

Chapter 3

Minkowski spacetime

3.1 Events

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

3.2 Reference frames

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

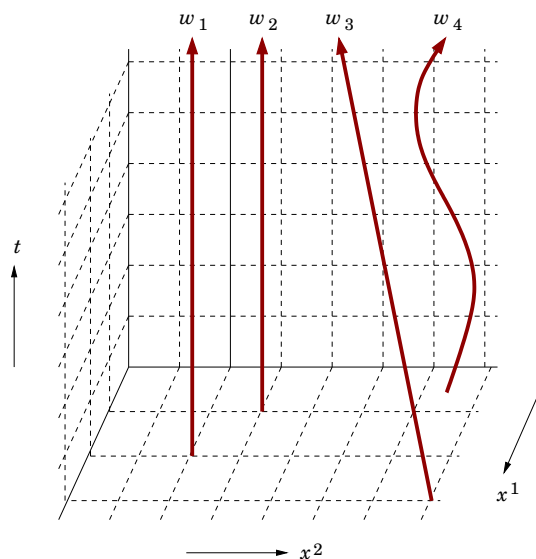


Figure 3.1: An inertial reference frame. Worldlines w_1 and w_2 represent observers at rest in this reference frame, w_3 is the spacetime trajectory of an inertial observer who is moving in this frame, and w_4 is the spacetime trajectory of a non-inertial object whose velocity and acceleration fluctuates.

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

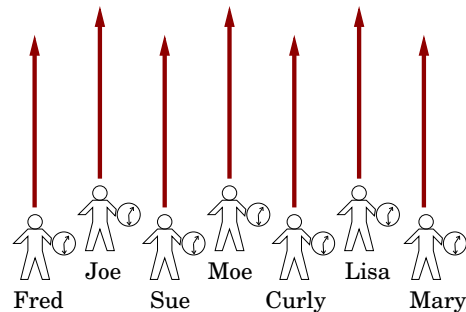


Figure 3.2: A family of inertial observers, all with synchronized watches and mutually at rest, defines an inertial reference frame.

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

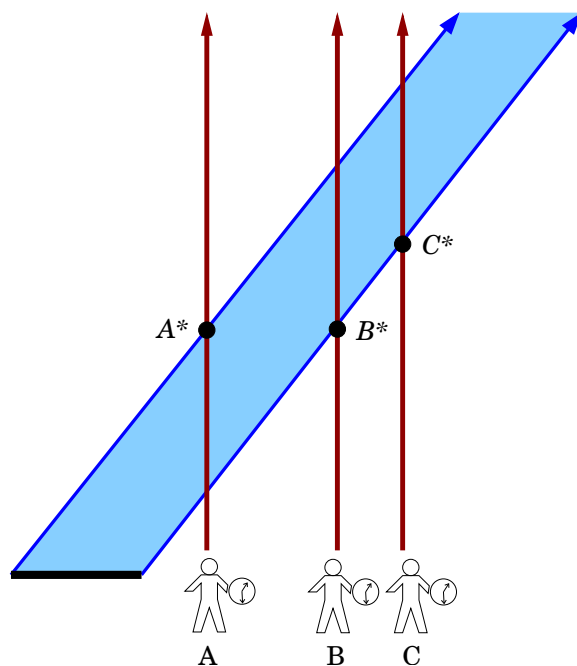


Figure 3.3: A moving rod which passes by three inertial observers who are at rest.

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

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

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

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

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

3.3 Lightcones

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

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

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

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

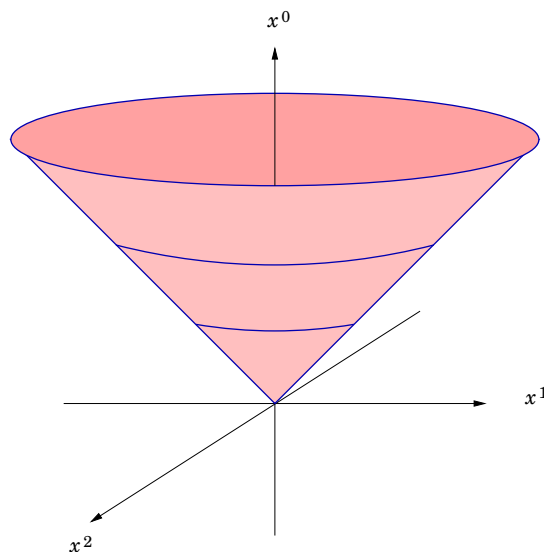


Figure 3.4: The “lightcone” of a flash of light emitted from the origin.

3.4 Simultaneity

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

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

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

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

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

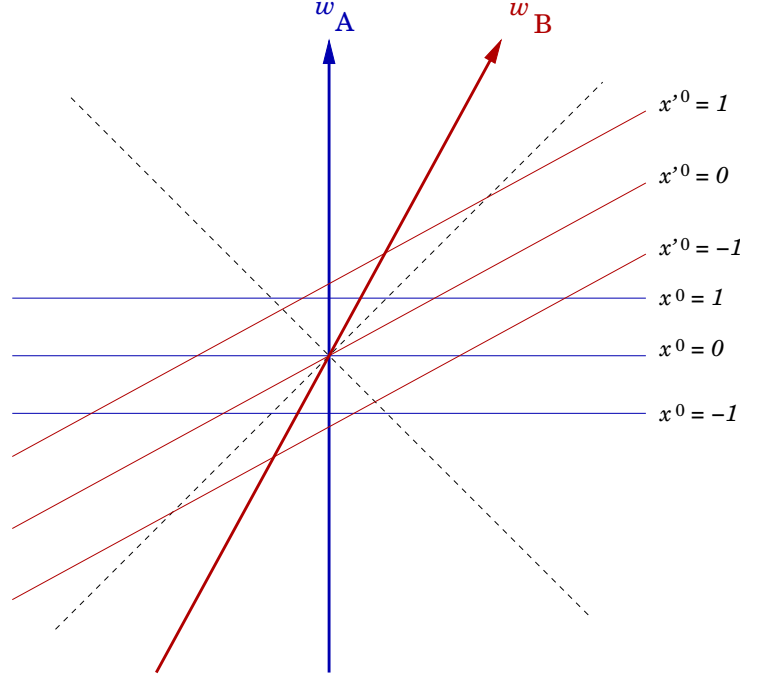


Figure 3.5: Worldline of two observers, and corresponding surfaces of simultaneity. The dashed lines show the lightcone of the origin.

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

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

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

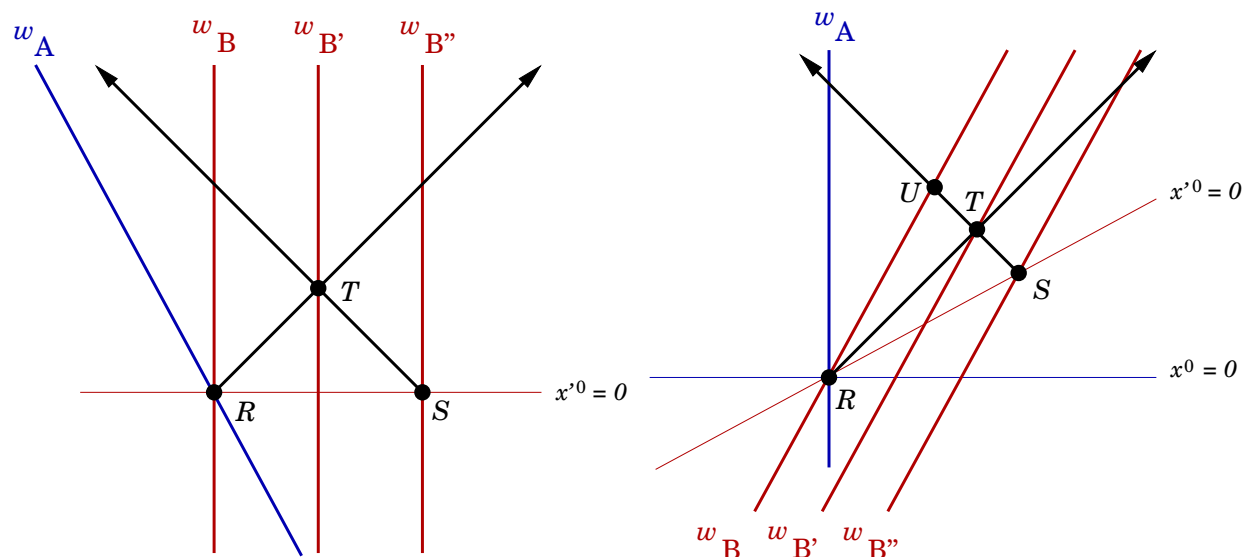


Figure 3.6: Two flashes of light emitted at events R and S which meet at event T , as described in two different frames.

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

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

3.5 Lorentz transformations

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

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

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

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

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

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

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

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

implying that the boost matrix Λ has the block diagonal form

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

3.6 Spacetime vectors

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Next recall that the dot or scalar product of two spatial vectors, $\vec{a} \cdot \vec{b}$, is defined geometrically, without reference to any coordinate system, as the product of the length of each vector times the cosine of the angle between them. One can then show that this is the same as the component-based definition, $\vec{a} \cdot \vec{b} = \sum_i a^i b^i$, for *any* choice of Cartesian coordinates. It is this frame or rotation *independence* that ensures this product is a scalar, *i.e.*, that it is not changed by a rotation.

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

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

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

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

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

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

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

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

Hence

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

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

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

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

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

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

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

(Note that in the East Coast definition the minus sign in the exponent becomes a plus sign.) The similarity between these two expressions already hints at the dual nature of particles and waves that is characteristic of quantum mechanics.

3.7 Units: What is large and what is small?

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

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

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

Further, as noted above, the actual magnitudes of the standard units were chosen to correspond to human scales (*e.g.*, the size of a king). These choices are, of course, unnatural for particle physics

applications. For example, the mass of a proton is 1.67×10^{-27} kg while the spatial “size” of a proton is measured in fermi’s (1 femtometer = 1 fm = 10^{-15} m), not meters. Likewise the lifetime of a typical particle that decays via the strong interactions is of order 10^{-23} s, which is the time for light to travel across a particle of size 1 fm. Since the particle physics we will discuss later in this course is “naturally” relativistic and quantum mechanical, we would like to make a different choice of scales so that the speed of light c and $\hbar$ are both of order 1. It turns out we can address all of the above issues by defining a new set of “particle physics units” such that both c and $\hbar$ are exactly equal to 1!!!! In the process we have reduced the number of types of dimensionfull quantities to 1. In these rather surprising “natural” units we have

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

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

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

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

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

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

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

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

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

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

3.8 Minkowski spacetime

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

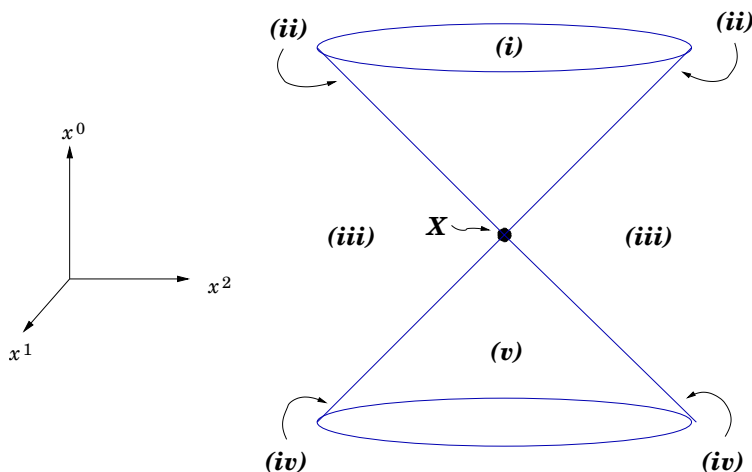


Figure 3.7: The past and future lightcones of an event X separate spacetime into those events which are: (i) timelike separated and in the future of X , (ii) lightlike separated and in the future of X , (iii) spacelike separated, (iv) lightlike separated and in the past of X , and (v) timelike separated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

3.9 The pole and the barn

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

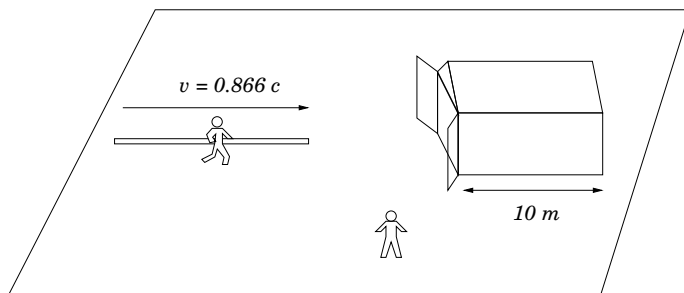


Figure 3.8: A relativistic runner, carrying a long pole, approaches and passes through a barn. Does the pole fit within the barn?

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

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

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

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

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

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

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

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

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

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

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

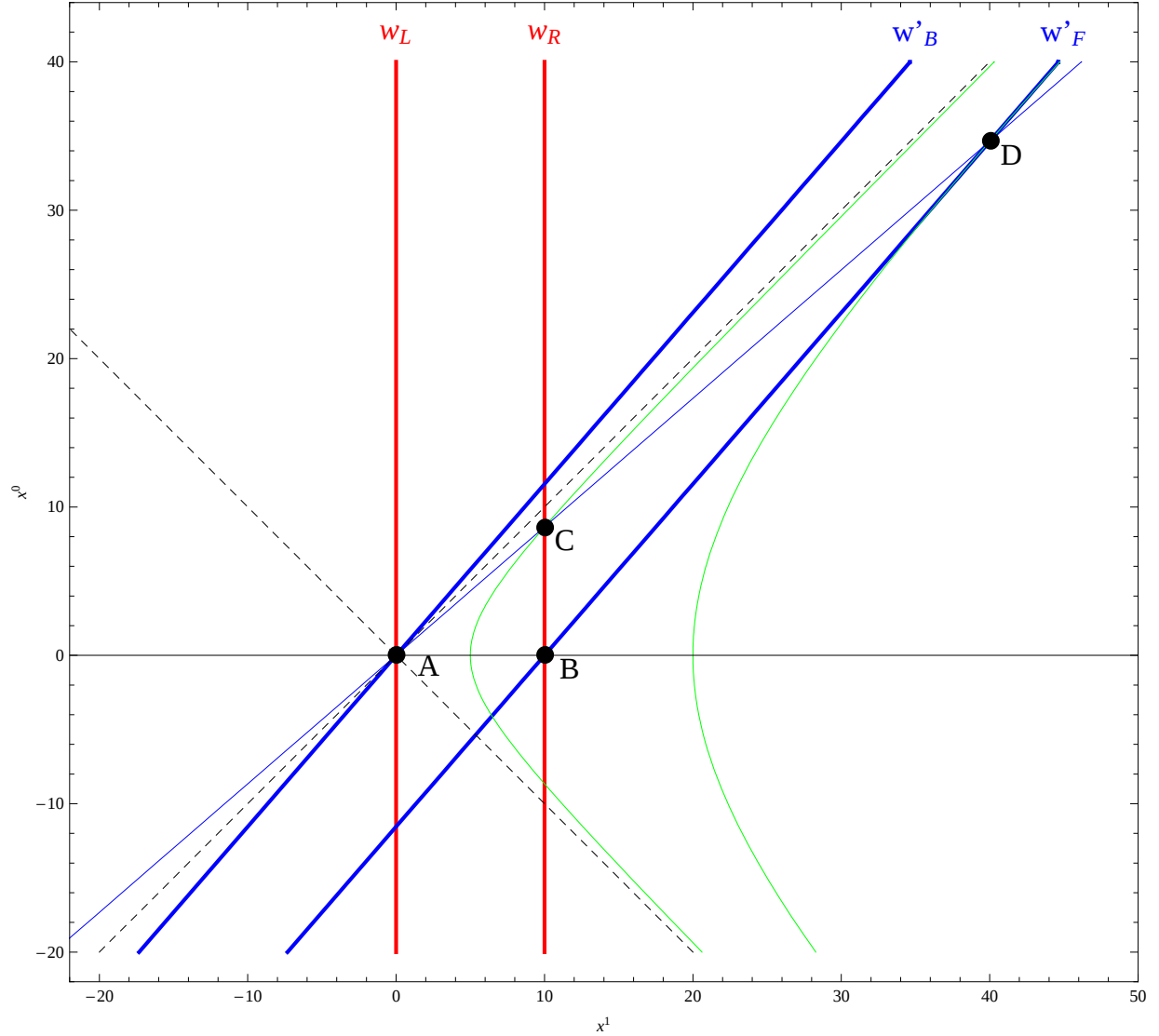


Figure 3.9: A spacetime diagram of the pole and the barn, showing events in the rest frame of the barn. The red vertical lines are the worldlines w_L and w_R of the left and right ends of the barn. The blue lines labeled w'_F and w'_B are the worldlines of the front and back of the pole, respectively.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

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

3.10 Causality

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

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

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

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

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

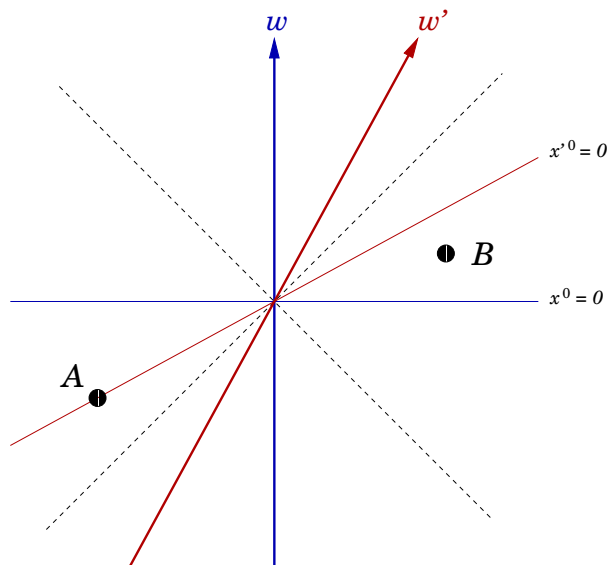


Figure 3.10: Two spacelike separated events A and B . In the unprimed frame, B is in the future of A , but in the primed frame B is in the past of A .

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

3.11 Example Problems

Kogut 2-6

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

Method #1: Thought-experiment

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

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

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

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

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

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

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

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

direction of motion in the two segments)

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

which is the usual time dilation result.

Method #2: Lorentz transformation

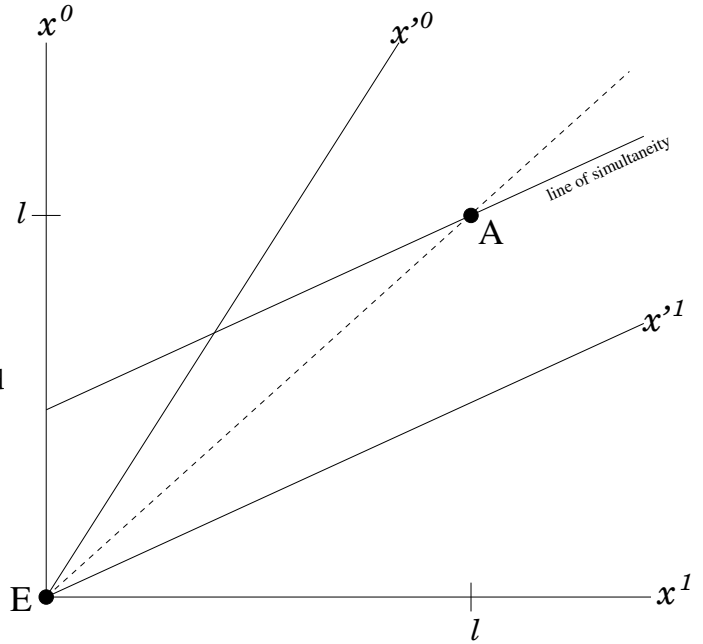
In the original frame, the emission event may be placed at the origin of the Minkowski diagram of spacetime. The absorption event then has coordinates $(x^0, x^1) = (\ell, \ell)$ which lies on the lightcone (since it describes the motion of light!). Under a boost, the origin is mapped to the origin so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

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

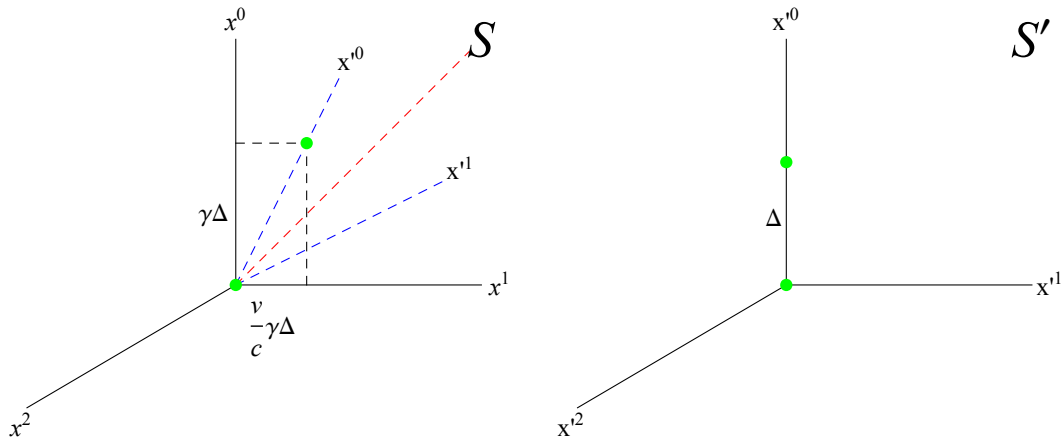
The spatial separation is given by $x'^1 = \gamma \ell (1 - v/c)$, which reduces to the same answer given above for ct' , i.e., $\ell \sqrt{(1 - v/c)/(1 + v/c)}$. Since the events lie on the lightcone, the time separation (times c) and spatial separation are equivalent.

Method #3: Spacetime diagram

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



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


 Figure 3.11: Timelike separation in S' frame.

More spacetime separation examples.

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

Timelike separation

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

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

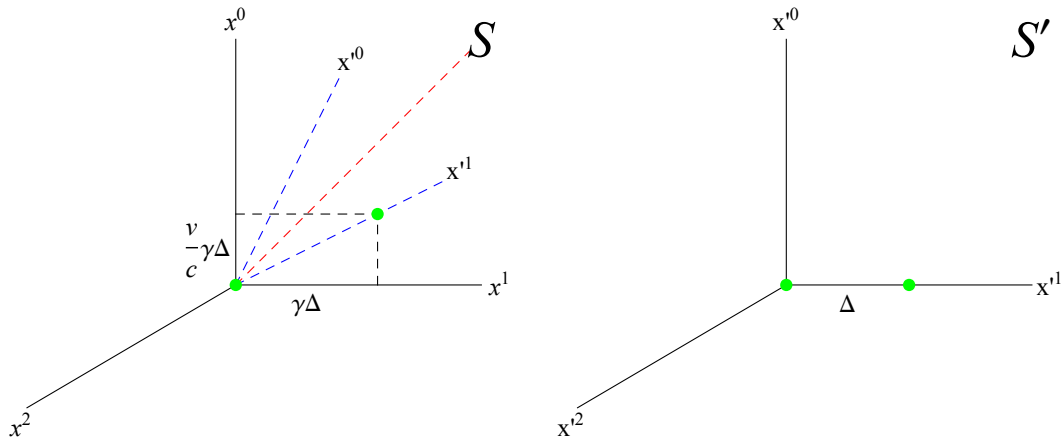
in agreement with the figure. The *invariant* separation squared is given by

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

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

Spacelike separation, $\hat{e}^1$ direction

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


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

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

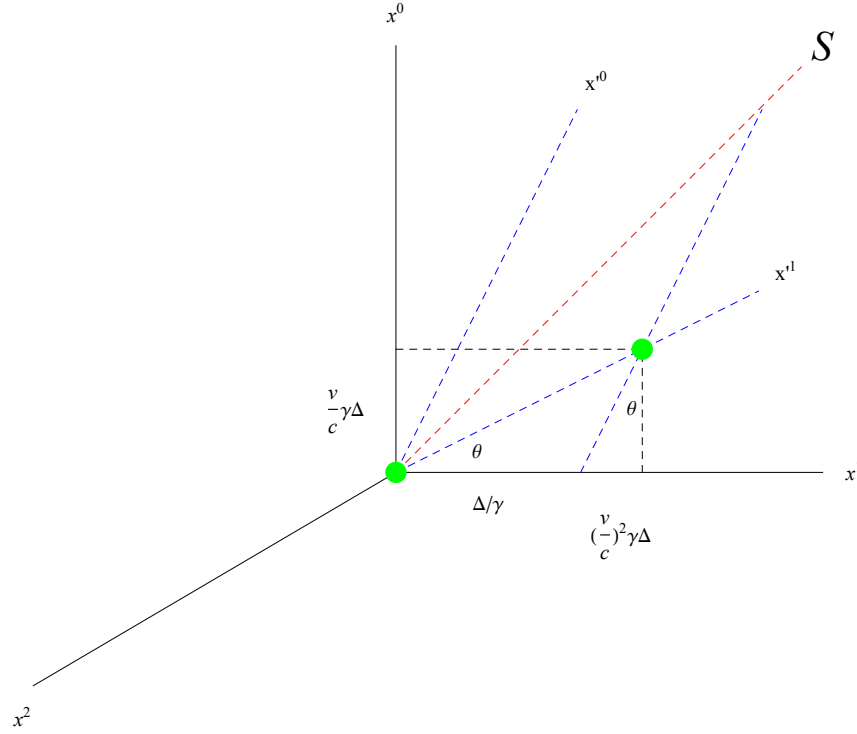
$$\Delta x' = \begin{pmatrix} 0 \\ \Delta \\ 0 \\ 0 \end{pmatrix}, \quad \Delta x = \Lambda \Delta x' = \begin{pmatrix} (v/c)\gamma\Delta \\ \gamma\Delta \\ 0 \\ 0 \end{pmatrix}, \quad (3.11.3)$$

while the invariant separation is

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

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

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


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

which is just the expected *contracted* length.

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

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

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

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

Lightlike separation, $\hat{e}^1$ direction

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

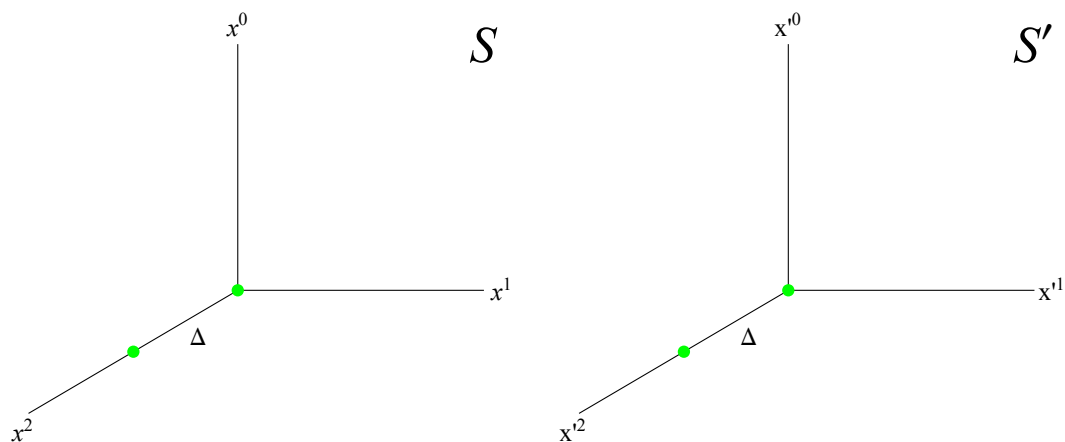


Figure 3.14: Spacelike separation (in x^2 direction) in S' frame (and S frame).

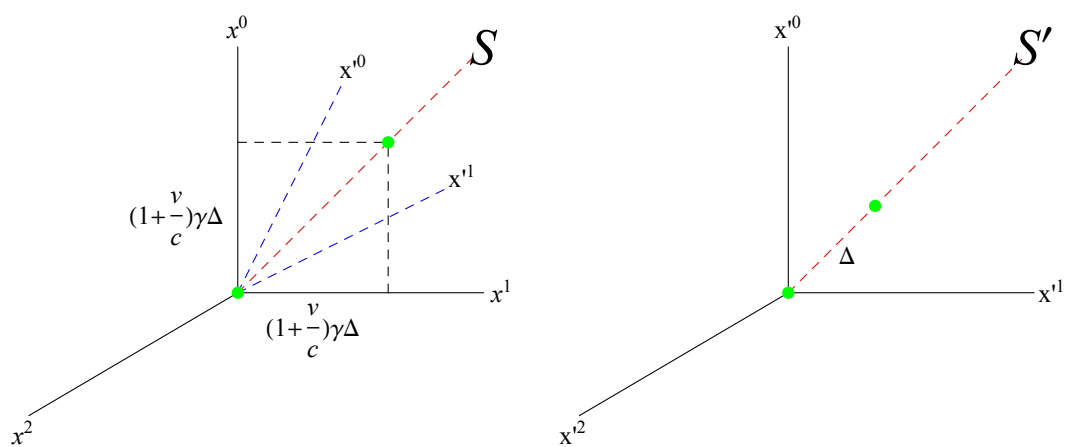


Figure 3.15: Lightlike separation (in x^0 and x^1 direction) in S' frame.

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

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

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

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

Lightlike separation, $\hat{e}^2$ direction

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

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

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

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

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