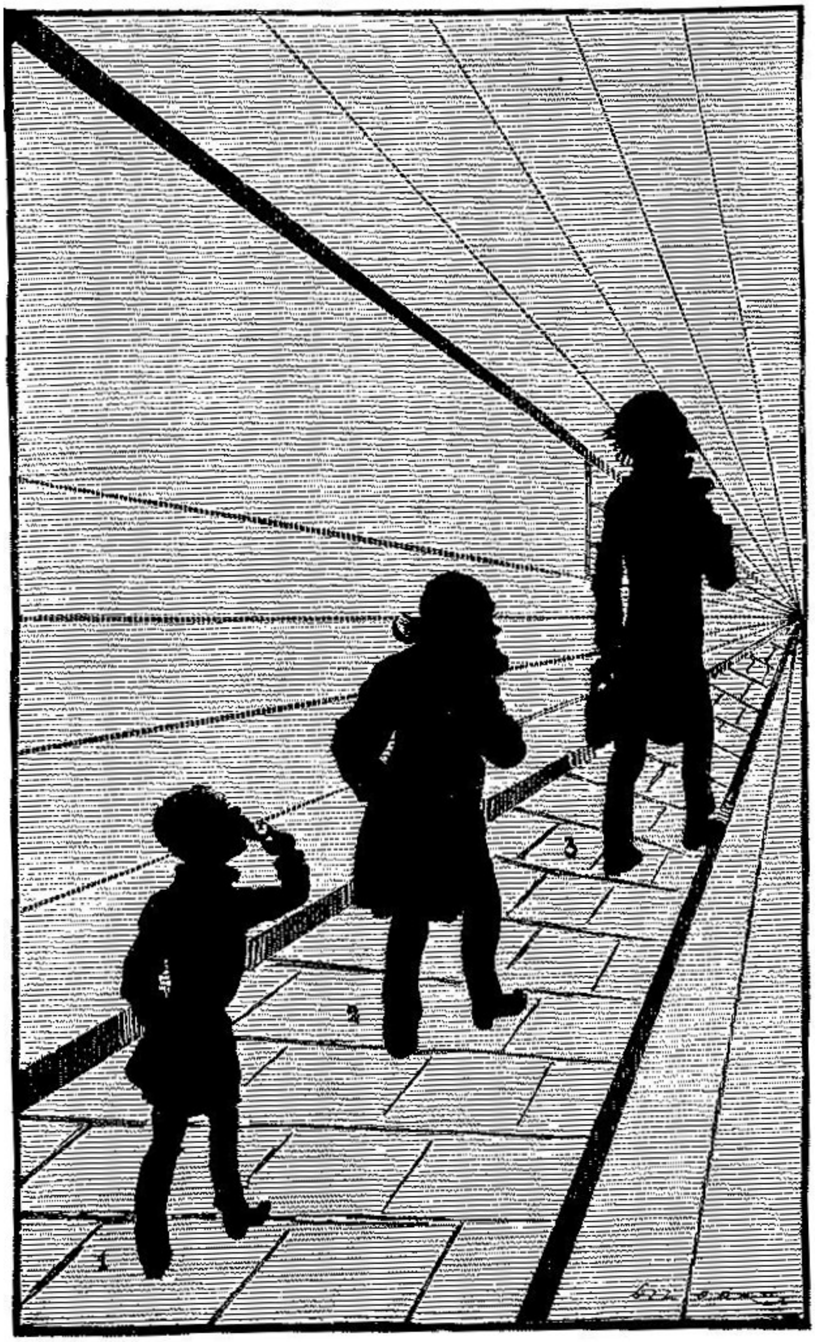
Special Relativity Primer

William L. Burke and Peter Scott



Here are 1. Lord Randolph Churchill,
2. Robert Arthur Salisbury, and
3. William Gladstone.

Which of them is the tallest?

About the cover. . .

Although intended originally as a simple optical illusion, the drawing* illustrates a key concept in Special Relativity Theory. The answer to the question “Which of them is the tallest?” depends upon your reference frame. If your reference frame is that of the tunnel, you get one answer. If it is that of the drawing itself, you get another. For further discussion, see page 30.

* Reprinted from Scientific American **53**, 82 (1885), whose editors write that it comes from an advertisement for an English soap manufacturer. The idea undoubtedly predates the nineteenth century.

Contents

Preface

In the early creative years here at UC Santa Cruz, Bill Burke, whose research interests focused on General Relativity and related topics, had the idea of incorporating a three-week segment on Special Relativity into the second quarter of our three-quarter introductory physics course sequence. Since treatments of Special Relativity in our introductory texts were then (and often still are) inadequate and confusing to students, Bill, in 1976, wrote up some notes. He gave these notes the title *Special Relativity Primer*.

In those notes, he focused on the then recent experimental basis for the behavior of *clocks* (decaying muons, for example), and how such behavior contradicted the Newtonian models, which treat the passage of time as absolute. He also emphasized the utility of *spacetime diagrams*, which brought focus to the hyperbolic geometry of spacetime and the role of invariants in that geometry.

I first used those notes when teaching the course in the Winter Quarter of 1977. Bill's treatment, which required only the understanding of the standard Newtonian mechanics introduced in the previous quarter, proved to be successful. Our students were more easily convinced that the behavior of mechanical systems, particularly those involving particles traveling at high velocity, was in fact well described by Special Relativity Theory.

Subsequently, I worked with Bill to rewrite and expand his notes so as to make them easier to follow, with defined sections, clear examples and many improved and simplified diagrams. We also added a section on units that are commonly used, especially by particle physicists.

In the Introduction to an earlier edition of our *Special Relativity Primer*, we wrote:

> In preparing these notes we have received considerable support. We thank the University of California Regents for providing a Summer Curriculum Improvement Grant to aid in the initial preparation of these notes in 1977. One of us (WLB) would like to express his appreciation to his relativity teacher Frank Estabrook, and to J.L. Synge's books for introducing him to spacetime diagrams. We also thank numerous students for their comments and suggestions, Patty Burns for typing the first edition of this manuscript, and Marilyn Stevens of the UC Santa Cruz Physics Office staff for invaluable assistance in preparing this revised edition.

Unfortunately these notes were never published, and so were available only to our UC Santa Cruz students. When in 1996, Bill Burke was killed in a tragic auto accident, it seemed worthwhile to transcribe the *Special Relativity Primer* into an easily portable electronic format so as to make it available to a wider audience. In the process, I have expanded the notes somewhat to update references, add some historical references, and in the process make use of the internet. I may also add an additional section describing briefly how the electromagnetic field may be dealt with using spacetime geometry, since this is a key part of Special Relativity Theory.

– Peter Scott, drip@ucsc.edu
February 2020

Introduction

Special Relativity Theory (SRT) provides a description for the kinematics and dynamics of particles in the four-dimensional world of spacetime. For mechanical systems, its predictions become evident when the particle velocities are high—comparable with the velocity of light. In this sense SRT is the physics of high velocity. As such it leads us into an unfamiliar world, quite beyond our everyday experience. What is it like? It is a fantasy world, where we encounter new meanings for such familiar concepts as time, space, energy, mass and momentum. A fantasy world, and yet also the real world: We may test it, play with it, think about it and experience it. These notes are about how to do those things.

SRT was developed by Albert Einstein in 1905. Einstein felt impelled to reconcile an apparent inconsistency. According to the theory of electricity and magnetism, as formulated in the latter part of the nineteenth century by Maxwell and Lorentz, electromagnetic radiation (light) should travel with a velocity c, whose measured value should not depend on the velocity of either the source or the observer of the radiation. That is, c should be *independent* of the velocity of the reference frame with respect to which it is measured. This prediction seemed clearly inconsistent with the well-known rule for the addition of velocities, familiar from mechanics. The measured speed of a water wave, for example, will depend on the velocity of the reference frame with respect to which it is measured. Einstein, however, held an intuitive belief in the truth of the Maxwell-Lorentz theory, and worked out a resolution based on this belief. He was driven to SRT mostly by aesthetic arguments, that is, by arguments of simplicity.

In his famous 1905 paper,[1] Einstein lays out the essential features. His paper contains an introduction, five sections on kinematics followed by five sections on electrodynamics, no references, and one acknowledgement. His development is based entirely on two postulates, which he termed the Principle of Relativity and the Principle of Constancy of the velocity of light. Here they are as he wrote them, translated from the German:

1. The laws by which the states of physical systems undergo change are not affected, whether these changes of state be referred to the one or the other of two systems of coordinates in uniform translatory motion.

2. Any ray of light moves in the "stationary" system of co-ordinates with the determined velocity c, whether the ray be emitted by a stationary or by a moving body. Hence

$$\text{velocity} = \frac{\text{light path}}{\text{time interval}}$$

   where time interval is to be taken in the sense of the definition in §1.

[1] A. Einstein, "Zur Electrodynamik bewegter Körper" (On the electrodynamics of moving bodies), *Annalen der Physik* **17**, 891 (1905). (A downloadable pdf containing a fine translation of this paper may be found here: https://www.fourmilab.ch/etexts/einstein/specrel/specrel.pdf.) This volume of *Annalen der Physik* contains, in addition to the above paper on special relativity, Einstein's original papers on the equivalence of mass and energy (see https://www.fourmilab.ch/etexts/einstein/E_mc2/e_mc2.pdf, the photoelectric effect (for which he was awarded the Nobel Prize), and on Brownian motion. If you look on the shelves of McHenry Library on the UCSC campus, you will find, where Vol. 17 should be, a block of wood. The volume is in Special Collections. It is indeed special. Look at it if you have a chance.

In modern language,[2] these two postulates may be stated as follows:

1. The laws of physics take the same form in all inertial reference frames. [The Principle of Relativity].
2. In empty space, light travels with a velocity c that is independent of the velocity of either the source or the observer. [The Principle of the Constancy of the velocity of light].

The development of SRT in most textbooks follows that of Einstein. From the two postulates, the Lorentz transformation equations may be derived, and from them a variety of strange new predictions, like how moving meter sticks are shorter, how moving clocks tick more slowly, and how nothing can go faster than the speed of light.

A pedagogical difficulty with this approach stems from the fact that only the second of the two postulates lends itself to direct experimental verification (the Michelson-Morley experiment, for example). The first postulate is difficult to justify except on the basis of one's belief in the simplicity of nature. A result is that suspicion develops among students (and occasionally instructors) about the validity of the various conclusions. Lengthy discussions commonly ensue about whether moving clocks "really" tick more slowly, and whether moving sticks "really" are shorter. In general, clear resolution of the common paradoxes of special relativity is difficult, the "twin paradox" being the chief example.

These notes represent an alternative approach to learning and understanding SRT. To start with, we do not rely on the "Principle of Relativity". Instead, we focus on the concept of a *clock*, and a description of the observed behavior of real clocks, and how such clocks are observed to tick more slowly when they are moving. Second, our approach is highly graphic, taking great care to understand the four-dimensional spacetime geometry within which SRT takes place, and making liberal use of spacetime diagrams to describe, illustrate and clarify each concept and observation. With this geometrical approach based on solid experimental evidence, coupled with definitions of certain primitive concepts such as *event*, *light signal* and *inertial reference frame*, it is easy to discuss the meaning of a spacetime interval, the meaning of simultaneity, the meanings of length contraction and time dilation, and finally to derive the Lorentz transformation equations. Descriptions of the various "strange" phenomena follow hard on the experimental evidence, and the arguments are clear and easily defended. The "twin paradox" in particular follows with clarity; indeed, it becomes difficult to see that there is even a paradox involved.

The discussion of relativistic dynamics is likewise clarified and simplified though the use of the geometrical approach, with a focus on the use of momentum-energy diagrams. With this technique, it becomes possible to present clear discussions of particle interactions, including pair annihilation, pair production and the Compton effect. Inherent in this treatment is the concept of the *4-vector*, and in particular, the *4-momentum* and the observation that 4-momentum is conserved. Clarity also ensues with the abandonment of the useless concept of "relativistic mass" often referred to in

[2] The expression "inertial reference frame" was not used by Einstein in his 1905 paper. Not until several years later did it come into vogue.

textbooks. In our treatment the "rest mass" is the only mass that appears. Finally, our approach lends clarity to the relations between energy and mass, a matter which is often somewhat obscure in the traditional approach.

We originally wrote these notes for use during a three-week segment of the second quarter of our three-quarter sequence entitled "Introduction to Physics". The material is preceded only by standard Newtonian mechanics, which is covered during the first quarter.

SRT deals with the most basic physical concepts. In the grandest sense, it provides us with a unifying framework encompassing not only mechanics, but also electricity and magnetism and quantum mechanics. Hence, it makes sense to put SRT near the beginning of our physics education. It's helpful to lay down the framework before building the rest of the structure. Perhaps the strongest reason, however, lies in the notion that a study of SRT can teach all of us something at a more basic level—a way of *thinking*. To start with the most elementary concepts we know, such as time, distance, velocity, momentum and energy, and to have our understandings of them drastically altered by the observational evidence, can only expand the mind, loosen up the brain, and cause us to distrust and re-evaluate all sorts of basic premises. In short, it teaches us to distrust our "common sense"—that set of predictions based on our past everyday experience. Since this kind of creative skepticism lies at the heart of the study of physics, it is appropriate that SRT be studied early on in the physics curriculum.

A great many people have written about special relativity. On the shelves of our library there are over 40 books with the words "special relativity" in the title, and numerous textbooks contain sections devoted to SRT. A thorough bibliography, though by no means encompassing all of the literature, appears in Gerald Holton's "Resource Letter SRT-1 on Special Relativity Theory", published in January 1962 in the American Journal of Physics. In addition, a collection of 16 reprints of journal articles listed in this bibliography has been compiled and published (along with the Resource Letter) by the American Institute of Physics under the title *Special Relativity—Selected Reprints*. Excellent translations of a number of Einstein's papers (including his famous paper of June, 1905), along with translations of other papers by Lorentz, Minkowski and Weyl, are contained in *The Principle of Relativity*, A. Sommerfeld, Ed. (Dover, 1923). A comprehensively annotated version of a portion of an earlier translation of the June 1905 paper is included in that marvelous collection entitled *Great Experiments in Physics*, edited by Morris H. Shamos (Holt, Rinehart and Winston, 1959).

Those wishing to trace the historical roots of SRT should take a look at Pais' marvelous biography of Einstein, *Subtle is the Lord...* (Oxford University Press, 1982). In addition, a delightfully readable, less technical but thoroughly-researched biography of Einstein by Walter Isaacson is now available. It's entitled *Einstein: His Life and Universe* (Simon & Schuster, 2007).

If you're looking for a book to supplement these notes with further explanations, exercises, problems, stories and paradoxes, the best source is *Spacetime Physics*, by Edwin F. Taylor and John Archibald Wheeler (W.H. Freeman and Co., 1992). This second edition has been extensively rewritten since it first appeared in 1966. Some say it has too many words, but we'll leave that for you to decide. However there is one gem from this book that we recommend. We reprint it here. It's from Page 20 of the new edition:

WHEELER'S FIRST MORAL PRINCIPLE: *Never make a calculation until you know the answer.* Make an estimate before every calculation, try a simple physical argument (symmetry! invariance! conservation!) before every derivation, guess the answer to every paradox and puzzle. Courage: No one else needs to know what the guess is. Therefore make it quickly, by instinct. A right guess reinforces this instinct. A wrong guess brings the refreshment of surprise. In either case life as a spacetime expert, however long, is more fun!

Another more traditional treatment, from which some of the ideas for these notes were taken, is *Einstein's Theory of Relativity*, by Max Born (Dover Publications, 1962). Gerald Holton calls it (prior to the existence of Taylor and Wheeler) "without doubt the best elementary account. Thoroughly works out everything, from how to plot graphs through Maxwell, Minkowski, to GRT, with very few rabbits being pulled out of the hat." It is indeed a very nice little book. An older, but highly readable volume is *Relativity, Thermodynamics and Cosmology*, by Richard Tolman (Oxford University Press 1934). It provides a good review of four-dimensional spacetime, with explanations of four-vectors and four-tensors that are helpful for those wishing to explore concepts in General Relativity Theory.

A fourth book, *Relativity—Special, General, and Cosmological*, by Wolfgang Rindler (Oxford University Press 2006), lays considerable stress on the geometrical aspects of relativity, and includes many problems and exercises, some of which we have incorporated into these notes. Both Born and Rindler also provide excellent introductions to General Relativity for those who may be inclined to make further explorations in that direction.

I. Spacetime and Spacetime Diagrams

In our geometrical approach to SRT, we shall emphasize the concept of *spacetime*, customarily represented by a *spacetime diagram*. The idea is a simple one, meant to draw attention to the intimate connection between the space and time coordinates that we use to describe the motion of particles in the universe.[3]

Spacetime is the non-Euclidean four-dimensional world delineated by the three spatial axes x, y, and z, and the time axis t. A spacetime diagram is a graphical representation of spacetime. The most convenient such representation is a two-dimensional cross-section of spacetime consisting (by convention) of a plot of the time coordinate t vs. a single space coordinate such as x. It is also possible to represent a three-dimensional cross-section, in which t is plotted vs. two spatial coordinates. The actual four-dimensional world of spacetime is impossible to represent on a piece of paper, or even in three-dimensional coordinate space.

A point in spacetime is characterized by a particular position at a particular time, and is called an *event*. We often label a particular *event* by its position coordinate x and its time coordinate t like this: (x, t).

In spacetime, the existence of a particle is represented by a sequence of events, called a *world line*, which we often abbreviate by *WL*.

For most of our examples we shall consider particles moving in only one spatial dimension, which we take to be the x-direction. For such motion, spacetime is two-dimensional, and can be easily drawn on a flat piece of paper. Examples of such diagrams are shown in Fig. 1.

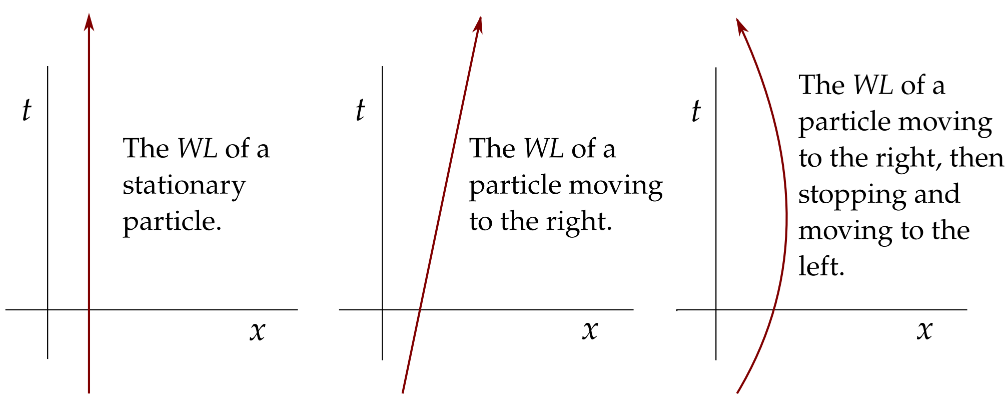


Figure 1: Three Simple Spacetime Diagrams.
Note that the time-axis is traditionally taken to be vertical.

[3] Spacetime diagrams were first introduced by Hermann Minkowski in 1908. They are therefore often referred to as *Minkowski diagrams*.

Later in our discussion we shall occasionally consider particles moving in two dimensions, so that the spacetime diagram becomes three dimensional:

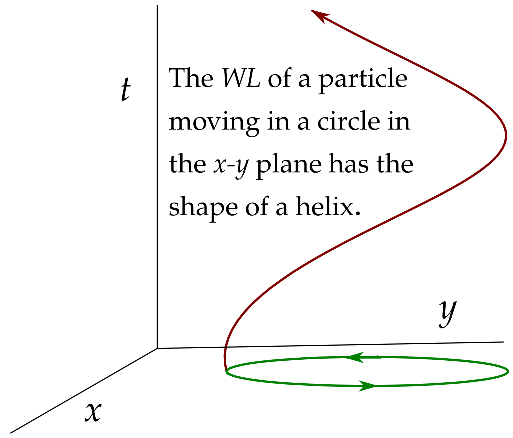


Figure 2: A Three-dimensional Spacetime Diagram

Of course, we may also consider particles moving in three spatial dimensions, but as mentioned above, the spacetime diagram becomes four dimensional and impossible to represent on a sheet of paper. Fortunately, nearly all of the interesting behavior of SRT can be found in a two-dimensional spacetime diagram.

As can be seen from the examples in Fig. 1, the greater the angle between the t-axis and the world line, the greater the speed of the particle. To represent this speed quantitatively, it is necessary to scale the position and time axes appropriately. In the Newtonian picture, where space and time are not related to each other, we may freely scale the space and time axes to suit the problem at hand. For example, when considering the motion of a golf ball, we might choose position in meters and time in seconds. In SRT, however, the speed of light will play a crucial role.

The light velocity provides a close connection between space and time, and it is helpful to scale the space axes so that distance is measured in units of time. Such scaling is already familiar to astronomers who often measure time in years, and distances in *light years*. Alternatively, if time is in seconds, distance will be in *light seconds*, or simply *seconds*. With the light velocity $c \approx 3 \times 10^8$ meters/sec, this means that 1 light second is 3×10^8 meters.

A foot is about a nanosecond (10^{-9} sec), the length of a football field is about 1/3 of a microsecond, the moon is about 1.3 seconds away, and the sun is about 500 seconds away. In our diagrams, we'll make this explicit by plotting t vs. x/c, rather than t vs. x. In such diagrams, a light signal must travel along either the line $t = x/c$ or $t = -x/c$, that is, along lines at 45° to the t-axis. We shall commonly draw such *light lines* as dashed lines as shown in Fig. 3, shown on the next page.

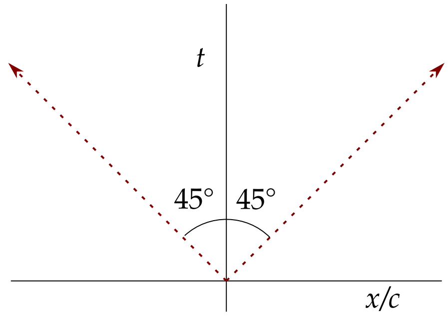


Figure 3: Light signals always go at 45 degrees to the t-axis.

Should we have occasion to consider motion in two spatial dimensions, where the spacetime diagram becomes three-dimensional, light signals must travel along the surface of a *cone* making an angle of 45° with the t-axis. This cone is commonly called the *light cone*:

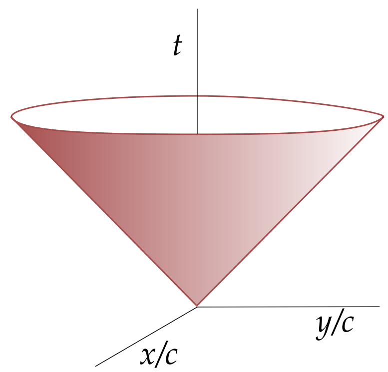


Figure 4: The Light Cone.

Note that in a spacetime diagram, a particle traveling at a constant velocity v will be represented by a world line making an angle θ with the t-axis, where $\tan\theta = v/c$, as shown in Fig. 5:

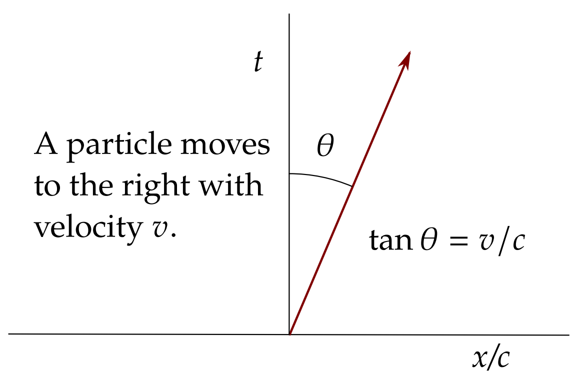


Figure 5: The World Line of a Moving Particle.

Later on in our discussion of SRT, we shall find it cumbersome to keep writing "c" in the denominator every time we wish to express a length or velocity. At that time we shall set $c = 1$, and write x instead of x/c, and v instead of v/c. After all, c is simply 1 light-second per second, and it is usually redundant to carry "c" along in all our formulas. In setting $c = 1$, we imply that x is measured in seconds and v is measured as a fraction of the light velocity. For example, for a particle traveling from the earth to the moon at half the speed of light, we would mean that it would travel a distance $x = 1.3$ seconds at a speed $v = 0.5$.

However, most of us find it confusing to be suddenly forced to think in terms of measuring distances in seconds. If I am asked "How tall are you?", it is not likely that I would say "6.3 nanoseconds", and I might be confused by a highway sign like that shown in Fig. 6, where the distances are in microseconds.[4] For a brief exercise, translate the distances shown in Fig. 6 to the more familiar miles.

Figure 6: A sign on Highway 1 at the northern boundary of Santa Cruz.

Hence for the next few sections, we shall put "c" in explicitly. This will serve to emphasize the important role played by the light velocity, as well as making it evident that it is x/c that is measured in seconds and v/c that is a fraction of the light velocity.

[4] We do, however, sometimes use units of time to specify distances, when we say, *e.g.*, "Oh, San Francisco is about an hour and a half north of Santa Cruz". In such statements, we imply that we are traveling in a car at some typical speed v in miles per hour. In other parts of the world, there are often signs on footpaths that specify distances in minutes or hours. In such cases a typical walking speed is implied.

II. Clocks, Clock Rates and Time Dilation

The notion of a *clock* is basic to our discussion of SRT. In its simplest form, we take it to be a single particle that either undergoes a periodic motion, like a vibrating molecule, or else goes "on" at one time and "off" at a later time, like a meson that is formed and later decays.[5] We shall assume that such clocks "tick" uniformly, that is, that they neither slow down nor speed up, that the mean lifetime of a pi meson is the same this year as it was last year, or that the vibrational frequency of an ammonia molecule is the same now as it was a thousand years ago. We also assume that clocks tick homogeneously, that is, that a clock's ticking rate does not depend on its spatial coordinates. Hence if we bring any two clocks together their relative rates will not depend upon where or when we bring them together.

We can use the language of a spacetime diagram and the concept of a clock to describe the behavior of moving clocks. This description will provide us with an essential key to the understanding of SRT.

First we imagine the following idealized experiment: Take a number of identical clocks, say a number of hypothetical mesons, each of which exists for exactly τ seconds. (We ignore the uncertain nature of the decay process). Suppose they are all born simultaneously at the origin of a horizontal ruler, or x-axis, and then move along the axis away from the origin, some to the left and some to the right, at various different constant velocities. One of the mesons could even remain right at the origin; it would have therefore zero velocity. In general, a meson moving at velocity v will move a distance $x = vt$ away from the origin after a time t, where t is the time measured by a clock at the origin of the x-axis. Now each meson will subsequently die (or decay); the death of each meson constitutes a distinct event with spacetime coordinates $(x/c, t)$. Here, we have arranged our experiment so that all the mesons are born at the event whose spacetime coordinates are $(0, 0)$.

Now we ask a question: How is the spacetime coordinate t for each of the decay events related to the meson lifetime τ? "Huh?" we say, "What do you mean? Since each meson lives for τ seconds, $t = \tau$ for each decay event." This is what Newton would have said too: This is the Newtonian prediction for clock rates. The entire experiment may be illustrated on a spacetime diagram, shown in Fig. 7 on the next page. Here we have drawn the world line for each clock, and have indicated each supposed decay event by a little dot on the diagram. A line drawn through all such events we call the line of "τ-second ticks".

[5] It may seem at first glance that a real particle that is formed and later decays does not constitute an accurate clock, because of the uncertain nature of the decay process. Given a number of particles, some will decay at times less than the mean life, some will decay at times greater than the mean life, and in general it is impossible to predict exactly when any given particle will decay. However, it is possible to determine the *mean* lifetime of a number of particles to any desired accuracy simply by observing a sufficient number of such particles, and in this sense, decaying particles are clocks that are just as good as vibrating molecules. Indeed, for a vibrating molecule it is necessary to observe it for a large number of cycles in order to determine its frequency precisely; this is analogous to observing a large number of decays in an exponentially decaying system.

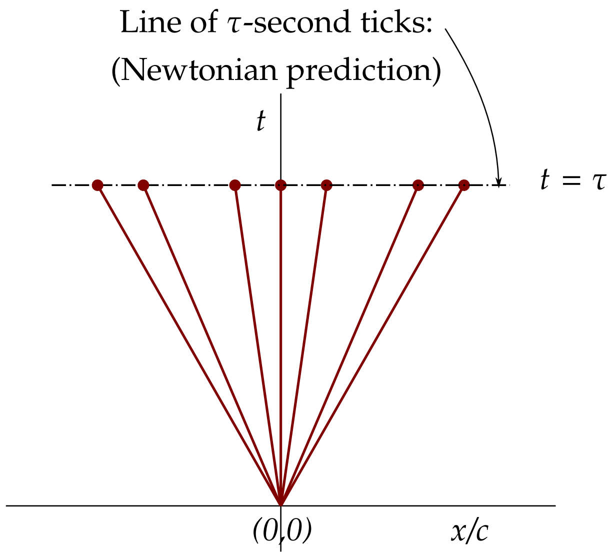


Figure 7: Newtonian Spacetime.

The surprising result is that this prediction is *wrong*. It is approximately true *only* for clocks that move at speeds much less than the light velocity. When we examine the behavior of real clocks moving at speeds comparable with the light velocity, we find that the line of "τ-second ticks" is not a straight line, but is curved. The form of the curve, while originally predicted by Einstein, is also based on experiments using high speed mesons, experiments we discuss in the next section. The result is that the curve is described by the hyperbola:

$$t^2 - x^2/c^2 = \tau^2$$

This curve is shown on a spacetime diagram in Fig. 8, where we have added dashed world lines for light pulses going left and right from the origin.

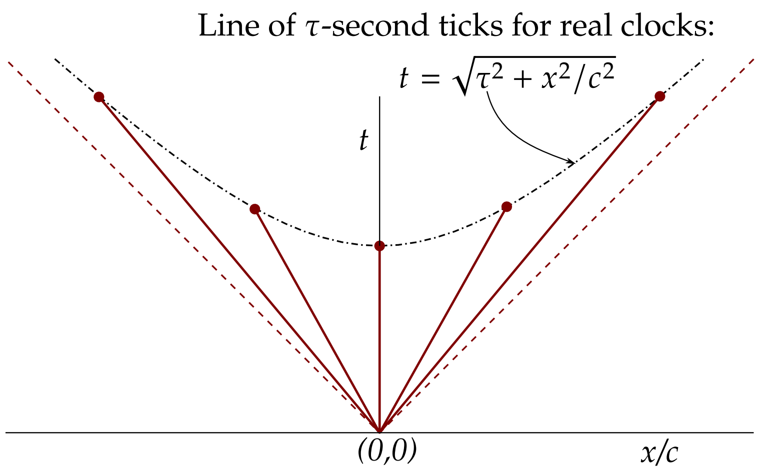


Figure 8: Real Spacetime.

Note that for $x/c \ll t$, that is, for $v \ll c$, we may neglect the term x^2/c^2 in the equation for the hyperbola, and $t^2 \approx \tau^2$, or $t \approx \tau$. This is the region of spacetime where the Newtonian concept of absolute time is approximately correct.

Exercise

The Newtonian prediction for clock rates is sufficiently accurate if the clock velocity is small enough. What is the maximum clock velocity allowed such that clock rates be accurately predicted by Newtonian theory to within 1 per cent? To within 0.01 percent? (Hint: Make use of the binomial theorem. Answer: $0.14c$, $0.014c$).

The clocks used in the experiments leading to Fig. 8 are typically mu mesons (muons), for which $\tau = 2.2$ microseconds. The experiment may also be done with pi mesons (pions), for which $\tau = 0.027$ microseconds. In both cases, the locus of decay events is given by the hyperbola. The hyperbola thus represents a description of the experimental data for real clocks, and is not subject to dispute. On the basis of these experimental data (see the following section for details), we are led to believe that *any* clock moving from the origin of our spacetime diagram to an event with spacetime coordinates $(x/c, t)$ will record a time interval τ where

$$\tau^2 = t^2 - x^2/c^2$$

One must distinguish between t and τ. Here t is called the *coordinate* time, or the time recorded by a clock at a fixed value of x, say at $x = 0$. τ, on the other hand, is the elapsed time recorded by an observer traveling with the clock. It is called the *proper* time. The proper time is characteristic of the clock, and does not depend on how the clock moves.

Note that the parameter c, the light velocity, enters into the equation for the hyperbola. This is an indication that the light velocity is a fundamental quantity that must enter into the theory of high-velocity motion of particles. This also provides us with the motivation for scaling our axes as we have discussed earlier in connection with Fig. 3.

In actual experiments, not all clocks will pass through the same event. Our assumptions that all clocks tick uniformly and homogeneously will allow us to compare clocks ticking at different times and at different places.

We may generalize the above relation: If a clock is carried at constant velocity from the event $(x_1/c, t_1)$ to the event $(x_2/c, t_2)$, it will record a time interval given by $\Delta\tau$, where

$$(\Delta\tau)^2 = (t_2 - t_1)^2 - \frac{1}{c^2}(x_2 - x_1)^2 \equiv (\Delta t)^2 - \frac{1}{c^2}(\Delta x)^2$$

This forms the basic rule for thinking about spacetime geometry, and forms a concrete basis for SRT. It is an experimental basis, a way of summarizing the description of real observations. The quantity $\Delta\tau$ has a name. It is called the *spacetime interval*, separating the spacetime events $(x_1/c, t_1)$ and $(x_2/c, t_2)$. The quantitiy $(\Delta\tau)^2$ also has a name. It is called the *metric*, a quantity that becomes particularly useful when one considers General Relativity Theory.[6]

This expression for the spacetime interval looks a lot like the Pythagorean theorem, *except for the minus sign*. That minus sign is *crucial*. It is the single feature leading to the surprising differences between spacetime and ordinary Euclidean geometry.

[6] Note that $(\Delta\tau)^2$ can be positive, negative or zero, depending on the relative values of the spacetime coordinates of the two events. The ramifications of this will become clear in subsequent sections. By $\Delta\tau$ we mean $(\pm(\Delta\tau)^2)^{1/2}$, where the sign is chosen to make $\Delta\tau$ real.

Example

Consider a clock that is being carried from the origin at E_1, where $t = 0$ and $\tau = 0$, to a distant point E_2 (at $x/c = 6$ seconds) where the clock's carrier notes a clock reading $\tau = 8$ seconds. What will be the time t, recorded by a stationary observer remaining at the origin?[7] The answer will be given by

$$t^2 = \tau^2 + \frac{x^2}{c^2} = 8^2 + 6^2 = 100, \qquad \text{so } t = 10 \text{ seconds!}$$

Note the obvious: 10 seconds is *more* than 8 seconds. We can illustrate with a spacetime diagram:

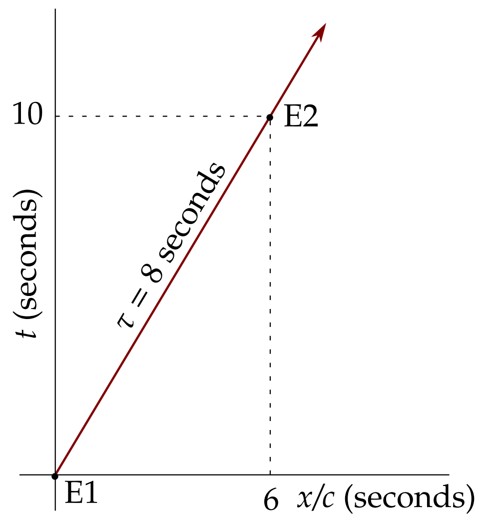


Figure 9: The time on a clock at E_2.

If our clock were carried faster from the origin so as to reach a distance of $x/c = 15$ seconds, still by the time $\tau = 8$ seconds of the clock as observed by the clock's carrier, the time t measured by the stationary observer would be 17 seconds, over twice the reading of the clock.

This is strange stuff indeed. It can be made a bit less strange, however, if we realize two things: First, note that the distances of 6 and 15 light-seconds are extremely large—several times as far away as the moon—so that for a clock to travel that far in only a few seconds it must travel at a very high velocity, much higher than we are accustomed to, even with the highest speed spaceship we could make. (Note that for the case illustrated in Fig. 9, our clock must move at $v = 0.6c$ in order to pass through the two events). So perhaps we shouldn't be too astonished if we encounter strange results.

Second, although we represent our spacetime by a drawing on a piece of paper, so that it *looks* like ordinary Euclidean space, it is *not* a representation of Euclidean space, and the rule for measuring distance in Euclidean space (the Pythagorean theorem) does not apply. We must use the spacetime interval rule instead.

[7] In Section 5 we shall see how this stationary observer can measure t using light signals.

The behavior of moving clocks is called the relativistic *time dilation* for the following reason: Consider Fig. 10 (a generalization of Fig. 9), in which a clock moves with constant velocity from the event $(0, 0)$ to the event $(x/c, t)$:

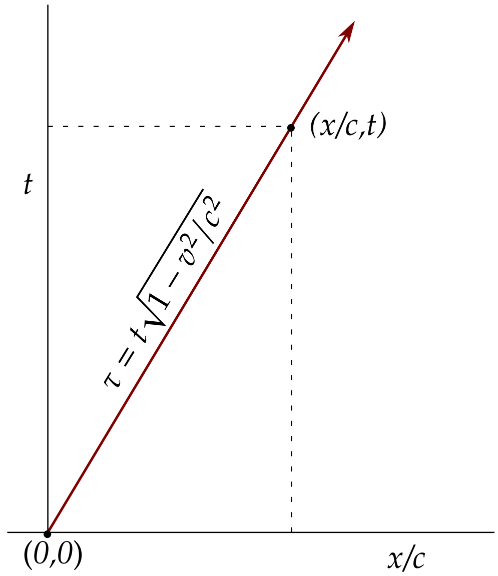


Figure 10: The time on a clock traveling at speed v.

What will this moving clock read when it passes the second event? That is, what will be the elapsed time as recorded by this clock? Since the clock moves at speed v, $x = vt$, and we may calculate:

$$\tau^2 = t^2 - x^2/c^2 = t^2 - (v^2/c^2)t^2 = t^2(1 - v^2/c^2), \quad \text{or} \quad \tau = t\sqrt{1 - v^2/c^2}$$

The coordinate time t, which may be measured by a stationary observer, is therefore related to the proper time interval τ recorded by the moving clock:

$$t = \frac{\tau}{\sqrt{1 - v^2/c^2}}$$

The ratio of the clock readings depends only on the relative velocity v. This particular function of v appears so often that it is represented a special symbol, γ:

$$\gamma \equiv \frac{1}{\sqrt{1 - v^2/c^2}}$$

Since v/c is always observed to be less than 1, we see that γ is always greater than 1. Hence, as measured by the stationary observer's clock, the time intervals of the moving clock are lengthened, or *dilated*, by the quantity γ: $t = \gamma\tau$. This is the meaning of the expression *time dilation*.

An additional implication resulting from this behavior of clocks is that moving unstable particles will travel farther than we might think before decaying. Newtonian theory would predict that a moving particle whose lifetime was τ would move a distance of $v\tau$ before decaying. The SRT rule for clocks says that the moving clock appears to run

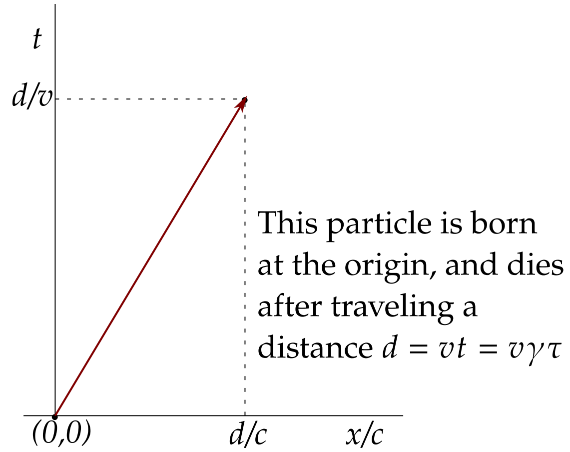


Figure 11: The distance traveled by a clock.

slowly, and so the particle can actually cover a distance $d = vt = v\gamma\tau$ before decaying. This effect may also be illustrated with a spacetime diagram as shown in Fig. 11:

Again, suppose that the particle decays after a time interval τ during which time it travels a distance d. We have

$$\tau^2 = \frac{d^2}{v^2} - \frac{d^2}{c^2} = \frac{d^2}{v^2}\left(1 - \frac{v^2}{c^2}\right), \quad \text{or} \quad d = v\gamma\tau$$

This result plays an important role in the experiments on real clocks, discussed in the following section.

So far, we have considered only clocks moving at constant speeds relative to each other. How do we treat a clock that is moving on a *curved* world line, that is, a clock that is undergoing *acceleration*? Contrary to the opinions of some effervescent mythologists, General Relativity need not be invoked. SRT gives the complete description. We merely consider an infinitesimal interval of elapsed time along the world line of the clock:

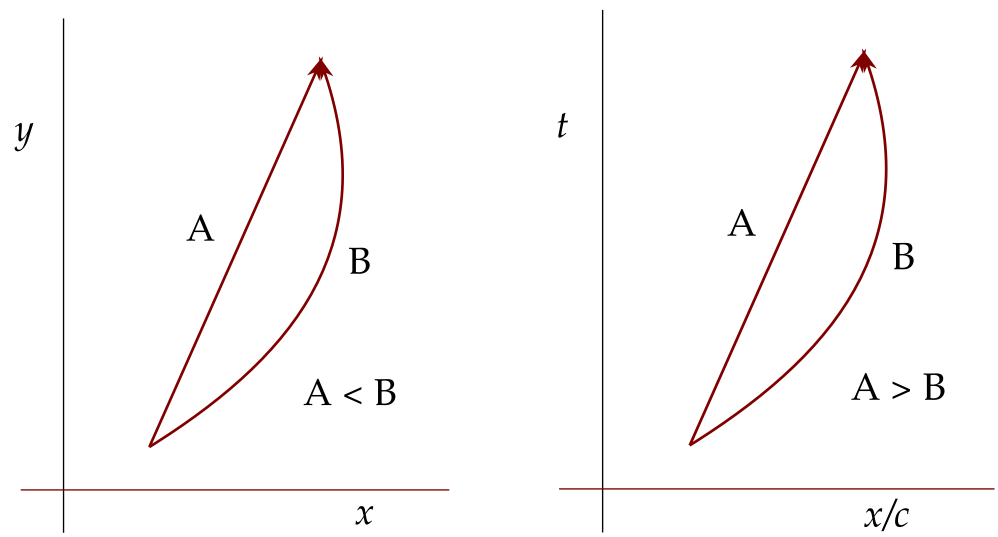
$$d\tau = (dt^2 - \frac{1}{c^2}\,dx^2)^{1/2}$$

To calculate a total elapsed time for any given clock between any two specified events, we simply add up all the infinitesimal spacetime intervals, that is, we perform a line integral along the world line:

$$\tau = \int_{E_1}^{E_2} d\tau = \int_{E_1}^{E_2} \left(dt^2 - \frac{1}{c^2}\,dx^2\right)^{1/2} = \int_{t_1}^{t_2} \left[1 - \frac{1}{c^2}\left(\frac{dx}{dt}\right)^2\right]^{1/2} dt$$

Now, to perform this integral, we need to know the equation of the world line, that is, we must know how x depends on t, so that we can determine how dx/dt depends on t and do the integral. The result we get for τ depends not only upon the spacetime coordinates of E_1 and E_2, but also upon the equation of the world line connecting E_1 with E_2.

Now for a surprise: For spacetime geometry, the *greatest* elapsed time recorded by a clock carried between any two events occurs when the world line connecting the two events is *straight*. All other world lines yield *shorter* elapsed times. Contrast this with Euclidean geometry, where the *shortest* distance between two points is along the straight line connecting the points. We compare the two geometries in Fig. 12. Note the differences in the coordinate labels.



(a) Euclidian Cartesian Geometry

(b) Hyperbolic Spacetime Geometry

Figure 12: Two Geometries

Problem

For this exercise, assume that the light velocity is 5 miles per hour. Dave starts from home at 6 AM and walks down a long straight road at 1 mile per hour. His friend Erin starts (from the same home) 9 hours later (at 3 PM) and follows Dave, walking at 2 miles per hour.

Draw their world lines (to scale) on a suitable spacetime diagram, and determine graphically the coordinates of the event E: Dave and Erin meet. Their dog Fido leaves home just when Erin does, pursuing Dave at 4 miles per hour, meets Dave, reverses direction and returns to Erin (also at 4 miles per hour), reverses to Dave, etc., until the event E. How far does Fido walk? Now Dave, Erin and Fido each carry ordinary clocks, all of which have been synchronized at 6 AM, the moment when Dave leaves. What are the readings of each of the 3 clocks at the event E, when they are all back together again? (Partial answer: Erin's clock reads 11:15 PM)

III. An Experimental Basis for Clock Behavior

The first experiment to verify the behavior of clocks moving at high speed is described by B. Rossi and D. B. Hall in *Physical Review*, **59**, 223 (1941). A more recent experiment of the same type is described in lucid detail by D. H. Frisch and J. H. Smith in the *American Journal of Physics*, **31**, 342 (1963). Both of these experiments deal with the behavior of the mu-mesons (muons) that are produced near the top of the earth's atmosphere by the interaction of cosmic ray protons with the gas molecules in the atmosphere. Some of these muons then travel toward the earth's surface where they may be detected using appropriate detecting and counting apparatus.

Now the mean lifetime of a muon at rest has been measured to be 2.195 microseconds, or about 2.2 microseconds. Consider the plight of a muon produced at the top of the atmosphere such that it heads straight down toward the Earth's surface. According to pre-relativistic notions, it would travel for a distance equal to the product of its speed times its lifetime before decaying. Thus, even if it traveled at the speed of light, it would travel for an average distance

$$x = (2.2 \times 10^{-6} \text{ sec})(3 \times 10^{8} \text{ m/sec}) = 660 \text{ meters}$$

Since all the muons are produced at the top of the atmosphere, roughly 60,000 meters up, we would expect, if the muons behaved like Newtonian clocks, to see almost no flux of muons at the earth's surface, since they all would have decayed long before they reached the earth. More accurately, we would expect less than 1 in 10^{40} of the mesons produced to reach the earth.

In fact, the observed flux of muons at the earth's surface is quite large, and one must conclude that either they travel much faster than the light velocity on their way down, or else that they live for much longer than 2.2 microseconds. The experiments described in the above papers, in which the muon flux is measured at two different elevations and compared, determine that the moving muons must have a longer mean life, as measured by an observer on earth. For example, Frisch and Smith, in one of their runs on muons moving at $.995c$, measured the muon flux at the top of Mt. Washington in New Hampshire to be 563 counts/hour through their detector, while the flux at Cambridge, Massachusetts, 1907 meters lower, was 408 counts/hour. From these numbers they determined that the muons they observed must live for an average of about 19 microseconds (as measured by Frisch and Smith on earth), or about 9 times longer than the mean life for muons at rest.

Over the past several years, a variety of additional experiments have been performed on decaying particles, all of which confirm the relativistic behavior of fast-moving clocks. In particular, the lifetimes of pi-mesons (26.6 nanoseconds), and K-mesons (12.4 nanoseconds), have both been observed to lengthen appropriately at increased velocities.

A particularly elegant experiment on muons is described in a paper by J. Bailey, *et al.*, in *Nature*, **268** (1977). These experimenters looked at muons, both positively and negatively charged, in the CERN Muon Storage Ring, moving at $.9994c$. The observed lifetime for these muons was about 64.4 microseconds, or about 29.3 times the rest lifetime. Furthermore, since the muons were traveling in a circular orbit, they experienced large transverse accelerations (of the order of 10^{18} g). This acceleration appeared to have no observable effect on the lifetime. Their results were in complete agreement with SRT, making it clear that such observations have nothing to do with General Relativity.

We can summarize the results of all of these experiments by stating that the hyperbolic form of the locus of "τ-second ticks" on a spacetime diagram has been well-verified, at least for τ lying between 10^{-5} and 10^{-9} seconds. To date, there has been no reason to believe that the locus of "τ-second ticks" deviates from the hyperbolic form for values of τ outside this range.

It is this behavior of real clocks that we use as one of the starting points for our development of SRT.

IV. The Twin Paradox

The Twin Paradox, or Clock Paradox as it is sometimes called, is usually stated in terms of two twins, say Arlo and Bob. Each has a watch, which they synchronize as they stand beside each other on Earth. They are, let us say, 20 years old. Arlo then climbs into a rocket ship and zips off into outer space at high velocity, while Bob stays behind on Earth. The years go by until one day Arlo returns to reunite with his twin brother.

Now from Bob's point of view, it is Arlo's watch that has been moving, and therefore running slowly. Bob is now a man of 50, with gray hair beginning to show at the temples, and having learned a smattering of SRT as a student, he expects to see his twin brother to have aged very little, still with dark hair and lots of vim and energy, alighting from his rocket ship and rushing over to greet him with open arms, joyfully exulting at finally returning home.

Arlo, on the other hand, who also recalls delving into a description of SRT as a student, views the scene from his rocket ship and reasons that from his point of view it is his brother Bob who zipped away and returned. Arlo also knows about time dilation, and reasons that his brother's watch will have run more slowly than his. Arlo thus expects, upon reuniting with Bob, to find a vigorous young man much younger than himself, now perhaps just beginning his career and looking forward to a long and productive life.

Hence the paradox: Arlo's line of reasoning leads to the conclusion that Arlo is older than Bob, while Bob's line of reasoning leads to the opposite conclusion that Bob is older than Arlo. Which is correct? It is natural to wish to compromise, and by invoking common sense deduce that somehow neither one will have aged more than the other. After all, aren't they still twins?

By now, we should have no trouble answering this question and hence resolving the paradox. To gain clarity of thought, we may find it helpful to consider an analogous situation, relating to paths taken in Euclidean space: Suppose that Arlo flies on an airplane from San Francisco to Los Angeles by way of Hawaii, while Bob, who is of more modest means, takes the direct flight to Los Angeles. The question is: Who travels the longer distance?

We may suppose, for the benefit of this example, that each takes the same amount of time to fly from San Francisco to Los Angeles so that as they leave (simultaneously) from San Francisco, each can keep the other's plane in view. Now Bob will look out his window at Arlo, and since he sees Arlo's plane at first going away and then coming back, he will conclude that Arlo has traveled the longer distance. Arlo, on the other hand, looking out his window at Bob, will see Bob's plane first going away and then coming back, and so naturally will conclude that Bob has traveled the longer distance. Hence here too there is an apparent paradox. Of course, if each one carries a device to record the number of air miles traveled, they can compare their findings when they meet again in Los Angeles and of course (of course?) the result will be that Arlo will have recorded the longer distance. Think: What qualitative criterion allows you (or Arlo or Bob) to determine immediately which path is longer?

It is a curious thing that the twin paradox has received far more attention than any other aspect of SRT. Well over 50 papers have been published on the subject, all in reputable journals, and countless chapters have appeared in books on relativity. To cite one example: Out of the American Institute of Physics collection of 16 reprints of journal articles on SRT (mentioned in the introduction), 10 are devoted specifically to the twin paradox, leaving only 6 for everything else. No doubt the notion that we might fly away in a spaceship and return to find ourselves younger than our children is a conclusion most of us find pretty hard to accept, a true space child's fantasy.

V. Spacetime Coordinates, Simultaneity and Inertial Reference Frames

In our earlier discussion of clock rates, we imagined a situation in which a moving clock travels between two stationary clocks, and noted that the moving clock lags behind the stationary clocks by an amount depending upon the velocity of the moving clock—the *time dilation* effect (see Fig. 10). There we stated that we can use the pair of synchronized stationary clocks to measure the spacetime coordinates of the event at which the moving clock passes the second stationary clock. How do we synchronize the stationary clocks? If we put them side by side, set them to the same time and then move one of them to a new spatial position, they will no longer be synchronized because the one we move will lag behind. Perhaps we could move both clocks symmetrically, one to the right and one to the left, but this will lead to difficulties when we try to extend the process. What we need is a clean unambiguous procedure for synchronizing a number of clocks at different spatial positions, a procedure that doesn't involve moving clocks around. We describe such a procedure now, or more generally, a procedure for measuring the spacetime coordinates of a distant event. In the process we shall be led to a unique definition of ısimultaneity—what we mean when we say that two spatially separated events are simultaneous.

Our procedure involves the use of only a single clock and some *light signals*, along with a means of sending and receiving such light signals. By a light signal we mean a little pulse of light, such as might be generated by blinking a flashlight once. So far we have discussed the role of light only peripherally, pointing out only that it plays an essential role in linking space and time. Now we shall begin to examine this role more carefully.

Suppose we send out a light signal to a distant mirror, where it is reflected and returned to us, and using our clock, we record the clock reading when the signal is sent (t_1) and again when it is received (t_2). The reflection of the light signal from the mirror constitutes an event, whose spacetime coordinates we wish to determine. Figure 13 shows an appropriate spacetime diagram:

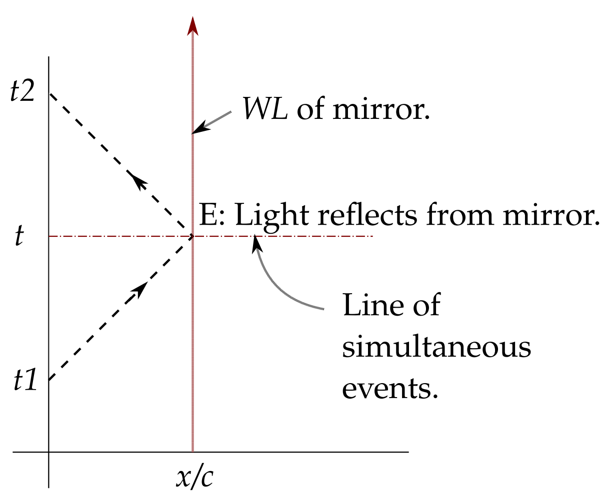


Figure 13: The Spacetime Coordinates of an Event E.

What does our clock read when the light signal strikes the mirror? There is only one simple answer: Half way between t_1 and t_2. That is, the time coordinate of the event E is given by

$$t = \frac{1}{2}(t_1 + t_2)$$

Another way of stating this is to say that the reflection of the light signal from the mirror and the reading of time t on our clock are ısimultaneous events. We have drawn a line on our spacetime diagram through these two events; clearly all the points on this line are simultaneous with the clock reading of t. We therefore call such a line a ıline of simultaneous events.

How far away is the mirror when the light signal is reflected from it? In other words, what is the space coordinate of the event E? Clearly, since the light signal goes out and back at velocity c, the distance to the mirror is given by

$$\frac{x}{c} = \frac{1}{2}(t_2 - t_1)$$

Thus we have outlined a procedure for determining the spacetime coordinates of the event E. Note that the line of simultaneous events is parallel to the space axis.

One of the advantages of using clocks and light signals to determine spacetime coordinates is that they can be just as easily understood by a moving observer as by one who is stationary. What if the spacetime coordinates of the event we have just considered (the reflection of a light signal from a distant mirror) are determined by an observer moving along with velocity v? Figure 14 shows a spacetime diagram, drawn to include such a moving observer:

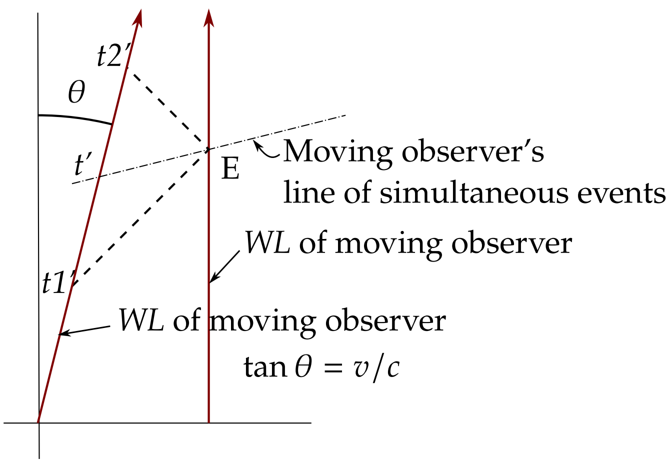


Figure 14: A Moving Observer Measures Spacetime Coordinates of an Event E.

The light signal is sent by the moving observer at time t'_1 (measured by his clock), and is received by him, after being reflected from the mirror, at time t'_2. What is the time coordinate of E as recorded by the moving observer? Here again, there is only one simple answer: Halfway between t'_1 and t'_2. That is, the moving observer's time coordinate of the event E is given by

$$t' = \frac{1}{2}(t'_1 + t'_2)$$

This is a powerful statement. It says that the reflection of the light signal from the mirror and the reading of time t' on the observer's clock are simultaneous events for the moving observer. To emphasize this, we have drawn another *line of simultaneous events* on our spacetime diagram, this time for the moving observer. It is clear from our diagram that events that are simultaneous for the moving observer are *not* simultaneous for the stationary observer. For the *stationary* observer, the event E occurs later than the reading of time t' on the moving observer's clock, while for the *moving* observer, the event E occurs earlier than the reading of t on the stationary observer's clock.

Exercise

Verify the above statement by drawing appropriate light signal world lines on Fig. 14.

Thus we have abandoned the Newtonian idea of absolute simultaneity. Clearly simultaneity is a relative concept. Whether or not two spatially separated events are simultaneous depends upon the velocity of the observer, *i.e.*, upon the direction of the observer's world line.

To gain an appreciation for this definition of simultaneity, it is helpful to dwell for a moment on the rather extraordinary but simple behavior of light signals. When we think of a little pulse of light traveling through space, we are apt to draw analogies from our past experience with moving things, such as a golf ball, a water wave, or even an elementary particle traveling along from one point to another. We often think, for example, of a light signal as a particle-like photon, which we suspect ought to behave like a moving electron. Alternatively, since we normally think of light as a wave, we might think that a light pulse behaves like a pulse of sound or a pulse of water waves, such as might be formed by dropping a pebble into a still pond.

In fact, light pulses are not completely analogous either to moving particles or to other moving wave pulses familiar to us. In the first place, light pulses are unlike particles in that their velocity is independent of the motion of the source. For example, some unstable particles emit a pulse of light as they decay; such a pulse will travel no faster if the decaying particle is moving, even if the particle moves at speeds comparable to the light velocity. In this respect, light pulses are like sound or water wave pulses: A pulse of noise given off in the forward direction by a moving (subsonic) airplane travels along through the air ahead of the plane no faster than if the plane were stationary. Similarly, a stone thrown into a quiet pond from a moving boat makes a wave pulse that travels with a velocity independent of the boat's speed.

There is, furthermore, something else. The speed of a sound pulse is measured with respect to the air through which it travels, and the speed of a water wave pulse is measured with respect to the water. Each requires the presence of a medium for its propagation. Sound waves require a gas, a liquid or a solid, and we can't have water waves without the water. Light wave pulses are different. They travel through empty space with no trouble at all. In fact, the emptier the space, the better the propagation of the light pulses. No medium is required for their propagation.

While this observation was hard to accept—a large number of attempts have been made to detect the so-called "luminiferous ether" as a supposed medium required for the propagation of light—the lack of such a medium provides a great simplification when we discuss the propagation of light signals. It implies, in particular, that since there is no medium with respect to which the speed of light is to be measured, there are only two possible directions for light signal world lines on our spacetime diagrams: Left-going and right-going, lines which for simplicity we have taken to be at ±45° relative to the spacetime axes. A light signal world line in any other direction would imply either the existence of a medium or that the light signal velocity depended upon the velocity of the source. In short, the abandonment of the notion of a "luminiferous ether", along with the observation that light travels with a speed independent of its source, leads to the description of nature that is simple, beautiful and in consonance with experiment.

For us, this constancy of the light velocity, together with the curious behavior of clocks, provide a firm basis from which all of SRT may be developed.

Now, returning to the determination of the spacetime coordinates of the event E by our moving observer, we conclude that the space coordinate of the event E must be given by

$$\frac{x'}{c} = \frac{1}{2}(t'_2 - t'_1)$$

just as would be determined for the stationary observer. Both observers use the same procedure for determining the spacetime coordinates of a distant event.

The space coordinate of E can be scaled off on an axis parallel to the moving observer's "line of simultaneous events". This is hence the x'-axis. Thus we are able to construct spacetime axes, x' and t', that are appropriate to the moving observer. Furthermore, by examining our spacetime diagram, we can show that the world line of the observer (the t'-axis) and his line of simultaneous events (the x'-axis) make equal Euclidean angles with the light signal world line. That is, we can show that the x'-axis is tilted inwards from the x-axis by the same angle (θ) that the t'-axis is tilted inwards from the t-axis and that this angle is such that $\tan\theta = v/c$.

To show this, we redraw, in Fig. 15, our spacetime diagram, taking t'_1 for convenience at the origin of our diagram.

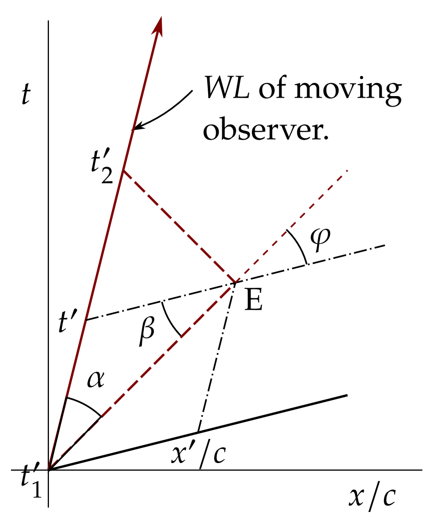


Figure 15: Determining the x'-axis.

Since t' bisects the interval $t'_1 \to t'_2$, $(t'_1 \to t') = (t' \to t'_2)$. Also $(t'_1 \to t') = (t' \to E) = x'/c$, that is, $t'_1 \to t' \to E \to t'_1$ is an isosceles triangle in the Euclidean sense, and $\alpha = \beta = \varphi$. Our assertion is proved. The angle between the t and the t' axis is equal to the angle between the x and the x' axis. Both are equal to θ, where $\tan\theta = v/c$.

The result of this bit of geometrical analysis is that we now have a quick method for drawing spacetime diagrams for moving observers as well as for stationary observers. We simply draw the spacetime axes for the moving observer at an angle $\theta = \tan^{-1}(v/c)$ from the (perpendicular) axes for the stationary observer.

Note that although we have termed one of our observers "stationary" and the other "moving", we have introduced no basis for determining which is which. We could as well have thought of the second observer as "stationary" and the first one as "moving". It is only the *relative* motion of one observer with respect to the other that matters. We shall soon dwell further on this point, when we discuss a procedure by which such relative motion can be measured. Before doing so, however, we shall introduce the notion of an *inertial reference frame*.

We are accustomed to thinking of a *reference frame* (RF) as a set of spatial coordinate axes (usually Cartesian) with respect to which we can measure the position of a particle. We frequently envision a laboratory, a moving train, an elevator, a spaceship, the earth, or the system of "fixed" stars as examples of such reference frames. Further, we term a RF *inertial* if it is not accelerated, that is, one in which Newton's First Law (the "Law of Inertia") holds: Particles upon which no forces act are observed to travel in straight lines at constant velocities.

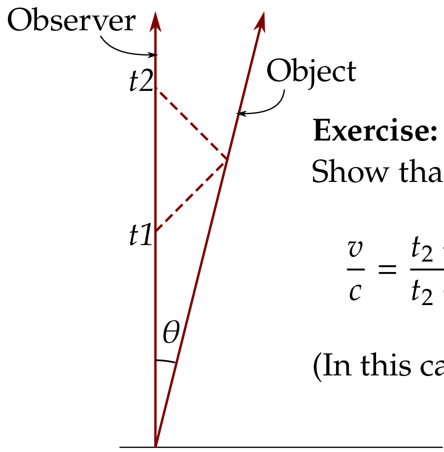
In our discussion of SRT we expand upon this concept to include the time dimension: By a reference frame, we mean a set of spacetime axes, with respect to which the coordinates of events can be measured. Thus, when we draw a set of axes on a spacetime diagram we imply the existence of a reference frame. Further, an inertial reference frame is one for which the world lines of free particles (particles upon which no forces are acting) are straight lines. In our discussions, we shall not normally consider non-inertial frames, and so will usually omit the word "inertial". We may think of a RF as a classical Newtonian frame with a clock added. Often, of course, there is an observer, which we may think of as traveling along in his own RF, equipped with a clock and a means of sending and receiving light signals—the tools for measuring the spacetime coordinates of events.

Problem

If two events occur at the same spatial position in some reference frame S, prove that their temporal sequence is the same in all reference frames, and that the least time separation is assigned to them in S. This illustrates that local *causality* is not violated by SRT. Solve this problem geometrically, using a spacetime diagram.

VI. The Velocity of a Moving Object

In the previous section we discussed our method for using clocks and light signals to determine the space and time coordinates for any event. A slight extension of this technique allows us to determine the velocity of a moving object in any reference frame. Suppose, for example, that at $t = 0$ an object passes an observer. Then at some later time the observer can bounce a light signal off the object, determine how far away it is, and hence determine its velocity. Here is a possible construction, first for a "stationary" observer (Fig. 16) and then for a "moving" observer (Fig. 17):

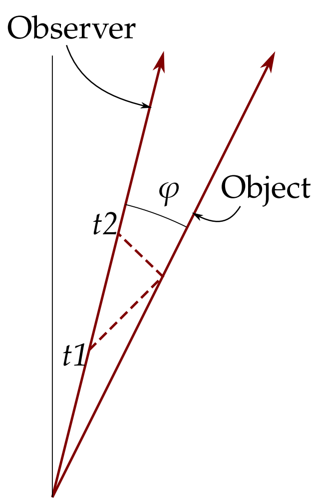


Exercise:
Show that

$$\frac{v}{c} = \frac{t_2 - t_1}{t_2 + t_1}$$

(In this case, $\tan\theta = v/c$)

Figure 16: The Velocity of an Object in a "Stationary" Frame.



Here also,

$$\frac{v}{c} = \frac{t_2 - t_1}{t_2 + t_1}$$

However, the Euclidean angle φ bears no simple relation to the velocity v of the object relative to the observer.

Figure 17: The Velocity of an Object in a "Moving" Frame.

In each case, v is the velocity of the object with respect to the observer. How would you draw the observer's space axis in Fig. 17?

Although certain quantities may be represented with greater clarity in a spacetime diagram whose space and time axes are at right angles to each other (for example in Fig. 16, $v/c = \tan\theta$), there is nothing physically special about such a spacetime diagram. Either Fig. 16 or Fig. 17 can be used to provide a complete description of an object moving at velocity v relative to an observer.

VII. The Einstein Formula for Velocity Addition

We are familiar with the old "rowboat-on-the-river" problem: Suppose a boat is rowed down a river such that its velocity with respect to the river bank is $v_1 = 8$ miles per hour, while the water in the river flows with velocity $v_2 = 3$ miles per hour, also with respect to the river bank. With what velocity v_{12} does the boat move with respect to the water? Following Newton (and our natural intuition), we simply subtract the two velocities: $v_{12} = v_1 - v_2 = 5$ miles per hour. While this calculation is sufficiently accurate for these low velocities, it becomes increasingly incorrect for velocities approaching the velocity of light. The following exercise will serve to illustrate the problem:

Exercise

Let one particle move to the right while another moves to the left, each with speed $c/2$. How fast does the first particle move with respect to the second? Newton would have said "c". Show that this prediction is wrong by constructing a spacetime diagram with the world lines of the two particles. You can do this by using light signals as in Fig. 16, choosing t_1 and t_2 such that $(t_2 - t_1)/(t_2 + t_1) = 1/2$. Do it carefully, so the slopes are accurate. Now from your diagram, estimate graphically how fast the first particle is moving with respect to the second. You should get $(4/5)c$.

This is an example of the Einstein law for velocity addition. If one particle moves with a velocity v_1, while another moves with velocity v_2, the first particle moves with respect to the second with a speed v_{12}, where

$$v_{12} = \frac{v_1 - v_2}{1 - v_1 v_2/c^2}$$

At low velocities, the extra term in the denominator may be neglected, and we get the Newtonian formula:

$$v_{12} \approx v_1 - v_2$$

Exercise

Sometimes the Einstein formula is stated in a slightly different form: Suppose a particle moves with velocity v_{12} with respect to a reference frame (a railroad car, say) that itself is moving with velocity v_2. The velocity of the particle with respect to a stationary frame (the ground) is then

$$v_1 = \frac{v_2 + v_{12}}{1 + v_2 v_{12}/c^2}$$

Show that this formula is equivalent to the preceding one.

Another exercise

We can use the form of the Einstein formula stated in the previous exercise to show that no particle can move faster than the light velocity. Let $v_2 = v_{12} = \alpha c$, where α is some constant number, possibly greater than 1. Show that the maximum value for v_1 is obtained for $\alpha = 1$.

It is not too difficult to derive the Einstein formula. Here is a problem that will illustrate how it can be done:

Problem

Derive the Einstein formula using our spacetime diagram method for determining particle velocities, as discussed in the previous section. Start by sketching a spacetime diagram on which are drawn the world lines of two particles, one moving at speed v_1, and the other at speed v_2, where $v_2 < v_1$. Our job is to determine v_{12}, the velocity of the first particle with respect to the second, in terms of v_1 and v_2. Each velocity can be expressed in terms of clock readings using appropriate light signals. Invocation of the spacetime interval rule will then lead, with some algebra, to the desired result.

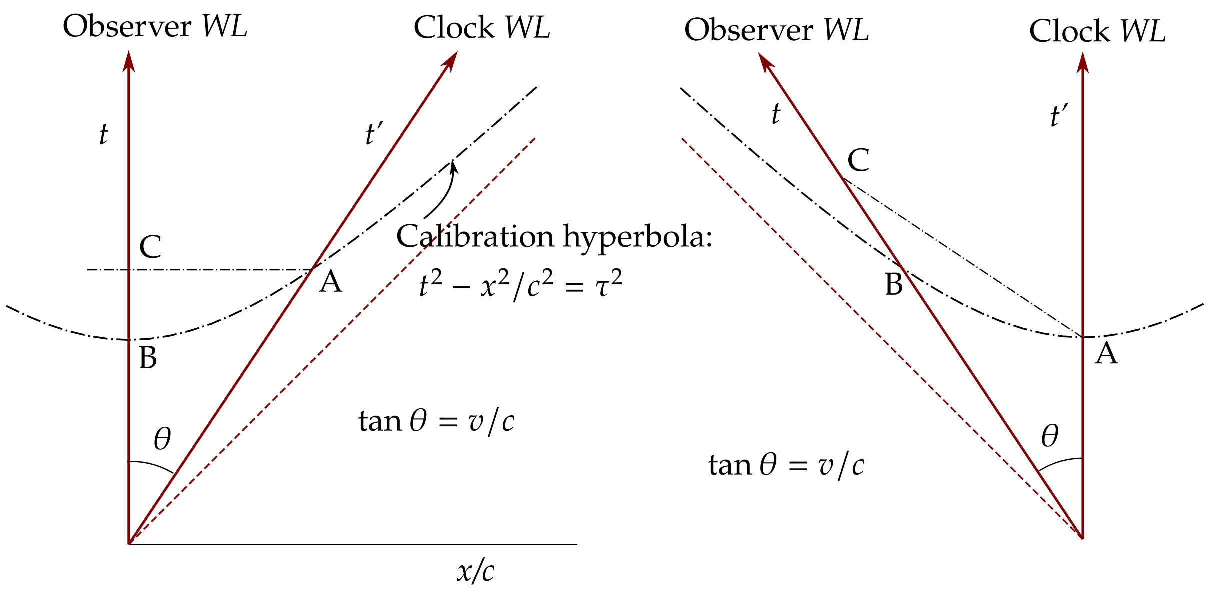
We shall discuss the Einstein formula further in Section IX, in the context of the Lorentz Transformation equations.

VIII. Moving Sticks and the Lorentz Contraction

In our discussion of clocks, we saw that a clock whose period is τ in its own frame is observed to have a period

$$T = \gamma\tau = \frac{\tau}{\sqrt{1 - v^2/c^2}}$$

when it is moving with respect to an observer at a uniform velocity v. Although we most easily imagine the observer to be stationary and the clock to be moving past with velocity v, we could equally well take the clock to be stationary and the observer to be moving past with velocity $-v$. It makes no difference; all that matters is that the *relative* velocity between the two is v. Figure 18 shows spacetime diagrams for the two cases.



(a) With a "Stationary" Observer (b) With a "Moving" Observer

Figure 18: Time Dilation

Four exercises

(a) The letters "A", "B" and "C" label events in each diagram of Fig. 18. Describe in words what each of these events is. Why have we labeled them?

(b) Where should the observer's space axis appear in the right diagram of Fig. 18?

(c) Draw a spacetime diagram showing a clock moving to the right with velocity u and an observer moving to the left with velocity u and show how the observer measures the period of the clock to be $T = \gamma\tau$, where

$$\gamma = \frac{1 + u^2/c^2}{1 - u^2/c^2}$$

(Use the Einstein formula to get γ.)

(d) By interchanging clock and observer, demonstrate the symmetry of time dilation—that each observer measures the other's clock to be running slowly.

A similar phenomenon occurs for sticks. Sticks are, however, more complex than clocks. A clock requires but a single world line to represent its spacetime path, whereas a stick, because it consists of a number of particles, is correctly represented by a family of world lines, one for each particle of the stick. Of this family, it is sufficient to draw two, one for the left end of the stick, and one for the right:

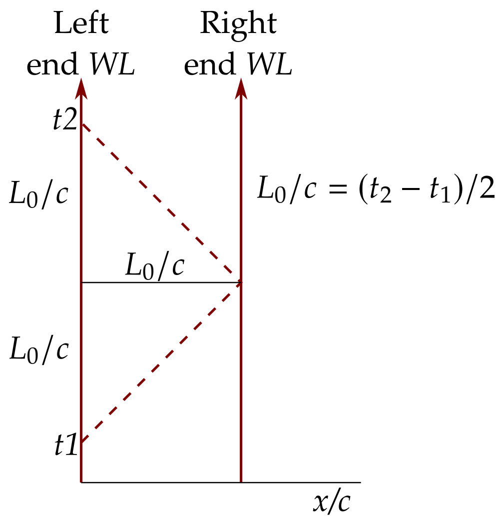
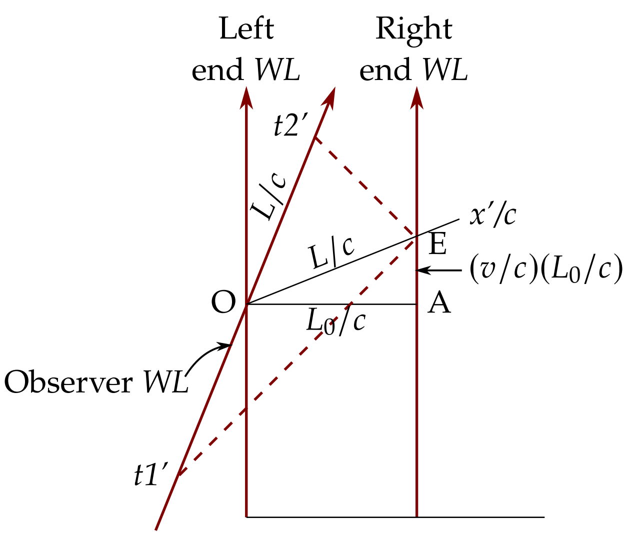


Figure 19: Measuring the Length of a Stick.

An observer sitting at one end of the stick can measure its length by bouncing a light signal from a mirror at the other end of the stick: $L_0/c = (t_2 - t_1)/2$. With this technique he can measure the spatial positions of the two ends of the stick *simultaneously* in his reference frame. L_0/c is called the *proper length* of the stick, just as τ is called the *proper time* recorded by a clock.

If an observer happens to be moving past a stationary stick, he can also measure its length by emitting a light signal so that it will bounce off a mirror at the right end of the stick just as he passes the left end. Again the two events (E and O) must be *simultaneous* in the observer's Reference Frame. Then $L/c = (t_2' - t_1')/2$:



For the triangle OEA:

$$\frac{L_0}{c^2} - \frac{v^2}{c^2}\frac{L_0}{c^2} = \frac{L^2}{c^2}$$

$$\text{or } L = L_0\sqrt{1 - v^2/c^2} = \frac{L_0}{\gamma}$$

Figure 20: A Moving Observer Measures a Stationary Stick.

Since γ is always greater than 1, L must always be less than L_0. This phenomenon is known as the *Lorentz Contraction*, after Hendrik A. Lorentz, who first proposed the contraction as an explanation for the lack of an observed fringe shift in the Michelson-Morley experiment. (Kennedy and Thorndike, in their later experiment with unequal arms in the interferometer, showed that Lorentz's "explanation" must be incorrect.)

Of course, the effect is symmetrical: A stationary observer will conclude that a moving stick is also shortened by the factor $1/\gamma$. All that matters is that the observer and the stick are moving with speed v relative to each other. Figure 21 shows a spacetime diagram representing a stick which is moving past a stationary observer:

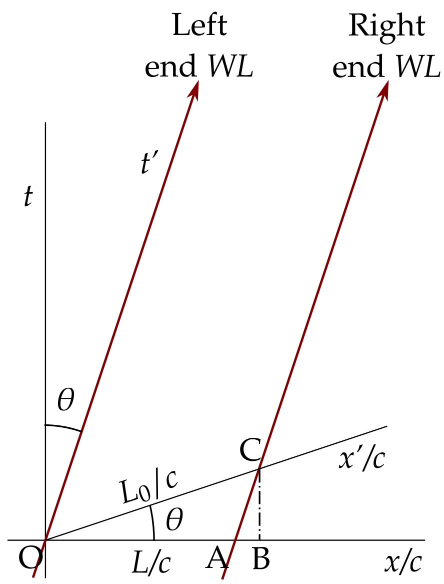


Figure 21: A Stationary Observer Measures a Moving Stick.

In this figure, the observer, who has spacetime coordinates x/c and t, sees the left end of the stick passing by at O (the origin), where both x/c and t are zero. Simultaneous with that event is the event at A, also at time $t = 0$, so he measures the length of the stick to be L/c. He could do this by bouncing a light pulse off a mirror attached to the right end of the stick.

An observer (a smart little bug?) riding the stick at its left end has spacetime coordinates x'/c and t', and can measure (the *proper*) length of the stick as L_0/c—the spacetime interval O C. (It may, if it wishes, bounce a light signal from the stick's right-end mirror. O and C will be simultaneous in the bug's reference frame.)

Exercise:

By using the spacetime interval rule and the fact that $\tan\theta = v/c$, show that $L = L_0/\gamma$ for this case also. (Hint: make use of the little triangle A B C in your geometrical analysis).

It is often said, somewhat carelessly, that “moving clocks run slowly” and that “moving meter sticks are shortened”. It is more accurate to say that moving clocks are *measured* to run more slowly, and that moving meter-sticks are *measured* to be shorter, where in each instance the measurement is made by a stationary observer, or by an observer moving with a different velocity.

The effects are similar to the distortions occurring with perspective: If we were to measure the heights of objects by holding a vertical measuring rod at arm’s length, the measured height of an object would depend upon the observer’s distance from the object. A tall tree will seem only an inch high by this method if we are far enough away from it. The "proper" height of the tree can be determined either by moving to the tree and putting the measuring rod right next to it, (we will need a long measuring rod) or else by making a triangulation calculation, as a surveyor might do.

In a similar fashion we can determine the proper time interval for a clock or the proper length for a stick, either by moving into the same RF as the clock or stick, or else by doing a calculation using SRT. The main difference for us is that perspective is an everyday experience, while clocks and sticks moving at high velocities are not. SRT would be part of our common sense and everyday experience if the velocity of light were 10 miles per hour instead of 186,000 miles per second.

Having established the principle that clocks and light signals may be used to measure lengths, it becomes possible to simplify the representation of lengths on a spacetime diagram. Suppose we imagine a number of observers moving with different velocities. For each observer we may draw a “line of simultaneous events”, or space axis. If a length L_0 is scaled off on each of these space axes (we can suppose that each observer carries a stick of proper length L_0), we get a hyperbola of “L_0-meter lengths”:

$$\frac{x^2}{c^2} - t^2 = \frac{L_0^2}{c^2}$$

This hyperbola, which is analogous to the hyperbola of “τ-second ticks” discussed earlier, can be used as a calibration curve to measure lengths in any given reference frame.

Figure 22 on the next page illustrates a situation in which five identical sticks (each has a proper length L_0) travel at different velocities past a stationary observer. The diagram can be a little confusing. Although the stick lengths appear to *increase* the faster they travel, they actually are measured by the observer to *decrease*! Hey, spacetime is hyperbolic.

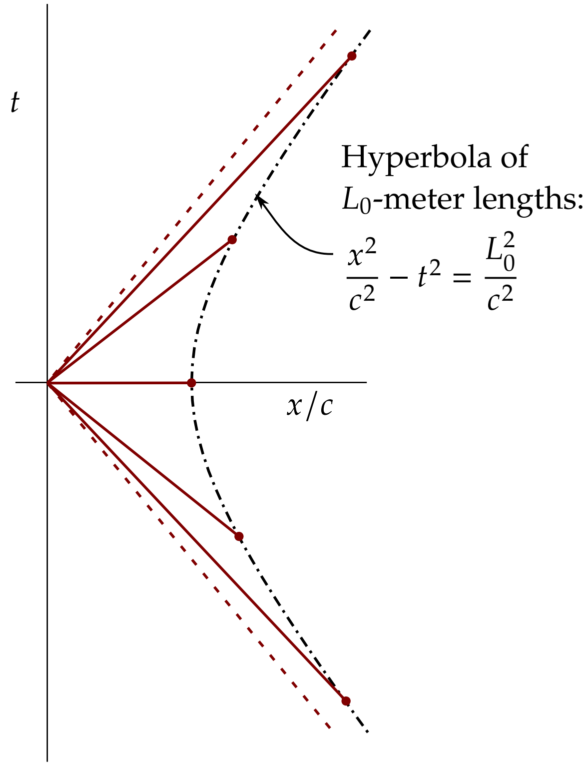


Figure 22: A calibration hyperbola for sticks.

Exercise:

Use a spacetime diagram to represent a stationary stick of length L_0, an observer moving with velocity $v \approx 0.58c$, ($\theta \approx 30$ degrees) and the stick length L as measured by the observer, thus reproducing the essentials of Fig. 20. Now add the hyperbola of "L_0-meter lengths", and thus show that L must be less than L_0. For this velocity, $\gamma \approx 1.22$. Is L less than L_0 by the correct amount?

Another Exercise:

Draw a spacetime diagram representing two sticks of proper length L_0, one going at $c/\sqrt{3}$ to the right and the other at $c/\sqrt{3}$ to the left. An observer moves along with each rod.

(a) What is the speed of the first observer with respect to the second?

(b) Show that each observer measures the other's rod to be only $L_0/2$ in length.

IX. Lorentz Transformations

Now we are ready to ask the general question: How are the spacetime coordinates of an event E in a moving RF related to the spacetime coordinates of the same event in a stationary RF? In other words, how are x' and t' related to x and t? The answer to this question can be formulated as a set of equations, known as the *Lorentz transformation equations*.

Before we answer the question, we are going to issue a warning: From now on, we are going to set $c = 1$. This means that velocities such as v will always be measured as a fraction of the light velocity, and all distances such as x, y, or z will be measured in seconds. While this may take some getting used to, it will result in greater elegance, less ink on the paper, and possibly greater insight. In fact, as we shall see in our later discussions of various applications of SRT, units with $c = 1$ are the natural units for us to use. If we find it troublesome, we can always put the c's back in at the end, simply by dividing every length and every velocity by c.

Now back to our question. We can work out the answer fairly simply, in geometrical fashion, using only clocks and light signals. The diagram drawn in Fig. 23 illustrates the process. In this diagram, the red dashed lines are light signals sent and received by each observer.

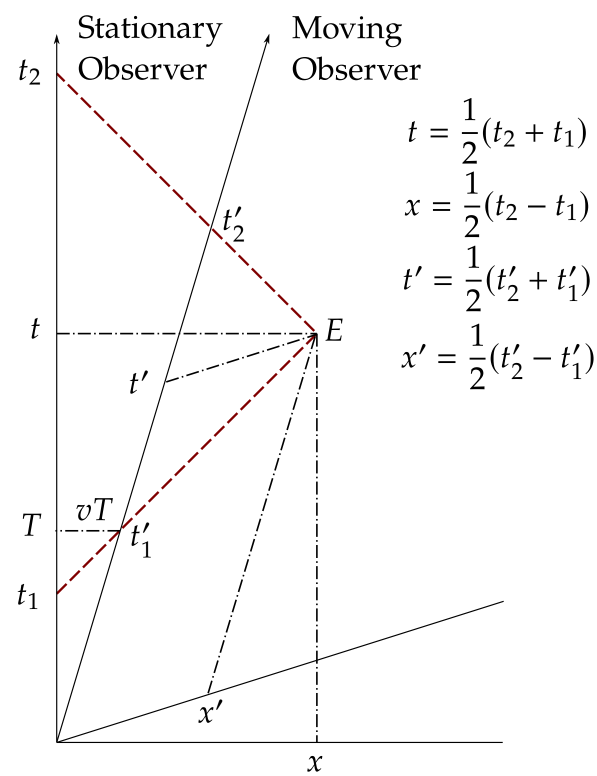


Figure 23: An event E measured by two observers.

As can be seen from the figure, the spacetime coordinates of E can be expressed simply in terms of each observer's clock readings. All we have to do is to relate the clock readings of the moving observer to those of the stationary observer. The relationship will involve the relative velocity v between the observers, that is, between their reference frames. The necessary algebra is worked out on the next page.

First we note that

$$vT = T - t_1\,, \quad \text{or} \quad T = \frac{t_1}{1-v}$$

Now we make use of the spacetime interval rule:

$$t_1'^2 = T^2 - v^2T^2 = (1-v^2)T^2,$$

and substituting from the first equation for T, we obtain

$$t_1'^2 = (1-v^2)\frac{t_1^2}{(1-v)^2} = \frac{1+v}{1-v}t_1^2, \quad \text{or} \quad t_1' = \left(\frac{1+v}{1-v}\right)^{1/2} t_1$$

In similar fashion, we obtain

$$t_2' = \left(\frac{1-v}{1+v}\right)^{1/2}$$

Finally, we can substitute these values into the expressions for x' and t' to obtain

$$x' = \gamma(x - vt) \quad \text{and} \quad t' = \gamma(t - vx), \quad \text{where} \quad \gamma \equiv \frac{1}{\sqrt{1-v^2}}$$

For completeness' sake, we may include the other two dimensions, which are unaffected if v is parallel to x and x':

$$\begin{aligned} x' &= \gamma(x - vt) \\ y' &= y \\ z' &= z \\ t' &= \gamma(-vx + t) \end{aligned}$$

These are called the *Lorentz transformation equations*. In keeping with tradition, we have put the equation for t' last. It is in this sense that time is often referred to as the "fourth dimension". These equations allow us to determine completely the spacetime coordinates of an event in one RF in terms of the spacetime coordinates of the same event in another RF which is moving with respect to the first RF with relative velocity v.

These equations may be inverted, that is, they may be solved to give x, y, z, and t in terms of x', y', z', and t'. The result is

$$\begin{aligned} x &= \gamma(x' + vt') \\ y &= y' \\ z &= z' \\ t &= \gamma(vx' + t') \end{aligned}$$

These equations have the same form as the original equations, except that the sign of v is reversed. Of course this must be true: If the primed RF is moving at $+v$ with respect to the unprimed RF, then the unprimed RF must be moving at $-v$ with respect to the primed RF. The relative velocity is all that matters.

Exercise:

Invert the original equations to find x, y, z, and t in terms of x', y', z' and t', that is, prove the above statement.

Such a transformation of spacetime coordinates is analogous in some ways to the transformation of ordinary Cartesian coordinates under a rotation in Euclidean space:

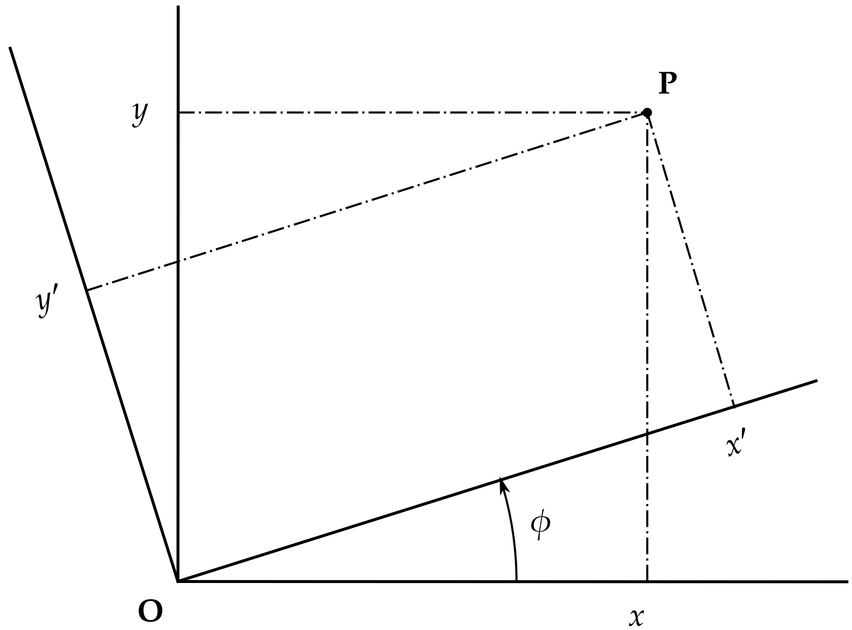


Figure 24: A rotation in Euclidean space.

Here in this space we can write the analog of the Lorentz transformation equations:

$$\begin{aligned} x' &= x\cos\phi + y\sin\phi \\ y' &= -x\sin\phi + y\cos\phi \end{aligned}$$

Is there an invariant, *i.e.*, a quantity that does not change its form when the coordinate system is rotated? Yes! It is the square of the distance between the origin **O** and the point **P**:

$$\begin{aligned} x'^2 + y'^2 &= x^2\cos^2\phi + 2xy\cos\phi\sin\phi + y^2\sin^2\phi \\ &\quad + x^2\sin^2\phi - 2xy\cos\phi\sin\phi + y^2\cos^2\phi \\ &= x^2 + y^2 \end{aligned}$$

Note the circular functions $\sin\phi$ and $\cos\phi$. The "locus of one-meter-marks" is the circle of radius 1:

$$x^2 + y^2 = 1$$

In spacetime, where the calibration curves are hyperbolas, we might expect to see the circular functions replaced by the hyperbolic functions $\cosh\phi$ and $\sinh\phi$. It is true. We may set $\cosh\phi = \gamma$, and $\sinh\phi = v\gamma$. In this case, the invariance of the spacetime interval boils down to the identity

$$\cosh^2\phi - \sinh^2\phi = 1$$

Although it is insightful to pursue this analogy, we shall not explore it further. It is discussed at considerable length in *Spacetime Physics*, by Taylor and Wheeler.

Having now derived the general Lorentz transformation equations, we may use them to yield the Einstein formula for the addition of velocities. Suppose a particle moves with velocity $u' = x'/t'$ with respect to an RF which itself is moving with velocity v with respect to a fixed RF.

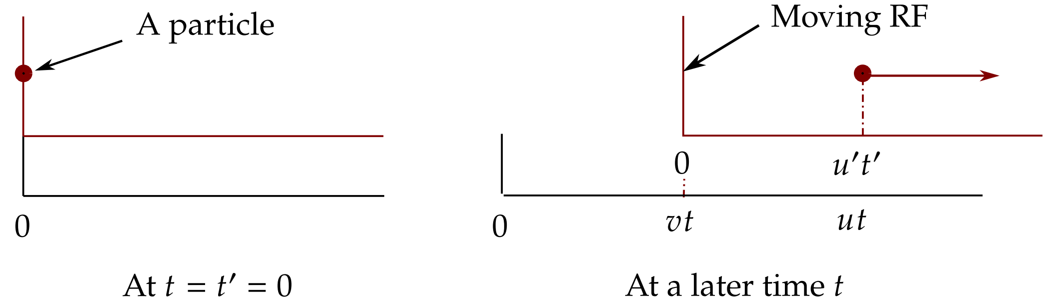


Figure 25: Checking the Einstein formula.

With what velocity u is the particle moving with respect to the fixed RF? That is, what is $u = x/t$? For simplicity, we shall assume that the particle is at the origin of both RFs when they coincide at $t = t' = 0$. Using the Lorentz transformation equations, we have

$$u = \frac{x}{t} = \frac{x' + vt'}{t' + vx'} = \frac{x'/t' + v}{1 + vx'/t'} = \frac{u' + v}{1 + u'v}$$

Note that the existence of the term $u'v$ in the denominator implies that u cannot be greater than 1, that is, that u cannot be greater than c.

Example:

Suppose that a particle moves with speed $\frac{3}{4}c$ in an RF which is itself moving with velocity $\frac{3}{4}c$ with respect to a stationary RF. (For this example, we've put "c" back in to emphasize its relevance). How fast is the particle moving with respect to the stationary RF? Newton would predict $\frac{3}{2}c$. Actually, we have

$$u = \frac{3/4 + 3/4}{1 + 9/16}c = \frac{3/2}{25/16}c = \frac{48}{50}c = 0.96c$$

Exercise:

Consider two events whose coordinates (x, y, z, t) relative to some reference frame S are $(0, 0, 0, 0)$ and $(2, 0, 0, 1)$ in units with $c = 1$. Find the speeds of reference frames moving in the x-direction for which

(a) the two events are simultaneous, and

(b) the second event *precedes* the other by one unit of time. Is there an RF for which the two events occur at the same spatial position? Does this show that causality is violated by SRT? Discuss.

Another exercise:

A rod of proper length 8 cm moves longitudinally at speed $0.8c$ in a reference frame S. It is passed by a particle moving at speed $0.8c$ through S in the opposite direction. In S, how long does the particle take to pass the rod? (Answer: 10^{-10} seconds.)

X. Lorentz Invariance

In the previous section we demonstrated that the Lorentz transformation equations may be derived using only light signals and the observed behavior of clocks. A key feature is that $x^2 - t^2$ (or $x^2 + y^2 + z^2 - t^2$ if we include all four spacetime coordinates) is an *invariant*. It has the same numerical value for all observers. That is, $x'^2 - t'^2 = x^2 - t^2$, where (x', t') are the spacetime coordinates for an event in one RF, and (x, t) are the spacetime coordinates for the same event in another RF. This is an example of *Lorentz invariance*.

Exercise:

Starting from the Lorentz transformation equations, show that

$$x'^2 - t'^2 = x^2 - t^2$$

Geometrically, Lorentz invariance has the following implications:

First, the light signal world lines are invariant. These lines are described by the equations $x = \pm t$, or $x^2 - t^2 = 0$. Invariance implies that $x'^2 - t'^2 = 0$ also, or that $x' = \pm t'$, so that the light signal world lines are described by the same equation in all reference frames, or for all observers. The light signal world lines transform into themselves.

Second, the clock calibration hyperbolas are invariant. In this case, $x^2 - t^2 = \tau^2$ describes these hyperbolas, and invariance implies that these hyperbolas also transform into themselves.

Third, the length calibration hyperbolas are invariant. In this case, $x^2 - t^2 = L_0^2$ describes these hyperbolas, and invariance again implies that these hyperbolas transform into themselves.

All of these conclusions can be illustrated by drawing a spacetime diagram (see Fig. 26), where we include light signal world lines for left-and right-going light signals, clock calibration hyperbolas for $\tau = \pm 1$ second, and length calibration hyperbolas for $L_0 = \pm 1$ light-second. The light signal world lines and the calibration hyperbolas apply equally well to all reference frames, no matter what their relative velocity. Unlike the spacetime axes, they do not need to be redrawn for each new RF.

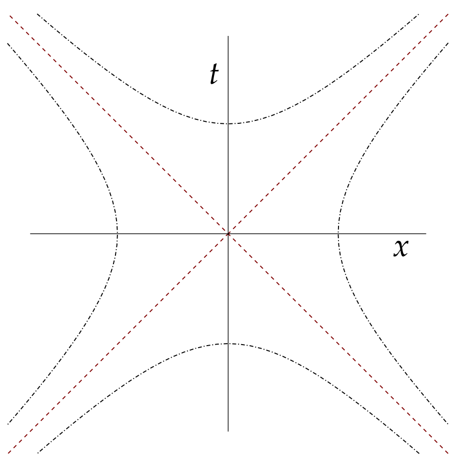


Figure 26: Spacetime invariants.

These observations of Lorentz invariance suggest the following new representation of spacetime: Since the light signal world lines are always represented by a pair of perpendicular lines, no matter what the RF, why not use them as new spacetime axes? We can. The new coordinates are simply

$$t_1 = t - x, \quad \text{and} \quad t_2 = t + x. \quad \text{When inverted:} \quad t = \frac{1}{2}(t_2 + t_1) \quad \text{and} \quad x = \frac{1}{2}(t_2 - t_1)$$

We have seen such relationships before. t_1 and t_2 are the same as those shown in Fig. 23 (page 33) for the sending and receiving of a light signal to an event. We call them *light-signal coordinates*. Our spacetime diagram will look like this:

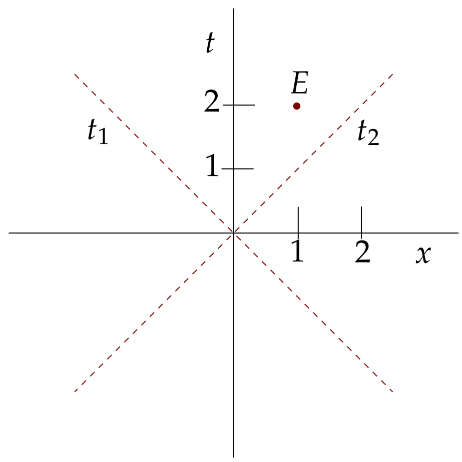


Figure 27: t_1 and t_2 as Light signal coordinates.

Exercise:

In Fig. 27 there is an event E, which could be the ticking of a moving clock (a decaying meson, perhaps?), at $t = 2$ seconds and $x = 1$ light-seconds. Show where its light-signal coordinates are on the diagram. Also, if E lies on a clock's world line whose beginning is at the origin, what would be the values of the light-signal coordinates as observed from the clock's reference frame, and where would they be on the t_1 and t_2 axes (which are also the t'_1 and t'_2 axes).

Note that a change of reference frames does not mean redrawing a new set of axes, but rather rescaling the old ones:

$$t'_1 = \left(\frac{1+v}{1-v}\right)^{1/2} t_1 \quad \text{and} \quad t'_2 = \left(\frac{1-v}{1+v}\right)^{1/2} t_2$$

These are the Lorentz transformation equations for light-signal coordinates. Note that the axes become unequally scaled for a RF moving at velocity v with respect to the original one, but that the reciprocity of the transformation becomes obvious. Lorentz invariance becomes similarly obvious: $t'_1 t'_2 = t_1 t_2$. Light signals through the origin satisfy $t_1 t_2 = 0$ and the calibration hyperbolas are described by $t_1 t_2 = -L_0^2$.

Light-signal coordinates are not much used, but they provide some insight for thinking about invariant quantities.

XI. Four-Vectors

First: a review of 3-vectors:

In Euclidean 3-space, we often use a vector to represent, say, the *displacement* of a particle from one point to another. We can represent such a vector by a bold character, which indicates that it has both a *length* and a *direction*. Thus the *displacement* of a particle through a distance d in some direction we can represent by the symbol **d**.

It is often convenient to represent such a vector in the context of a Cartesian coordinate system or Euclidean reference frame. Here is a representation of the displacement of a particle from the origin to a point in space:

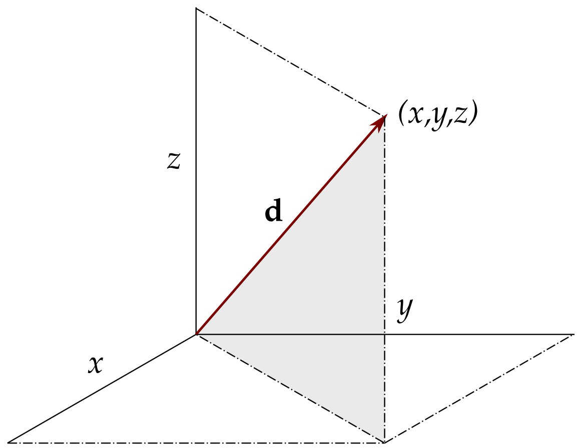


Figure 28: A vector in a 3-dimensional Euclidean reference frame.

In this reference frame, we can label the components of **d** by the symbols x, y and z. (x, y and z are an ordered triple of scalar quantities, which is a better way of defining a vector in 3-dimensional space.) If the reference frame is rotated about the origin, the vector itself will not change, but its components will change to a new set: x', y', z'. Traditionally[8], we do the following three steps. In each step, imagine looking along the relevant axis toward the origin:

1. rotate counter-clockwise about the z-axis by φ,
2. rotate counter-clockwise about the new x-axis by θ,
3. rotate counter-clockwise about the new z-axis by ψ.

The new components x', y', and z' will then be related to the original set by these messy transformation equations:

$$
\begin{aligned}
x' &= (\cos\psi\cos\varphi - \cos\theta\sin\varphi\sin\psi)x + (\cos\psi\sin\varphi + \cos\theta\cos\varphi\sin\psi)y + (\sin\psi\sin\theta)z \\
y' &= (-\sin\psi\cos\varphi - \cos\theta\sin\varphi\cos\psi)x + (-\sin\psi\sin\varphi + \cos\theta\cos\varphi\cos\psi)y + (\cos\psi\sin\theta)z \\
z' &= (\sin\theta\sin\varphi)x + (-\sin\theta\cos\varphi)y + (\cos\theta)z
\end{aligned}
$$

[8] See Goldstein's *Classical Mechanics* (3rd ed.) Page 151ff. The angles φ, θ and ψ are usually called the Euler angles.

More succinctly, we may write, where **R** denotes a Rotation matrix:

$$\mathbf{d}' = \mathbf{R}\mathbf{d}$$

where

$$\mathbf{d}' \equiv \begin{bmatrix} x' \\ y' \\ z' \end{bmatrix}, \quad \mathbf{R} \equiv \begin{bmatrix} r_{11} & r_{12} & r_{13} \\ r_{21} & r_{22} & r_{23} \\ r_{31} & r_{32} & r_{33} \end{bmatrix} \text{ and } \quad \mathbf{d} \equiv \begin{bmatrix} x \\ y \\ z \end{bmatrix}$$

Thus, following the rules for multiplying a vector by a matrix, we may write the transformation equations as

$$\begin{aligned} x' &= r_{11}x + r_{12}y + r_{13}z \\ y' &= r_{21}x + r_{22}y + r_{23}z \\ z' &= r_{31}x + r_{32}y + r_{33}z \end{aligned}$$

where the r_{ij} are the messy angular expressions shown on the previous page.

Note that the scalar product of **d** with itself ($\mathbf{d}\cdot\mathbf{d} = x^2 + y^2 + z^2$) is an *invariant* quantity under Euclidean rotations, that is, it has the same value in all reference frames. It is just the square of the magnitude of the displacement.

The displacement vector **d** is not the only 3-vector with which we are familiar. Velocity, acceleration, force and momentum are all examples of 3-vectors. It is a useful property of any 3-vector that its components transform under Euclidean rotations just like the components of the displacement. For example, suppose that **v** is a vector representing the velocity of a particle. Then if we rotate our reference frame as before, the components of **v** will change, as before, to a new set given by $\mathbf{v}' = \mathbf{R}\mathbf{v}$, where **R**, with its matrix elements r_{ij}, is the same as with the displacement vector **d**.

It is also a useful property of 3-vectors that the scalar product of any two 3-vectors has the same value in all reference frames. For example, if **F** is a 3-vector representing the force on a particle, and if **v** is the velocity of the particle, then $\mathbf{F}\cdot\mathbf{v}$, which is the work done per second on the particle by the force **F**, has the same value in all reference frames. We say that $F_x v_x + F_y v_y + F_z v_z$ is an *invariant* quantity.

Now about 4-vectors:

Four-vectors, or 4-vectors for short, are analogous to ordinary Euclidean 3-vectors in many ways. In fact, 4-vectors are to spacetime as 3-vectors are to Euclidean space. Just as a displacement may be represented by a 3-vector **d** in Euclidean space, an event may be represented in spacetime by the 4-vector **E**. In some reference frame, this event 4-vector will have 4 spacetime coordinates, x, y, z, and t. In another reference frame, say one moving along the x-axis with velocity v, it will have 4 different spacetime coordinates: x', y', z' and t', which are of course related to x, y, z and t by the Lorentz transformation equations:

$$\begin{aligned} x' &= \gamma x - \gamma v t \\ y' &= y \\ z' &= z \\ t' &= -\gamma v x + \gamma t \end{aligned} \qquad \text{where} \quad \gamma \equiv \frac{1}{\sqrt{1 - v^2}}$$

Analogously with rotations in Euclidean 3-space, we may write, where **L** denotes a Lorentz transformation matrix:

$$\mathbf{E}' = \mathbf{LE}$$

where

$$\mathbf{E}' \equiv \begin{bmatrix} x' \\ y' \\ z' \\ t' \end{bmatrix}, \quad \mathbf{L} \equiv \begin{bmatrix} l_{11} & l_{12} & l_{13} & l_{14} \\ l_{21} & l_{22} & l_{23} & l_{24} \\ l_{31} & l_{32} & l_{33} & l_{34} \\ l_{41} & l_{42} & l_{43} & l_{44} \end{bmatrix} \text{ and } \mathbf{E} \equiv \begin{bmatrix} x \\ y \\ z \\ t \end{bmatrix}$$

The general Lorentz transformation thus appears like this:

$$\begin{aligned} x' &= l_{11}x + l_{12}y + l_{13}z + l_{14}t \\ y' &= l_{21}x + l_{22}y + l_{23}z + l_{24}t \\ z' &= l_{31}x + l_{32}y + l_{33}z + l_{34}t \\ t' &= l_{41}x + l_{42}y + l_{43}z + l_{44}t \end{aligned}$$

Note that if our new reference frame is one moving only to the right along the x-axis with velocity v (as on the previous page), the matrix **L** becomes simplified:

$$\begin{bmatrix} l_{11} & l_{12} & l_{13} & l_{14} \\ l_{21} & l_{22} & l_{23} & l_{24} \\ l_{31} & l_{32} & l_{33} & l_{34} \\ l_{41} & l_{42} & l_{43} & l_{44} \end{bmatrix} \xrightarrow{\text{becomes}} \begin{bmatrix} \gamma & 0 & 0 & -\gamma v \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ -\gamma v & 0 & 0 & \gamma \end{bmatrix}$$

It is useful to *define* the scalar product of **E** with itself:

$$\mathbf{E}\cdot\mathbf{E} \equiv x^2 + y^2 + z^2 - t^2$$

We recognize this as our old friend the Lorentz Invariant. It has the same value, $\mathbf{E}\cdot\mathbf{E} = -\tau^2$, in all reference frames. Of course, the minus sign is essential to the scalar invariance.

As in Euclidean 3-space, we may define other 4-vectors, such as a velocity 4-vector, a momentum 4-vector, an acceleration 4-vector or a force 4-vector. It is a useful property of such 4-vectors that their components transform under Lorentz transformations just like the components of the event 4-vector. That is, if we have some 4-vector **A**, we must have

$$\begin{aligned} A'_x &= \gamma A_x - \gamma v A_t \\ A'_y &= A_y \\ A'_z &= A_z \\ A'_t &= -\gamma v A_x + \gamma A_t \end{aligned}$$

It is also a useful property of 4-vectors that the scalar product of any 4-vector with itself, or of any 4-vector with another 4-vector, has a value that is independent of the reference frame. If **A** and **B** are 4-vectors, then

$$\mathbf{A}\cdot\mathbf{A} = A_x^2 + A_y^2 + A_z^2 - A_t^2 \quad \text{and}$$
$$\mathbf{A}\cdot\mathbf{B} = A_x B_x + A_y B_y + A_z B_z - A_t B_t$$

have the same value in all reference frames.

Now the game is to look for 4-vectors. For example, what do we mean by a velocity 4-vector? Or a momentum 4-vector? In the following sections, we shall explore these questions, and in the process of our search we shall discover the world of relativistic dynamics.

XII. The Four-Velocity

So far we have discussed only one 4-vector, the event 4-vector **E**: (x, y, z, t). It is natural to see whether we can find a 4-vector that is related to the velocity of a particle. If we look at $u_x = dx/dt, u_y = dy/dt, u_z = dz/dt$ and $u_t = dt/dt$, we see that such quantities *cannot* transform properly, and hence cannot represent the components of a 4-vector. We get into trouble because although dx, dy, dz, and dt can be taken as 4-vector components, when we divide these by the quantity dt it doesn't work because dt itself is not an invariant quantity. In fact, if we work out how these ordinary velocities transform, we get:

$$u'_x = \frac{dx'}{dt'} = \frac{dx - vdt}{-vdx + dt} = \frac{u_x - v}{1 - u_x v}$$
$$u'_t = \frac{dt'}{dt'} = 1$$

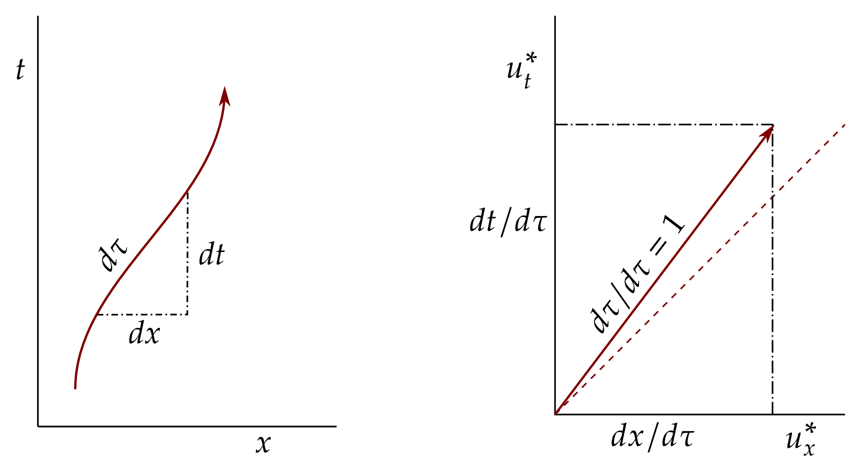
so that the transformation is clearly not of the proper form. If instead we divide by $d\tau$, the incremental *proper* time interval, we can define a velocity whose components transform correctly:

$$u^*_x = \frac{dx}{d\tau}, \quad u^*_y = \frac{dy}{d\tau}, \quad u^*_z = \frac{dz}{d\tau}, \quad u^*_t = \frac{dt}{d\tau}$$

Because $d\tau$ is a Lorentz invariant, the components of $\mathbf{u}^*$ transform like the components of a 4-vector. That is,

$$\begin{aligned} \frac{dx'}{d\tau} &= -\gamma\frac{dx}{d\tau} - v\gamma\frac{dt}{d\tau} \\ \frac{dt'}{d\tau} &= -v\gamma\frac{dx}{d\tau} + \gamma\frac{dt}{d\tau} \end{aligned} \qquad \text{or} \qquad \begin{aligned} u^{*\prime}_x &= \gamma u^*_x - v\gamma u^*_t \\ u^{*\prime}_t &= -v\gamma u^*_x + \gamma u^*_t \end{aligned}$$

We have marked the components of this special velocity with an asterisk (*) in order to distinguish them from the ordinary observed velocity components. Just what is this peculiar velocity? Figure 29 shows a pair of diagrams.



(a) A spacetime diagram.

(b) A 4-velocity diagram.

Figure 29: Comparing diagrams.

Look carefully at Fig. 29. The left diagram shows the World Line of a particle. We can think of the 4-velocity as the velocity of this particle along its World Line, *i.e.*, in the particle's own reference frame. The right diagram is new—it's a 4-velocity diagram. The slope of the 4-velocity vector must be equal to the slope of the World Line in the left diagram, since $d\tau$ is a Lorentz invariant. Hence, $(dt/d\tau)/(dx/d\tau) = dt/dx$.

Note that since $dt > dx$, all possible 4-velocity vectors must lie within the light cone. However, the components of $\mathbf{u}^*$ are *not* the observed particle velocities, but since we know how to convert $d\tau$ into dt for a particle in any given reference frame, we can always get the observed particle velocities.

Here's an interesting and useful thing: Since the components of the 4-velocity transform like the components of an event vector (the 4-velocity is a 4-vector), the scalar product of $\mathbf{u}^*$ with itself is a Lorentz invariant. What is the value of this invariant? Consider a particle moving at constant velocity v with respect to a stationary reference frame. Then

$$u_x^* = \frac{dx}{d\tau} = \gamma\frac{dx}{dt} = \gamma v, \qquad \text{and} \qquad u_t^* = \frac{dt}{d\tau} = \gamma\frac{dt}{dt} = \gamma$$

so that

$$\mathbf{u}^*\cdot\mathbf{u}^* = u_x^* u_x^* - u_t^* u_t^* = v^2\gamma^2 - \gamma^2 = -\gamma^2(1-v^2) = -1$$

The square of the 4-velocity vector is a *constant*, equal to -1. This may be contrasted with the square of the *event* 4-vector, which is $-\tau^2$. This also means that in a 4-velocity diagram, the end points of *all* 4-vector velocities lie on the hyperbola $u_x^{*2} - u_t^{*2} = -1$.

By the way, there is a clever trick that we might have used to calculate the scalar product of the 4-velocity with itself: Since this quantity is an invariant, it is simplest to calculate it in the reference frame of the particle. In that reference frame, it is clear that $u_x^* = 0$, and that $u_t^* = dt/d\tau = \gamma = 1$, so that $\mathbf{u}^*\cdot\mathbf{u}^* = -\mathbf{1}$. Very simple.

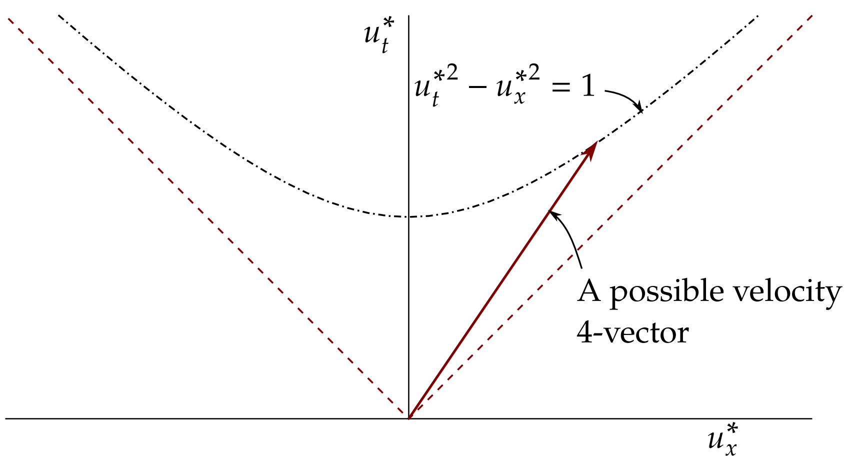


Figure 30: Every velocity 4-vector ends on the hyperbola.

Exercise:

What would be the value of $(\mathbf{u}^*\cdot\mathbf{u}^*)$ if we had not set $c = 1$?

XIII. The Four-Momentum

In Newtonian theory there exists a quantity called the momentum, which for a single particle is simply its mass times its velocity. We usually think of the momentum as a vector quantity, or more precisely, as a 3-vector, with x, y and z components. Two or more particles may be thought of as a *system*. We can calculate the total momentum for a system of particles by adding, vectorially, the momenta of each of the particles. This total momentum is useful in part because it is a constant of the motion, that is, it is *conserved*. At low velocities the principle of *conservation of momentum* is well established.

At high velocities, where we hope that SRT will describe the behavior of mechanical systems, we can formulate a 4-vector **P**, which we shall call the *4-momentum*. We simply multiply the 4-velocity of a particle by its mass m:

$$\mathbf{P} = m\mathbf{u}^*$$

If we think of a particle that is moving in a stationary reference frame with a 3-velocity **u** the components of this 4-momentum will be

$$P_x = mu_x^* = m\gamma u_x = \frac{mu_x}{\sqrt{1-u^2}}, \qquad P_y = mu_y^* = m\gamma u_y = \frac{mu_y}{\sqrt{1-u^2}}$$
$$P_z = mu_z^* = m\gamma u_z = \frac{mu_z}{\sqrt{1-u^2}}, \qquad P_t = mu_t^* = m\gamma \quad = \frac{m}{\sqrt{1-u^2}}$$

What happens to these 4-momentum components in the classical limit, where $u \ll 1$? We can expand the square roots using the Binomial theorem.[9] Thus in the classical limit, we may express the components of the 4-momentum as

$$P_x \approx mu_x, \quad P_y \approx mu_y, \quad P_z \approx mu_z \quad \text{and} \quad P_t \approx m + \frac{1}{2}mu^2$$

Exercise:

What would each component look like if we had not set $c = 1$?

---

[9] The Binomial theorem states that

$$(1+x)^k = 1 + kx = \frac{k(k-1)}{2!}x^2 + \frac{k(k-1)(k-2)}{3!}x^3 + \cdots$$

If k is a positive integer, the series ends with the $(k+1)$st term. Otherwise the series has an infinite number of terms and converges for $-1 < x < 1$. It is most useful when $x \ll 1$, in which case we need take only the first two or three terms to obtain a sufficiently accurate approximation.

In the practical applications of SRT, we often deal with situations where $u \ll 1$ (the classical limit) or where $u \approx 1$ (the relativistic limit). Thus in the classical limit, a useful approximation for γ becomes

$$\gamma = (1-u^2)^{-1/2} \approx 1 + \frac{1}{2}u^2$$

In the relativistic limit, where $u \approx 1$, it is often helpful to observe that

$$1 + u^2 = (1+u)(1-u) \approx 2(1-u)$$

Now, for any isolated system of particles, each of the three momentum components, the total kinetic energy, and the total mass are all observed to be constants of the motion, at least for elastic interactions between the particles of the system. Thus at low velocities and in interactions where the kinetic energy is conserved, the 4-momentum is observed to be a constant of the motion.

Now: What happens to the 4-momentum components when the velocities are large—approaching the light velocity? The experimental observation is that *each component of the 4-momentum is a constant of the motion, even when kinetic energy is not separately conserved*.

In the simple example of a collision between two particles, this implies that

$$P_x^a = P_x^b\,, \qquad P_y^a = P_y^b\,, \qquad P_z^a = P_z^b\,, \qquad P_t^a = P_t^b$$

Here, the superscripts "b" and "a" mean *before* and *after* the collision, respectively.

A striking example of the conservation of 4-momentum occurs for nuclear fission. Consider a nucleus of mass M, initially at rest, which splits up into two lighter mass fragments m_1 and m_2 which are observed to recoil, each carrying off some kinetic energy. The two fragments are observed to travel off in opposite directions; we take their line of motion to be the x-axis for simplicity. Thus we need consider only P_x and P_t.

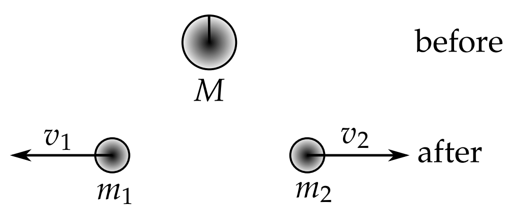


Figure 31: Nuclear fission.

Before the fission, we have

$$P_x = 0 \qquad \text{and} \qquad P_t = M$$

After the fission, we have

$$P_x = m_2 v_2 \gamma_2 - m_1 v_1 \gamma_1 \qquad \text{and} \qquad P_t = m_2\gamma_2 + m_1\gamma_1$$

In these expressions,

$$\gamma_1 \equiv \frac{1}{\sqrt{1-v_1^2}} \qquad \text{and} \qquad \gamma_2 \equiv \frac{1}{\sqrt{1-v_2^2}}$$

Conservation of 4-momentum means that *each* component is conserved. For the t-component P_t we have

$$m_1\gamma_1 + m_2\gamma_2 = M$$

Since the γ-factors are always greater than unity, this relation can be satisfied only if $m_1 + m_2 < M$. This is in fact what is observed. For low velocities, we can use our approximate expressions for γ_1 and γ_2 so that conservation of P_t can be written

$$m_1 + m_2 + \frac{1}{2} m_1 v_1^2 + \frac{1}{2} m_2 v_2^2 = M$$

We see that the mass which appears to have been lost in the fission process appears as kinetic energy of the fragments. This is the basis of what happens in a nuclear reactor (or in a nuclear explosion), where a nucleus such as U^{235} splits into fragments (including some neutrons) whose total mass is less than the original mass of the U^{235} nucleus. The fragments fly off with considerable kinetic energy—energy that can by converted into heat by stopping the fragment in an absorbing material such as water. The water heats up, perhaps turning to steam that can be used to drive a turbine and hence do work, usually to generate electrical power.

Having established the observed relevance of the 4-momentum in describing particle interactions, we return to discuss the individual components in greater detail. For a single particle moving along the x-axis with velocity u, we have

$$P_x = \frac{m}{\sqrt{1-u^2}}\, u \qquad \text{and} \qquad P_t = \frac{m}{\sqrt{1-u^2}}$$

A remark about "relativistic mass":

Historically, the expressions for the spatial components of the 4-momentum were forced into the Newtonian language by calling $m/\sqrt{1-u^2}$ the "mass" of the particle, so that one could still say that momentum is mass times velocity. A large number of textbooks continue this practice, often referring to this quantity as the "relativistic mass", thinking that a moving particle's mass increases as its velocity increases. We shall *not* follow this practice, which has never turned out to be useful. Indeed it can be quite confusing. For us there is only one mass, which we call "m". It is not dependent upon the velocity of the particle. Sometimes we call it the *rest mass*, or as we shall see, the *rest energy*, just to be certain.

It is a new idea that the spatial momentum P_x is not simply proportional to the velocity of the particle. It is only approximately proportional at low velocities, and becomes infinitely large as the velocity approaches the light velocity. The Newtonian theory is only a low velocity approximation.

The time component P_t is something new. We saw that at low velocities it may be approximately written as $m + mu^2/2$, which suggests that it has something to do with the energy of the particle. As a quantity that is observed to be exactly conserved in particle interactions, it is clearly a significant and useful parameter, and it has become customary to *define* P_t as the *total energy* E for the particle. That is,

$$E \equiv P_t$$

It is also useful at times to separate out the part of the total energy E that is independent of the motion of the particle. This is the mass of the particle, sometimes (as we have mentioned) also called the *rest energy*.[10] The remaining part, which contains all the dependence on the particle's motion, is the *kinetic energy* K. We can see how this is done by expanding the γ-factor in its binomial series:

$$E = m + \frac{1}{2}mu^2 + \frac{3}{8}mu^4 + \frac{5}{16}mu^6 + \cdots = m + K$$

Thus

$$K = E - m = m\gamma - m = m(\gamma - 1) = \frac{1}{2}mu^2(1 + \frac{3}{4}u^2 + \frac{5}{8}u^4 + \cdots)$$

We see that only at low velocities is the kinetic energy K given by $mu^2/2$. At higher velocities, we must use the expression

$$K = E - m$$

So far, we have identified P_x with the spatial momentum and P_t with the total energy for our particle. We can go even further by writing down the Lorentz invariant for the 4-momentum:

$$\mathbf{P}\cdot\mathbf{P} = P_x^2 - P_t^2 = -m^2$$

If we include the other two spatial momentum components, we have

$$\mathbf{P}\cdot\mathbf{P} = P_x^2 + P_y^2 + P_z^2 - P_t^2 = -m^2$$

or

$$P^2 - E^2 = -m^2$$

In this equation, $P^2 = P_x^2 + P_y^2 + P_z^2$ is just the square of the magnitude of the ordinary spatial momentum for the particle.

In calculating the magnitude of the Lorentz invariant, we have again used the trick of calculating it in a reference frame in which our particle is at rest. In that reference frame, the spatial components all vanish, and the time component is just the mass m.

Exercise:

Verify the Lorentz invariance of $\mathbf{P}\cdot\mathbf{P}$ by calculating it in a reference frame moving at the particle's constant velocity u.

The relation above regarding the Lorentz invariant for the 4-momentum is reminiscent of the relation $x^2 - t^2 = -\tau^2$, the hyperbola of "τ-second ticks" in a spacetime diagram. It too is a hyperbola, the hyperbola of "mass m particles" on a *momentum-energy* diagram, shown in Fig. 32 on the next page.

[10] As we shall see, particularly in the following section, when we re-insert the light velocity c, the *rest energy* appears as mc^2—Einstein's most well-known formula.

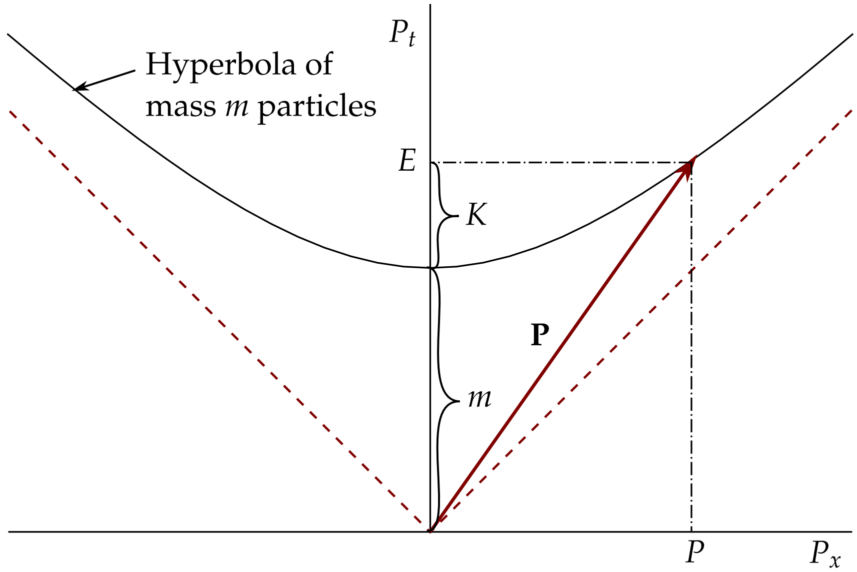


Figure 32: The momentum-energy diagram.

In the diagram above, we have shown a momentum 4-vector for a particle of mass m traveling at high velocity in the x-direction. Note that P/E is the velocity of the particle, so that the slope of the momentum 4-vector is the same as the slope of the world line for this particle on a spacetime diagram. Note also that we can see graphically that $K = E - m$.

It is also interesting to note that as $m \to 0$, the hyperbola of mass m particles becomes the red dashed line. A 4-momentum having its end point on this line must represent a particle of zero mass traveling with the light velocity, such as a photon or neutrino. Such zero mass particles must also have $P = E$.

In a momentum-energy diagram representing two spatial momentum components, P_x and P_y, the dashed line becomes a conical surface. Since points on this surface may be used to represent photon 4-momenta, it is appropriate to refer to it as the "light cone", just as on a spacetime diagram.

Energy-momentum diagrams may often be used to provide a means of illustrating particular kinds of processes involving interactions between particles, as we shall see later on in our discussions of these processes. Such diagrams often allow us to see at a glance whether any particular process is "allowed" or not, that is, whether or not it conserves 4-momentum.

We have now developed SRT to the point where we can use it in a practical way to predict or describe the dynamical behavior of systems of particles for a variety of physical phenomena. By far the most frequent applications of SRT occur in the physics of elementary particles and nuclei—how electrons and protons behave in particle accelerators, and how collections of two or three particles can interact with one another to produce other particles. So far we have alluded to only one real example, that of nuclear fission. In the following sections we discuss a variety of other phenomena.

It is in the world of such particles that velocities approaching the speed of light are realized. None of us has ever observed an automobile, a rocket, or even a golf ball traveling at velocities approaching c, and although we often use such familiar objects in our discussions of relativistic behavior, the real experimental observations relating to them can be perfectly well accounted for using classical Newtonian theory.

Problem:

Suppose an elementary particle of rest mass M decays from rest into a photon and a new particle of rest mass $M/2$. Draw a momentum-energy diagram (to scale) to illustrate this process, and find, either graphically or algebraically, the velocity and kinetic energy of the new particle, and the energy of the photon. (Partial answer: $u = 0.6c$).

XIV. Units

For the past several sections we have been using a system of units in which $c = 1$, with the result that our diagrams have remained uncluttered and our equations have remained clean and elegant. That's very nice, but what about the applications? After all, we are used to measuring mass in kilograms, energy in joules, and momentum in whatever the SI unit of momentum is. If in fact we want to work in such SI units, then we should express our equations such that they contain "c" explicitly. For example, the two most useful equations resulting from our discussion of the previous section can be written incorporating the c's:

$$E^2 - P^2c^2 = m^2c^4, \qquad \text{and} \qquad E = mc^2 + K$$

This is the way these equations are written in most textbooks. Let us use them to answer a question: What is the momentum of an electron that has been accelerated through a potential difference $\Delta V = 10^5$ volts?

First of all, such an electron will possess a kinetic energy K, equal to its charge times the potential difference:

$$K = e\Delta V = (1.6 \times 10^{-19}\,\text{Coulombs})(10^5\,\text{Volts}) = 1.6 \times 10^{-14}\,\text{Joules}$$

If we eliminate E from the two equations at the start of this section, we obtain an expression for P directly in terms of K:

$$P^2c^2 = E^2 - m^2c^4 = (mc^2 + K)^2 - m^2c^4 = K^2 + 2mc^2K$$

so that

$$P^2 = \frac{K^2}{c^2} + 2mK$$

Now the electron's rest mass is about 0.911×10^{-30} kg, so

$$P^2 \approx \left(\frac{1.6 \times 10^{-14}}{3 \times 10^8}\right)^2 + 2(0.911 \times 10^{-30})(1.6 \times 10^{-14}) \approx 3.2 \times 10^{-44}$$

or

$$P \approx \sqrt{3.2 \times 10^{-44}} \approx 1.79 \times 10^{-22}\,\text{kg-m/sec}$$

This turns out to be a very clumsy way to make such calculations. Not only must one deal with large powers of 10, but such units simply do not provide much insight into the physics of any given situation. What does it *mean* to say that an electron having a kinetic energy of 1.6×10^{-14} Joules has a momentum of 1.79×10^{-22} SI units? It's hard to make sense out of such a statement. Moreover, these are not the units used by those who design and use particle accelerators. In what can be viewed as an unusual and happy circumstance, it turns out that the accelerator designers use units that are essentially equivalent to those we have been using, that is, units for which $c = 1$.

In such units, the basic unit of energy is the electron-volt, or eV. This is the amount of work done on an electron in accelerating it through a potential difference of one volt, and is therefore also the resulting kinetic energy of the electron. It is a tiny amount of energy:

$$1\text{eV} = (1\text{ electron charge})(1\text{ volt}) = (1.6 \times 10^{-19})(1) = 1.6 \times 10^{-19}\,\text{Joules}$$

In the example above, we would say simply that the kinetic energy of the electron is one hundred thousand electron-volts, or 100 keV. ($k \equiv$ "kilo" $= 10^3$).

Now the rest energy of the electron turns out to be 511 keV, or .511 MeV. ($M \equiv$ "mega" $= 10^6$). This is mc^2 for the electron, and is the same for every electron. In units where $c = 1$, we could also say that the electron's *mass* is 511 keV. However, in order to distinguish units of mass from units of energy, it is common to use the *name* "eV/c^2" (speak "ee-vee over c-squared") when referring to a unit of mass. Thus we say that the electron has a mass of 511 keV/c^2.

Exercise:

How many kilograms is 1 eV/c^2?

In a similar manner, in units where $c = 1$, we could also measure momentum in eV, since it is obvious from our equations that energy, mass and momentum can all be measured in the same units. Here again, however, in order to distinguish units of momentum from units of energy, we use the *name* "eV/c" when referring to a unit of momentum.

In our example, with $K = 100$ keV and $m = 511$ keV/c^2, we have

$$P^2 = K^2 + 2mK = (100)^2 + (2)(511)(100) = 1.122 \times 10^5 \ (\text{keV/c})^2,$$

or

$$P = \sqrt{1.122 \times 10^5} \approx 335 \ \text{keV/c}$$

Now we can begin to see the sense of using these units. We have an electron having a kinetic energy of 100 keV, a rest energy of 511 keV, and find that its momentum is 335 keV/c. The units are all directly comparable, and the physics takes on greater meaning.

To fix these ideas even more firmly, we illustrate with a momentum-energy diagram. Such a diagram, drawn to scale to represent the momentum 4-vector for this electron, is shown in Fig. 33 on the next page. Note, by the way, that the velocity of the electron is easily determined:

$$u = \frac{P}{E} = \frac{335}{611} = 0.548, \qquad \text{that is,} \qquad u = 0.548c$$

The value of γ is also evident:

$$\gamma = \frac{E}{m} = \frac{611}{511} = 1.20$$

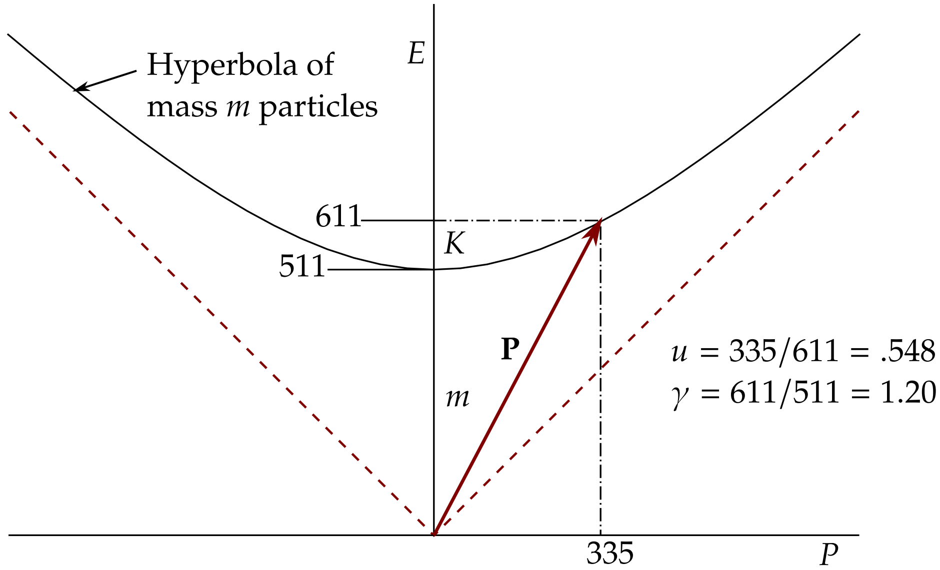


Figure 33: The 4-momentum vector for a 100 keV electron.

One advantage of using these new names for the units of energy, mass and momentum is that we can just as easily work with equations in which c is explicitly present:

$$\begin{aligned} P^2 &= \frac{K^2}{c^2} + 2mK \\ &= \frac{1}{c^2} 10^4 \, (\mathrm{keV})^2 + (2)(511 \, \frac{\mathrm{keV}}{\mathrm{c}^2})(100 \, \mathrm{keV}) = 1.122 \times 10^5 \, (\mathrm{keV})^2/\mathrm{c}^2 \end{aligned}$$

or, taking the square root,

$$P = \sqrt{1.122 \times 10^5} \, \frac{\mathrm{keV}}{\mathrm{c}} \approx 335 \, \frac{\mathrm{keV}}{\mathrm{c}}$$

Exercise:

What is the momentum of an electron having a kinetic energy of 1 eV? Of 1 GeV? (G = giga = 10^9) What is the velocity of the electron in each case, and what are the values of γ?

In applying SRT to problems involving elementary particles, it is useful (and easy) to memorize the rest energies (or rest masses) of a few particles. We have already mentioned that the electron has a rest energy of 511 keV = .511 MeV, or about half an MeV. A proton has a rest energy of 938 MeV, and a neutron a rest energy which is slightly larger at 939 MeV. One atomic mass unit (or amu, which is 1/12 the mass of C^{12}) has a rest energy of 931 MeV. All are approximately 1 GeV, an easy number to remember.

XV. Applications

Pair Annihilation

Perhaps the most dramatic illustration of the conversion of mass to energy occurs in the process of *pair annihilation*, in which a particle and its antiparticle, each of mass m, react to annihilate each other, leaving behind only massless photons. For example, an electron and a positron, initially next to each other and at rest, will react to produce two photons, each of energy 511 keV, like this:

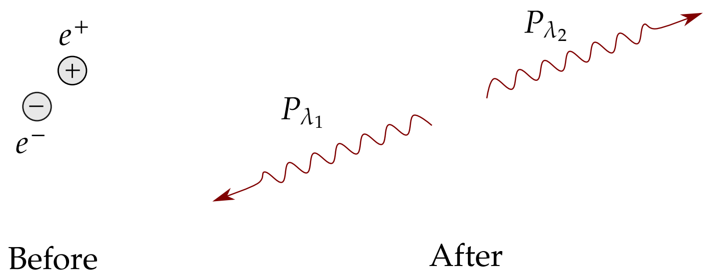


Figure 34: Pair annihilation.

The process can be illustrated using a momentum-energy diagram:

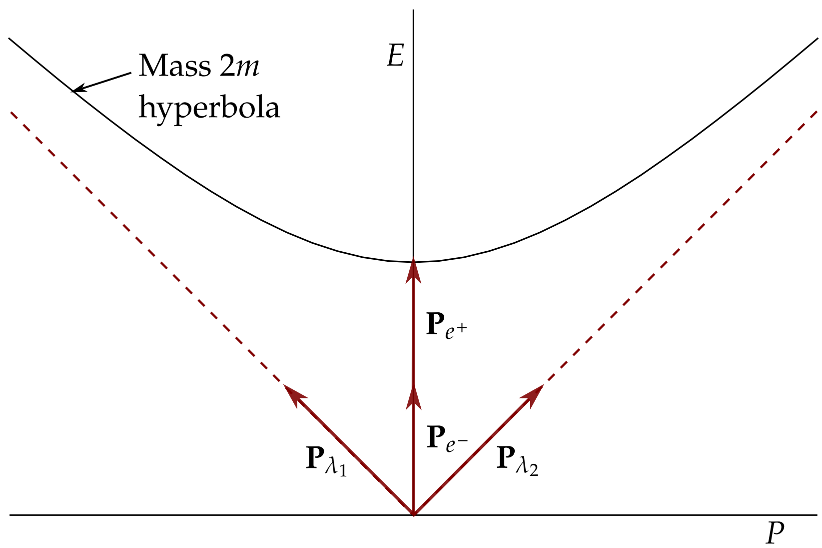


Figure 35: Pair annihilation on a momentum-energy diagram.

Here the initial 4-momentum is just the vector sum of $\mathbf{P}_{e^+}$ and $\mathbf{P}_{e^-}$, and so the total initial 4-momentum has $P = 0$ and $E = 2m$. Since 4-momentum is conserved, this must also be the total final 4-momentum. The two photons that are produced must therefore have equal and opposite spatial momenta, and each photon must have an energy equal to the electron rest mass, or 511 keV. Note also that at least two photons must be produced. A process involving a single final photon is not allowed since in that case 4-momentum would not be conserved. Photon 4-momenta must lie on the surface of the light cone.

Exercise:

What if the electron and the positron were moving toward each other, each with a kinetic energy of 100 keV, prior to annihilation? What would be the energy of the photons produced?

Pair Production

A process that is close to the inverse of pair annihilation, in which a high-energy photon is transformed into an electron-positron pair, is called *pair production*. At first glance we would predict that such a process could not occur, since 4-momentum cannot be conserved in a process in which a single photon disappears and a pair of particles, each of mass m, is created. In fact that is true, but if the photon passes near a massive nucleus or other massive particle the process can occur. Then the excess momentum can be carried off by the massive nucleus.

Here is a spatial picture of the process:

Figure 36: Pair production.

In the next figure, the process is represented on a momentum-energy diagram.

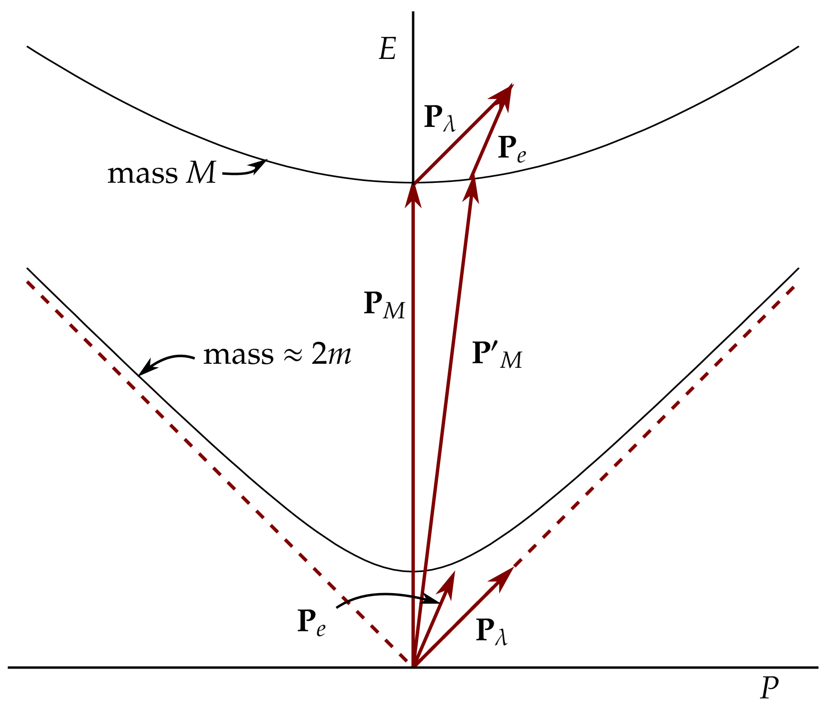


Figure 37: Pair production on a momentum-energy diagram.

In the diagram of Fig. 37, a photon of 4-momentum $\mathbf{P}_\lambda$ is converted into an electron-positron pair whose total (summed) 4-momentum is $\mathbf{P}_e$. The production of the pair takes place in the vicinity of a heavy nucleus of mass M and 4-momentum $\mathbf{P}_M$. Each member of the pair has a mass $m = 511\,\text{keV}$. (For the purpose of this discussion, we have drawn the diagram so that $M/m \approx 10$. In fact, M/m will be several thousand or more. The 4-momenta of both the photon and the pair are shown at both the bottom and the top of the diagram, so as to illustrate how the total 4-momentum is conserved.)

Two features are evident from looking at the diagram. First, it is clear that there is a threshold energy for the incident photon: It must have an energy that is at least equal to the total rest energy of the pair, or an energy greater than 1.02 MeV. If it has an energy greater than that, then the resulting pair will possess some kinetic energy; if the incident photon energy is less than this threshold, no pair will be produced. Second, nothing much happens to the nucleus. It gains a momentum comparable with the momentum of the incident photon, but gains almost no kinetic energy because of its large mass, and so is given a very small velocity.

Exercise:

Suppose a photon of energy only slightly greater than the threshold energy of 1.02 MeV produces a pair in the vicinity of a C^{12} nucleus. (The rest energy of a C^{12} nucleus is about 11.2 GeV.) What will be the recoil kinetic energy of the nucleus, and what will be its velocity? (Partial answer: about 47 eV.)

The Compton Effect

In this process, a beam of short wavelength photons, each characterized by a wavelength λ and a 4-momentum $\mathbf{P}_\lambda$, is incident upon matter containing electrons—a piece of graphite, for example. A photon can scatter elastically from an electron, giving up energy and momentum to the electron, and will come off at some angle φ to the original photon beam. With appropriate detectors we can observe the scattered photon and measure its wavelength λ', or more generally its 4-momentum $\mathbf{P}'$. We wish to predict by how much the photon wavelength (or photon momentum) will be shifted for any particular scattering angle. A spatial diagram of the process is shown in Fig. 38:

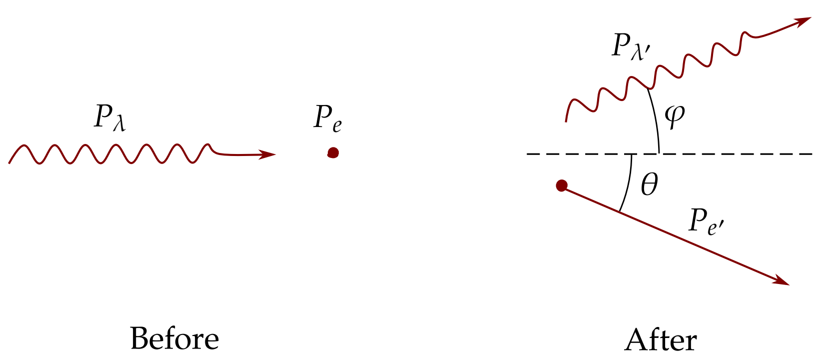


Figure 38: The Compton effect.

It is possible to treat this collision in a straightforward manner by writing down three equations, one for conservation of P_x, one for conservation of P_y, and one for conservation of the total energy E:

$$\begin{aligned}
P_\lambda &= P_{\lambda'}\cos\varphi + P_{e'}\cos\theta \\
0 &= P_{\lambda'}\sin\varphi - P_{e'}\sin\theta \\
P_\lambda + m &= P_{\lambda'} + E
\end{aligned}$$

Also the electron's total energy E is related to its momentum P:

$$E^2 = P^2 + m^2$$

These equations may be manipulated to yield

$$\frac{1}{P_{\lambda'}} - \frac{1}{P_\lambda} = \frac{1}{m}(1-\cos\varphi)$$

Problem:

Derive the above relation by manipulating the equations.

Now the photon wavelength λ is related to its momentum P_λ via the de Broglie formula: $P_\lambda = h/\lambda$ where h is Planck's constant. Thus the above equation may be rewritten to yield the wavelength shift:

$$\lambda' - \lambda = \frac{h}{m}(1-\cos\varphi)$$

or in units where $c \neq 1$, as

$$\lambda' - \lambda = \frac{h}{mc}(1-\cos\varphi)$$

This is the *Compton scattering formula*, and the method of derivation is identical to that outlined in numerous textbooks. If you will try the above problem, you will see that it lacks a certain amount of elegance. You just write down the equations and crank away.

Now we repeat the derivation, using techniques we have learned. We begin by writing down the 4-momentum vectors $\mathbf{P} : (P_x, P_y, P_z, P_t)$ for both the photon and the electron, both before and after the collision:

$$\text{Before:} \quad \begin{cases} \mathbf{P}_\lambda : (P_\lambda,\ 0,\ 0,\ P_\lambda) \\ \mathbf{P}_e : (0,\ 0,\ 0,\ m) \end{cases}$$

$$\text{After:} \quad \begin{cases} \mathbf{P}_{\lambda'} : (P_{\lambda'}\cos\varphi,\ P_{\lambda'}\sin\varphi,\ 0,\ P_{\lambda'}) \\ \mathbf{P}_{e'} : (P\cos\theta,\ -P\sin\theta,\ 0,\ E) \end{cases}$$

Conservation of 4-momentum requires that

$$\mathbf{P}_\lambda + \mathbf{P}_e = \mathbf{P}_{\lambda'} + \mathbf{P}_{e'}, \qquad \text{or} \qquad \mathbf{P}_\lambda - \mathbf{P}_{\lambda'} = \mathbf{P}_{e'} - \mathbf{P}_e$$

while conservation of P_t, or total energy, requires that

$$P_\lambda + m = P_{\lambda'} + E$$

Now we may take the scalar product of the first equation with itself:

$$(\mathbf{P}_\lambda \cdot \mathbf{P}_\lambda) + (\mathbf{P}_{\lambda'} \cdot \mathbf{P}_{\lambda'}) - 2(\mathbf{P}_\lambda \cdot \mathbf{P}_{\lambda'}) = (\mathbf{P}_{e'} \cdot \mathbf{P}_{e'}) + (\mathbf{P}_e \cdot \mathbf{P}_e) - 2(\mathbf{P}_{e'} \cdot \mathbf{P}_e)$$

Each of the first two terms on the left is zero, since every photon 4-momentum vector has zero Lorentz length. Likewise, each of the first two terms on the right is equal to $-m^2$, where m is the rest mass of the electron.

Thus:

$$(\mathbf{P}_\lambda \cdot \mathbf{P}_{\lambda'}) = m^2 + (\mathbf{P}_{e'} \cdot \mathbf{P}_e)$$

The cross-product terms may be developed by substituting the appropriate 4-momentum components and using the rule for taking scalar products of 4-vectors:

$$\begin{aligned}(\mathbf{P}_\lambda \cdot \mathbf{P}_{\lambda'}) &= P_\lambda P_{\lambda'} \cos\varphi - P_\lambda P_{\lambda'} = P_\lambda P_{\lambda'}(\cos\varphi - 1)\\ (\mathbf{P}_{e'} \cdot \mathbf{P}_e) &= -mE\end{aligned}$$

Thus we have

$$P_\lambda P_{\lambda'}(1 - \cos\varphi) = mE - m^2$$

Substituting for E from the equation requiring conservation of total energy, we obtain

$$P_\lambda P_{\lambda'}(1 - \cos\varphi) = m(P_\lambda - P_{\lambda'} + m) - m^2 = m(P_\lambda - P_{\lambda'})$$

which results, after inserting the de Broglie wavelength for each photon and inserting $c \neq 1$, in the *Compton scattering formula*:

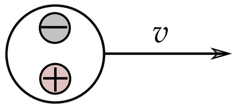
$$\lambda' - \lambda = \frac{h}{mc}(1 - \cos\varphi)$$

This is an illustration of the power of using the 4-momentum vector to answer questions about the dynamics of particle interactions. While such techniques do not always yield solutions to problems so quickly and easily, it is useful to have a thorough working knowledge of 4-vectors in our bag of tricks.

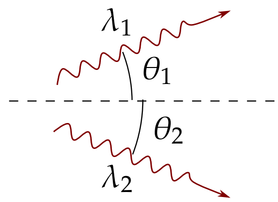
One might ask whether we might have made use of a momentum-energy diagram in our discussion of the Compton Effect, much as we used such diagrams in discussion pair annihilation and pair production. After all, such geometrical interpretations often lend considerable insight. In this case, however, the process requires a three-dimensional momentum-energy diagram, since the scattering takes place in a plane requiring two spatial dimensions. Thus the diagram is difficult to draw and to visualize. Moreover, while it is possible to interpret the scalar products of the 4-vectors geometrically, this interpretation sheds no additional light on the nature of the process, and so in this instance is not very useful.

Problem:

A *positronium* atom consists of an electron and a positron orbiting each other, like a hydrogen atom whose proton is replaced by a positron. Suppose a positronium atom is moving along with velocity v:



Suddenly, the electron and the positron annihilate each other, producing two photons of wavelengths λ_1 and λ_2:



(a) If the electron and the positron each have a rest mass m, what are the components of the 4-momentum of the pair prior to annihilation? (For this problem, you may ignore any orbital momentum of the particles within the positronium.)

(b) What are the components of each photon 4-momentum after the annihilation?

(c) The 4-momentum is conserved in the process:

$$\mathbf{P}_{e-} + \mathbf{P}_{e+} = \mathbf{P}_{\lambda_1} + \mathbf{P}_{\lambda_2}$$

By taking the scalar product of this equation with itself, show that

$$\sin\frac{\theta}{2} = \frac{m}{\sqrt{P_{\lambda_1} P_{\lambda_2}}} = \frac{m}{h}\sqrt{\lambda_1 \lambda_2}$$

where h is Planck's constant and $\theta = \theta_1 + \theta_2$ is the angle between the two photons.

(d) Oh, and by the way, if the above formula is written in units where $c \neq 1$, where should c appear in the formula?

Another problem:

Suppose a part m of the Earth's total rest mass M is somehow transmuted into photons and radiated directly ahead into the Earth's orbital direction, so that by reaction the remaining Earth's mass comes to a complete halt (and thereupon falls into the sun). If the Earth's orbital speed is u (about 18.5 miles/sec) show that

$$\frac{m}{M} = 1 - \frac{\sqrt{1 - u/c}}{\sqrt{1 + u/c}}$$

Show further that this reduces to the approximate relation $m/M \approx u/c$, so that about 1/10,000 of the earth's mass must be converted to accomplish this. Show also that if mass is ejected in the form of matter rather than light, a greater proportion of rest mass must be sacrificed. (Hint: Use the conservation of 4-momentum.) Finally, sketch a momentum-energy diagram (not to scale) illustrating this remarkable process.

Conclusion

These notes represent only the merest introduction to SRT, and to a way of thinking about the universe. To go further leads to new delights, with spacetime providing the basis for developing the techniques and applying them, as we cast more of our traditional classical physical concepts into proper relativistic form.

One of the most enlightening extensions, for example, occurs when SRT is used to describe electromagnetic theory, where we discover that a Lorentz-transformed static electric field becomes a magnetic field and vice-versa. For this, we must develop a clear understanding not only of 4-vectors, but also of 4-tensors. Much as space and time or momentum and energy coalesce to form the components of a 4-vector, so the electric and magnetic fields coalesce to form the components of a 4-tensor. This is where Einstein began in his original development of SRT. It is in Maxwell's formulation of electromagnetic theory that the light velocity c appears as an apparently universal physical constant, so it is not surprising that SRT and electromagnetic theory are intimately related.

Further, we expect that other areas of physics must be reconcilable with SRT. Quantum Theory in particular ought to be expressible in relativistic form. It can be, as exemplified most clearly in the Dirac theory of the electron.

The extension of SRT to include gravitation leads us to General Relativity, in which free particles move along the "straightest possible" lines in curved spacetime. Therein lies still another fascinating transformation of the classical concepts of Newton toward a more modern synthesis involving a strong focus on the basically *geometrical* properties of the universe.

So is that it? Has it all been done? Happily the answer is no. Cosmological theories of the universe, which must be intimately related to General Relativity, are neither wholly understood nor tested, and the reconciliation of General Relativity and Quantum Theory has yet to be accomplished. In short, when we reach the edges of our understandings, they turn out to be rough, ragged and perplexing. Perhaps a smoothing of these ragged edges will lead us once again back to question, to distrust and to re-evaluate some of the basic premises and foundations upon which our current knowledge of the universe is based.