Chapter 2

Introduction to Particle Physics

If I could remember the names of all the particles, I'd be a botanist.
— Enrico Fermi

Matter

At its most basic level, all matter consists of combinations of 12 elementary particles, which are listed in Fig. 2.1. They can be classified into two groups, leptons and quarks: quarks interact via the strong force but leptons do not. Both types of particles interact gravitationally (i.e., they all have mass) and via the weak force. Finally, all but the neutrinos interact electromagnetically because neutrinos are electrically neutral. The original motivation for the classification of leptons in 1947 was that the electron (the only known lepton at that time) was less massive than the proton and neutron (the only known nucleons—later determined to consist of quarks), and "lepton" is from a Greek word that means small or light. (See page 20.) Of course, after the discovery of the tau lepton in 1975 and the observation that it was almost twice as massive as a proton, the original reason no longer made sense. However, with the discovery of quarks and the fact that they are the only particles to interact via the strong force, the division into leptons and quarks is appropriate, albeit for reasons that have to do with forces rather than mass.[1]

Amazingly, all natural matter that we observe in the world around us consists of only three of these particles: electrons, up quarks, and down quarks. The atoms in our bodies are comprised of electrons as well as protons and neutrons, but the proton is made up of 2 up quarks and 1 down quark (commonly written 'uud'), while the neutron is 2 down quarks and 1 up quark (commonly written 'udd'). In this sense, the universe is very simple. There are only three particles, which combine in a myriad of ways to make up all the wonderful objects that we see: trees, rivers, oceans, mountains, planets, stars, and galaxies.

What are the intrinsic properties of these elementary particles? Two are very familiar, mass and electric charge, and three others, spin, magnetic moment, and color, are not as

[1] In addition to these 12 particles, there are the so-called "exchange particles," like the photon (denoted by the symbol γ), that mediate the four forces. These particles are also called "gauge bosons," or "intermediate vector bosons," and they are not normally considered to be matter. I will discuss them below on page 23.

e^-	**electron**	Leptons
ν_e	electron neutrino	
μ^-	muon (mu lepton)	
ν_μ	muon neutrino	
τ^-	tauon (tau lepton)	
ν_τ	tau neutrino	
u	**up quark**	Quarks
d	**down quark**	
c	charm quark	
s	strange quark	
t	top (truth) quark	
b	bottom (beauty) quark	

Figure 2.1: The twelve elementary particles that comprise all natural and man-made matter. The three particles in boldface — electron, up quark, and down quark — comprise all known natural matter. There are six leptons (three massive leptons and three massless neutrinos) and six *flavors* of quarks.

familiar. We will examine these five in detail in Sections 2.1 through 2.5. Of course, there are many others, such as strangeness, isotopic spin, lepton number, and baryon number, and we will investigate these in later chapters. The nomenclature of particle physics is very complicated, but if you remember to characterize particles based on their fundamental properties, like mass, charge, etc., it doesn't matter what they are called, you will be able to understand the physics of their interactions.

You may have noticed that I didn't mention size as an intrinsic property. The reason is that all of these elementary particles are thought to be point-like and have no size. For example, the size of an electron has been experimentally measured to be less than 10^{-22} meters![2] This simply means that the electric force that an electron feels is Coulombic (i.e., $\sim 1/r^2$) down to that distance, which means that there is no reason to think that electrons have any structure at any scale. Of course, when elementary particles combine to form protons, neutrons, atoms, and molecules, the physics of their interaction occurs on a spatial scale so that the conglomerations acquire a characteristic size and shape.

There is another characteristic of these particles that has no classical counterpart: they are identical and indistinguishable. Unlike our macroscopic world, where we can paint seemingly identical objects different colors to distinguish them (billiard balls, for example), in the microscopic world there is no way to tell two electrons apart. When a cue ball, say, collides with an eight-ball and they each move off in different directions, it is clear which ball is which after the collision. However, if two electrons collide and move off, the experimenter is not able to distinguish which electron is which after the collision. As we will see below, this fact has far-reaching implications on the allowable motions of these particles. The most well-known implication is the Pauli exclusion principle that is applied to electrons within atomic orbitals, which I will discuss in Chapter 4.

[2]Hans Dehmelt, "A Single Atomic Particle Forever Floating at Rest in Free Space: New Value for Electron Radius," *Phys. Scr.* **T22** 102-110 (1988)

e^+	positron (anti electron)	
$\overline{\nu}_e$	anti electron neutrino	
μ^+	anti muon (mu lepton)	anti Leptons
$\overline{\nu}_\mu$	anti muon neutrino	
τ^+	anti tauon (tau lepton)	
$\overline{\nu}_\tau$	anti tau neutrino	
$\overline{\mathrm{u}}$	anti up quark	
$\overline{\mathrm{d}}$	anti down quark	
$\overline{\mathrm{c}}$	anti charm quark	anti Quarks
$\overline{\mathrm{s}}$	anti strange quark	
$\overline{\mathrm{t}}$	anti top (truth) quark	
$\overline{\mathrm{b}}$	anti bottom (beauty) quark	

Figure 2.2: The twelve elementary antiparticles.

Antimatter

Antimatter is as much matter as matter is matter. — Abraham Pais[3]

For every particle, there is a corresponding "antiparticle," with the same mass, but opposite electric charge, and these are listed in Fig. 2.2. The antiparticles are denoted by an overbar, or sometimes by simply changing the sign, as with the positron. Do not ascribe any mysterious properties to antimatter. As Pais implies, from an antiparticle's point of view, *we* are made of "antimatter." In fact, current cosmological theories suggest that in the early universe, a short time after the Big Bang, there was approximately as much matter as antimatter. As the universe cooled, equal amounts of matter and antimatter were annihilated, and what was left over was the small amount of matter that makes up the visible universe. The question of why there was an asymmetry between the amounts of matter and antimatter (i.e., why there wasn't exactly the same amount of both kinds) is one that still has not been answered.

Why, then, does antimatter exist? No one knows, but that appears to be the way the universe is made. However, within the rules of our current structure of theoretical physics, antiparticles are a "necessary consequence of combining special relativity with quantum mechanics."[4] Paul Dirac [Nobel Prize, Physics, 1933] was the first to realize this fact when he attempted to construct a relativistic wave equation for the electron in 1928 (the Schrodinger equation was not relativistic). The mathematics implied the existence of positive electrons, which later turned out to be positrons.

[3] Abraham Pais is perhaps one of the foremost chroniclers of the story of modern physics. His writings, listed in the Bibliography, are all the more valuable because he was a practitioner — he worked on the front lines in 1940s through the 1970s — and he knew and collaborated with several of the key players personally, e.g., Bohr, Einstein, Heisenberg.

[4] Martin and Shaw, *Particle Physics*, p. 2.

2.1 Mass

A particle's mass indicates how strongly it interacts via the gravitational force. The mass of the electron is

$$m_e = 9.109\ 382\ 6\ (16) \times 10^{-31}\ \text{kg},$$

or, with our typical precision, $\boxed{m_e \approx 9.11 \times 10^{-31}\ \text{kg}}$. Rather than using the SI unit of kilogram, a common practice is to quote particle masses in terms of their "rest energy." Einstein's relativistic equivalence $E_0 = mc^2$ means that the electron's rest energy is $m_ec^2 \approx 8.19 \times 10^{-16}\ \text{J} \approx \boxed{0.511\ \text{MeV}}$. (Sometimes, physicists omit the factor c^2 because it is clear from the context that the mass is being quoted in energy units.) It is common to quote a particle's rest energy in millions of electron volts (MeV), rather than Joules. The other massive leptons, the muon and tauon, are *identical* to the electron, except for their mass: they interact in exactly the same manner. The accepted values of the lepton masses are

$$\begin{aligned} m(e^-) &= \quad 0.510\ 998\ 910(13)\ \text{MeV} \\ m(\mu^-) &= \quad 105.658\ 3668(38)\ \text{MeV} \\ m(\tau^-) &= \quad 1\ 776.99(29)\ \text{MeV} \end{aligned}$$

The first thing to notice is the progression of larger masses with the μ^- and τ^- leptons. This increasing mass is characteristic of the quarks and neutrinos as well. In fact, there are three "families" (or generations) of leptons and quarks, each composed of a lepton, its corresponding neutrino, and two quarks. The following table organizes them in this way.

Leptons	Quarks	
e^-	u	
ν_e	d	light
μ^-	c	↓
ν_μ	s	↓
τ^-	t	heavy
ν_τ	b	

The first family is the lightest, and each successive family is heavier than the previous. Similar to the leptons, the top and bottom quarks are the most massive, and the up and down quarks are the least massive. The neutrino masses also increase, with ν_e the lightest and ν_τ the heaviest. We will ignore the neutrino masses, however, because they are very small (on the order of a few eV). In fact, experiments are only able to set upper limits on their masses, and currently they are

$$\begin{aligned} m(\nu_e) &< \quad 2.2\ \ \text{eV} \\ m(\nu_\mu) &< \quad 170\ \ \text{keV} \\ m(\nu_\tau) &< \quad 15.5\ \ \text{MeV} \end{aligned}$$

In this book I will always assume these masses to be so small as to be ignorable in our calculations.[5] The quark masses are more problematic because quarks have never

[5]In 1998, the SuperKamiokande neutrino experiment determined that the different types of neutrinos can change into each other, which automatically implies that they must have mass. See Dennis W. Sciama, "Consistent neutrino masses from cosmology and solar physics," *Nature* **348**, 617-618 (13 December 1990) for an interesting discussion.

been observed in isolation, and therefore we can only infer their masses from theoretical arguments. That is, measurements of energy released in particle reactions must be used along with a theoretical structure, such as QCD (quantum chromodynamics), in order to predict the quarks' "free" mass.[6] For example, the up quark has a "free" mass of about 3 MeV/c^2, and the down quark about 6 MeV/c^2. The other quark masses are listed in the table below.

quark	mass (GeV/c^2)
u	0.003
d	0.006
c	1.5
s	0.5
t	175[7]
b	4.5

Keep in mind that the values of these masses have large error bars, and that it really only makes sense to talk about the mass of particles that can exist in isolation. Particles that can be isolated, such as protons and neutrons, have masses that *can* be experimentally measured:

$$\begin{aligned} m_p c^2 &= 938.272\,029(80)\ \text{MeV} \\ m_n c^2 &= 939.565\,360(81)\ \text{MeV} \end{aligned}$$

Usually, we will not need to express them so precisely, so we can use $\boxed{m_p c^2 \approx 938\ \text{MeV}}$ and $\boxed{m_n c^2 \approx 940\ \text{MeV}}$. However, we shall see that the mass *difference* between them is critical, so it's important to remember that while they are both approximately 2000 times more massive than the electron, the neutron is slightly heavier than the proton.

If you look at the free masses of the up and down quarks, it's clear that the masses of the proton and neutron are not simply the sums of the masses of their constituent particles. How can that be? The reason is because there is a significant amount of potential energy involved in assembling the proton and neutron from the quarks, and this fact highlights the need to discuss our first "modern" concept in detail, that of binding energy.

Binding Energy

The binding energy B of a compound particle of mass M is defined as the difference between the sum of the masses m_i of the individual constituent particles and the mass of the compound particle

$$B \equiv \left(\sum_i m_i - M \right) c^2. \tag{2.1}$$

[6]A quark's free mass is the mass we would theoretically expect it to have if it could be freed from the confines of the proton or neutron. However, because the quarks can't be isolated, their free mass depends sensitively on the theoretical assumptions made about the color force. The quarks's *constituent* masses, on the other hand, can be calculated in a straightforward manner using the concept of binding energy B, and ignoring any potential energy due to the strong force. See Problem 14.

The simplest example that illustrates binding energy is the deuteron (the nucleus of deuterium, ^{2}H, also known as heavy hydrogen), which consists of one proton and one neutron.[8] The deuteron's binding energy can be calculated from the measured rest energy of the deuteron and the (isolated) masses of the proton and neutron:

$$\begin{array}{rrl} m_p c^2 & 938.272\,029(80) & \text{MeV} \\ +m_n c^2 & 939.565\,360(81) & \text{MeV} \\ -m_D c^2 & 1\,875.612\,82(16) & \text{MeV} \\ \hline = B & 2.224\,57(20) & \text{MeV} \end{array}$$

This means that if we are able to combine a free proton and a free neutron to make a deuteron, we obtain ≈ 2.22 MeV of energy in return[9] — in the language of chemistry, it's an *exothermic* reaction. Where does the released energy go? It goes into the kinetic energy of the compound particle! In fact, combining two nucleons into a single nucleon is called *fusion*, so named because two or more particles "fuse" to form one particle. A more complicated fusion reaction occurs in the core of the sun, where four protons fuse to form one α particle (the nucleus of helium, ^{4}He).[10] That reaction, of course, is also exothermic, and is what powers the sun. These considerations lead us to the conclusion that mass is a form of potential energy:

$$\boxed{\text{Mass (and binding energy) is potential energy.}}$$

Another, more familiar, example is the case of the Earth and a 1-kg ball. If these two objects are infinitely far away from each other, they have well-defined masses, M_E and $m = 1$ kg, that can be measured precisely. As we bring the ball to the surface of the Earth, the Earth-ball system loses potential energy. The amount lost can be calculated from our knowledge of gravitational potential energy

$$\begin{aligned} \Delta U = \frac{G M_E m}{R_E} &= 6.26 \times 10^7 \text{ J} \\ &= 3.89 \times 10^{26} \text{ eV} \\ &= 6.93 \times 10^{-10} \text{ kg } c^2, \end{aligned} \tag{2.2}$$

[8]You might think that the proton or neutron would be simpler, but they are three-particle systems, not two, and more important, the binding energy is not well defined when the constituent particles cannot be isolated.

[9]There is another unit of mass that is commonly used when binding energy calculations are made, and that is the "atomic mass unit," or "u." Here, the carbon-12 atom sets the scale so that $m(^{12}\text{C}) \equiv 12.00$ u exactly, and the conversion to kilograms is 1 u = 1.660 538 86(28)$\times 10^{-27}$ kg $\approx 1.66 \times 10^{-27}$ kg. The atomic mass approximately measures the "atomic number" of the nucleus, i.e., the number of protons and neutrons. Working in atomic mass units, but keeping only six decimal places, the calculation of the deuteron's binding energy is

$$\begin{array}{rl} m_p & 1.007\,276 \text{ u} \\ +m_n & 1.008\,664 \text{ u} \\ -m_D & 2.013\,553 \text{ u} \\ \hline = B/c^2 & 0.002\,388 \text{ u} \end{array}$$

and converting to electron volts (1 u ≈ 931.494 MeV$/c^2$) I obtain $B \approx 2.22$ MeV.

[10]I am being a bit cavalier, in that Eq. (2.1) does not state *how* the constituent particles combine to form the compound particle; other laws of physics are needed to determine that.

where I used the constants $\boxed{G = 6.67 \times 10^{-11}\ \mathrm{Nm^2/kg^2}}$, $M_E = 5.98 \times 10^{24}$ kg, and $R_E = 6.37 \times 10^6$ m. Where did that energy (≈ 63 MJ worth) go? It went into heat, sound, etc. The Earth eventually radiated away the heat energy, and the sound energy also travels off. This leads me to make the following claim:

> CLAIM: The compound object (Earth and ball) has a smaller mass than the two separate objects combined!

The mass lost is exactly 6.93×10^{-10} kg, the mass equivalent of the potential energy difference. Of course, this mass difference is extremely tiny, and cannot be measured with present day experiments, but it must exist, nonetheless. If I were to separate the Earth and ball again, it would take 63 MJ of work, and when I measured their masses, they would "recover" their original masses, because I have put energy into the system with the work that I did to separate them.

While the underlying physics of binding energy and the mass of compound objects is identical in both the classical case (Earth and ball) and the subatomic case (proton and neutron), there are some subtle differences. In the classical case, the binding energy is small compared with the rest energies of the particles involved, and we tend to think of the constituent objects retaining their identity regardless of whether they are far apart or combined. However, with subatomic particles it is often the case that the binding energy is a significant fraction of the rest energies, and the compound object is usually considered to be a different object—the constituent particles lose their identity. For example, a proton "consists" of two up quarks and a down quark: uud. However, there is another compound particle, called Δ^+, which also consists of two up quarks and a down quark. But the mass of the Δ^+ is 1232 MeV, and it is considered to be a different particle from a proton. The mass is different because the three quarks are in a different quantum state than the proton (that is, they occupy a different energy level), which means that the proton and Δ^+ have different binding energies, and hence different masses.[11] On the other hand, when the 1-kg ball is on the surface of the Earth, we still consider the Earth and the ball to be separate, distinct, objects.

A final example of an interaction involving the mass-energy relationship (and anti-matter) is the decay of the neutron. A free neutron (not one that is bound in an atomic nucleus) spontaneously decays into a proton with a half-life of 10.23 minutes.[12] The reaction equation is

$$n \to p + e^- + \bar{\nu}_e, \tag{2.3}$$

where the electron and antineutrino must be part of the decay products in order to conserve both charge and the "lepton number," a quantity that is characteristic of the weak force. The lepton number is another quantum number that we will discuss later. Now, the neutron is NOT comprised of a proton and electron, so there is no binding energy, but we *can* calculate the energy released in this decay by computing the difference in rest energies before and after the decay

$$Q \equiv \left(\sum_{\text{initial}} m - \sum_{\text{final}} m \right) c^2. \tag{2.4}$$

[11]This difference in binding energies is really due to a different in spin – see Sec. 2.3.

[12]Radioactive decay and the concept of half-life is discussed in detail in Chapter 3.

The symbol Q (called "reaction energy") is used rather than B to denote that this is not a compound particle, but that there is some reaction energy that is released. If $Q > 0$ there is energy released (exothermic) and a spontaneous decay is energetically possible. However, if $Q < 0$ then simply because of energy conservation the decay is not allowed. In the case of the neutron decay, I obtain (and you should check the math) $Q \approx 0.782$ MeV.[13] What happens to this released energy? As before, it goes to the kinetic energy of the product particles, i.e., those on the right-hand-side of the reaction equation. In some sense, you can think of a neutron as being in a higher potential energy state than a proton (because it is more massive), and since objects like to lower their potential energy, the neutron wants to turn into a proton.[14]

> Why are neutrons in nuclei stable? If all neutrons were unstable to β decay, then there would be no heavy atoms, no life in the universe, and we would not exist. To understand the stability of nuclei, the concepts of binding energy and reaction energy give a simple explanation.
>
> Consider the case of a deuteron, the nucleus of deuterium (also known as ^{2}H, or heavy hydrogen). It consists of a proton and a neutron, and if the neutron decayed it would emit an electron and an antineutrino, leaving two protons. Two protons (also known as a "diproton") do not form a stable nucleus, so they immediately split into two separate protons. We therefore must look at the reaction energy of the following reaction
>
> $$d \to p + p + e^- + \bar{\nu}_e.$$
>
> The deuteron has a mass of 1875.613 MeV/c^2, while the combined masses of the products is 1877.055 MeV/c^2. This means that $Q = -1.442$ MeV. Thus the deuteron cannot decay spontaneously! At least 1.4 MeV must be added to "cause" this reaction. Note that this is a different calculation than the binding energy of the deuteron.
>
> Why is the deuteron stable but the diproton (^{2}He) and the dineutron are not? To answer this question satisfactorily requires a detailed knowledge of quantum mechanics and nuclear physics. However, at its most basic level the reason has to do with spin and the Pauli exclusion principle. These are the reasons why we exist!

Classification according to mass, and particle names

There were three known particles in the 1930s: the electron, proton and neutron. With a mass of about 0.5 MeV/c^2 the electron was the lightest, and with a mass of about 1000

[13] In calculating this value, I ignored the small neutrino mass. Since the upper limit on the rest energy of the electron neutrino (and antineutrino) is about 2.2 eV, it doesn't affect the calculation at this level of precision.

[14] As far as we know, the proton is a stable particle because there is no baryon that is less massive for it to decay into, although the possibility that the half-life for proton decay is so long that we haven't noticed it yet is an active area of research. Baryon number is another quantum number that must be conserved—we will investigate this concept later. The electron is the lightest lepton and hence it, too, is stable against spontaneous decay.

MeV/c^2 the nucleons (the common name for the proton and neutron) were heavy. With the discovery in 1937 of an intermediate mass particle, about 100 MeV/c^2, in cosmic rays, the particles were given "nicknames" according to their mass. Since the electron was light, it was called a "lepton," from the Greek word *leptos* ($\lambda\varepsilon\pi\tau o\zeta$) meaning "small." And, since the nucleons were massive, they were called "baryons," from the Greek *barys* ($\beta\alpha\rho\upsilon\zeta$) meaning "heavy." The cosmic ray particle was therefore called a "meson," from the Greek word *mesos* ($\mu\varepsilon\sigma o\zeta$) meaning "middle."[15] It wasn't realized until later that the intermediate mass cosmic ray particle was actually the μ^- lepton, although it was originally called the μ-meson.

Under our current naming scheme, however, a baryon has come to mean a particle that is made up of three quarks (such as a proton or neutron), a meson has come to mean a particle that is made up of a quark–anti-quark pair, and leptons are the elementary particles that do not interact via the strong force. Since any three of the six flavors of quarks can combine to form a baryon, there are 56 possible combinations, although there are more than 56 different baryons since it is possible for the same set of quarks to have different binding energies (see the comparison between the proton and Δ^+ above). For example, the sigma (Σ) baryons are combinations of one strange quark and two up or down quarks. Their quark content and masses are

Σ^+	uus	1189.4 MeV
Σ^0	uds	1192.5 MeV
Σ^-	dds	1197.4 MeV

The pi (π) mesons are composed of different combinations of up and down quarks and their anti-particles:

π^+	u$\bar{\text{d}}$	139.6 MeV
π^0	(u$\bar{\text{u}}$-d$\bar{\text{d}}$)/$\sqrt{2}$	135.0 MeV
π^-	d$\bar{\text{u}}$	139.6 MeV

Note that the π^0 meson is actually a *superposition* of quark states. This means that when an experimenter "looks" at a π^0 meson, 50% of the time they will "see" the u$\bar{\text{u}}$ combination, and the other 50% they will see d$\bar{\text{d}}$. This is just one of the strange implications of quantum mechanics. It should look somewhat familiar, however, because it is similar to the fact that the general solution of a linear second-order differential equation is a linear combination (or superposition) of two independent solutions. In the same way, some quarks (and baryons) can be linear combinations of two (or more) independent quark states. Also note that the π^+ and π^- are antiparticles of each other, and hence have the same mass, and that the π^0 is its own antiparticle.

The Σ and π particles are just a few of the possible baryon and meson combinations that can be constructed with the six known quark flavors. A short list, along with their quark constituents, are shown in Table 2.1. At this time, no other combinations of quarks

[15]Interestingly, Hideki Yukawa, who predicted the existence of an intermediate-mass particle in 1935, initially proposed to call it a "mesotron," in keeping with the name of the electron. However, Werner Heisenberg noted that the correct Greek word was *mesos* and it had no "tr."

baryon		mass (MeV/c^2)
p	uud	938.3
n	udd	940.6
Σ^+	uus	1189.4
Σ^0, Λ^0	uds	1192.5, 1115.7
Σ^-	dds	1197.4
Ξ^-	dss	1321.7
Ξ^0	uss	1314.9

meson		mass (MeV/c^2)
π^+	u$\bar{\text{d}}$	139.6
π^0	(u$\bar{\text{u}}$-d$\bar{\text{d}}$)/$\sqrt{2}$	135.0
π^-	d$\bar{\text{u}}$	139.6
K^0, $\bar{K}^0$	d$\bar{\text{s}}$, $\bar{\text{d}}$s	497.6
K^+, K^-	u$\bar{\text{s}}$, $\bar{\text{u}}$s	493.7
η	(u$\bar{\text{u}}$+d$\bar{\text{d}}$-2s$\bar{\text{s}}$)/$\sqrt{6}$	547.9
η'	(u$\bar{\text{u}}$+d$\bar{\text{d}}$+s$\bar{\text{s}}$)/$\sqrt{3}$	957.7

Table 2.1: Tables of the light (u, d, s quarks only), spin $\frac{1}{2}$ baryons and the light, spin 0 mesons. Note that the Σ^0 and Λ^0 have the same quark content but different masses. The heavier one, Σ^0, is an electromagnetic excited state and decays in about 7×10^{-20} s into the lighter one, Λ^0. Also note that the π^0, η, and η' are all neutral mesons, but are just different linear combinations of the same set of three quark—anti-quark pairs.

other than qqq and q$\bar{\text{q}}$ have been observed, although there have been searches for exotic combinations such as so-called "penta-quarks," made up of four quarks and one anti-quark: qqqq$\bar{\text{q}}$. In some sense this looks like a baryon and meson bound together. Either these do not exist, or their lifetimes are too short to measure.

2.2 Electric Charge

A particle's charge indicates how strongly it interacts via the electromagnetic force. In addition, however, charge is quantized; that is, it appears in nature only as integer multiples of the fundamental unit of charge, e, which happens to be the charge of the electron,

$$q_e = -e = -1.602\ 176\ 53(14) \times 10^{-19}\ \text{C},$$

or, for our purposes $\boxed{e \approx 1.60 \times 10^{-19}\ \text{C}}$.[16] The other massive leptons (muon and tau) have the same negative charge as the electron, and the neutrinos are neutral. In fact the word neutrino was proposed by Wolfgang Pauli in 1930, and means "little neutral one" in Italian.

What about the quarks? What are their charges? The quarks come with *fractional* charges, that is, submultiples of e! For example, the charge on the up quark is $q_u = +\frac{2}{3}e$, and that on the down quark is $q_d = -\frac{1}{3}e$. At first sight, this appears strange. How can any particle have a fractional charge? There are two ways to reconcile this with the proposed quantization of charge. First, all this really says is that the fundamental unit of charge is not e, but is $\frac{1}{3}e$. Charge is still quantized and all particles have integer multiples of this fundamental unit. Second, because quarks are never observed in isolation (they always appear in groups of 3 — baryons — or in a quark–anti-quark pair — mesons), the charges of particles that *can* exist in isolation must be a multiple of e. So the proton and neutron

[16]Keep in mind that e is a positive quantity, and that negative particles have charges that are integer multiples of $-e$.

have integer charges

$$\begin{aligned} q_p &= \left(+\frac{2}{3}+\frac{2}{3}-\frac{1}{3}\right)e = e \\ q_n &= \left(+\frac{2}{3}-\frac{1}{3}-\frac{1}{3}\right)e = 0. \end{aligned}$$

This second fact was helpful in convincing skeptics about the usefulness, and the ultimate reality, of quarks. The charges on the quarks are

u	c	t	$+\frac{2}{3}e$
d	s	b	$-\frac{1}{3}e$

One important fact about electric charge is that it is absolutely conserved. There is no way to transform charge into energy, as there is with mass, so the charge of compound particles is just the sum of the charges of the constituent particles. This conservation law is sometimes stated as

$$\boxed{\text{Electric charge is neither created nor destroyed.}}$$

Why? We don't know. All we know is that the violation has never been observed, so until then it remains a "law."

Interactions

Gravity and electromagnetism are the two classical (non-quantum) forces. The other two "forces," the weak and strong nuclear forces are inherently quantum mechanical in nature. For this reason, you won't be able to fully understand them in detail until *after* a thorough study of quantum mechanics; however, we can discuss them now using some of the classical concepts that you already know, such as energy and momentum. This quantum nature leads to a new way of describing and understanding these forces that is completely different from our previous descriptions. Previously, you have learned about forces in two different ways. First, as "action-at-a-distance," propounded by Newton with his Universal Law of Gravitation.[17] Second, utilizing the concept of a "field," devised by Faraday (and honed by Maxwell) to explain the electric and magnetic forces. Gravity can also be described in terms of the gravitational "field," both in the Newtonian limit and in general relativity. Due to the necessity of using quantum ideas to describe the weak and strong nuclear forces, we are forced to use quantum field theory, and this third description postulates the existence of exchange particles.

For example, two electrons repel each other not because of a mysterious action-at-a-distance Coulomb force, nor even the electric field, but by exchanging photons. Just like two ice skaters who, throwing a ball back and forth, appear to repel each other, electrons

[17] Newton had a philosophical objection to action-at-a-distance, which he expressed in a letter in 1692: "*That gravity should be innate, inherent and essential to matter, so that one body may act upon another at a distance through a vacuum, without the mediation of anything else by which their action and force may be conveyed from one to another, is to me so great an absurdity that I believe no man who has in philosophical matters a competent faculty of thinking can ever fall into it.*"

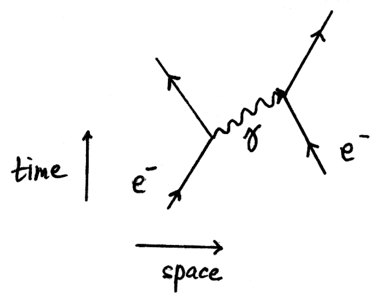


Figure 2.3: Feynman diagram depicting the electromagnetic interaction between two electrons. Only one spatial dimension is shown, and the time axis is conventionally drawn upward. Matter particles are depicted by straight lines, while the photon, an exchange particle, is shown as a wavy line and indicated by a γ.

exchange photons, which, due to the conservation of momentum, exert impulses on the electrons, and they appear to repel each other. The photon, therefore, is the exchange particle that "mediates" the electromagnetic force. The ice skater analogy does not work for particles that attract each other, but the concept is still valid. In his thinking that led to the proposal of the meson, the mediating particle that held the nucleons together in the nucleus, Hideki Yukawa [Nobel Prize, Physics, 1949] wrote

> If one visualizes the [nuclear] force field as a game of "catch" between protons and neutrons, the crux of the problem would be the nature of the "ball" or particle.

This view has three aesthetically pleasing features. First, all interactions are "local," which means that particles must be in the same location for any effect. Second, it nicely explains the $1/r^2$ nature of the electric and gravitational forces: the "density" of mediating particles must decrease as $1/r^2$ from the "source" particle, a simple geometrical effect. Finally, effects are not instantaneous, but take a finite time as the mediating particle traverses the intervening distance. A graphical method of describing interactions that automatically displays the first and third of these features is called a "Feynman diagram." A Feynman diagram of the electromagnetic interaction between two electrons is shown in Fig. 2.3. This is similar to a position-time graph from elementary mechanics, where the trajectories of all particles are shown. Note that the photon (γ) comes into and out of existence when it interacts with an electron, and each electron undergoes a momentum change. At each "vertex" all quantities such as charge and other quantum numbers are conserved, the only exception being energy. That is, energy is created when the photon is emitted by the first electron, and then lost when the photon is absorbed by the second electron. The time interval in which the photon exists (and during which energy conservation is violated) is short enough so that Heisenberg's uncertainty principle is not violated. A photon of this type is called a "virtual" photon, so that in this third picture of interacting electrons, they do not exert a Coulomb force (at a distance), nor do they create an electric field, but they exchange virtual photons in order to exchange momentum and repel each other.

To the extent that each of the fundamental forces can be described by a quantum field theory, each force must be mediated by an exchange particle. If the photon mediates the electromagnetic force, what particles mediate the other forces? They are listed below, along with their mass, charge, spin, and color.

		mass	charge	spin	color
gravity	graviton[18]	0	0	2	no
E&M	photon	0[19]	0	1	no
color	gluons	0	0	1	yes
weak	$W^{\pm}$	80.4 GeV/c^2	$\pm e$	1	no
	Z^0	91.2 GeV/c^2	0	1	no

Our "zoo" of particles is now complete. We have 12 particles of matter, 12 of anti-matter, and 13 "gauge bosons." (There are 8 types of gluons, which are distinguished because they also carry color.)

We will see later that when the mass of the mediating particle is zero, then the interaction is long range. This makes sense for gravity and electromagnetism, in that they both are $1/r^2$ forces which means that although they become weaker with distance, they never go to zero. The weak force, on the other hand, is extremely short range because the W and Z bosons are very massive. This means that the weak force is very "weak" (hence the name) and particles must be very close to interact in this manner. The color force is also long range, but it turns out to become stronger with distance rather than weaker. The strong force, which is the force between baryons and mesons, is a short range force that is the residual, or "leftover," color force between objects that are color neutral.[20]

2.3 Spin

Our third property of interest, after mass and electric charge, is spin. It is a property that does not relate specifically to one of the four fundamental forces, but is an inherently quantum property, and therefore nicely illustrates the differences between the quantum world and the classical world. A discussion of spin is a nice place to introduce the Heisenberg uncertainty principle and the Pauli exclusion principle so that you can see just how the quantum world operates and how it differs from the world you know.

Spin is also known as "intrinsic" angular momentum. The Earth, for example, has both orbital angular momentum due to its revolution about the sun, as well as "spin" angular momentum: it rotates on its axis once every 24 hours.[21] In the same way the elementary particles, such as electrons, have intrinsic angular momentum that appears to be due to their actual spin about an internal axis. There are two problems with this

[18] The graviton, while postulated to exist, has not yet been observed.

[19] The current upper bound for the photon mass is 1.2×10^{-51} g. (Luo *et al.*, Phys. Rev. Lett. **90** 081801, 2003)

[20] See Sec. 2.5 for more details.

[21] The "spin" angular momentum of the Earth about its geographic axis is about 7×10^{33} J s.

Chapter 5

Introduction to Special Relativity

Einstein's special theory of relativity is a description of kinematics and dynamics in four-dimensional "spacetime," and in particular, how to describe the motion of objects from different points of view, that is, by observers in different *reference frames*. It is essentially a reformulation of Newtonian mechanics, correct for all particle velocities. In fact, Newtonian mechanics turns out to be a special, limiting case, valid when velocities are small compared with the speed of light c.

Since its development by Einstein over a century ago, a virtual cottage industry has existed, consisting in the delineation and resolution of a multitude of counterintuitive paradoxes. Even though everyday Newtonian physics is contained within special relativity, the consequences of the principle of the constancy of the speed of light introduces profound effects that we do not experience in everyday life.

Common paradoxes, such as the "twin paradox," where a person's age depends on their speed of travel, and the "pole and barn paradox," where the observations of a pole fitting inside a barn depends on the observer's frame of reference, will be covered in this chapter. More exotic paradoxes, such as the appearance of a cube that is moving at near the speed of light, are beyond the scope of this book.

Einstein's own words

Einstein was led to his theory of relativity by the observation of certain electromagnetic effects. Specifically, how does a magnet induce current in a nearby conductor? The induced current depends only on the *relative* motion of the magnet and conductor, but the physical explanation — in terms of a magnetic field or an electric field — depends on which object is "actually" moving. This apparent contradiction led Einstein to abandon Newton's concepts of "absolute time" and "absolute space," and required replacing the Galilean transformation between two reference frames, moving with respect to each other, with the Lorentz transformation.

As with any other piece of fundamental physics, it is always useful to go first to the primary source; in this case, here are the first two paragraphs of Einstein's seminal paper on this subject.

> It is known that Maxwell's electrodynamics — as usually understood at the present time — when applied to moving bodies, leads to asymmetries which

> do not appear to be inherent in the phenomena. Take, for example, the reciprocal electrodynamic action of a magnet and a conductor. The observable phenomenon here depends only on the relative motion of the conductor and the magnet, whereas the customary view draws a sharp distinction between the two cases in which either the one or the other of these bodies is in motion. For if the magnet is in motion and the conductor at rest, there arises in the neighbourhood of the magnet an electric field with a certain definite energy, producing a current at the places where parts of the conductor are situated. But if the magnet is stationary and the conductor in motion, no electric field arises in the neighbourhood of the magnet. In the conductor, however, we find an electromotive force, to which in itself there is no corresponding energy, but which gives rise — assuming equality of relative motion in the two cases discussed — to electric currents of the same path and intensity as those produced by the electric forces in the former case.
>
> Examples of this sort, together with the unsuccessful attempts to discover any motion of the earth relatively to the "light medium," suggest that the phenomena of electrodynamics as well as of mechanics possess no properties corresponding to the idea of absolute rest. They suggest rather that, as has already been shown to the first order of small quantities, the same laws of electrodynamics and optics will be valid for all frames of reference for which the equations of mechanics hold good. We will raise this conjecture (the purport of which will hereafter be called the "Principle of Relativity") to the status of a postulate, and also introduce another postulate, which is only apparently irreconcilable with the former, namely, that light is always propagated in empty space with a definite velocity c which is independent of the state of motion of the emitting body. These two postulates suffice for the attainment of a simple and consistent theory of the electrodynamics of moving bodies based on Maxwell's theory for stationary bodies. The introduction of a "luminiferous ether" will prove to be superfluous inasmuch as the view here to be developed will not require an "absolutely stationary space" provided with special properties, nor assign a velocity-vector to a point of the empty space in which electromagnetic processes take place.[1]

The standard way to develop the principles of special relativity is to follow Einstein. This method starts with Einstein's two postulates mentioned above, and derives time dilation and length contraction using a light ray as a clock. This development is covered in Appendix J. Here, I want to take a different approach, one that is due to Peter Scott and Bill Burke. It starts with an experimental fact and deduces what theoretical conclusions are needed to explain the observation. The experimental fact that we will use is the observation of muons from cosmic rays. Appendix G describes the process of muon creation and observation, and constitutes good background reading at this time.

[1] From a translation by W. Perrett and G.B. Jeffery, *The Principle of Relativity*, 1923.

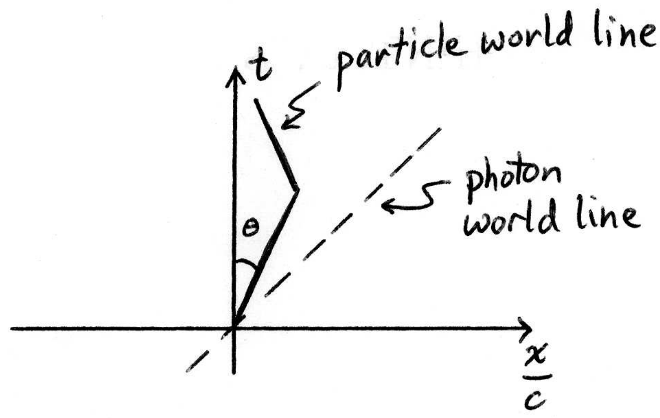


Figure 5.1: Spacetime diagram depicting the world line of a photon (dashed line) and the world line of a massive particle (solid line). The particle is shown starting at the origin, moving to the right and then moving to the left. The speed of any particle can be determined by the angle its world line makes with a vertical line parallel to the t axis. That is, $v = c \tan\theta$.

Spacetime diagrams

An extremely useful way to conceptually understand (as well as make quantitative calculations) special relativity is the *spacetime diagram*. A spacetime diagram is nothing more than a one-dimensional position-time graph, familiar from elementary mechanics, with two simple changes: (i) an interchange of the x and t axes, and (ii) a rescaling of the spatial axis so that it has dimensions of time. That is, rather than x versus t, a spacetime diagram is a plot of t versus x/c. The trajectory of a particle is called its "world line." A particle with constant speed v therefore has a world line that makes an angle θ with the positive t axis that is given by $v = c\tan\theta$. Velocities are also scaled by the speed of light, c, so that it is common to refer to $\beta \equiv v/c$, instead of v. In addition, a photon (that is, light) has a world line that makes a 45° angle with the positive t axis. In other words, it has a slope of unity. Figure 5.1 depicts these features.

Why do I rescale distances so that they have dimensions of time? It turns out that in relativity, space and time are treated on equal footing, so the fabric in which events occur is called *spacetime*. Because they have equal status, it makes sense to measure them in the same units. In a similar vein, it would be strange to measure north-south distances in meters and east-west distances in feet. From this point of view, the speed of light is just a conversion factor between our usual distance units and our usual time units. For example, a "light-year" is a unit of length equal to the distance that light travels in one year. Hence, the speed of light can be written as

$$c = 3 \times 10^8 \text{ m/s} = 1\,\frac{\text{light-year}}{\text{year}}. \tag{5.1}$$

Therefore, if $x = 5$ light-years, say, then

$$\frac{x}{c} = \frac{5 \text{ light-years}}{1 \text{ light-year/year}} = 5 \text{ years}. \tag{5.2}$$

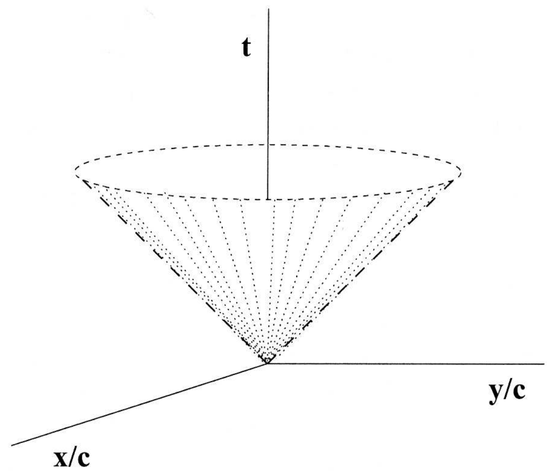


Figure 5.2: Spacetime diagram with two spatial dimensions, x and y. Since light travels at speed c, the set of all possible photon world lines forms a cone with its vertex at the origin. This is called the "light cone." All massive particles must have world lines that are "within" this cone, or closer to the t axis. From Scott and Burke, *Special Relativity Primer*, Figure 4.

Why do I scale velocities to the speed of light? Since c appears to be a cosmic speed limit, it makes sense to compare all speeds to c. More fundamentally, however, we'll see that it is impossible to determine how fast you are traveling in an absolute sense. You can only tell how fast you are traveling with respect to another object. In addition, all observers, regardless of their motion, measure light to travel at the same speed, c. For these reasons, the speed of light is the natural speed in the universe. If you tell someone how fast you are moving, and they reply with the query, "Compared to what?", the speed of light is the only possible comparison.

Both of these rescalings compel some physicists to set $c \equiv 1$ so that it does not appear in any equations. However, in this chapter I will leave it explicitly in the equations, but it is important to remember that it is simply a scaling factor.

It is possible to portray two spatial dimensions (and time, of course) using a perspective portrayal of three dimensions on two-dimensional paper, as shown in Fig. 5.2. All the essential physics, however, can be shown in a two-dimensional spacetime diagram with just one spatial dimension.

5.1 Time dilation

An investigation into the phenomenon of time dilation cuts to the heart of what special relativity is, and how it differs from classical (i.e., Galilean) relativity. The first concept to be very clear about is that of a CLOCK. We will assume that our clocks tick *uniformly* and *homogeneously*. What this means is that their ticking rate does not depend on time nor on the clock's location. Of course, since a clock is how we actually measure time, we cannot tell if it were not uniform, so we just have to assume it.

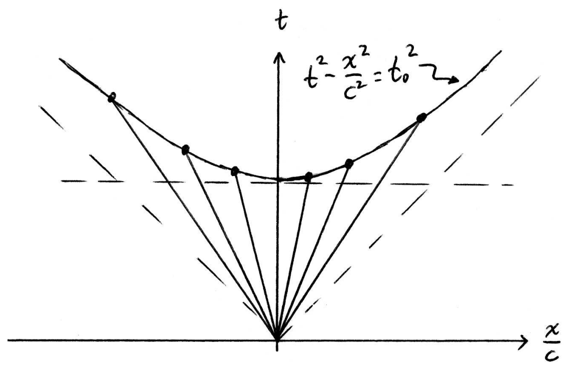


Figure 5.3: The locus of points in spacetime at which muons decay after moving a distance x/c at speed v. The horizontal dashed line is the Newtonian prediction $t = t_0$, and is a good approximation when $v \ll c$. The diagonal dashed lines are, of course, the world lines of photons, i.e., the light cone.

For our first thought experiment (*Gedanken* experiment), a muon will act as our clock.[2] The average lifetime of a muon, as measured by the muon itself, is $t_0 = 2.2\ \mu$s. Even though this is just an average, because some muons exist for longer time intervals and some shorter, it will be convenient to assume that they all live for exactly 2.2 μs before decaying into an electron. (Read App. G for an introduction to the production and observation of cosmic rays and muons.) Our thought experiment is given by the following scenario.

> Imagine that we have a number of identical clocks (muons). They are all created simultaneously at the origin, and subsequently they move away from the origin (in both the positive and negative x directions) at different constant speeds. (One spatial dimension is all that is needed for this experiment.) Let's now follow one of the muons. If it moves at speed v, it will have traveled a distance $x = vt$ in a time t, as measured by another clock that remains at the origin. How is t (the spacetime coordinate) related to t_0 (the muon's lifetime)?

"What?" you say. "Isn't it obvious? They are equal, $t = t_0$." This is exactly what Newton and Galileo would say. But they are wrong! Experiments show that

$$t_0^2 = t^2 - \frac{x^2}{c^2}. \tag{5.3}$$

A plot of the spacetime locations (x/c, t) of the decays of each of the muons results in a hyperbola, shown in Fig. 5.3. It seems rather strange that a moving clock, as represented here by a muon, and a stationary clock would measure different time intervals between the same two events (in this case the creation and decay of the muon), but as Scott and Burke state, "The hyperbola represents a description of the experimental data for real clocks,

[2]Initially, since the mass of the muon was intermediate between the electron and the proton, it was named a "mu meson." Even though we now realize that it is a lepton, not a quark—anti-quark pair, it is still sometimes referred to as a meson.

and is not subject to dispute."[3] This means that our task as physicists is not to mold this result into our pre-existing notions of space and time, but to *explain* this observation. In this case it requires a completely new and different conception of spacetime.

Muon experiments

As explained in App. G, particles from outside the Earth's atmosphere collide with the molecules in the atmosphere and create many secondary particles, some of which are muons. The altitude at which most of the muons are created in this process is near 20 km, and these muons are subsequently observed on the ground. Let's assume that the muons are of very high energy, and that they move almost at the speed of light. If this is true, then the maximum distance they could travel before decaying is

$$d = v\,t_0 = c\,t_0 = (3 \times 10^8 \text{ m/s})(2.2\,\mu\text{s}) = 660 \text{ m}, \tag{5.4}$$

which means that they would not reach the ground! They would decay, on average, far above the ground.

However, since they *are* observed reaching the ground, and since they do *not* travel faster than the speed of light c, we can determine what their average lifetime must be in order to explain the observations. A clock on the ground measures their lifetime to be

$$t = \frac{d}{c} = \frac{20 \text{ km}}{c} = 6.7 \times 10^{-5} \text{ s} = 30\,t_0. \tag{5.5}$$

This means that cosmic ray muons that are moving close to the speed of light — relative to the ground — survive about 30 times longer than stationary muons in the laboratory. This is what is meant by *time dilation*.

The spacetime interval

Can we draw a general conclusion from this result? Yes. First, it is important to distinguish the two times that we have been discussing. The first, t, is called the *coordinate time*. It is the time measured by clock tied to the coordinate system $(x/c,\,t)$ in which we measure the location of events. The second, t_0, is called the *proper time*.

> The *proper time* is the elapsed time between two events as measured by a clock that is at the same spacetime location as both events.

That is, a clock always measures its own proper time, because it is always at the origin of its own frame of reference, its own coordinate system.

In addition, in special relativity we refer to events that occur at specific locations and specific times, i.e., at specific points in spacetime. For example, if a firecracker's fuse is lit at $(x_1/c,\,t_1)$ and the firecracker then explodes at $(x_2/c,\,t_2)$, then a clock attached to the firecracker will record a time interval Δt_0 between the two events, where

$$(\Delta t_0)^2 = (t_2 - t_1)^2 - \frac{(x_2 - x_1)^2}{c^2} = (\Delta t)^2 - \frac{(\Delta x)^2}{c^2}. \tag{5.6}$$

[3] Scott and Burke, *Special Relativity Primer*.

This quantity, $(\Delta t)^2 - (\Delta x)^2/c^2$, is called the *spacetime interval*, or just the interval. A clock tied to the coordinate system will record a time interval Δt *and* a space interval Δx between the two events.

The relativistic factor γ

More information can be gleaned from Eq. (5.6) by noting that, if the muon is traveling at a constant speed, which we are assuming, the coordinate distance it travels Δx is simply its speed v times the coordinate time interval Δt.[4] Replacing Δx by $v\Delta t$ in Eq. (5.6) and rearranging to solve for Δt gives

$$\Delta t = \frac{1}{\sqrt{1-\frac{v^2}{c^2}}}\Delta t_0 = \frac{1}{\sqrt{1-\beta^2}}\Delta t_0 \equiv \gamma\,\Delta t_0, \tag{5.7}$$

where γ is the relativistic factor introduced back in Chapter 1. This equation states the general conclusion that stationary clocks observe moving objects to last longer than expected, which means that stationary clocks observe the moving clocks to "run slow." The moving clocks thus experience a time dilation.

With this result, and our determination from Eq. (5.5), we can deduce how fast the muon was actually moving (relative to the ground). Since the muon's lifetime must increase by a factor of 30 in order to be observed at the ground, Eq. (5.7) states that this factor of 30 is nothing more than the relativistic factor γ. Setting $\gamma = 30$ and solving for β gives

$$\beta^2 = 1 - \frac{1}{900} \qquad \text{or} \qquad \beta \approx 0.9994 \tag{5.8}$$

A typical cosmic ray muon travels at 99.94% of the speed of light.

Inertial reference frames and the invariance of the interval

The spacetime interval, as defined by Eq. (5.6), is a quantity that is invariant for any observer. We have been talking about stationary reference frames (e.g., the frame tied to the Earth) and moving reference frames (e.g., the frame tied to the muon), but there really is no way to tell who is moving and who is not. We could just as correctly take the muon as stationary and the Earth as moving. On the other hand, it *is* possible to determine if a reference frame is accelerating, and this brings us to another definition:

> An *inertial reference frame* is one in which Newton's first law holds true.

That is, if you perform an experiment in which you release an object at rest, if it remains at rest then your reference frame is inertial. A frame that is moving at constant velocity is a natural choice for an inertial frame.[5] If we now consider two inertial reference frames that may be in relative motion, and we investigate how each of them measure events, we'll find that the spacetime locations of events (i.e., positions and times) will be different, but

[4] This is just the constant velocity relation that we misapplied in Eq. (5.4).

[5] Free fall in a gravitational field is *also* an inertial reference frame. But including gravity leads us into the domain of general relativity, which is beyond the scope of this chapter.

Figure 5.4: The "straight line" path between events 1 and 2 results in the longest elapsed proper time. Any other path results in a shorter elapsed proper time.

observers in the two frames will measure the same interval between two events. That is, if there are two events, then one observer will measure the coordinates (x_1/c, t_1) and (x_2/c, t_2) as before. But the other observer (who is in what we will call the "primed frame") measures the coordinates (x'_1/c, t'_1) and (x'_2/c, t'_2). Although $x_1 \neq x'_1$ in general, and the other coordinates will also have different values in the two frames, it will always be true that

$$(\Delta t)^2 - \frac{(\Delta x)^2}{c^2} = (\Delta t')^2 - \frac{(\Delta x')^2}{c^2}. \tag{5.9}$$

This invariance is very similar to the measurement of the distance between two spatial locations using two different coordinate systems. While the x and y coordinates of the two points as measured in the two systems will, in general, be different, the *distance* between the two points, ΔL, is invariant — measurements in *any* coordinate system will give the same result

$$(\Delta L)^2 = (\Delta x)^2 + (\Delta y)^2. \tag{5.10}$$

The minus sign in Eqs. (5.6) and (5.9) has far reaching consequences. First, it indicates that instead of the Euclidean geometry of flat space, spacetime is described by a hyperbolic geometry, also known as "Minkowski space," after Hermann Minkowski (1864-1909). Second, in the standard distance formula of Euclidean geometry, the plus sign implies that the shortest distance between two points is a straight line. This can be proved by integrating the differential form of Eq. (5.10)

$$dL = \sqrt{(dx)^2 + (dy)^2} \qquad \Rightarrow \qquad L = \int_1^2 \sqrt{1 + \left(\frac{dy}{dx}\right)^2}\, dx. \tag{5.11}$$

To actually evaluate L it is necessary to know $y(x)$, the shape of the curve connecting the two points. A technique from advanced calculus, called "calculus of variations," can be used to show that when $y(x)$ is a straight line, then L is a minimum. Similar mathematical steps can be used to show that the world line that joins two spacetime events with a constant slope results in the largest proper time. That is, integrating the differential form of Eq. (5.6) gives

$$dt_0 = \sqrt{(dt)^2 - \left(\frac{dx^2}{c^2}\right)} \qquad \Rightarrow \qquad \Delta t_0 = \int_1^2 \sqrt{1 - \left(\frac{1}{c}\frac{dx}{dt}\right)^2}\, dt. \tag{5.12}$$

Again, in order to evaluate the proper time t_0 that elapses on a clock traveling from event 1 to event 2, the function $x(t)$ must be known. The two integrals in Eqs. (5.11) and (5.12) are *not* path independent. This property is investigated mathematically in Problem 93 for a simple case where the different functions $x(t)$ are connected straight line segments. However, the result that the largest elapsed proper time is when $x(t)$ is a straight line can be seen graphically in Fig. 5.4. The "straight line" path between events 1 and 2 is indicative of a constant velocity, $dx/dt = v$, and a proper time that can be easily integrated to obtain $\Delta t_0 = \Delta t/\gamma$. Here, Δt is the elapsed coordinate time, and this result confirms what we have already deduced in Eq. (5.7).

5.2 Length contraction

An unescapable conclusion that is intimately connected to time dilation is the fact that objects are measured to be shorter when they are moving with respect to the measuring reference frame. Let's take the example of the cosmic ray muon again. In the rest frame of the Earth, the distance between the ground and the layer of the atmosphere in which most muons are created is 20 km. However, in the reference frame of the muon, the Earth is moving while the muon is stationary. In addition, the muon lives for only 2.2 μs, which means that even if the Earth is moving at close to the speed of light (with respect to the muon) only 660 m will have passed by in those 2.2 μs — this is the same calculation as in Eq. (5.4). Therefore, the 20 km object (the atmosphere) must have *contracted* to 660 m from the muon's point of view.

More rigorously, Newton would say that since the muon's lifetime is Δt_0 it must travel a distance $v\Delta t_0$. On the other hand, Einstein would say that moving clocks run slow, so that the muon travels a distance $v\Delta t = v\gamma\Delta t_0 > v\Delta t_0$. In other words, the distance traveled by the muon — measured in the Earth's frame — is[6]

$$\Delta x = \gamma v \Delta t_0. \tag{5.13}$$

From the muon's point of view the Earth is moving at speed v, which means that the distance the Earth has traveled is $v\Delta t_0 = \Delta x/\gamma < \Delta x$. This means that the muon thinks that the distance from the top of the atmosphere to the ground is *less* than 20 km, and therefore it is possible for it to cover that distance in only 2.2 μs.

How to measure lengths

How can you actually measure the length of a moving object? Before we consider this problem, we need to decide how to measure the length of a *stationary* object. Consider a rod of length L that is at rest in a particular coordinate system. You can mark the location at one end of the rod with an event in spacetime (position *and* time), and then wander over to the other end of the rod and do the same. But because the rod is at rest, there is no time dilation, which means that the spatial distance between the two events will not depend on the temporal distance between the two events. This length that you have just measured is called the *proper length*, L_0.

> The *proper length* of an object is the length measured in a reference frame in which the object is at rest.

[6]This can also be obtained directly from the invariance of the interval. Replacing Δt in Eq. (5.6) with $\Delta x/v$ results in

$$(\Delta t_0)^2 = \frac{(\Delta x)^2}{v^2}\left(1 - \frac{v^2}{c^2}\right)$$

which can be rearranged to give Eq. (5.13).

How do you measure the length of a *moving* rod? In this case, you don't have the luxury of marking the position at one end of the rod, and then wandering over to mark the position of the other end of the rod at a later time. The rod has moved in the time between the two events, so the distance you have measured is not the length of the rod. To account for this, there are two methods that can be used. First, you can measure the positions of the two ends of the rod *at the same time*. This means that you need a partner to mark the location of one end while you mark the location of the other. However, the two events, while they are simultaneous in your frame, they are *not* simultaneous in the frame that is moving with the rod.[7] To relate your length measurement, L, to the proper length, L_0, therefore, requires a transformation from your frame to the moving frame. This transformation is called the *Lorentz transformation*, and a correct analysis of this measurement method must wait until Section 5.3.2.

The second method is to remain at one location and mark the two times at which the two ends of the moving rod pass your location, and then multiply by the relative velocity of the two frames. This method requires the use of the invariance of the interval as stated in Eq. (5.9). The two events of interest are event A, when the front end of the stick passes you at $t = t_A$, and event B, when the back end of the stick passes you at $t = t_B$. The spatial distance between the two events is, of course, $\Delta x = 0$ (they occur at the same spatial location), and the temporal distance (what you measure with your clock) is $\Delta t = t_B - t_A$. By definition,

$$\Delta t = \frac{L}{v}. \tag{5.14}$$

What about the same quantities as measured by an observer traveling with the rod? Of course, the spatial distance is just the proper length, $\Delta x' = L_0$, because that observer does not care how much time passes between the two events, and the temporal distance in that moving frame, $\Delta t'$, is the length of the rod (this time the proper length) divided by the relative velocity, $\Delta t' = L_0/v$. Inserting these values into Eq. (5.9) and solving for L gives

$$L^2 = L_0^2\left(1 - \frac{v^2}{c^2}\right) \qquad \text{or} \qquad L = \frac{L_0}{\gamma}. \tag{5.15}$$

Since $\gamma > 1$, the length you measure is always *less* than the proper length. The rod's length has *contracted*.

5.3 Transformations between reference frames

A transformation between two coordinate systems, or reference frames, is simply a relation between the position and time coordinates of an event as measured by observers in the two frames using the two different coordinate systems. The form of the transformation is determined by the geometry of the space and time in which the measurements are made. The Galilean, or Newtonian, transformation assumes that both time and space are *absolute*. In addition, it is assumed that space itself is flat, or Euclidean. At speeds small compared to the speed of light, this is approximately true, and since c is so large

[7] As we will see below, the simultaneity of events depends on which reference frame is making the observation.

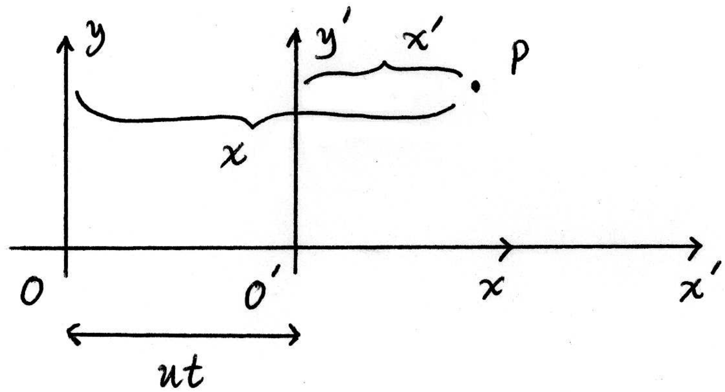


Figure 5.5: Geometry of a transformation between two coordinate systems. Reference frame O is stationary, while reference frame O' moves with speed u (relative to O) in the positive x direction. At $t = t' = 0$ both origins coincide, the y and y' directions are the same, and the z and z' directions are the same. In the figure, the z and z' axes are suppressed for clarity.

compared with everyday motion, it is not surprising that Galileo and Newton obtained this approximation to the correct, special relativistic, Lorentz transformation. Because the method of obtaining the transformation equations is identical in the two cases, I will develop the Galilean transformation in some detail so that any assumptions we make are clearly stated. Then, when I derive the Lorentz transformation equations, some of those assumptions will have to be modified, but the method will be exactly the same.

5.3.1 Galilean transformation

If two different observers measure the spatial location and the time of the same event, it is useful to know how the measurements of the two observers compare. That is, how can we *transform* the quantitative results of one observer to obtain those of the second observer? Galileo and Newton asked this question, and they were able to answer it in the following manner.

Galileo developed his understanding through his study of projectiles in the uniform gravitational field near the Earth's surface. He realized that any trajectory was parabolic, and, more important, that the horizontal and vertical motions were independent. This meant that if one observer saw a parabola, another observer, moving with the same (constant) horizontal velocity of the projectile, would see the projectile simply rise and fall, as if it had initially been thrown straight upward.

Consider Fig. 5.5, where and event P takes place at a specific position and time. An observer in the unprimed frame O (sometimes called the "lab frame") measures the event to take place at (x, y, z) and time t. An observer in the primed frame O' (sometimes called the moving frame or the "rocket frame," which is moving with speed u in the positive x direction relative to the lab frame[8]) measures the same event at (x', y', z') and time t'.

[8]I use u for the relative speed of the two frames because the point P may be moving, and its speed

How are the sets of coordinates related? The first, and most fundamental underlying idea is that

$$t = t'. \tag{5.16}$$

This relation means that time is absolute — all observers, regardless of their frame of reference, experience the same time. As Newton put it

> *Absolute, true and mathematical time, of itself, and by its own nature, flows uniformly on, without regard to anything external...*[9]

This proposition, which seems obvious, turns out not to be true. The second relation, also obvious, is that the coordinates perpendicular to the relative motion of the two coordinate systems are equal

$$y = y' \tag{5.17}$$
$$z = z'. \tag{5.18}$$

This will also be true relativistically, and can be shown quite easily by having someone in the O' frame drag a piece of chalk along a wall that is stationary in the O frame. This line on the wall *must* be the same distance from the x axis that the chalk is from the x' axis.

Finally, what about the x and x' coordinates? Since x' is the distance between the y' axis and point P (at time t'), and the distance between the y and y' axes increases uniformly with time, it must be true that

$$x = x' + ut'. \tag{5.19}$$

This just expresses the fundamental notion that the total length of a straight line is the sum of the lengths of the segments that make up the line. Of course, we could write $x = x' + ut$ since $t = t'$ by Eq. (5.16), but Eqs. (5.16)–(5.19) take the form of a transformation, in which the unprimed variables (x, y, z, t) are expressed as function of the primed variables (x', y', z', t'). In this way, knowing the space and time coordinates of an event as measured in one coordinate system, we can predict the coordinates of the same event as measured in another coordinate system.

How do velocities of objects, as measured by the two observers, transform? We should obtain, of course, the "relative velocity" formula that you have learned in elementary mechanics. If point P labels the position of an object that happens to be moving, then we can take its position as a function to time to be a series of events, each with a spatial and temporal location in each frame of reference. If, for example, the velocity components are known in O', then the Galilean transformation equations can be differentiated to give the velocity components in the O frame.

To show how this works, let's differentiate Eq. (5.17) with respect to t. This, of course, will give us the y-component of velocity in the O frame

$$\frac{dy}{dt} = \frac{dy'}{dt} = \frac{dy'}{dt'}, \tag{5.20}$$

will be denoted by v and v' as measured by observers in the two frames.

[9]Newton, *Principia*, 1687.

where the second equality holds because *time is absolute.* The left-hand-side of Eq. (5.20) is the definition of v_y, and the right-hand-side is the definition of v'_y. A similar analysis results in $v_z = v'_z$.

The x component of velocity is slightly trickier. Equation (5.19) differentiated gives

$$\frac{dx}{dt} = \frac{dx'}{dt} + \frac{d}{dt}(ut') = \frac{dx'}{dt'} + u, \tag{5.21}$$

where again, absolute time has been invoked. This last result can be written more concisely

$$v_x = v'_x + u, \tag{5.22}$$

which is nothing but the relative velocity formula.

The final transformation is that of the accelerations. Because the relative velocity of the two frames u is not a function of time, the accelerations are identical. A time derivative of Eq. (5.22) gives

$$a_x = a'_x, \tag{5.23}$$

with a similar transformation for the other coordinates. Since the accelerations are the same, and the force on an object is the same, then Newton's second law takes the same form in all reference frames! This is just Einstein's Principle of Relativity.

5.3.2 Lorentz transformation

How did Lorentz develop his transformation? Why was he unhappy with the Galilean transformation? Because, although Eqs. (5.16)–(5.19) correctly showed that Newton's laws had the same form in all reference frames, when he applied the Galilean transformation to Maxwell's equations of electrodynamics, the form of Maxwell's equations were modified, which was not in accordance with Einstein's Principle of Relativity. Even before Einstein's work in 1905, therefore, it was realized that Newton's laws of dynamics were not consistent with Maxwell's equations. So Lorentz developed a transformation between two coordinate frames that were moving at a constant velocity relative to each other, with the requirement that the form of Maxwell's equations was invariant (a slightly different requirement than the Galilean transformation). In other words, Lorentz proved a "theorem of corresponding states."[10] This theorem says that if $\vec{E}$ and $\vec{B}$ are the electromagnetic fields in a coordinate system ($\vec{x}$, t) that is *at rest relative to the ether*, then in a second coordinate system ($\vec{x}'$, t') moving with velocity $\vec{u}$ relative to the first, then, to first order in u/c, $\vec{E}'$ and $\vec{B}'$ are the same functions of ($\vec{x}'$, t') as $\vec{E}$ and $\vec{B}$ are of ($\vec{x}$, t), if the coordinates transform as Eqs. (5.24) and (5.25), below.

What is the ether? It was believed that electromagnetic waves must propagate in some medium — just like sound propagates in a gas or solid — and that medium was called the "luminiferous aether," or ether for short. The ether was supposed to be at rest with respect to Newton's absolute space, and it was thought that its existence allowed a measurement of the absolute velocity of light — with respect to the ether. However, Einstein showed in 1905 that it was possible to understand the Lorentz transformation in

[10]Pais, *Subtle is the Lord*, page 124.

a “relative” manner, such that the speed of light was constant in all reference frames, and no ether was required.

A simple method to obtain the Lorentz transformation is to assume a linear transformation between the primed and unprimed coordinates. The correct transformation might not be linear, but one should always make the simplest attempt first. The simplest “guess” is therefore

$$\frac{x}{c} = A\,\frac{x'}{c} + B\,t' \tag{5.24}$$

$$t = C\,\frac{x'}{c} + D\,t', \tag{5.25}$$

and our task is to determine A, B, C, and D. Of course, Eqs. (5.17) and (5.18) must still hold due to the chalk-on-the-wall argument. In addition, we require that in the limit of low speeds, $u/c \to 0$, the Lorentz transformation should be approximated by the Galilean transformation in Eqs. (5.16) and (5.19).

We can obtain B and D from the muon thought experiment by assuming that the muon remains at the origin (at rest, of course) of the primed frame, O'. Further mathematical simplification occurs if the muon is created at time $t = t' = 0$, when the two origins are co-located. This makes the coordinates of event 1 (muon creation)

$$\begin{array}{ll} x_1 = 0 & x_1' = 0 \\ t_1 = 0 & t_1' = 0 \end{array} \tag{5.26}$$

The coordinates of event 2 (muon decay) are, and I'll leave off the subscript 2 in the interest of notational simplicity,

$$x' = 0 \tag{5.27}$$

$$t' = t_0 = 2.2\ \mu\text{s}, \tag{5.28}$$

because in its own frame, the muon doesn't move and lives for 2.2 μs. What about the coordinates of event 2 in the unprimed frame? The time, as we discovered above, is just dilated,[11] $t = \gamma_r t'$, and the position is simply the position of the origin O' after a time t: $x = ut = u\gamma_r t'$, or

$$\frac{x}{c} = \beta_r\,\gamma_r\,t', \tag{5.29}$$

where

$$\gamma_r \equiv \frac{1}{\sqrt{1-\beta_r^2}} \qquad \text{and} \qquad \beta_r \equiv \frac{u}{c}, \tag{5.30}$$

are the usual relativistic factors for the *relative* velocity of the two frames. These two facts mean that

$$B = \beta_r\,\gamma_r \tag{5.31}$$

$$D = \gamma_r. \tag{5.32}$$

[11] Recall that we used the invariant interval to get

$$t'^2 = t^2 - \frac{x^2}{c^2} = t^2 - \frac{u^2}{c^2}t^2 = \frac{t^2}{\gamma_r^2}.$$

What about A and C? The muon thought experiment doesn't give us any information because $x' = 0$ for all time, so the coefficients of x' can take on any values and the transformation in Eqs. (5.24) and (5.25) will still correctly describe the muon creation and decay. One way to determine A and C is to utilize the invariance of the interval. Starting with the partially complete transformation equations that we have just determined

$$\frac{x}{c} = A\,\frac{x'}{c} + \beta_r\,\gamma_r\,t' \tag{5.33}$$

$$t = C\,\frac{x'}{c} + \gamma_r\,t', \tag{5.34}$$

we can insert the values for the unprimed coordinates $(x,\,t)$ into the right-hand-side of

$$t'^2 - \frac{x'^2}{c^2} = t^2 - \frac{x^2}{c^2}. \tag{5.35}$$

Note that this is identical to Eq. (5.9) since the coordinate values for event 1 are all zero. The final step is to require that the coefficients of both x' and t' match (see Problem 95). The final result is the complete Lorentz transformation

$$\frac{x}{c} = \gamma_r\,\frac{x'}{c} + \beta_r\,\gamma_r\,t' \tag{5.36}$$

$$t = \beta_r\,\gamma_r\,\frac{x'}{c} + \gamma_r\,t'. \tag{5.37}$$

A final check is to make sure that these reduce to the Galilean transformation in the limit of small velocities. Approximating the relativistic factor by $\gamma_r \approx 1 + \beta_r^2/2$, and keeping terms of order β_r or smaller, I find that Eqs. (5.36) and (5.37) become

$$\frac{x}{c} \approx \frac{x'}{c} + \frac{ut'}{c} \tag{5.38}$$

$$t \approx \frac{ux'}{c^2} + t'. \tag{5.39}$$

The first equation is identical to Eq. (5.19), but the second equation has an extra term. We can show that ux'/c^2 is actually of order β^2 (and therefore can be ignored) in the following way. Any object represented by the position x' must be moving slow compared with the speed of light, so I can write $x' = v't'$, where v' is its speed as measured in the primed frame. Using this replacement, Eq. (5.39) can be written

$$t \approx t'\left(1 + \frac{uv'}{c^2}\right). \tag{5.40}$$

Since the second term has two powers of velocity, $\beta_r\beta'$, it is of order β^2, as claimed.

5.4 Paradoxes

There are several situations in which the special relativistic result appears initially incorrect or inconsistent. A closer look, however, reveals that there is no inconsistency, and that the relativistic result is correct and can be confirmed by experiment.

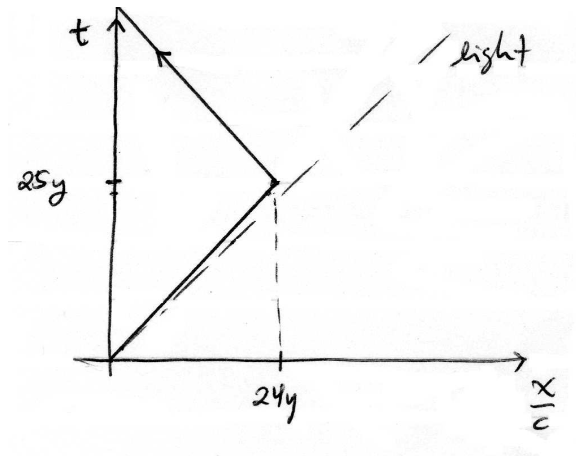


Figure 5.6: The spacetime diagram depicting the twin paradox. Paul's world line is just the t axis — he remains at the origin. Peter's world line consists of two segments, each inclined at an angle θ with the t axis, where $\tan\theta = v/c = 0.96$, or $\theta = 43.83°$.

5.4.1 The twin paradox

The twin paradox is usually stated as follows:

> On their twenty-first birthday, Peter leaves his twin Paul behind on the earth and goes off in the x direction for seven years of his time at $(24/25) = 0.96$ the speed of light, then reverses direction and in another seven years of his time returns at the same speed. (a) What is Peter's age on his return? (answer: 35 years) (b) How old is Paul at the moment of reunion? (answer: 71 years)
>
> When Peter returned from his fourteen years of traveling he was still young enough to learn some relativity. But the more he studied the more puzzled he became. He and his brother Paul, being in relative motion, "each should see the other's clocks running slow." This simple slogan, put in Paul's mouth, made it easy enough to understand why Peter's clocks—and Peter's aging process—ran slow, so that Peter was the younger of the two on his return. "But if the slogan is valid," Peter asked. "then would not *I*—if I had investigated—have found *Paul's* clocks running slow? So how did he age more than I?" What is the way out of Peter's difficulties?[12]

Figure 5.6 shows the world lines for both Peter and Paul. The seeming paradox is resolved by noting that the world lines of the two twins are not identical. In fact, in this inertial reference frame, Paul's world line is straight, while Peter's world line is curved. Given that they start and end at the same spacetime location, it's a direct consequence of the hyperbolic geometry that the clock which traveled the straight line (Paul's) recorded the greatest elapsed time. Another way of "proving" that Peter was the twin that left and came back is the fact that he must have accelerated. As we have already discussed, it is trivial to decide if your reference frame is accelerating or not. The fact that Peter

[12] Taylor and Wheeler, *Spacetime Physics*, pages 71, 94.

accelerated means that his inertial reference frame on the outbound trip was different from his inertial reference frame on the inbound trip. Finally, besides the observations of cosmic ray muons, there is concrete experimental verification of this resolution of the paradox: clocks have been flown around the Earth, traveling fast by terrestrial standards, but slow compared with the speed of light, and their elapsed time has been compared with stationary clocks. Result: the moving clocks run slow.

Let's analyze the twins' aging mathematically using our knowledge of time dilation and length contraction effects. (A similar case is in Problem 93.) Peter's age is quite simple, since the problem states that he travels for a total of 14 years *of his time*. This means he must have aged 14 years and so is 35 years upon his return. Paul's age is slightly trickier. The 14 years that Peter's clock measured we can take to be the "proper time" of the journey. Peter thus acts like the cosmic ray muon, traveling at 96% of the speed of light. It is necessary first to calculate the relativistic factor γ for Peter. Since $\beta = 24/25$, you can show that $\gamma = 25/7 \approx 3.57$.[13] Equation (5.7) then tells us how the coordinate time t compares with the proper time t_0

$$\Delta t = \gamma\,\Delta t_0 = \frac{25}{7}\,14 \text{ years} = 50 \text{ years}. \tag{5.41}$$

Therefore Paul is 71 years old.

What is the distance that Peter traveled? From Paul's point of view — Paul is in the "stationary" reference frame, so he measures the proper length — his brother traveled

$$\frac{\Delta x}{c} = \frac{L_0}{c} = \frac{u\Delta t}{c} = \frac{24}{25} \times 25 \text{ years} = 24 \text{ years} \tag{5.42}$$

each way for a total of 48 years. Or, converting to meters gives 2.27×10^{17} m each way. How far did Paul *think* he traveled? Since Paul was moving past the 'object,' he saw it length-contracted by a factor γ, which means that he thought his trip distance was $L/c = L_0/c\gamma = 24$ years$(7/25) = 6.72$ years, or 6.35×10^{16} m each way. This is just like the muon that observed the 20-km distance between the top of the atmosphere and the ground to contract to 660 m.

Can this twin paradox analysis be applied to astronauts on the Space Shuttle or the ISS? Do those astronauts (since they are accelerating, they take the place of Peter in our story) age more slowly that their friends they left behind on Earth? Yes, but by how much? Assuming they are in LEO (low Earth orbit, altitude $\approx$ 300 km), their speed (relative to the Earth) is 7.7×10^3 m/s. This gives them $\beta = 2.58 \times 10^{-5}$ and a relativistic factor of $\gamma = 1 + 3.32 \times 10^{-10}$. The time dilation effect means that their clocks run more slowly than clocks on Earth by this factor, or, for a mission of 1 year, the astronaut 'twin' is younger by 10 ms![14]

5.4.2 Einstein's *Gedanken* experiment on simultaneity

One of the perplexing new truths of special relativity is the fact that events that are simultaneous is one frame of reference are not simultaneous in another frame. This, of

[13] I have used the fact that a 7-24-25 triangle is a right triangle.

[14] I used the approximation $\gamma \approx 1 + \beta^2/2$ and $\gamma^{-1} \approx 1 - \beta^2/2$, where $\beta^2/2 = 3.32 \times 10^{-10}$. Multiplying this by 1 year $\approx \pi \times 10^7$ s gives 10 ms.

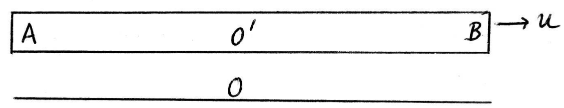


Figure 5.7: The geometry of Einstein's simultaneity experiment.

course, cuts at the heart of Newton's concept of absolute time. If there truly is a cosmic timekeeper, making sure that time everywhere in the universe flows "uniformly on," then any event should be observed simultaneously by all observers. However, the fact that the speed of light is not infinite, but finite (albeit large), implies directly that information takes a finite amount of time to propagate from the event to the observer. Since each observer is a different distance from the event, and perhaps moving relative to the event as well, they each will observe that the event occurred at a different time. If there are two events, they may be simultaneous for some observers and distinct for others.

In 1917, Einstein developed a *Gedanken* experiment that explored — and elucidated — this strange notion of non-simultaneity. He envisioned a train moving to the right with speed u, with three men riding on the train, one at the front (point B), one at the rear (point A), and one in the middle. There is also a man on the train platform (i.e., in a "stationary" frame of reference). He then asked the following question

> Are two events which are simultaneous *with reference to the railway embankment [platform]* also simultaneous *relatively to the train*? We shall show directly that the answer must be in the negative.[15]

The specific event geometry is as follows (see Fig. 5.7): The two men at A and B flash lights toward the center of the train. At the instant that the man at the center of the train (the origin of the O' coordinate system) passes the man on the platform (the origin of the O coordinate system), the two flashes of light reach both men. Since by definition that is the origin of both time coordinates also, it means that at $t = t' = 0$ the two flashes of light arrive at O and O' from A and B.

Who emitted their signal first, A or B?

The answer depends on who you ask. From the point of view of the man on the train, in the O' frame, everyone is at rest, A and B are equidistant and therefore they emitted the light *simultaneously.* From the point of view of the man on the platform, in the O frame, the flashes must have been emitted *before* O' reached O, and since at that time B was closer to O than A was, A must have sent the signal first.[16]

Who *really* emitted their signal first?

The Newtonian worldview states that there is a reality that is independent of any observation, so regardless of the fact that different observers make different measurements, there must be a way to definitively state what the reality of the situation is. Unfortunately, there is not. As Comstock states

[15] Einstein, *Relativity*, page 25. The italics are Einstein's.

[16] An observer in a third frame of reference, O'', moving to the right *faster* than the train would conclude that B sent his signal first.

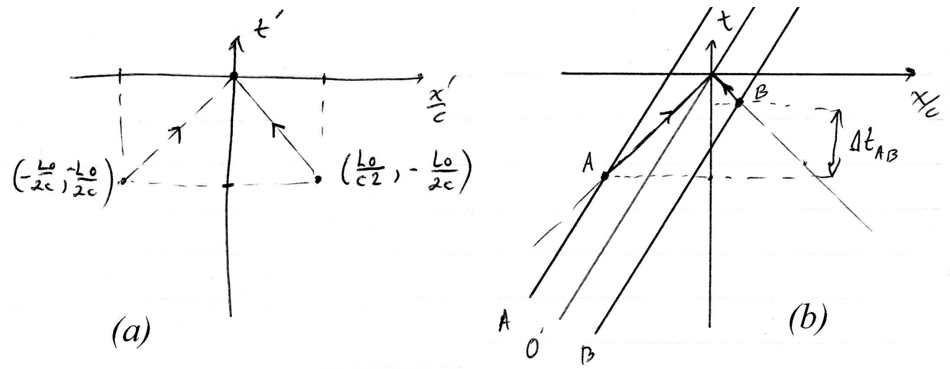


Figure 5.8: The two spacetime diagrams depicting the train and light rays from both reference frames. (a) The moving frame O' In this frame, the emission of the light rays is simultaneous. (b) The stationary frame O. Here, A and B label the world lines of the ends of the train, and O' labels the world line of the center of the train. In this frame, the man at the rear of the train (A) emits his flash of light before the man at the front of the train (B).

> We are, therefore, forced to the conclusion that, unless we discard one of the two relativity postulates, the simultaneity of two distant events means a different thing to two different observers if they are moving with respect to each other.[17]

We can solve this problem quantitatively with the tools we now have at our disposal. Figure 5.8(a) shows the spacetime diagram as viewed in the O' frame. The light signals meet at (0, 0), and since they must have traveled along the past light cone from points that were equidistant, they were emitted at the same time

$$t'_A = -\frac{L_0}{2c} \qquad t'_B = -\frac{L_0}{2c}, \tag{5.43}$$

where L_0 is the proper length of the train, and A and B denote the times that the flashed were emitted by the two sources. What about the observer on the ground? Figure 5.8(b) depicts the situation from his point of view. As can be seen graphically, A indeed emitted his signal first. By how much? Since we know the coordinates of each of the events (emission of light signals) in the O' frame, we can use the Lorentz transformation equations to calculate the coordinates of those same two events in the O frame.

Emission of light by A: The coordinates of this event in the primed frame are $t' = -L_0/2c$, $x'/c = -L_0/2c$. Of course they are identical because they are on a line that passes through the origin O' and has a slope of 1. The Lorentz transformation gives

$$\frac{x}{c} = \gamma_r\left(-\frac{L_0}{2c}\right) + \beta_r\,\gamma_r\left(-\frac{L_0}{2c}\right) = \left(-\frac{L_0}{2c}\right)\gamma_r(1+\beta_r) \tag{5.44}$$

$$t = \beta_r\,\gamma_r\left(-\frac{L_0}{2c}\right) + \gamma_r\left(-\frac{L_0}{2c}\right) = \left(-\frac{L_0}{2c}\right)\gamma_r(1+\beta_r). \tag{5.45}$$

[17]D. H. Comstock, "The principle of relativity," Nature **31**, 767-772 (1910).

Of course they have the same values for the same reason as the other reference frame. Most important, however, is the fact that the quantity $\gamma_r(1+\beta_r)$ is greater than unity (when $u > 0$), and it is equal to unity when $u = 0$. The proof is straightforward since $\gamma_r \geq 1$ and $\beta_r \geq 0$. This means that it always takes longer for the flash from A to arrive for the man on the platform than it does for the man on the train.

Emission of light by B: Using the coordinates of this event in the primed frame, $t' = -L_0/2c$, $x'/c = L_0/2c$, gives the time of this event in the unprimed frame as

$$t = \left(-\frac{L_0}{2c}\right)\gamma_r(1-\beta_r). \tag{5.46}$$

You can show (Problem **??**) that the factor $\gamma_r(1-\beta_r) \leq 1$, the equality holding when $u = 0$, as before. Our final result, therefore, is that the time difference between the emission of the flashed from A and B as viewed by the man on the platform is

$$\Delta t_{AB} = \left(\frac{L_0}{c}\right)\gamma_r\beta_r > 0, \tag{5.47}$$

which is positive definite. The man on the platform always thinks A emitted the flash first.

This result can be obtained in another, slightly more elegant manner. The Lorentz transformation equations are not just valid for coordinates of an event, they are also valid for intervals between events. In particular, for this interval between the flashes of light we have from the time equation, Eq. (5.37),

$$\Delta t = \gamma_r \Delta t' + \gamma_r\beta_r\frac{\Delta x'}{c}. \tag{5.48}$$

Since the observer on the train sees the flashes to be simultaneous, $\Delta t' = 0$, and the spatial separation to be just the proper length of the train, $\Delta x' = L_0$, the transformation equation gives the same result as Eq. (5.47), $\Delta t_{AB} = L_0\gamma_r\beta_r/c$.

5.5 Addition of velocities

Another strange result of Einstein's two postulates (one of many) is that no object with a nonzero mass can travel at (or faster than) the speed of light c. This result is independent of the frame of reference! It means that velocities do not add the way Galileo thought they should — Eq. (5.22). That equation states that if you are in a "stationary" frame of reference, standing on the side of the highway, for example, watching your friend drive by (in a moving frame of reference) at speed $u = 60$ mph, and if your friend throws a ball forward at speed $v'_x = 10$ mph (relative to the car), the you will observe the ball traveling at a speed $v_x = 70$ mph (relative to the ground). That statement (that velocities just add) is so patently obvious it is no wonder that Galileo believed it to be true. It turns out, though, that you observe the speed of the ball to be slightly *less* than 70 mph. If we take our example to the extreme and consider the situation where your friend's relative speed is $0.9c$, and he throws the ball at $0.9c$ forward, it can't be true that you observe the ball traveling at $1.8c$.

To determine the correct relative velocity formula we must differentiate the Lorentz transformation Eqs. (5.36) and (5.37) in the same manner as the Galilean transformation. We will assume the simplest case, namely, that the relative velocity of the two frames u is constant (which implies that β_r and γ_r are both constant). The differentiation process is a little bit trickier this time, since time is not absolute. The velocities in the two frames are defined as

$$v_x \equiv \frac{dx}{dt} \qquad \text{and} \qquad v'_x \equiv \frac{dx'}{dt'}, \tag{5.49}$$

where v'_x must be a derivative of x' *with respect to* t'. The simplest method is to differentiate Eq. (5.36) with respect to t'

$$\frac{1}{c}\frac{dx}{dt'} = \beta_r\gamma_r + \frac{\gamma_r}{c}\frac{dx'}{dt'} = \beta_r\gamma_r + \frac{\gamma_r}{c}v'_x. \tag{5.50}$$

The left-hand-side is *not* v_x because the derivative is with respect to t', not t. However, we can use the chain rule to obtain

$$v_x = \frac{dx}{dt} = \frac{dx}{dt'}\frac{dt'}{dt}, \tag{5.51}$$

which means that we need to evaluate dt'/dt, or in other words, the derivative of Eq. (5.37)

$$\frac{dt'}{dt} = \gamma_r + \frac{\beta_r\gamma_r}{c}\frac{dx}{dt} = \gamma_r + \frac{\beta_r\gamma_r}{c}v_x. \tag{5.52}$$

Now, inserting Eqs. (5.50) and (5.52) into the right-hand-side of Eq. (5.51) gives

$$\beta_x \equiv \frac{v_x}{c} = \left(\beta_r\gamma_r + \frac{\gamma_r}{c}v'_x\right)\left(\gamma_r + \frac{\beta_r\gamma_r}{c}v_x\right). \tag{5.53}$$

Unfortunately, there's a factor of v_x on both sides of the equation, so we need to rearrange, solve for β_x and simplify

$$\beta_x = \frac{\beta_r + \beta'_x}{1 + \beta_r\beta'_x}, \tag{5.54}$$

where $\beta'_x = v'_x/c$. Multiplying through by a factor of c gives a more familiar result

$$v_x = \frac{u + v'_x}{1 + \frac{uv'_x}{c^2}}. \tag{5.55}$$

Now the logic is clear: if both speeds, u and v'_x, are small compared with c, then the denominator is approximately unity, and the velocities add in a Galilean manner. However, as the speeds increase and become a sizable fraction of c, the extra term in the denominator keeps the "sum" from exceeding c. For example, if $\beta_r = 0.5$ and $\beta'_x = 0.5$ (your friend is driving at half the speed of light and throws a ball at half the speed of light) then the speed of the ball as observed by you is only $0.8c$. Another important example is that if your friend "throws a photon," which means $v'_x = c$, then no matter what the value of u is, Eq. (5.55) gives $v_x = c$ (you should show this). Photons travel at the speed of light *as measured by any observer*, confirming Einstein's second postulate.

5.6 Relativistic dynamics

In the preceding sections, we have focused on constant velocity kinematics. In that restricted case, when you are measuring phenomena in your own reference frame, e.g., velocities of particles, you don't see any relativistic effects. However, we have seen that when comparing your measurements of length and time with those of an observer in another reference frame you will detect that something is amiss. Each of you will conclude that the same laws of physics apply, but not agree on the detailed distances and times. However, we have not yet determined the form of those laws. They can't be the ones we are familiar with (i.e., $\vec{F} = m\vec{a}$) because we have just shown that particles can't travel faster than c, and $\vec{F} = m\vec{a}$ implies that by applying a constant force to a particle we should be able to accelerate it to any velocity whatsoever (if we apply the force for a long enough time). But we know that is incorrect. So we must use the relativistically correct laws.

It turns out that the familiar form of Newton's 2nd Law, $\vec{F} = m\vec{a}$ does *not* hold relativistically. The relativistically correct form is the one in which Newton first stated it

$$\vec{F} = \frac{d\vec{p}}{dt} \tag{5.56}$$

where

$$\vec{p} = \gamma m\vec{v}. \tag{5.57}$$

Why is this definition of momentum the correct definition? Because we still require that momentum must be conserved. If you analyze high velocity collisions between particles, the requirement of momentum conservation will lead you to Eq. (5.57). Some authors define a "relativistic mass" to equal to γm, where m is the "rest mass." This implies that the mass of an object increases as its speed increases. Currently, however, the more rigorous viewpoint is that the mass m is an invariant property of a particle, and that it is simply the momentum that increases as the speed increases so that it takes a larger and larger force to change a particle's momentum (as its speed becomes close to the speed of light).[18]

What about energy? You would think (and you'd be correct) that we should require that energy still be conserved, too. For this to be true, the energy of a particle must take the form

$$E = \gamma mc^2. \tag{5.58}$$

This is the total energy, kinetic plus rest, $E = E_0 + K$, where $E_0 = mc^2$ is the rest energy of the particle, as we've seen many times before.

Actually, there is another requirement that both the energy and momentum must satisfy (besides conservation), and that is they must reduce to the nonrelativistic forms when $v \ll c$. Since $\gamma \to 1$ in that limit, it is clear that $\vec{p} \to m\vec{v}$ in the low-velocity limit. If we let $\gamma = 1$ in Eq. (5.58), we would obtain simply $E = E_0$. This is, of course, true when $v = 0$, but what if v is small, but not zero? As we saw in Chapter 1, we can expand γ for small values of β,

$$\gamma = \left(1 - \beta^2\right)^{-1/2} \approx 1 + \frac{\beta^2}{2}. \tag{5.59}$$

[18] See L. B. Okun, "The concept of mass," Phys. Today, **42**, 31-36 (June 1989), for a more complete discussion.

Therefore, the total energy is approximately

$$E \approx \left(1 + \frac{\beta^2}{2}\right) mc^2 = mc^2 + \frac{1}{2}mv^2 \tag{5.60}$$

which is just the rest energy plus the (nonrelativistic) kinetic energy.

In contrast with elementary mechanics, where the mass and velocity of a particle are considered to be the fundamental quantities (from which the energy and momentum can be calculated), the fundamental quantities describing a particle in relativistic dynamics are its total energy E and its momentum $\vec{p}$. As we derived in Chapter 1, the relationship between the energy and the momentum obtained from their definitions is

$$E^2 = (pc)^2 + (mc^2)^2. \tag{5.61}$$

In the nonrelativistic limit, $pc \ll mc^2$, the energy can be approximated as

$$E = \sqrt{(pc)^2 + (mc^2)^2} = mc^2\sqrt{1 + \left(\frac{p}{mc}\right)^2} \approx mc^2 + \frac{p^2}{2m}. \tag{5.62}$$

We now have three nonrelativistic approximations that are all essentially equivalent. The first is that of low speed, $v \ll c$; the second is low kinetic energy, $K \ll E_0$, and the last is low momentum, $p \ll mc$.

In the ultrarelativistic regime, on the other hand (large kinetic energy and large momentum), the energy of a particle can be approximated as

$$E \approx pc. \tag{5.63}$$

For photons, of course, this is an exact equality, $E = pc$, not just an approximation, since photons have zero rest mass. Even though they have zero rest mass, they do have energy and momentum.[19]

Kaon decay

The classic situation in which the conservation of relativistic energy and momentum must be applied is that of Compton scattering, described in App. F. There, a photon interacts with a free electron, and one way of explaining the experimental results is to assume that the photon is a massless particle with energy $E = h\nu$, and since it is ultrarelativistic, it has momentum $p = E/c$. The experiment is not consistent with the picture of an electromagnetic wave interacting with a classical electron. (However, see the discussion at the end of App. C concerning what minimum assumptions are needed to predict the experimental results.)

Another process which illustrates the necessity of the relativistic conservation equations is that of kaon decay. A neutral kaon K^0 is a heavy meson ($E_0 = 498$ MeV) that decays into two pions with a mean lifetime of 9×10^{-11} s

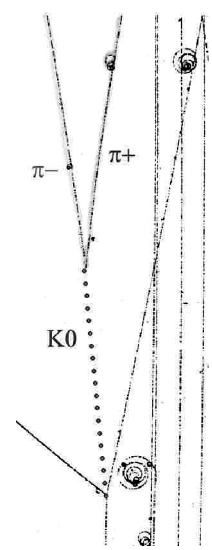
$$K^0 \rightarrow \pi^+ + \pi^-. \tag{5.64}$$

The rest energies of the pions are 140 MeV each, so that in the frame of reference in which the kaon is at rest, the reaction energy of the decay is $Q = 218$ MeV.

[19] Recall the discussion in Chapter 4 regarding the wavelength and frequency of light, and their relation to a photon's energy and momentum.

Figure 5.9: Schematic of the kaon decay in the laboratory frame. On the left is the situation *before* the decay, where the K^0 is traveling to the right with an energy E and momentum p. On the right are the particles *after* the decay, with the pions also moving in the x direction with energies $E_\pm$ and momenta $p_\pm$.

In this frame, the momentum of the kaon is zero, so that the total momentum of the pions must also be zero: they travel away from the location of the decay with equal kinetic energies ($K = 109$ MeV each) and oppositely directed momenta.[20]



The laboratory is not usually in this reference frame, however, since the kaon is typically the result of another interaction and therefore is moving when it decays. (The figure to the right shows a bubble chamber with tracks of various particles. Since the K^0 is neutral it does not ionize particles in the chamber and hence leaves no track, but its existence is inferred from the presence of the two pions.) To illustrate how either frame can be used to analyze such a decay, let's consider the situation where a kaon has 325 MeV of kinetic energy. This value is chosen because it is neither nonrelativistic nor ultrarelativistic, and no approximations can be made. The full relativistic expressions must be used.

Laboratory frame: Without any loss of generality, let's take the kaon's velocity to be in the positive x direction. To make the analysis simple, we'll assume that the resulting pions also move in the x direction (allowing them to have a nonzero y component of velocity is only slightly more difficult algebraically). Referring to Fig. 5.9, the K^0 has an energy E and momentum p, while the resulting pions $\pi^\pm$ have energy and momenta $E_\pm$ and $p_\pm$.

The first step is to analyze the motion of the kaon. Its relativistic factor is

$$\gamma = \frac{E}{E_0} = \frac{823\text{ MeV}}{498\text{ MeV}} = 1.65 \tag{5.65}$$

and hence its speed is $\beta = 0.796$. We know its total energy $E = E_0 + K = 823$ MeV, but its momentum must be obtained from Eq. (5.61)

$$pc = \sqrt{E^2 - (Mc^2)^2} = 655\text{ MeV}, \tag{5.66}$$

where $Mc^2 = 498$ MeV.

[20]The neutral kaon is a linear combination of the quark–anti-quark pairs $d\bar{s}$ and $s\bar{d}$. The combination I'm discussing here is called "K-short," because its lifetime is short. Another combination, "K-long," typically decays into three pions witha mean lifetime of 5×10^{-8} s.

The next step, just like in elementary mechanics, is to apply the conservation laws and obtain the pion energies and momenta. Since the reaction is one dimensional, there is only one component of momentum to worry about, and hence there are two conservation equations

$$E = E_+ + E_- \tag{5.67a}$$
$$pc = p_+c + p_-c. \tag{5.67b}$$

It appears that there are four unknowns, $E_\pm$ and $p_\pm$, but the energy and momentum of each the pions are related by Eq. (5.61), which gives us two more equations. Expressing the energies in terms of their respective momenta allows us to replace Eq. (5.67a) with

$$E = \sqrt{p_+^2c^2 + (mc^2)^2} + \sqrt{p_-^2c^2 + (mc^2)^2} \tag{5.68}$$

where $mc^2 = 140$ MeV. The simplest way to solve these two equations for the two unknowns, $p_\pm$, is to solve Eq. (5.67b) for p_+ and substitute that into Eq. (5.68). Solving then for p_- results in the quadratic equation for p_-c (see the box on page 178)

$$(p_-c)^2 - (pc)(p_-c) + \left(\frac{E^2(mc^2)^2}{(Mc^2)^2} - \frac{(Mc^2)^2}{4}\right) = 0, \tag{5.69}$$

where $(Mc^2)^2 = E^2 - p^2c^2$ is the square of the rest energy of the kaon.

Evaluating this equation for the current initial conditions of the kaon gives

$$p_-c = 667.9 \text{ MeV}, \qquad -12.7 \text{ MeV} \tag{5.70}$$

Which solution is the one we want? Both of them! Since Eq. (5.68) is symmetric in p_+ and p_-, we would have obtained the same result had we solved for p_+. This means that one of the pions is emitted traveling forward with momentum 667.9 MeV/c, and one is emitted backward with momentum 12.7 MeV/c, for a net momentum of 655 MeV/c, the initial momentum of the kaon. It is always a good idea to confirm energy conservation as well. The energies of two pions can be found from Eq. (5.61), 682.4 MeV and 140.6 MeV. These add, of course, to 823 MeV, which was the initial energy of the kaon.

Kaon rest frame: A simpler way to analyze this decay is in the rest frame of the kaon. Let's make a Lorentz transformation to the 'primed' frame that is moving with the kaon. The speed of the primed frame relative to the unprimed (laboratory) frame is just the speed of the kaon, $\beta_r = 0.796$. In this primed frame, the analysis is simple, as mentioned above: the pions each have a total energy $E' = E_0 + K' = 249$ MeV, and since they move in opposite directions, their momenta in the x' direction is

$$p'_\pm c = \pm\sqrt{E'^2 - (mc^2)^2} = \pm 205.9 \text{ MeV}. \tag{5.71}$$

In order to be able to transform *back* to the laboratory frame, and calculate the energy and momentum of the pions as measured by an observer in *that* frame, we need to calculate the relativistic factor γ' and the velocity β'_x of each pion as measured in the primed frame,

$$\gamma' = \frac{E'}{E_0} = \frac{249 \text{ MeV}}{140 \text{ MeV}} = 1.78 \tag{5.72}$$

and

$$\beta'_x = \pm\sqrt{1 - \frac{1}{\gamma'^2}} = \pm 0.827. \tag{5.73}$$

Now that we have the velocities in the primed frame, we can use Eq. (5.54) to solve for the velocities of the pions in the *unprimed* frame

$$\frac{v_\pm}{c} = \frac{\beta_r + \beta'_x}{1 + \beta_r \beta'_x} = \frac{0.796 \pm 0.827}{1 + (0.796)(\pm 0.827)} = \begin{cases} +0.979 \\ -0.090 \end{cases} \tag{5.74}$$

As expected, one pion has a large positive velocity, and the other has a small negative velocity. Do these velocities agree with the momenta calculated in Eq. (5.70)? The momentum pc is a function of the velocity β in the following way

$$pc = (\gamma m v)c = \gamma\beta\, mc^2. \tag{5.75}$$

Using our knowledge of $\beta_\pm$ these momenta are just 667.9 MeV/c and -12.7 MeV/c, exactly as calculated previously. You can also confirm that the energies are also identical.

We have shown that although the energies and momenta are measured to be different by observers in different frames of reference, Newton's laws (e.g., conservation of energy and momentum) still hold. Key to showing this is the Lorentz transformation, which was obtained from Einstein's two postulates: relativity and the constancy of the speed of light.

Algebra of kaon decay: Setting $c = 1$ changes Eq. (5.67b) to

$$p_+^2 = (p - p_-)^2,$$

and plugging this into Eq. (5.68) results in

$$E = \sqrt{(p - p_-)^2 + m^2} + \sqrt{p_-^2 + m^2}.$$

In this equation, all quantities are measured in energy units. The easiest way to solve this is to bring the second term on the ride side of the equation over to the left side, and then square both sides

$$\left(E - \sqrt{p_-^2 + m^2}\right)^2 = (p - p_-)^2 + m^2.$$

Expanding the squared terms on both sides gives

$$E^2 - 2E\sqrt{p_-^2 + m^2} + p_-^2 + m^2 = p^2 - 2pp_- + p_-^2 + m^2,$$

and the last two terms on each side cancel. Again, isolating the square root on one side, and then squaring gives

$$\left(E^2 - p^2 + 2pp_-\right)^2 = \left(2E\sqrt{p_-^2 + m^2}\right)^2 = 4E^2(p_-^2 + m^2).$$

Here, we can make a simplification by noting that $E^2 - p^2 = M^2$, where M is the mass of the kaon. Grouping terms results in the quadratic equation

$$0 = 4M^2 p_-^2 - 4M^2 pp_- + (4E^2 m^2 - M^4),$$

or

$$0 = p_-^2 - pp_- + (E^2 m^2/M^2 - M^2/4).$$

Restoring the factors of c gives

$$(p_- c)^2 - (pc)(p_- c) + \left(\frac{E^2(mc^2)^2}{(Mc^2)^2} - \frac{(Mc^2)^2}{4}\right) = 0.$$

The two solutions are 667.9008142 MeV and -12.68284657 MeV.