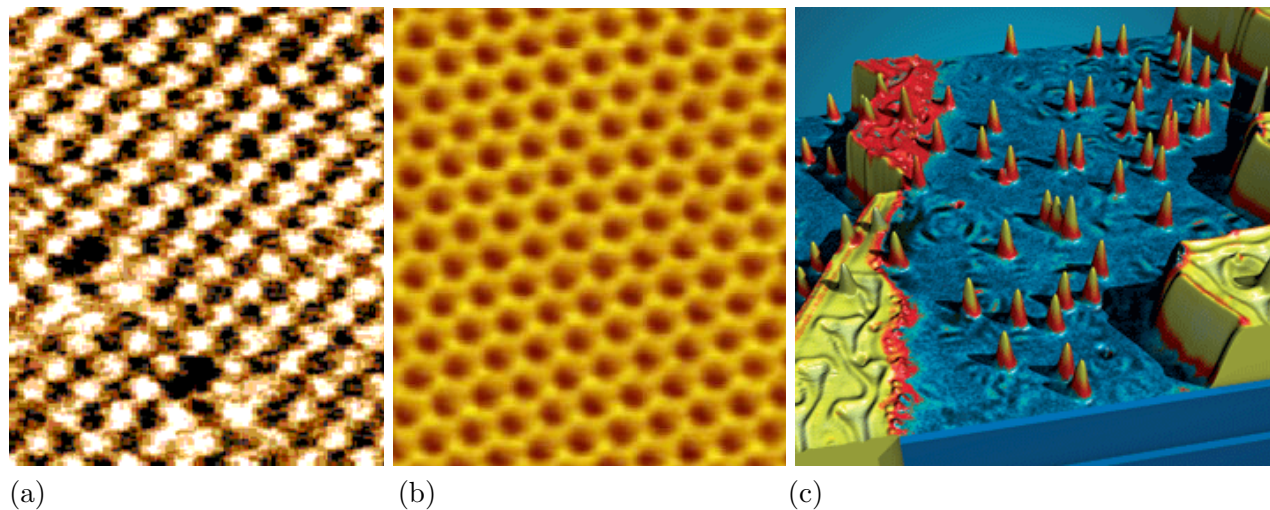


Figure 4.1: Three examples of modern atomic scale imaging. Image (a) shows the surface of sodium chloride, imaged by atomic force microscopy (AFM). Note the two surface defects. Image (b) (courtesy of E. Andrei) is a scanning tunneling microscope (STM) image of a freely suspended graphene sheet — a single 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 4.1.

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 (the precursor of televisions) when a suitably large voltage is applied. Thomson found that these particles have a mass to charge ratio which is independent of the type of material forming the cathode or the gas in the tube, and this ratio is about 2000 times smaller than the mass to charge ratio of a hydrogen ion. Measurement of the charge-to-mass ratio involves observing the deflection of a moving particle produced by a magnetic field. The charge of a single electron can be measured using the approach of Millikan and Fletcher's famous oil drop experiment. Based on refinements of such measurements, the magnitude of the electron charge is now known to a precision of a few parts in 10^8 ,

$$|-e| = 1.602\,176\,487\,(40) \times 10^{-19} \text{ C}. \quad (4.1.1)$$

(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 $6.241\,509\,65\,(15) \times 10^{18}$ electron charges. The mass of the electron is also known to a similar precision,

$$m_e = 0.510\,998\,910\,(13) \text{ MeV}/c^2 = 9.109\,382\,15(45) \times 10^{-31} \text{ kg}. \quad (4.1.2)$$

One MeV ($= 10^6$ eV) is the energy acquired by an electron passing through a potential difference of one million volts.

A few angstroms ($1 \text{ \AA} = 10^{-10} \text{ m}$) is the size of individual atoms, whereas nuclear sizes are naturally measured in units of the fermi (or femtometer), where $1 \text{ fm} = 10^{-15} \text{ m}$. (In natural units, $1 \text{ fm} \approx 5 \text{ GeV}^{-1}$ while $1 \text{ \AA} \approx 500,000 \text{ GeV}^{-1}$.) Direct evidence of the size of atomic nuclei comes from scattering experiments, specifically measurements of the momentum dependence of the scattering cross section. This 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.

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 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 4.2 shows a plot of known nuclear species (or *nuclides*), color coded

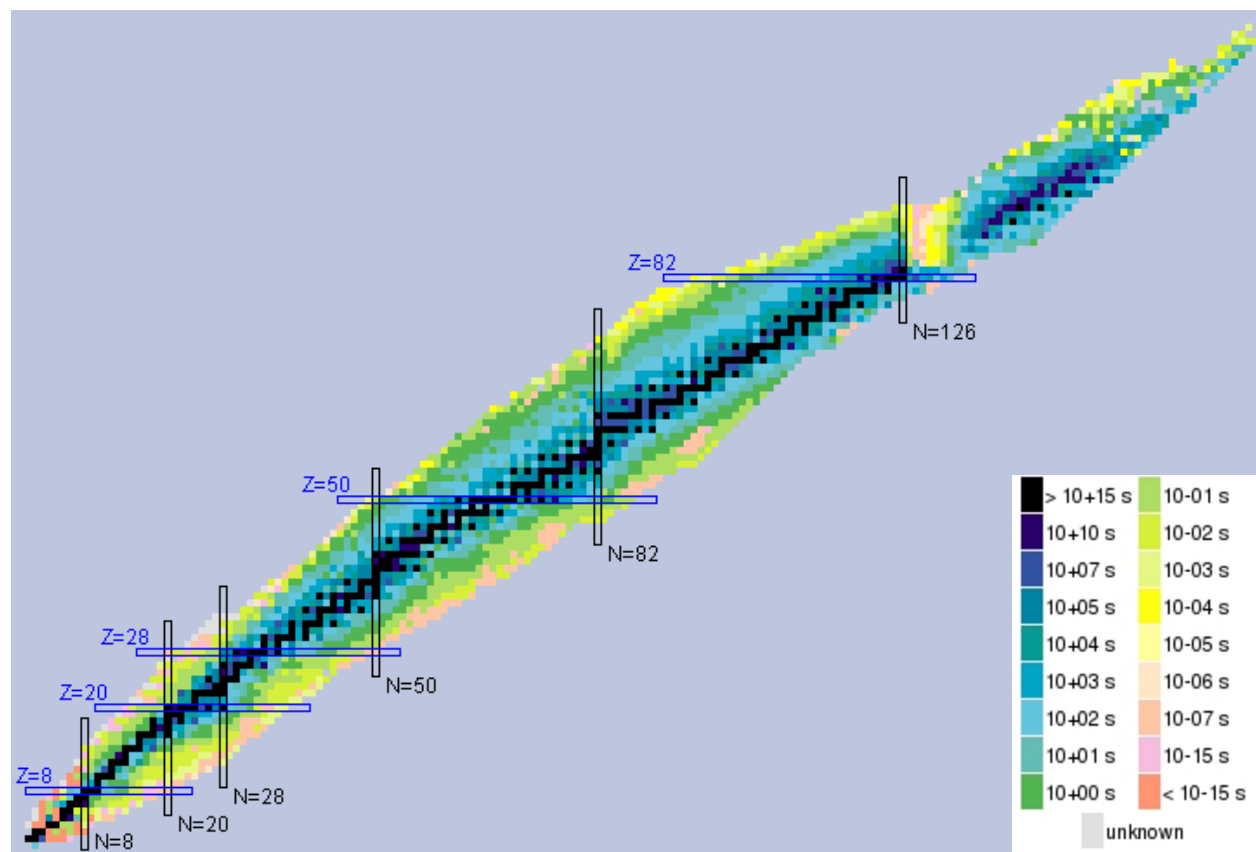


Figure 4.2: Chart of the nuclides (from the National Nuclear Data Center at Brookhaven National Laboratory). The number of protons, Z , is plotted vertically and the number of neutrons, N , horizontally. The color coding indicates the lifetime, with stable nuclei in black and lighter colors corresponding, 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.

according to their stability. Useful interactive online versions may be found at www.nndc.bnl.gov and atom.kaeri.re.kr.¹

Protons have charge $+e$, equal in magnitude but opposite in sign to the electron. This apparent *exact* equality, except for sign, between the proton’s electric charge and that of the electron has been tested to a precision of better than one part in 10^{21} and is a crucial feature of our universe. Because atoms have zero net charge, electrostatic repulsion does not prevent the assembly of macroscopic objects, such as your body or the Earth.²

¹Note 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.

²The observed exact equality, up to sign, of the electron and proton charge calls out for some more fundamental explanation — which is not currently known. Some hypothetical theories (called “grand unified theories” or GUTs) of possible physics beyond the Standard Model provide such an explanation by postulating a “grand” underlying symmetry which relates the electron and the constituents of the proton (*i.e.*, leptons and quarks, discussed below). The existence of such an underlying symmetry would lead to new interactions, beyond those discussed in the next section, which would allow the proton to decay at a very slow rate. Although the search continues, no such decay has yet been observed.

The mass of a proton is measured to be

$$m_p = 0.938\,272\,013(23) \text{ GeV}/c^2. \quad (4.1.3)$$

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,

$$m_n = 0.939\,565\,36(8) \text{ GeV}/c^2. \quad (4.1.4)$$

Neutrons, protons, and electrons are all spin 1/2 particles, where the spin is measured in units of the fundamental quantum of spin, \hbar .

Protons and neutrons are collectively referred to as *nucleons*. Nucleons are known to have internal structure: they may be regarded as bound states of three *quarks*. We will later be discussing quarks, and 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.³ Consequences of this peculiar experimentally-observed phenomena will be examined in the next chapter.

Today, the story of successive levels of microscopic structure — molecules made of atoms, atoms made of electrons and nuclei, nuclei made of nucleons, nucleons made of quarks — ends here. No evidence for internal structure within quarks, or electrons, has yet been found. If in the future these particles are discovered 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 scale of nucleons. This limit on the length scale is set by the corresponding energy scale (TeV) of experimental measurements at the LHC and the Fermilab Tevatron particle accelerators, which have not yet exhibited any evidence of internal structure for quarks or electrons.

4.2 Known interactions

Four different fundamental types of interactions, or forces, between particles are known today: *strong*, *weak*, *electromagnetic*, and *gravitational*. The relative strength and range of these interactions is summarized in Table 4.1: All particles participate in gravitational interactions, but this interaction is extremely weak and will largely be ignored in this class.⁴ Particles with a nonzero electric charge participate in electromagnetic interactions. Nearly all particles participate in the weak interactions, which are stronger than gravity but weaker than electromagnetic and strong interactions. Weak interactions are responsible for some forms of radioactivity and many nuclear and particle decays.

³Scattering experiments which attempt to disrupt a proton and liberate its constituents produce nucleons, pions, kaons and other particles we will discuss in the next chapter, but not free quarks. On spatial resolution scales larger than a fermi, quarks always appear “confined” inside nucleons or other strongly interacting particles. Understanding how this phenomena emerges from the mathematical details of the Standard Model has been a major intellectual challenge in theoretical particle physics.

⁴The assertion that gravity is weak may seem surprising since, on scales of the solar system and larger, gravity is the dominant interaction. This reflects the fact that strong and weak interactions act only over short ranges, and electromagnetic forces between electrically neutral objects fall off with distance much faster than does gravity. Hence, even though gravity is much weaker than the other interactions in terms of its effects on fundamental particles, it becomes the dominant interaction at large distances because its effects are long-ranged and add coherently in bulk matter.

Force	Relative Strength	Range
Strong	1	≈ 1 fm
Weak	10^{-6}	$\approx 10^{-3}$ fm
Electromagnetic	10^{-2}	Infinite ($\propto 1/r^2$)
Gravitational	10^{-43}	Infinite ($\propto 1/r^2$)

Table 4.1: Known forces of the Standard Model. The second column shows the (approximate) relative strength of the respective forces when acting between fundamental charged particles such as protons.

Finally, nucleons and other particles made from quarks (collectively known as *hadrons*) participate in the strong interactions. These are the interactions which cause quarks to become bound into composite particles. While gravitational and electromagnetic interactions (between charged particles) decrease slowly with distance like $1/r^2$, both weak and strong interactions are short-ranged, becoming negligible beyond a characteristic distance. This distance is about 1 fm for strong interactions, and roughly a thousand times shorter, 10^{-3} fm, for weak interactions.

4.3 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, will it eventually, spontaneously fall apart? Phrased in this manner, this is a “bad” question. It is fundamentally unanswerable — because feasible experiments must necessarily last 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,

$$\text{proton lifetime } \tau_p > 2.1 \times 10^{29} \text{ yr}, \quad (4.3.1)$$

$$\text{electron lifetime } \tau_e > 4.6 \times 10^{26} \text{ yr}. \quad (4.3.2)$$

You should be impressed with these limits, considering that they vastly exceed the age of the Earth (a mere 4.5 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 impossible. 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.⁵ 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.⁶

⁵The 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 $N(t_1) \approx N_0 - 1$.

⁶The development of such large detectors, essentially instrumented large tanks of water, has not yet led to any

Next, consider neutrons, the other basic constituent 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 $\bar{\nu}_e$,

$$n \longrightarrow p + e^- + \bar{\nu}_e. \quad (4.3.3)$$

This decay is referred to as a *beta decay*,⁷ and is a consequence of *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, and exhibit interesting quantum-mechanical phenomena which we will 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, 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 4.2) up to bismuth-209 (${}^{209}_{83}\text{Bi}$) which is the heaviest (known) nucleus that is essentially stable.⁸

4.4 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, somewhat more neutrons than protons). 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,

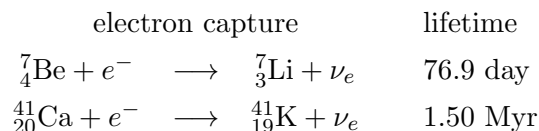
β decay	lifetime
${}^3_1\text{H} \longrightarrow {}^3_2\text{He} + e^- + \bar{\nu}_e$	17.8 yr
${}^6_2\text{He} \longrightarrow {}^6_3\text{Li} + e^- + \bar{\nu}_e$	1.16 s
${}^{10}_4\text{Be} \longrightarrow {}^{10}_5\text{B} + e^- + \bar{\nu}_e$	2.18 Myr
${}^{14}_5\text{B} \longrightarrow {}^{14}_6\text{C} + e^- + \bar{\nu}_e$	18.0 ms

observation of proton decay. But it has resulted in detectors capable of measuring neutrinos emitted from our Sun or supernovae occurring in our galaxy (or neighboring dwarf galaxies), as well as terrestrial nuclear power plants. This is a nice example of the synergies that drive science. Experiments and detector technology that was developed to observe or improve limits on proton decay played a critical role in the unexpected and fundamental discovery that neutrinos are not exactly massless.

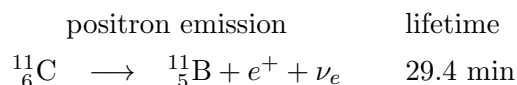
⁷This is a historical name which dates from the early 1900s, when three distinct types of radioactive decay, called α , β , and γ , had been identified. The different decay types were distinguished by the degree to which the particles emitted in the decay could penetrate ordinary matter. Alpha decays 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.

⁸In fact, bismuth-209 has recently been found to alpha decay with a lifetime of 2×10^{19} yr.

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,



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. The carbon-11 nucleus preferentially decays via positron emission even when it has an orbital electron it could otherwise capture,



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.

4.5 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” of the electromagnetic interaction. (One may regard electromagnetic interactions between charged particles as occurring via the exchange of photons.) Photons have no rest mass, unlike the other particles we have discussed so far. This means that the energy of a photon is directly proportional to its momentum, $E = c|\vec{p}|$. This energy can be arbitrarily small, unlike a massive particle whose energy is always greater than (or equal to) its rest energy mc^2 . The de Broglie wavelength of any particle is inversely related to its spatial momentum and thus, for photons, wavelength and energy (or frequency) are inversely related,

$$\lambda = \frac{2\pi\hbar}{|\vec{p}|} = \frac{2\pi\hbar c}{E} = \frac{2\pi c}{\omega} = \frac{c}{\nu}. \quad (4.5.1)$$

The familiar breadth of the classical electromagnetic spectrum, ranging from arbitrarily low frequency (radio waves) to extremely high (gamma rays) is a direct reflection of the fact that photons, as massless particles, can have energies which are arbitrarily small or large.

Photons are spin-1 particles, meaning that they carry one unit of angular momentum (in units of \hbar).

In everyday life, quantum aspects of the electromagnetic field are not readily apparent. For a great many applications, a classical treatment of electromagnetism suffices. This is the case when the number of photons is enormous and one 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.⁹

4.6 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 both annihilate and produce 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,

$$p + \bar{p} \longrightarrow \gamma + \gamma,$$

$$n + \bar{n} \longrightarrow \gamma + \gamma.$$

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 not known when the neutrino, the other fundamental electrically neutral particle, is its own antiparticle or not.¹⁰

Although antimatter is not present in everyday life (why this is so is another mystery), 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 this further in a later chapter.

4.7 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

⁹Human 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.

¹⁰If neutrinos are their own antiparticle, then certain rare nuclear processes known as neutrino-less double beta decay are possible. Such decays have not yet been observed, but quite a few experiments aiming to detect neutrino-less double beta (including CUORE, EXO, SNO+, MAJORANA and others) are currently running or under development.

particle	rest energy	lifetime	dominant decay	charge	L
ν_e	< 2 eV	\approx stable	—	0	1
ν_μ	< 2 eV	\approx stable	—	0	1
ν_τ	< 2 eV	\approx stable	—	0	1
e^-	0.511 MeV	stable	—	-1	1
μ^-	105.7 MeV	$2 \mu\text{s}$	$e^- \bar{\nu}_e \nu_\mu$	-1	1
τ^-	1777 MeV	0.3 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
$\bar{\nu}_\mu$	< 2 eV	\approx stable	—	0	-1
$\bar{\nu}_\tau$	< 2 eV	\approx stable	—	0	-1
e^+	0.511 MeV	stable	—	+1	-1
μ^+	105.7 MeV	$2 \mu\text{s}$	$e^+ \nu_e \bar{\nu}_\mu$	+1	-1
τ^+	1777 MeV	0.3 ps	$\pi^+ \pi^0 \bar{\nu}_\tau, e^+ \nu_e \bar{\nu}_\tau, \mu^+ \nu_\mu \bar{\nu}_\tau$	+1	-1

Table 4.2: Leptons and antileptons.

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 electromagnetic interactions. However, leptons do not participate in strong interactions; leptons 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 electric charge is (apparently) *identical* to that of the electron. Their antiparticles are the antimuon (μ^+) and antitau (τ^+). There are distinct neutrinos associated, via 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 ($\bar{\nu}_\mu, \bar{\nu}_\tau$). Why there are three different charged leptons, each with its corresponding neutrino, is another question for which no satisfying answer is currently known.

Basic properties of leptons are summarized in Table 4.2. 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 no one has yet succeeded). However, the observation of neutrino oscillations, which will be discussed in a later chapter, implies that neutrinos have non-zero masses. But at the moment only an upper bound on the actual values of the neutrino masses is known.

As indicated in Table 4.2, heavier (*i.e.*, more massive) leptons decay into lighter ones. These are weak interaction processes. 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 (two pions) and just a single light lepton, the tau neutrino. In all these processes *lepton number*, denoted L , is conserved. Lepton number is defined as the total number of leptons *minus* antileptons,

$$L \equiv (\# \text{ leptons}) - (\# \text{ antileptons}). \quad (4.7.1)$$

All known interactions conserve lepton number.¹¹ Nearly all processes (but not neutrino oscillations)

¹¹Actually, this is not quite true. The theory of weak interactions predicts extremely rare processes which change lepton number. However, the rate of these processes is so small that lepton number violation is completely unobservable.

conserve lepton number separately for each species or “flavor” of lepton: electron, muon, and tau. The electron-type lepton number L_e counts the number of electrons and electron neutrinos, minus the number of positrons and electron antineutrinos. Likewise, L_μ counts muons and muon neutrinos (minus their antiparticles), and L_τ counts taus and tau neutrinos (minus their antiparticles). In the decay $\mu^- \rightarrow e^- \bar{\nu}_e \nu_\mu$, for example, both initial and final states have $L_e = 0$ and $L_\mu = +1$. Similarly, each of the final states listed in Table 4.2 for the decay of a τ^- has $L_\tau = +1$ while $L_e = L_\mu = 0$.

4.8 Spin and statistics

Important attributes of particles, which will play essential roles in understanding possible interactions and decays, include their mass, spin, electric charge, lepton number, and a few other “quantum numbers” which will be introduced later. One very basic property concerns the value of a particle’s spin. As indicated above, many fundamental particles (including protons, neutrons, and all the leptons) have half-integer values of spin (in units of \hbar).¹² Such particles are called *fermions*. Other particles (such as the photon, and nuclei such as ${}^2_1\text{H}$, ${}^4_2\text{He}$ and ${}^{12}_6\text{C}$) have integer values of spin. These are called *bosons*.

The spin of a particle determines how its quantum states behave under rotations. (You should be reasonably familiar with the case of spin 1/2.) In particular, how a particle behaves under a 2π (*i.e.*, 360°) rotation is determined by whether its spin is an integer, or a half-integer. Bosons, with integer spin, are *unchanged* after a 2π rotation, the same as one would expect from classical physics. Fermions, with half-integer spin, have quantum states which, after a 2π rotation, come back to *minus* the initial state. In other words, the phase of the state of any fermion changes by π under a 360° rotation.

Of course, such an overall change of phase only matters in the context of quantum mechanics. The intrinsic difference between bosons and fermions is especially clear in quantum field theory (QFT), where one defines operators to represent the creation or removal of particles. For bosons, such operators commute ($[A, B] \equiv AB - BA = 0$, so $AB = BA$), while for fermions these operators must *anti-commute* ($\{A, B\} \equiv AB + BA = 0$, or $AB = -BA$). Consequently, when one builds multiparticle states out of fermions, the result must be *anti-symmetric* under the interchange of any pair of identical fermions. This immediately leads to the Pauli Exclusion principle — no two identical fermions can reside in the same quantum state, since such a situation would necessarily be symmetric under interchange. On the other hand, multiparticle states constructed from bosons must be symmetric under the interchange of any two identical bosons. Consequently, any number of identical bosons can reside in the same quantum state (and lasers really do produce beams of coherent photons). This connection between spin and interchange symmetry is referred to as the Spin-Statistics Theorem.

¹²“Half-integer” is shorthand for any integer plus 1/2.

Chapter 5

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 (or in some cases, α or 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 is called *quantum chromodynamics*, often abbreviated as QCD. 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. 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.

5.1 Quark flavor and color

Quarks are spin-1/2 particles (fermions) which come in various species, referred to as *flavors*. Different quark flavors have been given somewhat whimsical names, as shown in Table 5.1. In addition to the curious names, two other things in Table 5.1 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 $|e|$). The quark masses listed in this table must be interpreted with some care, as *isolated* quarks are never observed experimentally. The mass, or rest energy, of observed particles which are bound states of quarks (like the proton) largely reflects the binding energy of the quarks, and is not just the sum of the intrinsic quark masses. Nevertheless, it is remarkable that quark masses vary over such a wide range, from a few MeV to hundreds of GeV. The three lightest quark flavors, denoted u , d and s , have masses which are small relative to the proton mass; the three heavy flavors, c , b and t , have masses which are comparable or large relative to the proton mass.

Along with quarks, there are 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 $+1/3$.)

flavor	symbol	mass	charge
up	u	$\approx 2 \text{ MeV}/c^2$	$\frac{2}{3} e $
down	d	$\approx 5 \text{ MeV}/c^2$	$-\frac{1}{3} e $
strange	s	$\approx 95 \text{ MeV}/c^2$	$-\frac{1}{3} e $
charm	c	$\approx 1.3 \text{ GeV}/c^2$	$\frac{2}{3} e $
bottom	b	$\approx 4.2 \text{ GeV}/c^2$	$-\frac{1}{3} e $
top	t	$\approx 173 \text{ GeV}/c^2$	$\frac{2}{3} e $

Table 5.1: Known quark flavors

Quarks have an additional attribute, analogous to but different from electric charge, which is termed *color* charge. The color charge of a quark can have three possible values which may be denoted as ‘red’, ‘green’, or ‘blue’. These names are simply labels for different quantum states of the quark.¹ Antiquarks carry opposite electric and color charge as the corresponding quarks; color states of antiquarks can be called ‘anti-red’, ‘anti-green’ or ‘anti-blue’.

Since quarks (and antiquarks) have spin 1/2, so 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).

5.2 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 trapped inside bound states with other quarks and antiquarks.² Bound states produced by the strong interactions are called *hadrons* (*hadros* is Greek for ‘strong’).

Quantum chromodynamics 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.) 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

¹These 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!

²Except at sufficiently high temperatures. Above a temperature of $T_c \approx 2 \times 10^{12}$ K (or $kT \approx 170$ 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 of 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/heavy_ion.htm. There is an ongoing program studying heavy ion collisions both at the RHIC accelerator on Long Island, and at the LHC where some running time is devoted to colliding heavy nuclei rather than protons.

antiquarks have electric charges which are opposite to their partner quarks, they also have “opposite” color: anti-red, anti-blue, or anti-green. 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 (in the quantum superposition $r\bar{r} + g\bar{g} + b\bar{b}$ of correlated color states which is overall colorless).

Bound states of three quarks are called *baryons*, bound states of three antiquarks are called *antibaryons*, and quark-antiquark bound states are called *mesons*. Baryons and antibaryons, as bound states of three spin-1/2 quarks, necessarily have half-integer values of spin, and are fermions. Mesons, as bound states of two spin-1/2 constituents, have integer values of spin, and are bosons.

Strong interactions are effectively flavor-blind; except for the difference in mass, quarks of different flavors have identical strong interactions. Strong interaction processes cannot change the net number of quarks of a given flavor (*e.g.*, the number of up quarks minus up antiquarks, *etc.*)³

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. Since quarks come in multiple flavors, listed in Table 5.1, one can (and we will) enumerate the various possibilities.

The rest energy of a hadron (like any bound state) may be regarded as the sum of the rest energies of its constituents plus a binding energy. For some hadrons, including nucleons, we will see that most of their total energy comes from binding energy. But the masses of quarks also play a part. Looking at the quark masses listed in Table 5.1, 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 heavier. So it should not be surprising that the lightest hadrons will be those which are bound states of *u* and *d* quarks. 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. And hadrons containing the other quark flavors (*c*, *b*, or *t*) should be substantially heavier.

When enumerating possible combinations of quarks which could form hadrons, one must also think about spin and electric charge. Combining electric charge is easy: the charge of a hadron is just the sum of the charges of its constituent quarks. Combining the angular momentum of two spin 1/2 particles can yield either spin 1 or 0 (depending on whether the spin wavefunction is symmetric or antisymmetric). Three spin 1/2 particles can combine to form either spin 3/2 or 1/2.⁴ In addition to the combined total spin \vec{S} of the constituents, the total angular momentum \vec{J} of a multi-particle bound state can also receive a contribution from the orbital angular momentum \vec{L} which arises due to the internal motion of the constituents. So, in general, $\vec{J} = \vec{L} + \vec{S}$. For the lowest mass hadrons, of a given spin and flavor content, one may regard the quarks as nearly at rest (within the constraints imposed by quantum mechanics) with vanishing orbital angular momentum, $\vec{L} = 0$.

³In contrast, weak interactions (whose details will be discussed further in a later chapter) can change a quark of one flavor into a quark of a different flavor. Hence weak interactions need not conserve the net number of quarks of a given flavor.

⁴More generally, recall that when a system with spin S_1 is combined with a system with spin S_2 , the result can have a total spin which ranges from a minimum of $|S_1 - S_2|$ to a maximum of $S_1 + S_2$ in unit steps (when all spins are measured in units of \hbar). For any system with spin S , there are $2S + 1$ possible values for the projection $\vec{S} \cdot \hat{n}$ of the spin vector along some chosen spin quantization axis \hat{n} ranging from $-S$ to $+S$ in unit steps.

5.3 Mesons

Let us start with mesons and (for the moment) consider just the three lightest quark flavors, u , d , and s . Since a meson is a bound state of a quark and antiquark, there are nine different possible flavor combinations. Table 5.2 displays these possibilities arranged according to the resulting electric charge Q as well as the number of strange (or anti-strange) quarks. For each combination, there will be one spin zero state, and one spin one state. Reassuringly, the lightest observed mesons are either spin zero or spin one. Table 5.3 lists the light spin zero mesons, along with their dominant decay modes, while Table 5.4 does the same for light spin one mesons.⁵ These are the lightest mesons.

As Tables 5.3 and 5.4 show, the numbers and charges of the lightest spin zero and spin one mesons precisely match what is expected based on the possible combinations of a quark and antiquark. In general, mesons containing strange quarks are heavier than those without. But among the neutral mesons, it is noteworthy that it is certain linear combinations of the $Q = 0$ quark combinations listed in Table 5.2 which correspond to distinct particles. 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 simple 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.

5.4 Baryons

One can go through a similar exercise for baryons. Instead of dealing with two constituents, one is now dealing with three. Combining three spin $1/2$ objects can yield either spin $1/2$ or $3/2$. The lightest observed baryons are, in fact, either spin $1/2$ or spin $3/2$. Tables 5.5 and 5.6 list the lightest spin $1/2$ and spin $3/2$ baryons, respectively. As Table 5.6 shows, the ten light $J = 3/2$ baryons precisely match the ten possible choices of triples of quark flavors. From the masses listed in Table 5.6 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, as the binding dynamics due to the color interactions are independent of the quark flavors involved. The ten baryons shown in Table 5.6 are referred to as the spin $3/2$ baryon *decuplet*.

As Table 5.5 shows, there are only eight light $J = 1/2$ baryons, which are referred to as the spin $1/2$ baryon *octet*. This difference between $J = 3/2$ and $J = 1/2$ bound states may be understood as a consequence of the Pauli exclusion principle. Quarks, being fermions, must obey the Pauli exclusion principle. Wavefunctions of multi-quark states must be antisymmetric under the interchange of identical quarks. To see how this leads to the difference between the $J = 3/2$ decuplet and the $J = 1/2$ octet, we must examine baryon wavefunctions in more detail.

⁵Tables 5.3 and 5.4 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.

$(\#s) + (\#\bar{s})$	$Q = 1$	$Q = 0$	$Q = -1$
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 5.2: Possible light quark-antiquark combinations

meson	mass	lifetime	dominant decays	quark content
π^0	135.0 MeV	8×10^{-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_S^0	497.7 MeV	8.9×10^{-11} s	$\pi^+\pi^-, \pi^0\pi^0$	$\frac{1}{\sqrt{2}}(d\bar{s} - s\bar{d})$
K_L^0	497.7 MeV	5.1×10^{-8} s	$\pi^\pm e^\mp \nu_e, \pi^\pm \mu^\mp \nu_\mu$	$\frac{1}{\sqrt{2}}(d\bar{s} + s\bar{d})$
η	547.5 MeV	5×10^{-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	3×10^{-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 5.3: Light spin zero, parity odd mesons.

meson	mass	lifetime	dominant decays	quark content
ρ^+, ρ^0, ρ^-	775.5 MeV	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	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}	896.0 MeV	1.3×10^{-23} s	$K\pi$	$d\bar{s}, s\bar{d}$
ϕ	1019.5 MeV	2×10^{-22} s	$K^+K^-, K_L^0K_S^0, \pi\pi\pi$	$\approx s\bar{s}$

Table 5.4: Light spin one, parity odd mesons.

baryon	mass	lifetime	dominant decays	quark content
p	938.3 MeV	stable	—	uud
n	939.6 MeV	9×10^2 s	$pe^- \bar{\nu}_e$	udd
Λ	1116 MeV	2.6×10^{-10} s	$p\pi^-, n\pi^0$	uds
Σ^+	1189 MeV	0.8×10^{-10} s	$p\pi^0, n\pi^+$	uus
Σ^0	1193 MeV	7×10^{-20} s	$\Lambda\gamma$	uds
Σ^-	1197 MeV	1.5×10^{-10} s	$n\pi^-$	dds
Ξ^0	1315 MeV	2.9×10^{-10} s	$\Lambda\pi^0$	uss
Ξ^-	1321 MeV	1.6×10^{-10} s	$\Lambda\pi^-$	dss

Table 5.5: Light spin 1/2, parity even baryons.

baryon	mass	lifetime	dominant decays	quark content
$\Delta^{++}, \Delta^+, \Delta^0, \Delta^-$	1232 MeV	6×10^{-24} s	$p\pi, n\pi$	uuu, uud, udd, ddd
$\Sigma^{*+}, \Sigma^{*0}, \Sigma^{*-}$	1385 MeV	2×10^{-23} s	$\Lambda\pi$	uus, uds, dds
Ξ^{*0}, Ξ^{*-}	1530 MeV	7×10^{-23} s	$\Xi\pi$	uss, dss
Ω^-	1672 MeV	0.8×10^{-10} s	$\Lambda K^-, \Xi\pi$	sss

Table 5.6: Light spin 3/2, parity even baryons.

5.5 Baryon wavefunctions

To understand how the fermionic nature of quarks produces the observed pattern of spin and flavor for baryons, we must first return to the earlier assertion that only colorless bound states of quarks exist. 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,

$$\begin{pmatrix} \psi_\uparrow \\ \psi_\downarrow \end{pmatrix} = \begin{pmatrix} \langle \uparrow | \psi \rangle \\ \langle \downarrow | \psi \rangle \end{pmatrix}, \quad (5.5.1)$$

the wavefunction of a quark (of definite flavor and spin) is a three-component vector in “color” space,

$$\vec{\Psi} \equiv \begin{pmatrix} \psi_r \\ \psi_g \\ \psi_b \end{pmatrix} = \begin{pmatrix} \langle \text{red} | \psi \rangle \\ \langle \text{green} | \psi \rangle \\ \langle \text{blue} | \psi \rangle \end{pmatrix}. \quad (5.5.2)$$

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. Since the quark (color) wavefunction is a three-component vector, to build a colorless state from three quarks, A , B and C , one must combine the three color vectors describing the individual quarks, $\vec{\psi}_A$, $\vec{\psi}_B$, and $\vec{\psi}_C$, in such a way that the result is basis independent.

This may sound peculiar, but the mathematical problem is the same as asking how to build a rotationally-invariant scalar from three spatial 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 triple-product of the three vectors, $\vec{A} \cdot (\vec{B} \times \vec{C})$. This triple product may be expressed in several equivalent ways involving the determinant of components,

$$\vec{A} \cdot (\vec{B} \times \vec{C}) = \det \begin{vmatrix} A_1 & B_1 & C_1 \\ A_2 & B_2 & C_2 \\ A_3 & B_3 & C_3 \end{vmatrix} = \epsilon_{ijk} A_i B_j C_k. \quad (5.5.3)$$

In the last form, ϵ_{ijl} is the totally antisymmetric 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 interchanged; this is also encoded in the antisymmetry of the ϵ_{ijk} symbol. Consequently, the triple product is antisymmetric under any interchange of two of the vectors. The triple product (5.5.3) defines a rotationally invariant scalar, independent of the basis used to define vector components — and this is exactly the property we need when combining color wavefunctions of three quarks to produce a colorless result. Thus, the color part of the wavefunction for three quarks has the form

$$\Psi_{\text{color}} = \epsilon_{ijk} \psi_i^{q_1} \psi_j^{q_2} \psi_k^{q_3}, \quad (5.5.4)$$

where $\psi_i^{q_1}$ is the i 'th component of the color wavefunction (5.5.2) for quark 1, etc.

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 will be a product of a spatial wavefunction (depending only on the quark positions), a color wavefunction (depending only on the color vectors), and a flavor & spin wavefunction,

$$\Psi = \Psi_{\text{space}} \times \Psi_{\text{color}} \times \Psi_{\text{spin+flavor}}. \quad (5.5.5)$$

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. If this is not true then the wavefunction will have nodes across which it changes sign, and this increases the kinetic energy of the state. (In particular, this means the lightest hadrons will have no orbital angular momentum.) Because quarks are fermions, 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 and spin part must also be *symmetric* under permutations.⁷

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 $S_z = 3/2$ state,

⁶The geometric definition of a cross product only makes sense for real vectors, but the expressions (5.5.3) involving components of the vectors are equally well-defined for complex vectors.

⁷This was a significant puzzle in the early days of the quark model (circa 1970), when quarks had been postulated as constituents of baryons, but the role of a color quantum number was not yet understood. The spatial, flavor and spin wavefunctions that matched observed baryons were clearly symmetric, but quarks had spin 1/2 and hence were fermions. Was Pauli wrong? The additional attribute of color saved the spin-statistics theorem, and paved the way to the formulation of a microscopic theory of strong interactions, namely QCD.

the spin structure is also trivial, and totally symmetric, since all three quarks must individually have $S_z = 1/2$ if the total spin projection is $3/2$.⁸ Therefore the combined spin+flavor part of the wavefunction,

$$\Psi_{\text{spin+flavor}}^{\Delta^{++}} \sim (uuu) \times (\uparrow\uparrow\uparrow), \quad (5.5.6)$$

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 $3/2$ decuplet. Just like the example (5.5.6), these wavefunctions are simple products of a flavor part and a spin part, and are independently symmetric under permutations of quark spins, or permutations of quark flavors.

For $J = 1/2$ baryons, the situation is more complicated. To produce a spin $1/2$ state from three spin $1/2$ constituents, the spins of two constituents can be combined in a manner which produces spin zero (not spin one), so that adding the third constituent spin results in total spin $1/2$. But building a spin zero state out of two spin $1/2$ objects involves a spin wavefunction, $\uparrow\downarrow - \downarrow\uparrow$, which is antisymmetric under interchange of the two spins. Consequently, the spin wavefunction (alone) for a spin $1/2$ baryon cannot be totally symmetric under permutations of the spins. If we are to build a spin+flavor wavefunction which is symmetric under combined permutations of flavors and spins of quarks, then the flavor part of the wavefunction must also not be totally symmetric, and must compensate for the antisymmetry in the spin part of the wavefunction.

If all three quarks have the same flavor, this is not possible — like it or not, 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 $S = 0$ 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,

$$[(ud - du)u] \times [(\uparrow\downarrow - \downarrow\uparrow)\uparrow] = (udu - duu)(\uparrow\downarrow\uparrow - \downarrow\uparrow\uparrow), \quad (5.5.7)$$

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 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 for a spin-up proton, which is unique up to an overall normalization factor, is

$$\Psi_{\text{spin+flavor}}^{\text{proton}} = (udu - duu)(\uparrow\downarrow\uparrow - \downarrow\uparrow\uparrow) + (uud - udu)(\uparrow\uparrow\downarrow - \uparrow\downarrow\uparrow) + (uud - duu)(\uparrow\uparrow\downarrow - \downarrow\uparrow\uparrow). \quad (5.5.8)$$

⁸This ignores the possibility of further constituents in the baryon in addition to the three up quarks. Using an improved description of the structure of baryons does not change the essential conclusions of the following discussion.

One may write this wavefunction more explicitly with the various terms multiplied out,

$$\begin{aligned} \Psi_{\text{spin+flavor}}^{\text{proton}} = & [2u \uparrow u \uparrow d \downarrow - u \downarrow u \uparrow d \uparrow - u \uparrow u \downarrow d \uparrow \\ & + 2u \uparrow d \downarrow u \uparrow - u \downarrow d \uparrow u \uparrow - u \uparrow d \uparrow u \downarrow \\ & + 2d \downarrow u \uparrow u \uparrow - d \uparrow u \downarrow u \uparrow - d \uparrow u \uparrow u \downarrow] / \sqrt{18}, \end{aligned} \quad (5.5.9)$$

(where we have also normalized the result). You can, and should, check that this does satisfy the required condition of symmetry under interchange of spins and flavors of any pair of quarks. Similar constructions can be performed for all the other members of the spin 1/2 baryon octet.

One notable feature of the set of octet baryons, shown in Table 5.5, is the presence of two different baryons whose quark content is uds , specifically the Λ and the Σ^0 . This is not 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 two independent possibilities, completely consistent with the observed list of spin 1/2 baryons.

The mass values in Table 5.5 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 120–170 MeV.

5.6 Baryon number

Baryon number, denoted B , is defined as the total number of baryons minus the number of antibaryons, similar to the definition (4.7.1) of lepton number L . 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,

$$B = (\# \text{ baryons}) - (\# \text{ antibaryons}) = \frac{1}{3} [(\# \text{ quarks}) - (\# \text{ antiquarks})]. \quad (5.6.1)$$

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 $p + p \rightarrow p + p + n + n$.

5.7 Hadronic decays

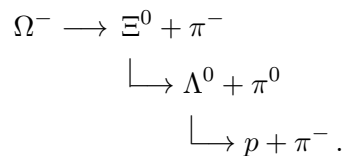
Turning to the decays of the various hadrons listed in Tables 5.3–5.6, it is remarkable how much can be explained using a basic understanding of the quark content of the different hadrons together with considerations of energy and momentum conservation. As an example, consider the baryons in the spin 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 pions, which are the lightest mesons. Consequently, a Δ baryon can decay to a nucleon plus a pion via strong interactions, which do not change the number of quarks minus antiquarks of each quark flavor. (Specifically, a Δ^{++} can decay

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.

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. (You can think of “strong” interactions as meaning rapid interactions.) This lifetime corresponds to a decay (or resonance) width $\Gamma_\Delta = \hbar/\tau \approx 120$ MeV, which is 10% of the rest energy of a Δ . Since light takes about 3×10^{-24} s to travel one fermi, a Δ baryon barely has time to “figure out” that it exists before it decays.¹⁰

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^* mass of 1385 MeV/ c^2 is larger than the 1116 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.

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 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 quark 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. The dominant decays involve the conversion of one strange quark into an up or down quark, leading to final states consisting of a Λ baryon plus K^- meson, a Ξ^0 baryon plus π^0 , or a Ξ^- plus π^+ . The Ω^- was first seen in bubble chamber photographs, and was relatively easy to discover due to its distinctive “cascade” decay,



So the overall process is $\Omega^- \rightarrow p\pi^-\pi^-\pi^0$ (with the pions eventually decaying to leptons and photons). Note that all final states of Ω^- decay 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 strong interaction decay, and is indicative of a weak interaction process.

Similar reasoning can be applied to the $J = 1/2$ baryons. The proton is stable (so far as we know), 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

¹⁰One might ask, “how long must some ‘particle’ live to justify calling it a particle?” With a lifetime under 10^{-23} s, a Δ baryon produced in some particle collision will never fly away from the interaction point and reach a particle detector, located some macroscopic distance away, before decaying. For such unstable particles, what is eventually detected are the decay products of the Δ baryon. Measuring the interaction *rate* as a function of energy in pion-nucleon scattering experiments, for example, one finds a resonance peak at the energy corresponding to production of Δ baryons. Because the Δ width is only 10% of its energy, this resonance peak is very recognizable.

lifetimes. (In a very real sense, electromagnetic interactions are stronger than weak interactions, but weaker than strong interactions.) The lifetimes of 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 kinetic energy of the decay products.

Similar observations apply to the decays of mesons. Just as most spin 3/2 baryons have strong interaction (*i.e.*, fast) decays into a spin 1/2 baryon plus a pion, the spin 1 (or “vector”) mesons of Table 5.4 all have strong decays into spin 0 (“scalar”) mesons with the same strange quark content, plus a pion. Lifetimes of these vector mesons are short, $10^{-22} - 10^{-24}$ s, indicative of strong interactions. Some aspects of these decays invite questions which we will consider in the next chapter. For example, the ρ^0 and ω are both neutral vector mesons with no strange quarks; why does the ω decay to three pions while the ρ^0 decays to just two pions? And in decays of ρ^0 , the actual two pion final state is $\pi^+\pi^-$, not $\pi^0\pi^0$. Why is that?

For the scalar mesons of Table 5.3, energy conservation rules out any strong interaction decays. The neutral π^0 , η and η' mesons all have electromagnetic decays with photons in the final state and lifetimes of order 10^{-20} s (similar to the Σ^0 lifetime). The K mesons (or “kaons”) can only decay via weak processes which turn a strange quark into an up or down quark (or \bar{s} into \bar{u} or \bar{d}). And the charged pions can only decay, via weak interactions, into leptons. All these weak decay lifetimes are in the $10^{-8} - 10^{-10}$ s range (a little slower or comparable to the Ω^- lifetime).

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.

5.8 Example problems

5.8.1 Neutron spin+flavor

Q: Find the spin+flavor wavefunction for a neutron.

A: This is most easily done by looking at the answer (5.5.8) for a proton, and just interchanging u and d quarks. So for a spin up neutron,

$$\Psi_{\text{spin+flavor}}^{\text{neutron}} = (dud - udd) (\uparrow\downarrow\uparrow - \downarrow\uparrow\uparrow) + (ddu - dud) (\uparrow\uparrow\downarrow - \uparrow\downarrow\uparrow) + (ddu - udd) (\uparrow\uparrow\downarrow - \downarrow\uparrow\uparrow).$$

This, as required for baryons, is symmetric under the combined interchange of spin and flavor of any pair of quarks.

5.8.2 Λ and Σ^0 spin+flavor

Q: Find two (mutually orthogonal) spin+flavor wavefunctions for $J = 1/2$ baryons with quark content uds . Can you deduce (or guess) which wavefunction represents the Σ^0 , and which represents the Λ ?

A: One $J = 1/2$ uds wavefunction can be found using exactly the same logic which works for the proton. Start with two quarks, say u and d , in the state $(ud - du) \times (\uparrow\downarrow - \downarrow\uparrow)$ which is symmetric under combined interchange of spins and flavors, and has the spins combined to form

$S = 0$. Add the strange quark with spin, say, chosen to be up, to obtain the $J = 1/2$ three quark state $[(ud - du)s] \times [(\uparrow\downarrow - \downarrow\uparrow)\uparrow] = (uds - dus)(\uparrow\downarrow - \downarrow\uparrow)$. To make this fully symmetric under combined interchange of spin and flavor of any pair of quarks, we need to add analogous terms related by cyclic permutations of the three quarks. This produces our first answer,

$$\Psi_{\text{spin+flavor}}^A = (uds - dus)(\uparrow\downarrow - \downarrow\uparrow) + (sud - sdu)(\uparrow\uparrow - \uparrow\downarrow) + (usd - dsu)(\uparrow\downarrow - \downarrow\uparrow).$$

To find a second orthogonal answer, one can start with a ud two quark in which the quark spins are combined to form $S = 1$ (instead of $S = 0$), and then add the third s quark in such a way that the resulting final spin is $S = 1/2$, not $S = 3/2$. This is a little bit trickier. The two spin wavefunction $\uparrow\uparrow$ has spin projection $S_z = 1$ and (necessarily) total spin $S = 1$. The two spin wavefunction $\uparrow\downarrow + \downarrow\uparrow$ has spin projection $S_z = 0$ but also has total spin $S = 1$. Either of these spin wavefunctions can be combined with the flavor wavefunction $ud + du$ to obtain a two quark spin+flavor wavefunction which is symmetric under combined interchange of spin and flavor. To each of these two quark states one may now add a strange quark, with its spin chosen in such a way that the resulting three quark state has $S_z = 1/2$, yielding $(uds + dus) \times (\uparrow\downarrow)$ or $(uds + dus) \times (\uparrow\downarrow + \downarrow\uparrow)$. By construction, each of these states has total $S_z = +1/2$, but neither state has definite total spin — each wavefunction describes a mixture of $S = 3/2$ and $S = 1/2$ states. We need to find the linear combination which describes a pure $S = 1/2$ state. The correct answer must be orthogonal to the wavefunction which describes three spins combined to form $S = 3/2$ with $S_z = 1/2$. That spin wavefunction is $\uparrow\uparrow\downarrow + \uparrow\downarrow\uparrow + \downarrow\uparrow\uparrow$, and is completely symmetric under interchange of spins (just like the $\uparrow\uparrow\uparrow$ spin state describing $S = 3/2$, $S_z = 3/2$). The linear combination of $\uparrow\uparrow\downarrow$ and $(\uparrow\downarrow + \downarrow\uparrow)$ which is orthogonal to this $S = 3/2$, $S_z = 3/2$ state is $2\uparrow\uparrow\downarrow - (\uparrow\downarrow + \downarrow\uparrow)$, up to an arbitrary overall normalization factor.

Therefore, the three quark uds spin+flavor wavefunction which is symmetric under combined interchange of spin and flavor of the first two quarks, has those quark spins combined to form $S = 1$, but then has all three quark spins combined in such a way that the result is pure $S = 1/2$, $S_z = 1/2$, is

$$(uds + dus) \times [2\uparrow\uparrow\downarrow - \uparrow\downarrow - \downarrow\uparrow].$$

To make the result symmetric under combined interchange of spin and flavor of any pair of quarks we must, as before, add circular permutations. The result is our second answer,

$$\begin{aligned} \Psi_{\text{spin+flavor}}^B &= (uds + dus) [2\uparrow\uparrow\downarrow - \uparrow\downarrow - \downarrow\uparrow] \\ &+ (sud + sdu) [2\downarrow\uparrow\uparrow - \uparrow\uparrow\downarrow - \uparrow\downarrow\uparrow] \\ &+ (dsu + usd) [2\uparrow\downarrow\uparrow - \downarrow\uparrow\uparrow - \uparrow\uparrow\downarrow]. \end{aligned}$$

The state $\Psi_{\text{spin+flavor}}^A$ describes the spin+flavor structure of the Λ , while state $\Psi_{\text{spin+flavor}}^B$ describes the spin+flavor structure of the Σ^0 . To see (or at least motivate) why, notice that under interchange of just the flavors of the two non-strange quarks, $\Psi_{\text{spin+flavor}}^A$ is *antisymmetric* while $\Psi_{\text{spin+flavor}}^B$ is *symmetric*. Contrast these behaviors with that of the wavefunction of a Σ^+ baryon: which looks just like the result (5.5.8) for the proton but with d everywhere replaced by s . The Σ^+ wavefunction, having two u quarks, is necessarily (and trivially) *symmetric* under interchange of just the flavors of the non-strange quarks. The Σ^0 differs from the Σ^+ just by replacing one up quark with a down quark. Up and down quarks have nearly identical masses and (reflecting this) so do the Σ^0 and Σ^+ . Therefore, it must be $\Psi_{\text{spin+flavor}}^B$ which describes the Σ^0 , as this state has the non-strange flavor interchange symmetry which matches the result for Σ^+ .

Chapter 6

Symmetries

6.1 Quantum dynamics

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, $\langle\psi|\psi\rangle = 1$.

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 ,

$$|\psi(t)\rangle = U(t) |\psi(0)\rangle. \quad (6.1.1)$$

Time evolution must map any properly normalized state at one time into a normalized state at another time. This implies that the time evolution operator is *unitary*,

$$U(t)^\dagger = U(t)^{-1}. \quad (6.1.2)$$

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,

$$i\hbar \frac{d}{dt} |\psi(t)\rangle = H |\psi(t)\rangle. \quad (6.1.3)$$

This is the (time dependent) Schrodinger equation. It is a linear first order differential equation, whose solution can be written immediately in terms of an exponential,¹

$$|\psi(t)\rangle = e^{-iHt/\hbar} |\psi(0)\rangle. \quad (6.1.4)$$

Comparing with the definition (6.1.1), one sees that this exponential of the Hamiltonian (times $-it/\hbar$) is precisely the time evolution operator,

$$U(t) = e^{-iHt/\hbar}. \quad (6.1.5)$$

¹This assumes that the Hamiltonian H does not, itself, depend on time. If this is false, then the laws of physics (*i.e.*, the form of the basic equations of motion) would change with time. For all theories of interest in this class, H will be time-independent.

The Hamiltonian must be Hermitian, $H^\dagger = H$, in order for $U(t)$ to be unitary.

We have just solved all quantum dynamics! Of course, evaluating this exponential of the Hamiltonian can be (and usually is) hard. A quantum system whose space of states is N -dimensional 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.

6.2 Symmetries

A linear transformation T which maps an arbitrary state $|\psi\rangle$ into some different state $|\tilde{\psi}\rangle$ is called a *symmetry* if T is unitary, $T^\dagger = T^{-1}$, and T commutes with the time evolution operator,²

$$TU(t) = U(t)T. \quad (6.2.1)$$

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 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 (6.2.1) 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 diagram³

$$\begin{array}{ccc} |\psi(t)\rangle & \xrightarrow{T} & |\tilde{\psi}(t)\rangle \\ U(t) \uparrow & & \uparrow U(t) \\ |\psi(0)\rangle & \xrightarrow{T} & |\tilde{\psi}(0)\rangle \end{array} \quad (6.2.2)$$

showing that $|\tilde{\psi}(t)\rangle$ can be constructed from $|\psi(0)\rangle$ by following either path.

The condition (6.2.1) that the transformation T commute with the time evolution operator is equivalent to the condition that T commute with the Hamiltonian,

$$[T, H] \equiv TH - HT = 0. \quad (6.2.3)$$

Symmetries have many useful consequences. One class of applications follows directly from the basic definition embodied in the diagram (6.2.2) — 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, turning a proton into an antiproton, a π^+ into a π^- , etc. Charge conjugation is a symmetry of strong and electromagnetic interactions. This symmetry implies that the rate at which a Δ^{++} baryon decays to a proton and π^+ (via strong interactions) is the same as the rate at which the $\bar{\Delta}^{--}$ antibaryon (the antiparticle of the Δ^{++}) decays to an antiproton and π^- . And it implies that the cross section for π^+ scattering on protons must be the same as the cross section for π^-

²This 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.

³Mathematicians call this a *commutative diagram*.

mesons to scatter on antiprotons. One can predict many aspects of strong interactions just from an understanding of the relevant symmetry properties, without knowing any details of the dynamics.

A second category of applications follows from the commutativity (6.2.3) 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 ,

$$H|\psi_n\rangle = E_n|\psi_n\rangle, \quad (6.2.4a)$$

$$T|\psi_n\rangle = t_n|\psi_n\rangle. \quad (6.2.4b)$$

The eigenvalue E_n of the Hamiltonian is the energy of the state $|\psi_n\rangle$ — Hamiltonian eigenstates are called energy eigenstates or *stationary states*. The latter name reminds one that energy eigenstates have simple harmonic time dependence; the time-dependent Schrodinger equation (6.1.3) implies that the time evolution of an energy eigenstate is just $|\psi_n(t)\rangle = e^{-iE_n t/\hbar}|\psi_n(0)\rangle$. Hence, an eigenstate of the Hamiltonian is also an eigenstate of the time evolution operator $U(t)$, with eigenvalue $e^{-iE_n t/\hbar}$.

Because the Hamiltonian is a Hermitian operator, its eigenvalues E_n must be real. Because the symmetry transformation T is a unitary operator, its eigenvalues t_n must be phase factors, $t_n = e^{i\phi_n}$ for some *real* phase ϕ_n .⁴ 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. This provides a more refined, and useful, labeling of energy eigenstates.

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 theory in which a parity transformation, $x \rightarrow -x$, is

a symmetry. 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, electric charge, baryon number, and lepton number. As we will discuss below, 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 are not eigenstates of the Hamiltonian. Such states will have non-trivial 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 of the initial and final states coincide.

⁴To show this, multiply each side of the eigenvalue condition (6.2.4b) 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 $|t_n| = 1$.

That is,⁵

$$\langle \psi_{\text{out}} | U(t) | \psi_{\text{in}} \rangle = 0 \quad \text{if} \quad t_{\text{in}} \neq t_{\text{out}}. \quad (6.2.5)$$

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 or decays, without detailed knowledge of the dynamics. As we discuss below, 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 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. 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 .⁶ It may be a linearly independent state — $|\tilde{\psi}_n\rangle$ need not be proportional to $|\psi_n\rangle$. Consequently, the existence of symmetries which do not mutually commute can lead to degenerate energy levels, *i.e.*, multiple linearly independent states with exactly the same energy. 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 angular momentum is $j\hbar$, then the multiplet will contain $(2j + 1)$ states, since the projection of the angular momentum along some 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.

6.3 Continuous symmetries

Continuous symmetries are symmetries which depend (continuously!) on some parameter which controls the magnitude of the transformation. Examples include translations and rotations. Let $T(a)$ denote a continuous symmetry 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 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 . This implies that $T(a)$ depends exponentially on the parameter a , so that one can write

$$T(a) = e^{iQa}, \quad (6.3.1)$$

for some operator Q , which is called the *generator* of the symmetry $T(a)$. In order for $T(a)$ to be unitary (as required), the generator Q must be Hermitian. Note that the relation between $T(a)$ and

⁵To see this, use the fact that T is unitary to write $1 = T^\dagger T$. Inserting the identity operator does nothing, so $\langle \psi_{\text{out}} | U(t) | \psi_{\text{in}} \rangle = \langle \psi_{\text{out}} | T^\dagger T U(t) | \psi_{\text{in}} \rangle = \langle \psi_{\text{out}} | T^\dagger U(t) T | \psi_{\text{in}} \rangle$. 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 $\langle \psi_{\text{out}} |$ to simplify $\langle \psi_{\text{out}} | T^\dagger U(t) T | \psi_{\text{in}} \rangle$. The result is $\langle \psi_{\text{out}} | U(t) | \psi_{\text{in}} \rangle = t_{\text{out}}^* t_{\text{in}} \langle \psi_{\text{out}} | U(t) | \psi_{\text{in}} \rangle$. 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 T are pure phases, $t_{\text{out}}^* = 1/t_{\text{out}}$. Hence the condition that $t_{\text{out}}^* t_{\text{in}} = 1$ is the same as the statement that t_{in} and t_{out} coincide.

⁶This follows from the given assumption that T_2 is a symmetry, so that it commutes with H . Consequently, $H|\tilde{\psi}_n\rangle = H(T_2|\psi_n\rangle) = T_2(H|\psi_n\rangle) = T_2(E_n|\psi_n\rangle) = E_n|\tilde{\psi}_n\rangle$.

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 (6.2.3) that $T(a)$ commute with H implies that the generator Q of any continuous symmetry must also commute with the Hamiltonian,

$$[Q, H] = 0. \quad (6.3.2)$$

Once again, this implies that Q and H are simultaneously diagonalizable.

Note that, given some continuous symmetry $T(a)$, 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 $Q = -i \frac{d}{da} T(a)|_{a=0}$. 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 (6.3.1). So one can regard either the generator Q , or the finite transformation $T(a)$, as defining a continuous symmetry.

Because the generator Q of a continuous symmetry commutes with the Hamiltonian [Eq. (6.3.2)], Q also commutes with the time evolution operator, $QU(t) = U(t)Q$. This shows that Q defines a *conserved quantity*. Any state which is an eigenstate of Q at some initial time must also be an eigenstate of Q , with exactly the same eigenvalue, at all other times.

6.4 Spacetime symmetries

Spacetime symmetries are symmetries which reflect 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. The additional discrete transformations of parity and time reversal are *not* exact symmetries, and will be discussed below as examples of approximate symmetries.

The total momentum operator \vec{P} (divided by \hbar) is the generator of spatial translations. 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,⁷

$$T_{\text{trans}}(\Delta\vec{x}) = e^{i\vec{P}\cdot\Delta\vec{x}/\hbar}. \quad (6.4.1)$$

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.

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

$$U(\Delta t) = e^{-iH\Delta t/\hbar}. \quad (6.4.2)$$

⁷Recall 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 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 $U(\Delta t)$.⁸ Defining $P^0 \equiv H/c$ allows one to write this as a single exponential of a Minkowski space dot product,

$$T_{\text{trans}}(\Delta x) = T_{\text{trans}}(\Delta \vec{x}) \times U(\Delta x^0/c) = e^{iP^\mu \Delta x_\mu/\hbar}. \quad (6.4.3)$$

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 angular momentum projection along \hat{n} ,

$$T_{\text{rot}}(\theta, \hat{n}) = e^{i\theta \hat{n} \cdot \vec{J}/\hbar}. \quad (6.4.4)$$

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,

$$T_{\text{boost}}(\eta, \hat{n}) = e^{i\eta \hat{n} \cdot \vec{G}/\hbar}. \quad (6.4.5)$$

The parameter η is precisely the *rapidity*, introduced in section 2.6. Recall that the rapidity determines the boost velocity via $v/c = \tanh \eta$. 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.⁹ Because of this, invariance under Lorentz boosts does not lead to any additional conserved quantities analogous to momentum or angular momentum.

6.5 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,

$$Q|\Psi\rangle = q_{\text{tot}}|\Psi\rangle \quad (6.5.1)$$

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 (*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.

⁸Because $T_{\text{trans}}(\Delta \vec{x})$ commutes with $U(\Delta t)$ (or equivalently, because \vec{P} commutes with H), the order in which one performs this product doesn't matter.

⁹The boost generators \vec{G} depend explicitly on time, and the required condition that they must satisfy turns out to be $\frac{\partial}{\partial t} \vec{G} + i[H, \vec{G}] = 0$.

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,

$$T_Q(\alpha) \equiv e^{i\alpha Q}, \quad (6.5.2)$$

with α an arbitrary real number. Applying this transformation to any state multiplies the state by a phase equal to its electric charge times the parameter α , $T_Q(\alpha)|\Psi\rangle = e^{i\alpha q_{\text{tot}}}|\Psi\rangle$. The conserved quantity associated with this continuous symmetry is just the total electric charge.

In precisely the same fashion, one may regard baryon number B and lepton number L as quantum operators which measure the total baryon number or lepton number, respectively. And one may exponentiate either of these operators to form continuous symmetry transformations generated by B and L ,

$$T_B(\alpha) \equiv e^{i\alpha B}, \quad T_L(\alpha) \equiv e^{i\alpha L}. \quad (6.5.3)$$

6.6 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, then the splitting induced by the field will be small (compared to the spacings between 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,

$$H = H_0 + \Delta H. \quad (6.6.1)$$

One can systematically calculate properties of the atom as a power series expansion in the size of ΔH (or more precisely, the size of ΔH divided by differences between eigenvalues of H_0). The starting point involves ignoring ΔH altogether and understanding the properties of H_0 . And when studying the physics of H_0 , one can use the full three dimensional rotation symmetry to characterize energy eigenstates.

In particle physics, exactly the same approach can be used to separate the effects of weak and electromagnetic interactions from the dominant influence of strong interactions.

6.7 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,

$$N_u = (\# u \text{ quarks}) - (\# \bar{u} \text{ quarks}), \quad (6.7.1a)$$

$$N_d = (\# d \text{ quarks}) - (\# \bar{d} \text{ quarks}), \quad (6.7.1b)$$

$$N_s = (\# s \text{ quarks}) - (\# \bar{s} \text{ quarks}), \quad (6.7.1c)$$

$$N_c = (\# c \text{ quarks}) - (\# \bar{c} \text{ quarks}), \quad (6.7.1d)$$

$$N_b = (\# b \text{ quarks}) - (\# \bar{b} \text{ quarks}), \quad (6.7.1e)$$

$$N_t = (\# t \text{ quarks}) - (\# \bar{t} \text{ quarks}), \quad (6.7.1f)$$

all commute with the 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,

$$B = \frac{1}{3} [(\# \text{ quarks}) - (\# \text{ antiquarks})] = \frac{1}{3} \sum_{f=u,d,s,c,b,t} N_f, \quad (6.7.2)$$

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,

$$S \equiv -N_s = (\# \bar{s} \text{ quarks}) - (\# s \text{ quarks}). \quad (6.7.3)$$

This definition assigns strangeness +1 to the K^+ meson (and this convention predates the development of QCD and the quark model of hadrons).

In a world without weak interactions, strangeness S , as well as all the net flavor numbers (6.7.1), would be conserved. In this hypothetical world, charged π mesons would be absolutely stable because there are no combinations of lighter particles with the same values of N_u and N_d which pions could decay into. 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 and electromagnetic interactions). Similarly, the Ω^- baryon, containing three strange quarks, would be stable because there is no other combination of particles, with lower total energy, having baryon number one and strangeness minus three.

Completely analogous arguments apply to hadrons containing heavier charm and bottom quarks. In the absence of weak interactions, there would be many additional stable hadrons containing net charm, or net “bottomness”.

6.8 Isospin

Figure 6.1 graphically displays the mass spectrum of light mesons and baryons. Looking at this figure, or the tables containing information about hadrons in section 5.2, many degeneracies or near-degeneracies are immediately apparent. For example, the masses of the π^+ and π^- mesons are the

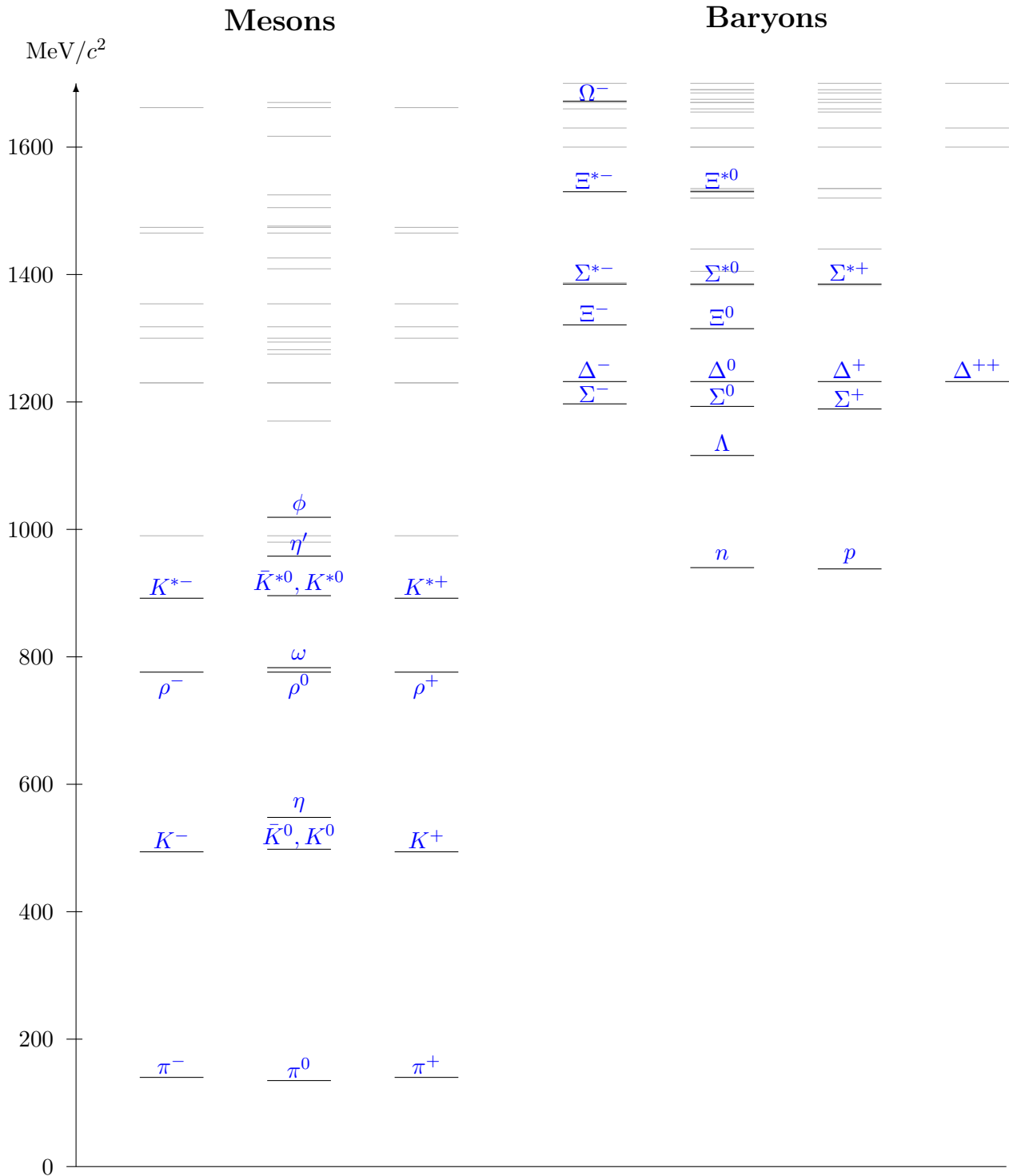


Figure 6.1: The mass spectrum of light mesons and baryons. Each column show mesons or baryons with a particular charge. Only meson states in the lightest $J = 0$ and $J = 1$ nonets, and baryons in the lightest $J = 1/2$ octet and $J = 3/2$ decuplet, are labeled. The unlabeled light grey lines show higher excited states. Many of these higher states are quite broad, with decay widths ranging from 50 MeV to several hundred MeV.

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 each other, but only slightly. And likewise for many other “multiplets” of mesons and baryons. What explains all these near degeneracies?

Comparing the quark content of various hadrons (and referring to Tables 5.3–5.6 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. 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 5.1) 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, the interactions of up and down quarks 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 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.

This should remind you of angular momentum multiplets. 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 $J = \frac{1}{2}\hbar$, whose multiplets contain two (linearly independent) states conventionally chosen to have angular momentum up or down along some given axis, $|\uparrow\rangle$ and $|\downarrow\rangle$. The action of a rotation corresponds to a linear transformation,

$$\begin{pmatrix} |\uparrow'\rangle \\ |\downarrow'\rangle \end{pmatrix} = M \begin{pmatrix} |\uparrow\rangle \\ |\downarrow\rangle \end{pmatrix}. \quad (6.8.1)$$

For a rotation about an axis \hat{n} through an angle θ , the matrix M has the form

$$M = e^{i(\theta/2)\hat{n}\cdot\vec{\sigma}} = \left(\cos \frac{\theta}{2} + i\hat{n} \cdot \vec{\sigma} \sin \frac{\theta}{2} \right), \quad (6.8.2)$$

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 $\sigma_3 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$. You can easily check that M is a unitary matrix with determinant equal to one. The space (or group) of such 2×2 matrices is called $\text{SU}(2)$.

When u and d quarks are degenerate (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,

$$\begin{pmatrix} |u'\rangle \\ |d'\rangle \end{pmatrix} = M \begin{pmatrix} |u\rangle \\ |d\rangle \end{pmatrix}, \quad (6.8.3)$$

where M is any 2×2 unitary matrix with determinant one. This symmetry is called *isospin* (or isotopic spin) rotation.

The mathematical structure of isospin rotations is identical to the structure of spatial rotations (although isospin has nothing to do with ordinary spatial rotations). 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),

$$[I_a, I_b] = i\epsilon_{abc} I_c. \quad (6.8.4)$$

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 $I_3 = +1/2$.¹⁰

When you 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. An antisymmetric combination of u and d quarks,

$$(ud - du), \quad (6.8.5)$$

gives $I = 0$, while a symmetric combination gives isospin one. Hence, the three $I = 1$ flavor states of two quarks are

$$(uu), (ud + du), (dd), \quad (6.8.6)$$

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 $3/2$.

Looking back at the nearly degenerate set of particles shown in Figure 6.1, 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 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 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 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. 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 Λ baryon, with 1690 MeV rest energy and quark content uds . Roughly 25% of the time, a $\Lambda(1690)$ decays to a Λ plus two pions. But it doesn’t decay to a Λ 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 Λ plus two pions are favored, note that the Λ baryon, and its excited states like the $\Lambda(1690)$ have $I = 0$, while pions have $I = 1$. So the decay $\Lambda(1690) \rightarrow \Lambda + \pi$ would have $\Delta I = 1$ — an initial state with isospin zero and a final state of isospin one. But in the final

¹⁰It is $\begin{pmatrix} -|\bar{d}\rangle \\ |\bar{u}\rangle \end{pmatrix}$ which transforms in the same manner as $\begin{pmatrix} |u\rangle \\ |d\rangle \end{pmatrix}$, namely $\begin{pmatrix} -|\bar{d}'\rangle \\ |\bar{u}'\rangle \end{pmatrix} = M \begin{pmatrix} -|\bar{d}\rangle \\ |\bar{u}\rangle \end{pmatrix}$.

state of the observed decay $\Lambda(1690) \rightarrow \Lambda + \pi + \pi$, the total isospin is the combination of two $I = 1$ pions plus the $I = 0$ Λ 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.

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.

6.9 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

$$P|\vec{p}\rangle = \eta_P |-\vec{p}\rangle, \quad (6.9.1)$$

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.¹¹

Applying two parity transformations amounts to reversing the directions of all spatial coordinate axes, and then reversing them all over again. This 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. (6.9.1) 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 $(-1)^\ell$ appears in the result of a parity transformation,¹²

$$P|\Psi_\ell^{A+B}\rangle = \eta_P^A \eta_P^B (-1)^\ell |\Psi_\ell^{A+B}\rangle, \quad (6.9.2)$$

where η_P^A and η_P^B are the intrinsic parities of the individual particles.

¹¹Recall 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} .

¹²This factor comes from the behavior of spherical harmonics under the transformation $\vec{x} \rightarrow -\vec{x}$, namely $Y^{lm}(-\hat{x}) = (-1)^l Y^{lm}(\hat{x})$.

Intrinsic parities can be assigned to particles in such a way that parity is a symmetry of strong and electromagnetic interactions. In particular, the light mesons in Tables 5.3 and 5.4 are all parity-odd (*i.e.*, they have negative intrinsic parity).¹³ The photon is also parity-odd. The baryons listed in Tables 5.5 and 5.6 are all parity-even.

Parity is *not* a symmetry of weak interactions. This will be discussed further in the next chapter. So parity is an approximate symmetry, useful for understanding strong or electromagnetic processes, but is not a true symmetry of nature.

6.10 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 phase factor,

$$C|A\rangle = \eta_C |A\rangle \quad (\text{self-conjugate particles}). \quad (6.10.1)$$

These phases, which depend on particle type, can be defined in such a way that charge conjugation is a symmetry of strong and electromagnetic interactions. 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 flavor quantum numbers (B , L , S , I_3) all have their signs changed by a 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. Pion decay is an electromagnetic process, which 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. 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

¹³All 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. Higher energy even-parity mesons do exist; these may be understood as bound states with non-zero orbital angular momentum.

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. 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. 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_z] is symmetric under interchange of the two spins. But the singlet spin wavefunction, $(\uparrow\downarrow - \downarrow\uparrow)$, is antisymmetric under interchange of spins.

The Pauli principle implies that a state of two fermions must be antisymmetric under interchange of the two fermions. This means interchanging the types, spins, and positions simultaneously. In contrast, charge conjugation in 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 wavefunction. In the $S = 1$ spin triplet (ortho-positronium), swapping the two spins also does not change the value of the wavefunction, since the spin wavefunction is symmetric. Hence, the action of charge conjugation in ortho-positronium is the same as completely interchanging the two particles (because the state is symmetric under interchange of positions and spins). And we know from the Pauli principle that a complete interchange of fermions must flip the sign of the state. Consequently, ortho-positronium must be charge-conjugation odd.

In contrast, 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.

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 positronium is 125 ps, while the lifetime of spin triplet positronium is 142 ns.

6.11 Time reversal and CPT

Time reversal, denoted T , is a transformation which has the effect of flipping the sign of time, $t \rightarrow -t$.¹⁴ 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.¹⁵

¹⁴Because 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, $\mathcal{O}(c_1|\Psi_1\rangle + c_2|\Psi_2\rangle) = c_1^*(\mathcal{O}|\Psi_1\rangle) + c_2^*(\mathcal{O}|\Psi_2\rangle)$.

¹⁵The Wikipedia entry on CPT symmetry has a nice sketch of the proof of the CPT theorem, together with a summary of its history.

Chapter 7

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.

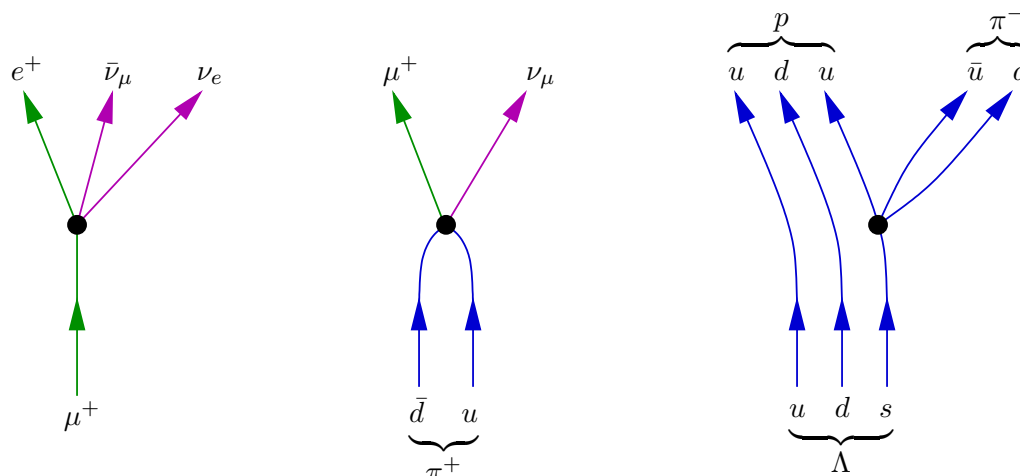


Figure 7.1: Depictions, at the level of quarks and leptons, of the weak decays $\mu^+ \rightarrow e^+ + \bar{\nu}_\mu + \nu_e$, $\pi^+ \rightarrow \mu^+ + \nu_\mu$, and $\Lambda \rightarrow p + \pi^-$.

Figures 7.1 and 7.2 depict, at the level of quarks and leptons, some of these weak interaction processes. As these figures illustrate, every weak interaction involves four fermions, either one fermion turning into three (as in muon decay) or two incoming fermions scattering and producing two outgoing fermions (as in neutrino scattering). As the above Λ baryon decay 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,

$$H = H_{\text{strong}} + H_{\text{EM}} + H_{\text{weak}} . \quad (7.0.1)$$

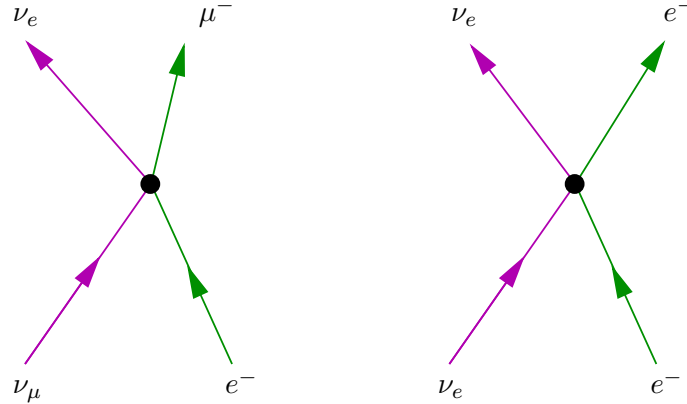


Figure 7.2: Left: inelastic neutrino scattering, $\nu_\mu + e^- \rightarrow \nu_e + \mu^-$. Right: elastic neutrino scattering, $\nu_e + e^- \rightarrow \nu_e + e^-$.

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.

7.1 Muon decay

Consider (anti)muon decay, $\mu^+ \rightarrow e^+ + \bar{\nu}_\mu + \nu_e$. 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

$$M \equiv \langle e(\vec{p}_e) \bar{\nu}_\mu(\vec{p}_{\bar{\nu}}) \nu_e(\vec{p}_\nu) | H_{\text{weak}} | \mu(\vec{p}=0) \rangle, \quad (7.1.1)$$

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,

$$\Gamma \sim \sum_{\text{final states}} |M|^2. \quad (7.1.2)$$

The amplitude M must vanish, due to momentum conservation, if $\vec{p}_e + \vec{p}_{\bar{\nu}} + \vec{p}_\nu \neq 0$. When momentum is conserved, $p_{\bar{\nu}}$ will equal $-(p_\nu + p_e)$, 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 negligible 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 guess turns out to work remarkably well. If the amplitude M is momentum independent then, with just a little calculation, one can perform the sum over final states in Eq. (7.1.2) and predict the muon decay spectrum as a function of positron energy. (That is, the fraction of decays in which the positron has energy between E and $E + dE$.) Figure 7.3 shows the comparison between experimental data for the decay spectrum and the result of this calculation. The agreement is excellent.

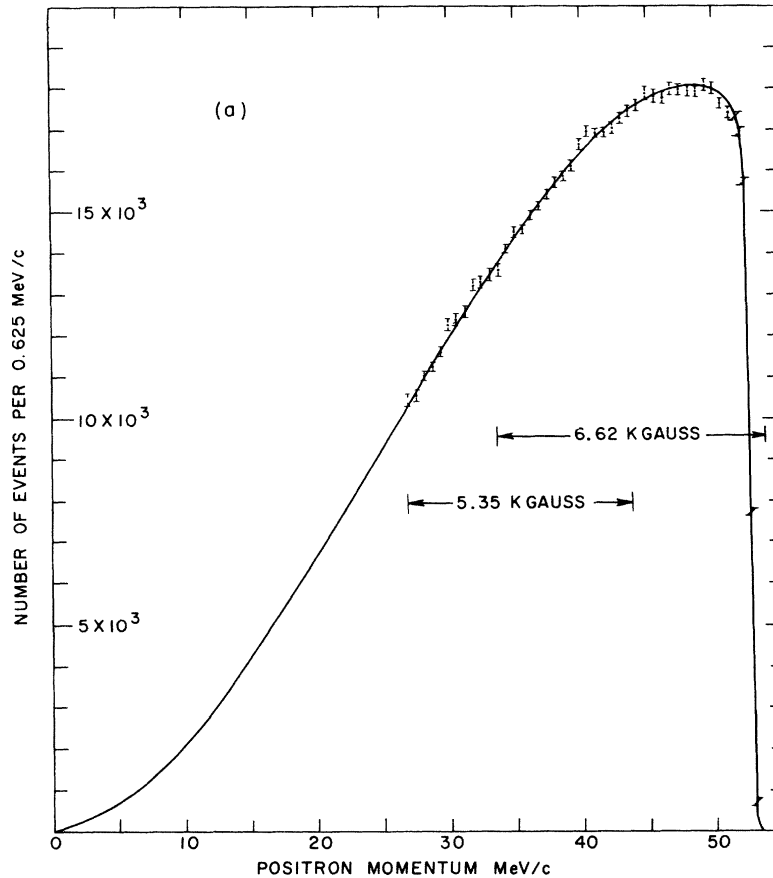


Figure 7.3: Energy spectrum of positrons emitted from decays of positively charged muons. The solid curve is the theoretical prediction; data points are shown with error bars. [From M. Bardon *et al.*, Phys. Rev. Lett. 14, 449 (1965)].

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 $\hbar = c = 1$. 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 , also have dimensions of energy, provided the states appearing in the matrix element are normalized.

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 $1 = \int_{\mathcal{V}} d^3x |\Psi(\vec{x})|^2$, where the integral only includes the interior of the region \mathcal{V} . For simplicity, suppose that this region is a cube of size L (and hence volume L^3). A normalized state describing a particle with momentum \vec{p} will thus have a wavefunction $\Psi(\vec{x}) = e^{i\vec{p}\cdot\vec{x}/\hbar}/L^{3/2}$. 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 $\psi_{\bar{\nu}_\mu}(\vec{x}) = e^{i\vec{p}_{\bar{\nu}}\cdot\vec{x}/\hbar}/L^{3/2}$, 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 which 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 $\psi_e(\vec{x})^* \psi_{\nu_{\mu}}(\vec{x})^* \psi_{\bar{\nu}_e}(\vec{x})^*$, giving the amplitudes for the created positron, neutrino, and antineutrino all to be at point \vec{x} , all times some overall constant which will control the rate of this process,

$$M = \left[\int_{\mathcal{V}} d^3x \psi_e(\vec{x})^* \psi_{\nu_{\mu}}(\vec{x})^* \psi_{\bar{\nu}_e}(\vec{x})^* \psi_{\mu}(\vec{x}) \right] \times (\text{const.}). \quad (7.1.3)$$

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,

$$\psi_e(\vec{x})^* \psi_{\nu_{\mu}}(\vec{x})^* \psi_{\bar{\nu}_e}(\vec{x})^* \psi_{\mu}(\vec{x}) = \frac{e^{-i(\vec{p}_e + \vec{p}_{\nu} + \vec{p}_{\bar{\nu}})\cdot\vec{x}/\hbar}}{(L^{3/2})^4} = L^{-6}, \quad (7.1.4)$$

provided the momenta satisfy conservation of momentum, $\vec{p}_e + \vec{p}_{\nu} + \vec{p}_{\bar{\nu}} = 0$. Integrating over the region \mathcal{V} thus simply yields a factor of the volume, L^3 , of this region. Hence, we find

$$M = \frac{G_F/\sqrt{2}}{L^3}. \quad (7.1.5)$$

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. (7.1.2) agree 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. (7.1.2) 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 $\Gamma = G_F^2 m_{\mu}^5/(192\pi^3)$. Equating this with the inverse of the observed decay rate and solving for G_F yields

$$G_F = 1.2 \times 10^{-5} \text{ GeV}^{-2} = 12 \text{ TeV}^{-2}. \quad (7.1.6)$$

7.2 Neutrino scattering

The significance of this determination of the Fermi constant 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

$$\nu_\mu + e^- \rightarrow \nu_e + \mu^-, \quad (7.2.1)$$

depicted in Fig. 7.2. 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 that¹

$$\sigma \propto |M|^2 \propto G_F^2. \quad (7.2.2)$$

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. One must be able to express the cross section in terms of Lorentz invariant quantities. A Lorentz invariant measure of the scattering energy is $s \equiv -(p_{\nu_\mu} + p_e)^2 = E_{\text{c.m.}}^2$. At low energies, the value of the cross section will also depend on the electron and muon masses. After all, if $E_{\text{c.m.}} < m_\mu c^2$, then the reaction $\nu_\mu + e^- \rightarrow \nu_e + \mu^-$ cannot possibly occur. It must be possible to express the cross section in the (dimensionally consistent!) form

$$\sigma = G_F^2 s \times f\left(\frac{m_e}{\sqrt{s}}, \frac{m_\mu}{\sqrt{s}}\right), \quad (7.2.3)$$

where f is some function of the dimensionless ratios $m_e/E_{\text{c.m.}}$ and $m_\mu/E_{\text{c.m.}}$. (This function will be non-vanishing only when both arguments are less than one.)

The simplest regime to consider is high energy relative to the muon mass, $E_{\text{c.m.}} \gg m_\mu c^2$. In this domain, the ratios $m_e/E_{\text{c.m.}}$ and $m_\mu/E_{\text{c.m.}}$ are both tiny. Since the cross section can be expressed in the form (7.2.3), 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.

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 $E_{\text{c.m.}}$). In the relativistic relation between energy and momentum, the zero mass limit is perfectly smooth, and just leads to the energy-momentum relation of a massless particle,²

$$E(\vec{p}) = \sqrt{\vec{p}^2 + m^2} = |\vec{p}| + \frac{1}{2} \frac{m^2}{|\vec{p}|} + \cdots \xrightarrow{m \rightarrow 0} |\vec{p}|. \quad (7.2.4)$$

¹In 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*.

²In contrast, the non-relativistic energy $E_{\text{NR}}(\vec{p}) = \vec{p}^2/(2m)$ is *not* well-behaved if $m \rightarrow 0$ for fixed momentum \vec{p} .

Similarly, the massless limit of the function $f(\frac{m_e}{\sqrt{s}}, \frac{m_\mu}{\sqrt{s}})$ appearing in the cross section (7.2.3) 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 (7.2.1), the number A is $1/\pi$. Therefore, the inelastic neutrino cross section is given by

$$\sigma_{\nu_e e^- \rightarrow \nu_e \mu^-} = \frac{G_F^2 E_{c.m.}^2}{\pi}, \quad (7.2.5)$$

when $E_{c.m.} \gg m_\mu c^2$. 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

$$\sigma_{\nu_e e^- \rightarrow \nu_e e^-} = 0.175 G_F^2 E_{c.m.}^2. \quad (7.2.6)$$

These predictions of rising neutrino cross sections with increasing energy have been confirmed experimentally for energies in the multi-MeV to multi-GeV range.³ 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

$$\sigma < \frac{\lambda^2}{4\pi} = \frac{\pi}{\vec{p}^2}, \quad (7.2.7)$$

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.

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 $E_{c.m.} \simeq 2|\vec{p}|$. Equating expression (7.2.5) for the neutrino cross section with the unitarity bound (7.2.7), one finds that the cross section (7.2.5) violates the unitarity bound when the center-of-mass energy exceeds

$$E^* \equiv \sqrt{\frac{2\pi}{G_F}} \approx 700 \text{ GeV}. \quad (7.2.8)$$

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.

7.3 Weak gauge bosons

In fact, 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

³See, for example, the plots of the (anti)neutrino-nucleon total cross section at the particle data group website. Note that for neutrino scattering on a nucleus, the lab frame energy is proportional to the square of the center-of-mass energy, $E_{\text{lab}} \propto E_{c.m.}^2$, when E_{lab} is large compared to the target mass. So the quadratic rise of the cross section with $E_{c.m.}$ is equivalent to linear growth as a function of E_{lab} .

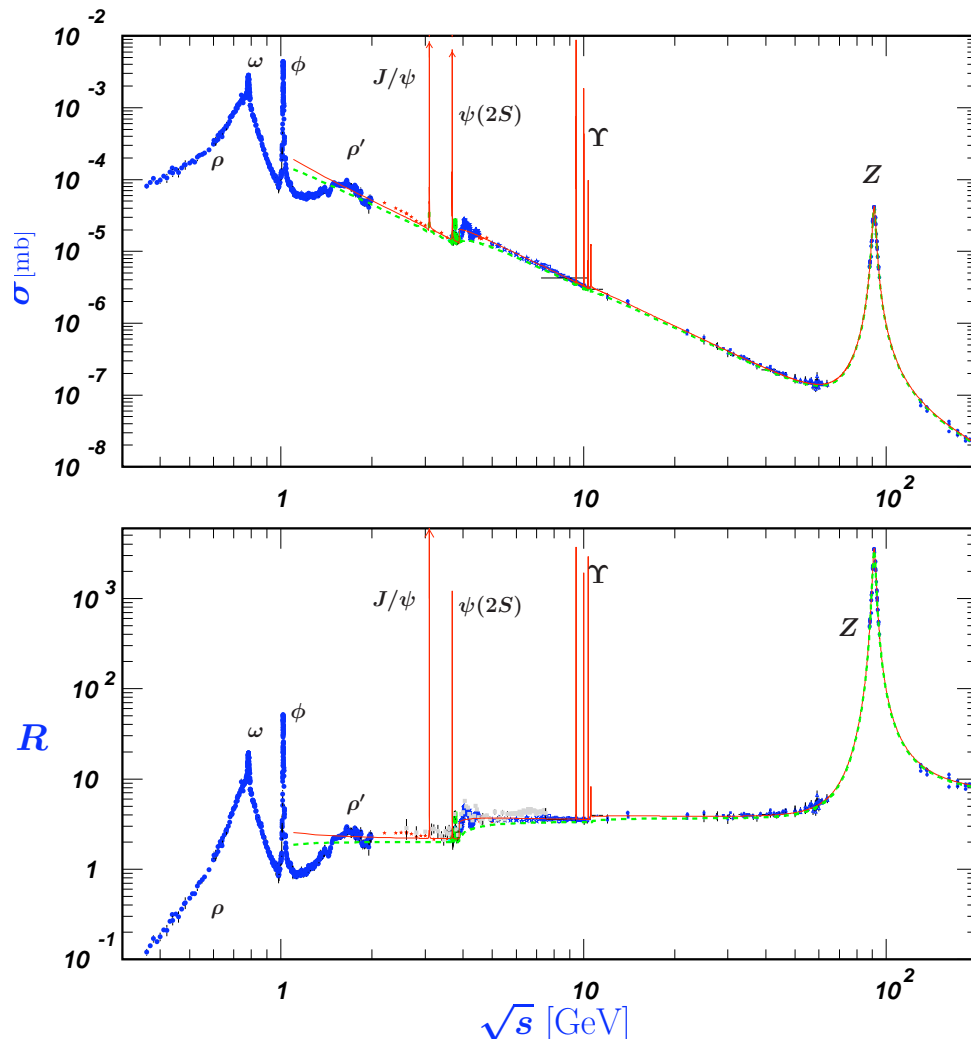


Figure 7.4: Top: Cross section for e^+e^- annihilation to hadrons as a function of $\sqrt{s} = E_{\text{c.m.}}$. Bottom: Ratio of cross sections for e^+e^- annihilation to hadrons versus annihilation to muon pairs, $R = \sigma_{e^+e^- \rightarrow \text{hadrons}} / \sigma_{e^+e^- \rightarrow \mu^+\mu^-}$. [From <http://pdg.lbl.gov/2007/reviews/hadronicrpp.pdf>.]

of both electromagnetic and weak interactions. This turns out to be true. Figure 7.4 shows the cross section for electron-positron annihilation into hadrons as a function of $\sqrt{s} = E_{\text{c.m.}}$. At energies below about 50 GeV, one sees that the cross section generally decreases with increasing energy (note the logarithmic scale), but is punctuated by various spin one, parity odd hadronic resonances — the broad ρ and ρ' , the narrower ω and ϕ , and the very narrow “spikes” associated with $c\bar{c}$ and $b\bar{b}$ heavy quark states. 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 big resonance which is something new. 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 7.4 because a single W^+ or W^- cannot result from e^+e^- annihilation — this would violate charge conservation!

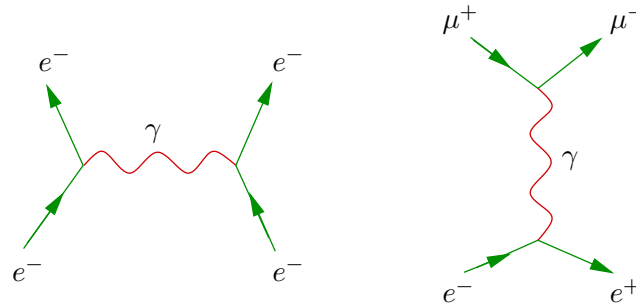


Figure 7.5: Feynman diagrams for Coulomb scattering: $e^-e^- \rightarrow e^-e^-$ (left), and electron-positron annihilation to muons: $e^+e^- \rightarrow \mu^+\mu^-$ (right).

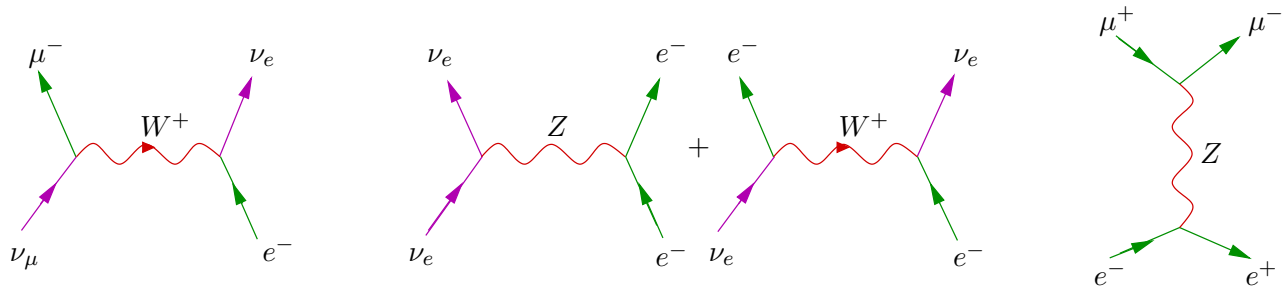


Figure 7.6: Feynman diagrams for inelastic neutrino scattering: $\nu_\mu + e^- \rightarrow \nu_e + \mu^-$ (left), elastic neutrino scattering: $\nu_e + e^- \rightarrow \nu_e + e^-$ (middle), and the weak interaction contribution to $e^+e^- \rightarrow \mu^+\mu^-$ (right).

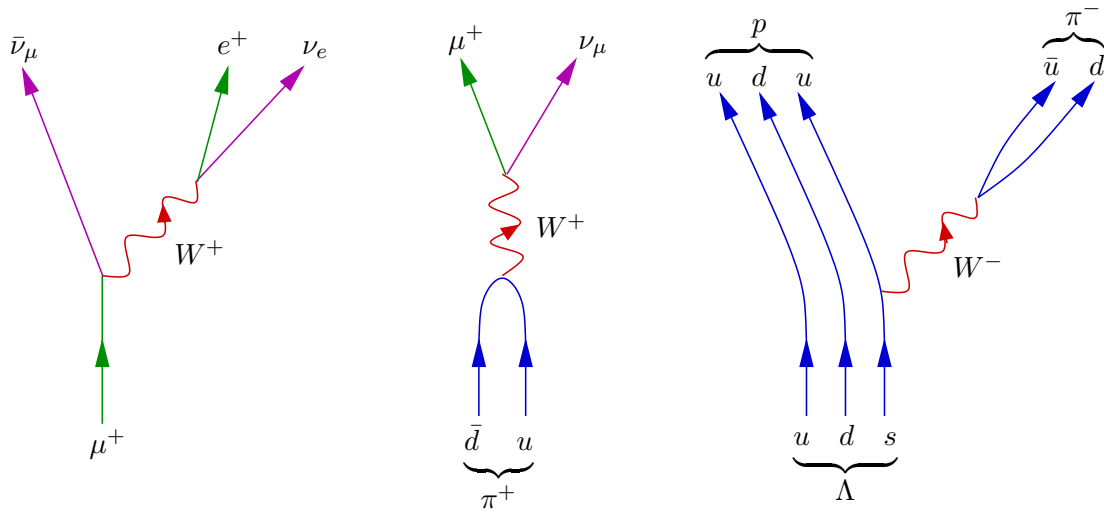


Figure 7.7: Depictions of the weak decays $\mu^+ \rightarrow e^+ + \bar{\nu}_\mu + \nu_e$ (left), $\pi^+ \rightarrow \mu^+ + \nu_\mu$ (middle), and $\Lambda \rightarrow p + \pi^-$ (right), showing the exchange of weak gauge bosons.

Together, the W^\pm and Z are known *weak gauge bosons*. They are spin one particles with masses

$$m_W = 80.4 \text{ GeV}, \quad m_Z = 91.2 \text{ GeV}. \quad (7.3.1)$$

These particles 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 photon, which lives only a very short time before converting into the final μ^+ and μ^- . The diagrams of Figure 7.5 depict these electromagnetic processes.

In the same fashion, weak interactions may be regarded as arising from the exchange of W and Z bosons. Figure 7.6 depicts the same weak interaction scattering processes illustrated in Figure 7.2, plus the weak interaction contribution to $e^+e^- \rightarrow \mu^+\mu^-$, showing the exchange of weak gauge bosons. Figure 7.7 does the same for the weak decays of Figure 7.1. The diagrams of Figures 7.5–7.6 are examples of *Feynman diagrams*. 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 will have to be left for a later class.

With this brief sketch of the current understanding of weak interactions, we must conclude our introduction to particles and symmetries. I hope it has whetted your appetite to learn more about this subject.

Appendix A

Basic mathematics

Elementary functions

The *exponential function* e^x (also denoted $\exp x$) is the unique solution to the differential equation $\frac{df}{dx} = f$ with initial condition $f(0) = 1$. Exponential identities include:

$$\frac{d}{dx}e^x = e^x, \quad e^{-x} = 1/e^x, \quad e^x e^y = e^{x+y}, \quad (e^x)^y = e^{xy}. \quad (\text{A.1})$$

If $z = x + iy$ is a complex number, then $e^z = e^x e^{iy}$.

The *natural logarithm* $\ln x$ is the inverse function of the exponential, and satisfies $\ln(e^x) = x$. Logarithm identities include:

$$\frac{d}{dx} \ln x = 1/x, \quad \ln(xy) = \ln x + \ln y, \quad \ln(x/y) = \ln x - \ln y, \quad \ln(x^y) = y \ln x, \quad x^y = e^{y \ln x}. \quad (\text{A.2})$$

For complex arguments the logarithm is multi-valued, if $z = r e^{i\theta}$ then $\ln z = \ln r + i(\theta + 2\pi n)$ for any integer n . When not specified, $\ln z$ generally refers to the *principal value* for which $-\pi < \text{Im}(\ln z) \leq \pi$.

The *trigonometric functions* $\sin x$ and $\cos x$ are linearly independent solutions of the differential equation $\frac{d^2 f}{dx^2} = -f$. The *sine* function, $\sin x$, is the solution with initial conditions $f(0) = 0$ and $f'(0) = 1$, while the *cosine* function, $\cos x$, is the solution with initial conditions $f(0) = 1$ and $f'(0) = 0$. The function $\sin x$ is an odd function of x , while $\cos x$ is an even function. Both $\sin x$ and $\cos x$ are periodic functions of their argument with period 2π . Basic identities satisfied by trigonometric functions include

$$e^{ix} = \cos x + i \sin x, \quad 1 = \cos^2 x + \sin^2 x, \quad (\text{A.3})$$

along with:

$$\frac{d}{dx} \sin x = \cos x, \quad \cos x = \frac{1}{2}(e^{ix} + e^{-ix}), \quad (\text{A.4})$$

$$\frac{d}{dx} \cos x = -\sin x, \quad \sin x = \frac{1}{2i}(e^{ix} - e^{-ix}), \quad (\text{A.5})$$

$$\sin(x + y) = \sin x \cos y + \cos x \sin y, \quad \sin 2x = 2 \sin x \cos x, \quad (\text{A.6})$$

$$\cos(x + y) = \cos x \cos y - \sin x \sin y, \quad \cos 2x = \cos^2 x - \sin^2 x = 1 - 2 \sin^2 x, \quad (\text{A.7})$$

$$\sin(x + n\pi) = (-1)^n \sin x, \quad \sin(x + \frac{\pi}{2}) = \cos x, \quad (\text{A.8})$$

$$\cos(x + n\pi) = (-1)^n \cos x, \quad \cos(x + \frac{\pi}{2}) = -\sin x. \quad (\text{A.9})$$

(In the periodicity relations (A.8) and (A.9), n must be an integer.) The auxiliary trigonometric functions *tangent*, *secant*, and *cosecant* are defined by

$$\tan x \equiv \sin x / \cos x, \quad \sec x \equiv 1 / \cos x, \quad \csc x \equiv 1 / \sin x, \quad (\text{A.10})$$

respectively.

The *hyperbolic* functions $\sinh x$ and $\cosh x$ are linearly independent solutions to the differential equation $\frac{d^2 f}{dx^2} = f$. The *hyperbolic sine* function, $\sinh x$, is the solution with initial conditions $f(0) = 0$ and $f'(0) = 1$, while the *hyperbolic cosine* function, $\cosh x$, is the solution with initial conditions $f(0) = 1$ and $f'(0) = 0$. The function $\sinh x$ is an odd function of x , while $\cosh x$ is an even function. Basic identities satisfied by hyperbolic functions include

$$e^x = \cosh x + \sinh x, \quad \cosh^2 x - \sinh^2 x = 1, \quad (\text{A.11})$$

along with:

$$\frac{d}{dx} \sinh x = \cosh x, \quad \cosh x = \frac{1}{2}(e^x + e^{-x}) = \cos(ix), \quad (\text{A.12})$$

$$\frac{d}{dx} \cosh x = \sinh x, \quad \sinh x = \frac{1}{2}(e^x - e^{-x}) = -i \sin(ix). \quad (\text{A.13})$$

$$\sinh(x + y) = \sinh x \cosh y + \cosh x \sinh y, \quad \sinh 2x = 2 \sinh x \cosh x, \quad (\text{A.14})$$

$$\cosh(x + y) = \cosh x \cosh y + \sinh x \sinh y, \quad \cosh 2x = \cosh^2 x + \sinh^2 x = 1 + 2 \sinh^2 x. \quad (\text{A.15})$$

The *hyperbolic tangent* $\tanh x \equiv \sinh x / \cosh x$. For real values of x , $\tanh x$ runs from -1 to $+1$ as x runs from $-\infty$ to ∞ .

Series expansions

$$\textit{Binomial} \quad (1 + x)^a = \sum_{k=0}^{\infty} \binom{a}{k} x^k = 1 + ax + \frac{1}{2}a(a-1)x^2 + \dots, \quad (\text{A.16})$$

$$\textit{Logarithmic} \quad \ln(1 + x) = \sum_{k=1}^{\infty} (-1)^k \frac{x^k}{k} = x - \frac{1}{2}x^2 + \frac{1}{3}x^3 - \dots, \quad (\text{A.17})$$

$$\textit{Exponential} \quad e^x = \sum_{k=0}^{\infty} \frac{x^k}{k!} = 1 + x + \frac{1}{2}x^2 + \frac{1}{3!}x^3 + \dots, \quad (\text{A.18})$$

$$\textit{Trigonometric} \quad \cos x = \sum_{k=0}^{\infty} (-1)^k \frac{x^{2k}}{(2k)!} = 1 - \frac{1}{2}x^2 + \frac{1}{4!}x^4 - \dots, \quad (\text{A.19})$$

$$\sin x = \sum_{k=0}^{\infty} (-1)^k \frac{x^{2k+1}}{(2k+1)!} = x - \frac{1}{3!}x^3 + \frac{1}{5!}x^5 - \dots, \quad (\text{A.20})$$

$$\textit{Hyperbolic} \quad \cosh x = \sum_{k=0}^{\infty} \frac{x^{2k}}{(2k)!} = 1 + \frac{1}{2}x^2 + \frac{1}{4!}x^4 + \dots, \quad (\text{A.21})$$

$$\sinh x = \sum_{k=0}^{\infty} \frac{x^{2k+1}}{(2k+1)!} = x + \frac{1}{3!}x^3 + \frac{1}{5!}x^5 + \dots. \quad (\text{A.22})$$

In these series expansions, x may be real or complex. The binomial and logarithmic series converge for $|x| < 1$, while the exponential, trigonometric and hyperbolic series converge for all x . If $|x| \ll 1$, then retaining only the first few terms in these series provides good approximations to the given functions, as successive terms in the series rapidly decrease. (In the binomial series (A.16), if the exponent a is a positive integer, then the expansion terminates after the term with $k = a$.)

Linear algebra

An $N \times N$ matrix M represents a *linear transformation* which may be applied to any N -component vector v . If M_{ij} denotes the row i , column j component of the matrix M , and v_j is the j 'th component of the vector v , then the linear transformation $u = Mv$ may be written explicitly in components as $u_i = M_{ij} v_j$ with an implied sum on the index j (with indices i and j running from 1 to N). One often writes $M = \|M_{ij}\|$ to indicate that M is the matrix constructed from the components M_{ij} , and similarly $v = \{v_j\}$. If A and B are both $N \times N$ matrices, then the matrix product $C = AB$ is equivalent to the component form $C_{ik} = A_{ij} B_{jk}$.

The *identity matrix* $I \equiv \|\delta_{ij}\|$ has components equal to the *Kronecker delta symbol* defined by $\delta_{ij} \equiv \begin{cases} 1 & i = j; \\ 0 & i \neq j, \end{cases}$ and represents the linear transformation which leaves every vector unchanged. The *inverse* of an $N \times N$ matrix M is denoted M^{-1} and, if it exists, satisfies both $M^{-1}M = I$ and $MM^{-1} = I$. The inverse M^{-1} exists provided the determinant of the matrix, denoted by $\det M$ or $|M|$, is non-zero. The linear equation $Mx = y$ has a unique solution given by $x = M^{-1}y$ provided $\det M \neq 0$. If $\det M = 0$ one says that the matrix M is *singular*. (A linear equation with a singular matrix may have zero solutions, or infinitely many solutions, depending on whether the vector y lies in the range of the matrix.)

Given some square matrix M , an *eigenvalue* λ and corresponding *eigenvector* v solve the *eigenvalue equation* $Mv = \lambda v$. The set of all eigenvalues equal the roots of the *characteristic equation* $\det(M - \lambda I) = 0$, which is an N 'th order polynomial in λ .

The *transpose*, *complex conjugate*, and *Hermitian conjugate* of a matrix M are denoted by M^T , M^* , and M^\dagger , respectively, with

$$(M^T)_{ij} \equiv M_{ji}, \quad (M^*)_{ij} \equiv (M_{ij})^*, \quad (M^\dagger)_{ij} \equiv (M_{ji})^*. \quad (\text{A.23})$$

A *symmetric* matrix is equal to its transpose, $M = M^T$. An *antisymmetric* matrix equals minus its transpose, $M = -M^T$. An *Hermitian* matrix is equal to its Hermitian conjugate, $M = M^\dagger$.

An *orthogonal* matrix O is a matrix whose inverse equals its transpose, so that $OO^T = 1$. A *unitary* matrix U is a matrix whose inverse equals its Hermitian conjugate, so that $UU^\dagger = 1$. A real symmetric matrix $M = M^T$ can be *diagonalized* by a real orthogonal transformation. In other words, there exists a real orthogonal matrix O such that $M = O\lambda O^T$ with λ a real diagonal matrix. The diagonal elements $\{\lambda_i\}$ are the eigenvalues of M , and the columns of O are the corresponding mutually orthogonal eigenvectors. Similarly, a complex Hermitian matrix $M = M^\dagger$ can be *diagonalized* by a unitary transformation. In other words, there exists a unitary matrix U such that $M = U\lambda U^\dagger$ with λ a diagonal matrix of real eigenvalues.

Matrix multiplication is non-commutative, meaning that $AB \neq BA$ for arbitrary matrices A and B . In other words, the *commutator* $[A, B] \equiv AB - BA$ is generally non-zero (but vanishes in special cases where the product is independent of order). Two Hermitian matrices A and B are *simultaneously diagonalizable* if and only if their commutator vanishes. If the condition $[A, B] = 0$ holds, then there exists a single unitary matrix U such that $A = U\lambda^A U^\dagger$ and $B = U\lambda^B U^\dagger$ with λ^A and λ^B both real and diagonal. Equivalently, each column of U is an eigenvector of both A and B , with eigenvalues for each matrix given by the corresponding diagonal elements of λ^A and λ^B .

Vector spaces

The above relations involving $N \times N$ matrices and N component vectors generalize in a natural fashion to arbitrary *vector spaces*. This name refers to any collection of objects (such as N component column vectors, geometric vectors, polynomials, or various classes of functions) in which it makes sense to add or subtract any two elements (or “vectors”) of the space, or multiply any element by an overall constant. Vector spaces can be real or complex, depending of whether it makes sense to multiply elements only be real numbers, or by arbitrary complex numbers. And vector spaces can be finite or infinite dimensional. Given some vector space, a *basis* for the space is a set of linearly independent elements, $\{\hat{e}_a\}$, (with $a = 1, 2, \dots$), such that any vector x in the space can be written as a linear combination of the basis elements,

$$x = \sum_a \hat{e}_a x_a, \tag{A.24}$$

for some set of coefficients $\{x_a\}$. If the vector space is N dimensional, with N finite, then a basis for the space will contain N basis vectors (so the index a labeling basis elements runs from 1 to N). If the vector space is infinite dimensional, then so is any set of basis elements. In this case, the index a labeling basis elements runs from 1 to ∞ .¹ (Or, in some circumstances, the natural label is a real number, in which case the sum (A.24) is replaced by an integral over this label.)

An *inner product* (or *dot product*) is some function which takes two vectors as arguments and returns a single number — a real number for real vector spaces, a complex number for complex vector spaces. The inner product of vectors x and y is commonly denoted as $\langle x, y \rangle$, (x, y) , or $x \cdot y$. Regardless of which notation is used, an inner product must satisfy:

$$\text{Symmetry} \quad \langle x, y \rangle = \langle y, x \rangle^*, \tag{A.25}$$

$$\text{Linearity} \quad \langle x, \alpha y \rangle = \alpha \langle x, y \rangle \quad \text{and} \quad \langle x, y+z \rangle = \langle x, y \rangle + \langle x, z \rangle, \tag{A.26}$$

$$\text{Positivity} \quad \langle x, x \rangle \geq 0, \tag{A.27}$$

$$\text{Non-degeneracy} \quad \langle x, x \rangle = 0 \text{ implies } x = 0. \tag{A.28}$$

Two vectors are *orthogonal* if their inner product vanishes. A basis is *orthonormal* if basis elements are mutually orthogonal, and every basis element is normalized so that its inner product with itself is unity. More succinctly, a basis is orthonormal if $\langle \hat{e}_a, \hat{e}_b \rangle = \delta_{ab}$.

A function T which acts on elements of a vector space and returns some element in the same vector space is called a *linear operator* if it satisfies the linearity conditions

$$T(\alpha x) = \alpha T(x), \quad T(x+y) = T(x) + T(y). \tag{A.29}$$

Here α is an arbitrary real number for real vector spaces, or arbitrary complex number for complex spaces. A linear operator T is *Hermitian* if $\langle x, Ty \rangle = \langle Tx, y \rangle$ for all vectors x and y .

Given an orthonormal basis, determining the expansion coefficients $\{x_a\}$ of an arbitrary vector x is easy: they are simply given by the inner product of x with each basis element, $x_a = \langle \hat{e}_a, x \rangle$. In other

¹In an infinite dimensional vector space, one may rightly ask whether the infinite sum (A.24) will converge for all vectors, or only for some vectors. A more formal mathematics class would carefully address this question, but for our purposes the claim that the sum will always make sense in physically sensible situations will have to suffice.

words, $x = \sum_a \hat{e}_a \langle \hat{e}_a, x \rangle$ for any vector x . This may be written without reference to any specific vector x as the *completeness relation*

$$I = \sum_a P_a, \tag{A.30}$$

where P_a is a *projection operator* onto vectors proportional to \hat{e}_a , and I is the identity operator which leaves all vectors invariant. (Explicitly, $P_a x \equiv \hat{e}_a \langle \hat{e}_a, x \rangle$ for any vector x .)

The above structure regarding abstract vector spaces, linear operators, and inner products is a natural generalization of N -component vectors, $N \times N$ matrices, and the usual definition of dot product. Definitions of eigenvectors and eigenvalues, and the above results on diagonalizability generalize directly from $N \times N$ matrices to arbitrary linear operators.

Appendix B

Quantum mechanics

Quantum kinematics

Physical states of a quantum system may be represented by elements of a complex linear vector space, (or more formally, a *Hilbert space*), and are symbolized by a *ket vector* $|\psi\rangle$. The corresponding *bra vector* $\langle\psi|$ denotes the Hermitian conjugate of $|\psi\rangle$. A bra vector $\langle\phi|$ may be applied to a ket vector $|\psi\rangle$ and defines the complex inner product of the two state vectors, $\langle\phi|\psi\rangle$. Properly normalized states satisfy $\langle\psi|\psi\rangle = 1$. If $|\phi\rangle$ and $|\psi\rangle$ are normalized states, then their inner product $\langle\phi|\psi\rangle$ may be interpreted as giving the projection of the state vector $|\psi\rangle$ onto the state $|\phi\rangle$. The funny “bra-ket” names and notation were introduced by Dirac, and give a distinctive and convenient notation for this physically important inner product. The representation of physical states by vectors in a linear space encodes the *superposition principle*: if $|\psi\rangle$ and $|\phi\rangle$ are two possible states of some physical system, then any complex linear combination $\alpha|\psi\rangle + \beta|\phi\rangle$ is also a possible physical state.

Physical observables — i.e., physical quantities which can, in principle, be measured — are represented by Hermitian linear operators. Suppose \hat{A} is such an operator.¹ Being Hermitian, the eigenvalues $\{\alpha_a\}$ of \hat{A} are real, and the corresponding eigenvectors (or “eigenstates”) $|v_a\rangle$ are mutually orthogonal (and may be chosen to be normalized). If a physical system is prepared in some state $|\psi\rangle$ and a measurement of the quantity A is performed, the result (of an ideal measurement with negligible experimental error) will yield one of the eigenvalues α_a . If one repeatedly prepares the state $|\psi\rangle$ and measures \hat{A} , individual measurements in identically prepared states may yield differing eigenvalues of \hat{A} . If this process is repeated many, many times, the fraction p_a of measurements which yield a particular eigenvalue α_a will be given by the absolute square of the projection of the initial state $|\psi\rangle$ onto the corresponding eigenstate $|v_a\rangle$ of the observable,

$$p_a = |\langle v_a|\psi\rangle|^2. \tag{B.1}$$

This is the probability that any single measurement of \hat{A} in the state $|\psi\rangle$ will yield the result α_a . The projection $\langle v_a|\psi\rangle$ is called a *probability amplitude*; physical probabilities are always given by the (absolute) square of the probability amplitude. If repeated measurements of identically prepared

¹Here and below, carets (\wedge) are placed over quantum operators to distinguish them from ordinary numbers — not to indicate unit vectors.

states are averaged together, the mean value of these measurements is

$$\langle \hat{A} \rangle \equiv \sum_a p_a \alpha_a = \sum_a \langle \psi | v_a \rangle \alpha_a \langle v_a | \psi \rangle = \langle \psi | \hat{A} | \psi \rangle. \quad (\text{B.2})$$

This is referred to as the *expectation value* of the operator \hat{A} in the state $|\psi\rangle$. Relations (B.1) and (B.2) are the fundamental predictions of quantum mechanics, and serve to connect the mathematical formalism to experimental observations. Quantum mechanics makes no prediction regarding the precise sequence of results obtained from repeated measurements of identically prepared systems, it only predicts the probabilities of different outcomes.

Depending on the system under consideration, the space of states may be finite dimensional or infinite dimensional. The simplest non-trivial quantum system is one with a two dimensional space of states. This can, for example, describe the spin state of a single spin-1/2 particle. Any such spin state can be written as a linear combination of two basis states which may be chosen to represent the spin pointing either up or down along some chosen axis. Denoting these basis states as $|\uparrow\rangle$ and $|\downarrow\rangle$, an arbitrary (normalized) spin state $|\psi\rangle = \alpha|\uparrow\rangle + \beta|\downarrow\rangle$ for some coefficients α and β satisfying $|\alpha|^2 + |\beta|^2 = 1$. Equivalently, the state $|\psi\rangle$ may be represented by the complex two-component vector $\begin{pmatrix} \alpha \\ \beta \end{pmatrix}$, so that $|\uparrow\rangle$ corresponds to $\begin{pmatrix} 1 \\ 0 \end{pmatrix}$ and $|\downarrow\rangle$ corresponds to $\begin{pmatrix} 0 \\ 1 \end{pmatrix}$.

If one chooses the z -axis as the spin quantization axis, then the operators representing the components of angular momentum (or “spin”) in the x , y , and z directions are given by $\hat{J}_1 = \frac{\hbar}{2} \sigma_1$, $\hat{J}_2 = \frac{\hbar}{2} \sigma_2$, and $\hat{J}_3 = \frac{\hbar}{2} \sigma_3$, respectively, where the $\{\sigma_i\}$ denote the 2×2 *Pauli matrices*,

$$\sigma_1 \equiv \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad \sigma_2 \equiv \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \quad \sigma_3 \equiv \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}. \quad (\text{B.3})$$

In this representation, \hat{J}_3 is diagonal and has eigenvalues $\pm\hbar/2$ with corresponding eigenvectors $\begin{pmatrix} 1 \\ 0 \end{pmatrix}$ and $\begin{pmatrix} 0 \\ 1 \end{pmatrix}$ representing the states $|\uparrow\rangle$ and $|\downarrow\rangle$ with definite projection of spin along the z -axis. Squaring any Pauli matrix gives the 2×2 identity matrix, $\sigma_i^2 = I$, so for any system with spin-1/2, the operator $\hat{J}^2 \equiv \sum_i \hat{J}_i^2 = \frac{3}{4} \hbar^2 \hat{1}$ (where $\hat{1}$ is the identity operator). In other words, every spin state of an object with spin 1/2 is an eigenstate of \hat{J}^2 with eigenvalue $\frac{3}{4} \hbar^2$.

The angular momentum of any quantum mechanical system is quantized in either integer or half-integer multiples of \hbar . A system with angular momentum j has a $2j + 1$ dimensional space of spin states. The projection of angular momentum along some chosen spin quantization axis, say the z -axis, can take on $2j + 1$ possible values: $j, j - 1, \dots, -(j - 1), -j$ (times \hbar). These are the eigenvalues of the angular momentum operator \hat{J}_3 . Different components of angular momentum do not commute, and hence are not simultaneously diagonalizable. Rather, they satisfy the commutation relation

$$[\hat{J}_i, \hat{J}_j] = i\hbar \epsilon_{ijk} \hat{J}_k, \quad (\text{B.4})$$

where the antisymmetric symbol ϵ_{ijk} is +1 if (ijk) is an even permutation of (123), -1 if (ijk) is an odd permutation of (123), and 0 otherwise. The square of the angular momentum $\hat{J}^2 \equiv \sum_i \hat{J}_i^2$ does commute with the individual angular momentum components. For a system with angular momentum j , \hat{J}^2 is proportional to the identity operator,

$$\hat{J}^2 = j(j + 1) \hbar^2 \hat{1}. \quad (\text{B.5})$$

Hence, specifying the spin j of a system is equivalent to specifying the eigenvalue of \hat{J}^2 .

In many circumstances (including multi-particle systems like atoms or molecules) one is interested in a quantum system containing two or more subsystems. Suppose a system has two subsystems, A and B . The angular momentum of the total system is the sum of the angular momenta of the subsystems, $\vec{J}^{\text{tot}} = \vec{J}^A + \vec{J}^B$. Suppose subsystem A is known to have spin j^A , and subsystem B has spin j^B . Hence $(\vec{J}^A)^2 = j^A(j^A + 1)$ and $(\vec{J}^B)^2 = j^B(j^B + 1)$. One can show that the total spin j^{tot} of the combined system must be one of the values $\{j_{\min}, j_{\min}+1, \dots, j_{\max}-1, j_{\max}\}$ where

$$j_{\min} = |j^A - j^B|, \quad j_{\max} = j^A + j^B. \quad (\text{B.6})$$

Any system which moves (in ordinary 3-dimensional space) will have an infinite dimensional space of states. In such systems, the position \vec{X} and momentum \vec{P} are basic quantum operators. Position operators mutually commute, $[\hat{X}_i, \hat{X}_j] = 0$, and hence may be simultaneously diagonalized. One may introduce a basis of position eigenstates $\{|\vec{x}\rangle\}$ satisfying the delta-function orthonormality relation,

$$\langle \vec{x} | \vec{y} \rangle = \delta^3(\vec{x} - \vec{y}), \quad (\text{B.7})$$

and corresponding completeness relation

$$\hat{1} = \int d^3x |\vec{x}\rangle \langle \vec{x}|. \quad (\text{B.8})$$

An arbitrary quantum state $|\psi\rangle$ (with no additional degrees of freedom) may be represented as a superposition of position eigenstates,

$$|\psi\rangle = \int d^3x |\vec{x}\rangle \Psi(\vec{x}). \quad (\text{B.9})$$

The amplitude $\Psi(\vec{x})$ is the *position space wavefunction*, equal to the projection of the given state $|\psi\rangle$ onto position eigenstates, $\Psi(\vec{x}) \equiv \langle \vec{x} | \psi \rangle$. Momentum operators also mutually commute, $[\hat{P}_i, \hat{P}_j] = 0$, and may be simultaneously diagonalized. It is convenient to define a basis of momentum eigenstates $\{|\vec{p}\rangle\}$ to satisfy the orthonormality relation

$$\langle \vec{p} | \vec{p}' \rangle = (2\pi\hbar)^3 \delta^3(\vec{p} - \vec{p}'), \quad (\text{B.10})$$

so the corresponding completeness relation takes the form

$$\hat{1} = \int \frac{d^3p}{(2\pi\hbar)^3} |\vec{p}\rangle \langle \vec{p}|. \quad (\text{B.11})$$

An arbitrary quantum state $|\psi\rangle$ may also be represented as a superposition of momentum eigenstates,

$$|\psi\rangle = \int \frac{d^3p}{(2\pi\hbar)^3} |\vec{p}\rangle \tilde{\Psi}(\vec{p}), \quad (\text{B.12})$$

where the *momentum space wavefunction* $\tilde{\Psi}(\vec{p})$ is the projection of the given state $|\psi\rangle$ onto momentum eigenstates, $\tilde{\Psi}(\vec{p}) \equiv \langle \vec{p} | \psi \rangle$. Corresponding components of position and momentum do not commute,

$$[\hat{P}_i, \hat{X}_j] = \frac{\hbar}{i} \delta_{ij}, \quad (\text{B.13})$$

and hence cannot be simultaneously diagonalized. This non-vanishing commutator leads to the *uncertainty relation* $\Delta x_i \Delta p_j \geq \frac{1}{2} \hbar \delta_{ij}$, where $\Delta x_i \equiv [\langle \hat{X}_i^2 \rangle - \langle \hat{X}_i \rangle^2]^{1/2}$ is the uncertainty (or standard deviation) of \hat{X}_i in some given quantum state, and $\Delta p_j \equiv [\langle \hat{P}_j^2 \rangle - \langle \hat{P}_j \rangle^2]^{1/2}$ is the analogous uncertainty of \hat{P}_j in the same state.

The overlap between position and momentum eigenstates is given by an exponential phase factor,

$$\langle \vec{x} | \vec{p} \rangle = e^{i\vec{p}\cdot\vec{x}/\hbar}. \quad (\text{B.14})$$

In other words, the position space wavefunction of the state $|\vec{p}\rangle$ with definite momentum is a plane wave, $e^{i\vec{p}\cdot\vec{x}/\hbar}$. More generally, position and momentum space wavefunctions of any state $|\psi\rangle$ are related via Fourier transforms,

$$\tilde{\Psi}(\vec{p}) = \int d^3x \langle \vec{p} | \vec{x} \rangle \langle \vec{x} | \psi \rangle = \int d^3x e^{-i\vec{p}\cdot\vec{x}/\hbar} \Psi(\vec{x}), \quad (\text{B.15})$$

$$\Psi(\vec{x}) = \int \frac{d^3p}{(2\pi\hbar)^3} \langle \vec{x} | \vec{p} \rangle \langle \vec{p} | \psi \rangle = \int \frac{d^3p}{(2\pi\hbar)^3} e^{i\vec{p}\cdot\vec{x}/\hbar} \tilde{\Psi}(\vec{p}). \quad (\text{B.16})$$

Quantum dynamics

The state of a quantum system will evolve with time. If $|\psi(0)\rangle$ denotes the state of some quantum system at time $t = 0$, then the state $|\psi(t)\rangle$ of the system at some later time t is the solution of the linear evolution equation

$$i\hbar \frac{d}{dt} |\psi(t)\rangle = \hat{H}(t) |\psi(t)\rangle, \quad (\text{B.17})$$

known as the *time dependent Schrodinger equation*, with the specified initial condition $|\psi(0)\rangle$. The linear operator $\hat{H}(t)$ is called the *Hamiltonian*; acting on any state at time t , $\hat{H}(t)$ yields the time derivative of the state (times $i\hbar$). The Hamiltonian must be Hermitian, $\hat{H}^\dagger = \hat{H}$, which implies that its eigenvalues are real. If the Hamiltonian has no explicit time dependence, $\hat{H}(t) = \hat{H}$, then the solution of the Schrodinger equation (B.17) involves an exponential of the Hamiltonian,²

$$|\psi(t)\rangle = e^{-i\hat{H}t/\hbar} |\psi(0)\rangle. \quad (\text{B.18})$$

The exponential $\hat{U}(t) \equiv e^{-i\hat{H}t/\hbar}$ is the *time evolution operator*; the Hermiticity of the Hamiltonian implies that (for real values of t) the time evolution operator is unitary, $\hat{U}(t)^{-1} = \hat{U}(t)^\dagger = \hat{U}(-t)$.

In addition to defining time evolution, the Hamiltonian is also the operator which represents the *energy* of a quantum system. Eigenvalues $\{E_a\}$ of the Hamiltonian are called *energy levels*. The corresponding eigenstates $\{|v_a\rangle\}$ satisfy the *time independent Schrodinger equation*

$$\hat{H} |v_a\rangle = E_a |v_a\rangle. \quad (\text{B.19})$$

If an initial state equals some eigenstate of the Hamiltonian, $|\psi(0)\rangle = |v_a\rangle$, then the state merely acquires an overall phase factor under time evolution,

$$|\psi(t)\rangle = e^{-iE_a t/\hbar} |v_a\rangle. \quad (\text{B.20})$$

Such states are called *stationary states*. The expectation value of any quantum operator in such a state is independent of time.

²The exponential of an Hermitian operator (or matrix) is defined by the exponential of its eigenvalues in a basis where the operator is diagonal. If an Hermitian operator \hat{A} has eigenvalues α_a and corresponding eigenvectors $|v_a\rangle$, then $e^{\hat{A}} \equiv \sum_a |v_a\rangle e^{\alpha_a} \langle v_a|$.

Index

- G_F , *see* Fermi constant
- γ , *see* Lorentz factor
- 3-velocity, 34
- 4-acceleration, 34
- 4-force, 34
- 4-momentum, 32
 - conservation, 40
- 4-vector, *see* spacetime vector
- 4-velocity, 31

- antimatter, 54
 - antibaryon, 59
 - antilepton, 54
 - antineutrino, 54

- baryon, 59
 - decuplet, 60
 - number, 65
 - octet, 60
 - wavefunction, 62
- basis, 96
- beta decay, 52
- boost, *see* Lorentz transformation
- boson, 56
- bra vector, 69, 98

- causality, 24
- charge conjugation, 81
- charge quantization, 58, 74
- CM frame, 40
- color, *see* quark, color
- commutator, 95
- conserved quantity, 73
- cosmic ray, 7
- CPT, 83

- Doppler shift, 4
- dot product, *see* inner product

- eigenvalue, 95

- electromagnetism, 39, 50
- electron, 47
- energy, 33
- event, 10

- Fermi constant, 87
- fermion, 56
- Feynman diagrams, 92
- field strength tensor, 39
- flavor, *see* quark, flavor
- force, 34
- frame
 - inertial, 3, 10
 - reference, 3, 10
 - rest, 10, 32
- function
 - elementary, 93
 - exponential, 93
 - hyperbolic, 94
 - logarithm, 93
 - trigonometric, 93

- Galilean relativity, 3
- gamma rays, 53
- GPS, 7
- gravity, 50

- hadron, 51, 57
 - decay, 65
 - spectrum, 77
- Hamiltonian, *see* quantum Hamiltonian

- ideal clock, 5
- implied summation, 17, 95
- inner product, 96
- interaction
 - electromagnetic, 50
 - gravitational, 50
 - strong, 50
 - weak, 50

- intrinsic parity, 80
- invariance, 3
- invariant interval, 20
- isospin, 76

- ket vector, 69, 98
- kinetic energy, 33
- Kronecker delta, 95

- lab frame, 40
- lepton, 54
 - number, 55
- lightcone, 12
- lightlike, 20
- linear
 - algebra, 95
 - operator, 96
- Lorentz
 - contraction, 7
 - factor, 6
 - force, 39
 - transformation, 14

- matrix, *see* linear algebra
- meson, 59
- Minkowski space, *see* spacetime
- momentum, 32
- muon, 55
 - decay spectrum, 86
 - shower, 7

- neutrino, 54
 - anti-, 52
- neutron, 48
- Newton's laws, 3
- nucleon, 50
- nucleus, 47
- nuclide, 48

- parity, 80
- Pauli principle, 56, 60
- photon, 53
- plane wave, 19, 37
- positron, 53
- proper time, 30
- proton, 48

- QCD, *see* quantum chromodynamics

- quantum
 - dynamics, 69, 101
 - Hamiltonian, 101
 - kinematics, 98
 - mechanics, 98
 - time evolution, 101
- quantum chromodynamics, 57
- quark, 50, 57
 - color, 57
 - flavor, 57
 - masses, 58

- rapidity, 16
- reference frame, *see* frame, reference
- relativity postulates, 5
- rest energy, 33
- rest frame, *see* frame, rest
- rest mass, 32

- scattering, 39
- Schrodinger equation, 69
- simultaneity, 12
- spacelike, 20
- spacetime, 10
 - dot product, 19
 - Minkowski, 19
 - vector, 17
- speed of light, 4
- spin-statistics theorem, 56
- strangeness, 76
- strong interaction, 50, 57
- surface of simultaneity, 11
- symmetry, 69
 - approximate, 75
 - continuous, 72
 - crossing, 88
 - generator, 72
 - spacetime, 73

- tau lepton, 55
- time dilation, 6
- time reversal, 83
- timelike, 20

- unitarity bound, 89
- units, 42
 - natural, 42, 86

vector space, 96

velocity

 3-velocity, 34

 4-velocity, 31

 relativistic addition, 44

weak gauge boson, 92

weak interaction, 50, 52, 84

worldline, 10

worldvolume, 10