# The Special Theory of Relativity

Albert Einstein

(1879 - 1955)

November 9, 2001

## Contents

In this chapter we depart temporarily from the study of electromagnetism to explore Einstein's special theory of relativity. One reason for doing so is that Maxwell's field equations are inconsistent with the tenets of "classical" or "Galilean" relativity. After developing the special theory, we will apply it to both particle kinematics and electromagnetism and will find that Maxwell's equations are completely consistent with the requirements of the special theory.

# 1 Einstein's Two Postulates

Physical phenomena may be observed and/or described relative to any of an infinite number of "reference frames;" we regard the reference frame as being that one relative to which the measuring apparatus is at rest. The basic claim (or postulate) of relativity, which predates Einstein's work by many centuries, is that physical phenomena should be unaffected by the choice of the frame from which they are observed. This statement is quite vague. A simple explicit example is a collision of two objects. If they are seen to collide when observed from one frame, then the postulate of relativity says that they will be seen to collide no matter what reference frame is used to make the observation.

## 1.1 Galilean Invariance

Given that one believes some version of the postulate of relativity, then that person should, when constructing an explanation of the phenomena in question, make a theory which will predict the same phenomena in all reference frames. The original great achievements of this kind were Newton's theories of mechanics and gravitation. Consider, for example, $\mathbf{F} = m\mathbf{a}$. If the motion of some massive object is observed relative to two different reference frames, the motion will obey this equation in both frames provided the frames themselves are not being accelerated. This qualification leads one to restrict the statement of the relativity principle to unaccelerated or *inertial* reference frames.

In order to test the postulate of relativity, one needs a transformation that makes it possible to translate the values of physical observables from one frame to another. Consider two frames $K$ and $K'$ with $K'$ moving at velocity $\mathbf{v}$ relative to $K$.

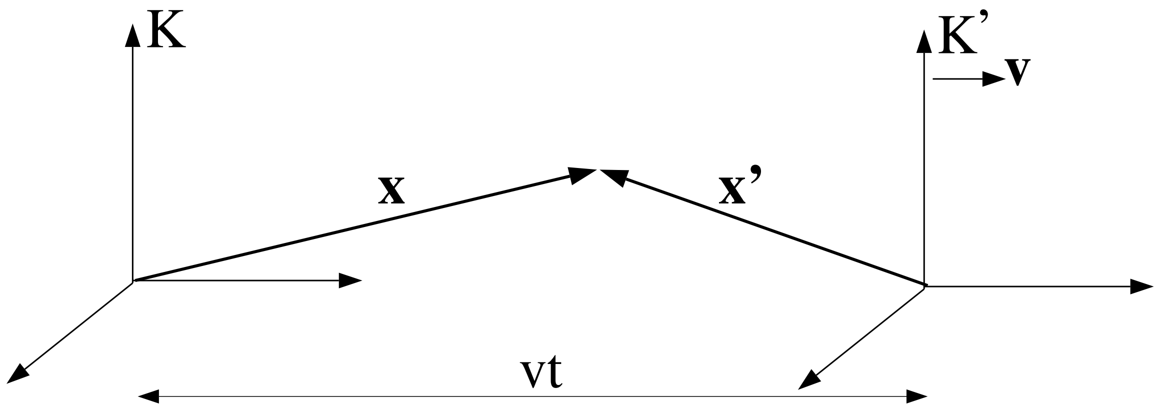


*Figure 1: Inerital frames $K$ and $K'$*

Then the (almost obvious) way to relate a space-time point $(t, \mathbf{x})$ in $K$ to the same point $(t', \mathbf{x}')$ in $K'$ is via the *Galilean transformation*

$$\mathbf{x}' = \mathbf{x} - \mathbf{v}t \qquad \text{and} \qquad t' = t, \tag{1}$$

or so it was believed up to the time of Einstein. Notice that the transformation is written so that the (space) origins coincide at $t = t' = 0$; we shall say simply that the origins (in space and time) coincide.

In what sense is Newton's law of motion consistent with the Galilean transformation? If his equation satisfies the postulate of relativity, then the motion of a massive object must obey it in both frames; thus

$$\mathbf{F} = m\mathbf{a} \qquad \text{and} \qquad \mathbf{F}' = m'\mathbf{a}' \tag{2}$$

where primed quantities are measured in $K'$ and unprimed ones in $K$. Now, experiments demonstrate (not quite correctly) that the force and mass are invariants, meaning that they are the same in all inertial frames, so if Newton's law is to hold in all inertial frames, then it must be the case that $\mathbf{a} = \mathbf{a}'$. The Galilean transformation provides a way of comparing these two quantities. In Eqs. (1), let $\mathbf{x}$ and $\mathbf{x}'$ be the positions of the mass at, respectively, times $t$ and $t'$ in frames $K$ and $K'$. Then we have

$$\frac{d\mathbf{x}'}{dt'} = \frac{d\mathbf{x}}{dt} - \mathbf{v} \tag{3}$$

and

$$\mathbf{a}' = \frac{d^2\mathbf{x}'}{dt'^2} = \frac{d^2\mathbf{x}}{dt^2} = \mathbf{a}, \tag{4}$$

assuming $\mathbf{v}$ is a constant. Thus we find that Newton's law, Galileo's transformation, and the observed motions of massive objects are consistent. [1].

## 1.2 The difficulty with Galilean Invariance

Now we come to the dilemma posed by Maxwell's equations. They are not consistent with the postulate of relativity if one uses the Galilean transformation to relate quantities in two different inertial frames. Imagine the quandary of the late-nineteenth-century physicist. He had the Galilean transformation and Newton's equations of motion, backed by enormous experimental evidence, to support the almost self-evident principle of relativity. But he also had the new - and enormously successful - Maxwell theory of light which was not consistent with Galilean relativity. What to do? One possible way out of the morass was easy to find. It was well-known that wavelike phenomena, such as sound, obey wave equations which are not properly "invariant" under Galilean transformations. The reason is simple: These waves are vibrational motions of some medium such as air or water, and this medium will be in motion with different velocities relative to the coordinate axes of different inertial frames. If one understood this, then one could see that although the wave equation takes on different forms relative to different frames, it did correctly describe what goes on in every frame and was not inconsistent with the postulate of relativity.

The appreciation of this fact set off a great search to find the medium, called the "luminiferous ether" or simply the *ether*, whose vibrations constitute electromag-

[1] Of course, they aren't consistent at all if one either makes measurements of extraordinary precision or studies particles traveling at an appreciable fraction of the speed of light. Neither of these things was done prior to the twentieth century.

netic waves. The search (i.e. Michelson and Morely) was, as we know, completely unsuccessful[2], as the ether eluded all seekers.

However, for Einstein, it was the Fizeau experiment (1851) which convinced him that the ether explaination was incorrect. This experiment looked for a change in the phase velocity of light due to its passage through a moving medium, in this case water.

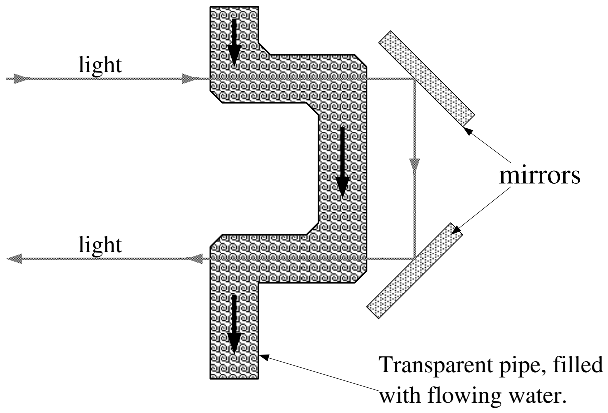


*Figure 2: Diagram of Fizeau experiment*

Fitzeau found that this phase velocity was given by

$$v_{phase} = \frac{c}{n} \pm v\left(1 - \frac{1}{n^2}\right) \quad \text{experiment}$$

where $n$ is the index of refraction of the water, and $v$ is its velocity. The plus(minus) sign is taken if the water is moving with(against) the light.

Lets analyze the experiment from a Galilean point of view. The dielecric water is moving in either the same or opposite direction as the light, and so acts as a moving source for the light with is refracted (i.e. reradiated by the water molecules). Nonrelativistically, we just add the velocity $v$ of the source to the wave velocity for the stationary source. Thus Galilean therory says

$$v_{phase} = \frac{c}{n} \pm v \quad \text{Galilean theory,}$$

[2] Or completely successful, if we adopt a somewhat different (Einstein's) point of view.

which is clearly inconsistent with experiment.

The stage was now set for Einstein who, in 1905, made the following postulates:

1. Postulate of relativity: The laws of nature and the results of all experiments performed in a given frame of reference are independent of the translational motion of the system as a whole.

2. Postulate of the constancy of the speed of light: The speed of light is independent of the motion of its source.

The first postulate essentially reaffirmed what had long been thought or believed in the specific case of Newton's law, extending it to all phenomena. The second postulate was much more radical. It did away with the ether at a stroke and also with Galilean relativity because it implies that the speed of light is the same in all reference frames which is fundamentally inconsistent with the Galilean transformation.

# 2 Simultaneity, Separation, Causality, and the Light Cone

## 2.1 Simultaneity

The second postulate - disturbing in itself - leads to many additional "nonintuitive" predictions. For example, suppose that there are sources of light at points A and C and that they both emit signals that are observed by someone at B which is midway between A and C.

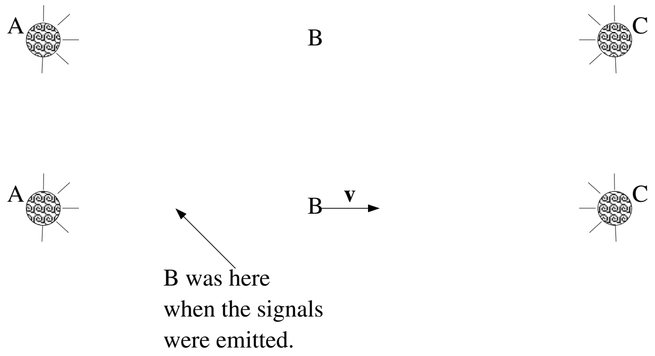


*Figure 3: Simultaneity depends upon the rest frame of the observer*

If he sees the two signals simultaneously and knows that he is equidistant from the sources, he will conclude quite correctly that the signals were emitted simultaneously. Now suppose that there is a second observer, B',who is moving along the line from A to C and who arrives at B just when the signals do. He will know that the signals were emitted at some earlier time when he was closer to A than to C. Also, since both signals travel with the same speed $c$ in his rest frame (because the speeds of the signals relative to him are independent of the speeds of the sources relative to him), he will conclude that the signal from C was emitted earlier than that from A because it had to travel the greater distance before reaching him. He is as correct as the first observer. Similarly, an observer moving in the opposite direction relative to the first one will conclude from the same reasoning that the signal from A was emitted before that from C. *Hence Einstein's second postulate leads us to the conclusion that events, in this case the emission of light signals, which are simultaneous in one inertial frame are not necessarily simultaneous in other inertial frames.*

## 2.2 Separation and Causality

If simultaneity is only a relative fact, as opposed to an absolute one, what about causality? Because the order of the members of some pairs of events can be reversed by changing one's reference frame, we must consider whether the events' ability to influence each other can similarly be affected by a change of reference frame. This question is closely related to a quantity that we shall call the *separation* between the events. Given two events A and B which occur at space-time points $(t_1, \mathbf{x}_1)$ and $(t_2, \mathbf{x}_2)$, we define the squared separation $s_{12}^2$ between them to be

$$s_{12}^2 \equiv c^2(t_1 - t_2)^2 - |\mathbf{x}_1 - \mathbf{x}_2|^2. \tag{5}$$

Let the two events be (1) the emission of an electromagnetic signal at some point in vacuum and (2) its reception somewhere else. Then, because the signal travels with the speed $c$, these events have separation zero, $s_{12}^2 = 0$. This result will be the same in any inertial frame since the signal has the same speed $c$ in all such frames.

Now, if we have two events such that $s_{12}^2 > 0$, then we have a "causal relationship" in the sense that a light signal can get from the first event to the place where the second one occurs before it does occur. Such a separation is called *timelike*. On the other hand, if $s_{12}^2 < 0$, then a light signal cannot get from the first event to the location of the second event before the second event occurs. This separation is called *spacelike*. A separation $s_{12}^2 = 0$ is called *lightlike*.

It is important to ask whether there is some other type of signal that travels faster than $c$ and which could therefore produce a causal relationship between events with a spacelike separation. None has been found and we shall assume that none exists. Consequently, we claim that events with a timelike separation are such that the earlier one can influence the later one, because a signal can get from the first to the location of the second before the latter occurs, but that events with a spacelike separation are such that the earlier one cannot influence the later one because a signal cannot get from the first event to the location of the second one fast enough.

## 2.3 The Light Cone

The question now is whether the character of the separation between two events, timelike, spacelike, or lightlike, can be changed by changing the frame in which it is measured. For simplicity, let the "first" event, A, occur (in frame $K$) at $(t = 0, \mathbf{x} = 0)$ while the second takes place at some general $(t, \mathbf{x})$ with $t > 0$. Further, let $ct$ be larger than $|\mathbf{x}|$ so that $s^2 > 0$ and A may influence B. Consider these same two events in another frame $K'$. By an appropriate choice of the origin (in space and time) of this frame, we can make the first event occur here, just as it does in frame $K$. The second event will be at some (t',$\mathbf{x}$').

We can picture the relative positions of the two events in space and time by using a *light cone* as shown.

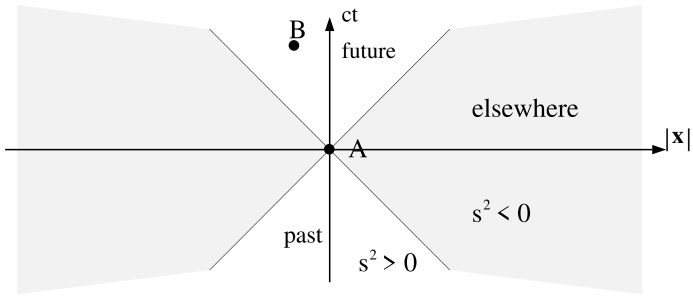


*Figure 4: The Light Cone*

The vertical axis measures $ct$; the horizontal one, separation in space, $|\mathbf{x}|$. The two diagonal lines have slopes $\pm 1$. The event B is shown within the cone whose axis is the $ct$ axis; any event with a timelike separation relative to the origin will be in here.

The question we wish to ask now is whether, by going to another reference frame, one may cause event B to move across one of the diagonal lines and so wind up in a place where it cannot be influenced by the event at the origin? The point is that A can influence any event inside of the "future" cone; it can be influenced by any event inside of the "past" cone; but it cannot influence, or be influenced by, any event inside of the "elsewhere" region. If an event B and two reference frames $K$ and $K'$ can be found such that the event when expressed in one frame is on the opposite side of a diagonal from where it is in the other frame, then we have made causality a frame-dependent concept.

Suppose that we have two such frames. We can effect a transformation from one to the other by considering a sequence of many frames, each moving at a velocity only slightly different from the previous one, and such that the first frame in the sequence is $K$ and the final one is $K'$.

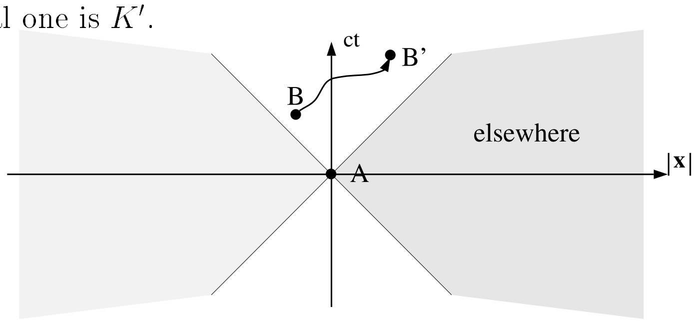


*Figure 5: Evolution of the event B as we evolve from frame $K$ to $K'$*

If we let this become an infinite sequence with infinitesimal differences in the velocities of two successive members of the sequence, then the positions of the event B in the sequence of light cones for each of the frames will form a continuous curve when expressed in a single light cone. If B crosses from timelike to spacelike in this sequence, then at some point, B must lie on one of the diagonals. For such an event, the separation from A is lightlike, or zero.

Consider now two events with zero separation $s_{12}^2 = 0$. These events can be coincident with the emission and reception of a light signal. But these events must be coincident with the emission and reception of the light signal in all frames, by the postulate of relativity (the first postulate), and so these events must have zero separation in all inertial frames because of the constancy of the speed of light in all frames. Consequently, what we are trying to do above is in fact impossible; that is, one cannot move an event such as B onto or off of the surface of the light cone by looking at it in a different reference frame, and for this reason, one cannot make it cross the surface of the light cone. An event in the future will be there in all reference frames. One in the past cannot be taken out of the past; and one that is "elsewhere" will be there in any reference frame.

## 2.4 The invariance of Separation

With a little more thought we can generalize the conclusion of the previous paragraph that two events with zero separation in one frame have zero separation in all frames. In fact, the separation, whatever it may be, between any two events is the same in all frames. We shall call something that is the same in all frames an *invariant*; the separation is an invariant. To argue that this should be the case, suppose that we have two events which are infinitesimally far apart in both space and time so that we may write $ds_{12}$ for $s_{12}$,

$$(ds_{12})^2 = c^2(t_1 - t_2)^2 - |\mathbf{x}_1 - \mathbf{x}_2|^2 \tag{6}$$

in frame $K$. In another inertial frame $K'$ we have separation $(ds'_{12})^2$, and we have argued that this is zero if $ds_{12}^2$ is zero. If $K'$ is moving with a small speed relative to $K$, the separations in the two frames must be nearly equal which means that they will be infinitesimal quantities of the same order, or

$$(ds_{12})^2 = A(ds'_{12})^2 \tag{7}$$

where $A = A(v)$ is a finite function of $v$, the relative speed of the frames. Furthermore, $A(0) = 1$ since the two frames are the same if $v = 0$. Now, if time and space are homogeneous and isotropic, then it must also be true that

$$(ds'_{12})^2 = A(v)(ds_{12})^2. \tag{8}$$

Comparing the preceding equations, we see that the only solutions are $A(v) = \pm 1$; the condition that $A(0) = 1$ means $A(v) = 1$. Hence

$$(ds'_{12})^2 = (ds_{12})^2 \tag{9}$$

which is a relation between differentials that may be integrated to give

$$(s'_{12})^2 = (s_{12})^2, \tag{10}$$

thereby demonstrating that the separation between any two events is an invariant. The locus of all points with a given separation is a hyperbola when drawn on a light cone (or a hyperboloid of revolution if more spatial dimensions are displayed in the light cone).

# 3 Proper time

Another related invariant is the so-called *proper time*. This is tied to a particular object and is the time that elapses in the rest frame of that object. If the object is accelerated, its rest frame in not an inertial frame.

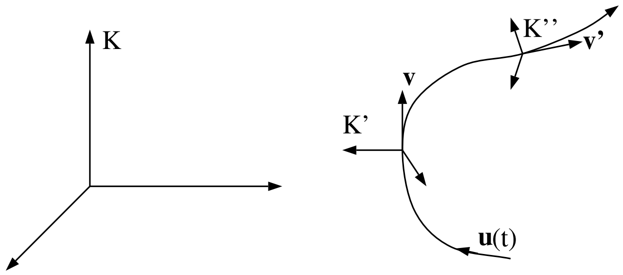


*Figure 6: Instaneous rest frames $K'$ and $K''$ of an object with velocity* $\mathbf{u}(t)$ *as measured in $K$*

It is then useful to make use of the "instantaneous" rest frame of the object, meaning an inertial frame relative to which the object is not moving at a particular instant of time. Thus, if in frame $K$ the object has a velocity $\mathbf{u}(t)$, its instantaneous rest frame at time $t$ is a frame $K'$ which moves at velocity $\mathbf{v} = \mathbf{u}(t)$. One may find the object's proper time by calculating the time that elapses in an infinite sequence of instantaneous rest frames.

Consider an object moving with a trajectory $\mathbf{x}(t)$ relative to frame $K$. Between $t$ and $t + dt$ it moves a distance $d\mathbf{x}$ as measured in $K$. Let us ask what time $dt'$ elapses in the frame $K'$ which is the instantaneous rest frame at time $t$. The one thing we know is that

$$(ds)^2 \equiv c^2(dt)^2 - (d\mathbf{x})^2 = (ds')^2 = c^2(dt')^2 - (d\mathbf{x}')^2 \tag{11}$$

where, as usual, unprimed quantities are the ones measured relative to $K$ and primed ones are measured in $K'$. Now, $d\mathbf{x}' = 0$[3] because the object is at rest in $K'$ at time $t$. Hence we may drop this contribution to the (infinitesimal) separation and solve for $dt'$:

$$dt' = \sqrt{(dt)^2 - (d\mathbf{x})^2/c^2} = dt\sqrt{1 - \frac{1}{c^2}\left(\frac{d\mathbf{x}}{dt}\right)^2} = dt\sqrt{1 - u^2/c^2} \tag{12}$$

where $\mathbf{u} \equiv d\mathbf{x}/dt$ is the object's velocity as measured in $K$. Now we may integrate from some initial time $t_1$ to a final time $t_2$ to find the proper time of the object which elapses while time is proceeding from $t_1$ to $t_2$ in frame $K$; that is, we are adding up all of the time that elapses in an infinite sequence of instantaneous rest frames of the object while time is developing in $K$ from $t_1$ to $t_2$.

$$\tau_2 - \tau_1 = \int_{t_1}^{t_2} dt\,\sqrt{1 - u^2(t)/c^2}. \tag{13}$$

## 3.1 Proper Time of an Oscillating Clock

As an example let the object move along a one-dimensional path with $u(t) = c\sin(2\pi t/t_0)$ with $t_1 = 0$ and $t_2 = t_0$. This velocity describes a round trip of a harmonic oscillator with a peak speed of $c$ and a period of $t_0$. The corresponding elapsed proper time is

$$\tau_0 = \int_0^{t_0} dt\,\sqrt{1 - \sin^2(2\pi t/t_0)} = \int_0^{t_0} dt\,|\cos(2\pi t/t_0)| = 2t_0/\pi. \tag{14}$$

This is smaller than $t_0$ by a factor of $2/\pi$ which means that a clock carried by the object will show an elapsed time during the trip which is just $2/\pi$ times what a clock which remains in frame $K$ will show, provided the acceleration experienced by the clock which makes the trip doesn't alter the rate at which it runs.

[3] More correctly $d\mathbf{x}'$ is a second order differential, and hence may be neglected

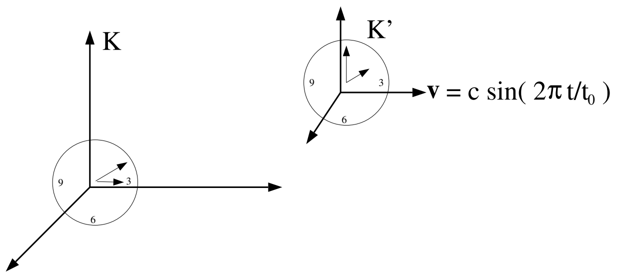


*Figure 7: The proper time for the oscillating frame is $2t/\pi$; which is less than the elapsed time in frame $K$.*

If this clock is a traveller, then the traveller ages during the trip by an amount which is only $2/\pi$ of the amount by which someone who stays at rest in $K$ ages. One may wonder whether, from the point of view of the traveller, the one who stayed at home should be the one who ages more "slowly." If the calculation is done carefully (correctly), one finds that the same conclusion is reached; the traveller has in fact aged less that the stay-at-home.

# 4 Lorentz Transformations

## 4.1 Motivation

So far we know the locations $(t, \mathbf{x})$ and $(t', \mathbf{x}')$ of a space-time point as given in $K$ and $K'$ must be related by

$$c^2t^2 - \mathbf{x} \cdot \mathbf{x} = c^2t'^2 - \mathbf{x}' \cdot \mathbf{x}', \tag{15}$$

given that the origins of the coordinate and time axes of the two frames coincide. This equation looks a lot like the statement that the inner product of a four-dimensional vector, having components $ct$ and $i\mathbf{x}$, with itself is an invariant. It suggests that the transformation relating $(t, \mathbf{x})$ and $(t', \mathbf{x}')$ is an orthogonal transformation in the four-dimensional space of $ct$ and $\mathbf{x}$. There is an unusual feature in that the transformation apparently describes an imaginary or complex rotation because the inner product, or length, that is preserved is $c^2t^2 - \mathbf{x} \cdot \mathbf{x}$ as opposed to $c^2t^2 + \mathbf{x} \cdot \mathbf{x}$. Recall that a rotation in three dimensions around the $\boldsymbol{\epsilon_3}$ direction by angle $\phi$ can be represented by a matrix

$$a = \begin{pmatrix} \cos\phi & -\sin\phi & 0 \\ \sin\phi & \cos\phi & 0 \\ 0 & 0 & 1 \end{pmatrix} \tag{16}$$

so that

$$x_i' = \sum_{j=1}^{3} a_{ij} x_j; \tag{17}$$

that is,

$$\begin{aligned} x_1' &= \cos\phi\, x_1 - \sin\phi\, x_2 \\ x_2' &= \sin\phi\, x_1 + \cos\phi\, x_2 \\ x_3' &= x_3. \end{aligned} \tag{18}$$

For an imaginary $\phi$, $\phi = i\eta$, $\cos\phi \to \cosh\eta$, and $\sin\phi \to -i\sinh\eta$. Further, let us reconstruct the vector as $\mathbf{y} = (x_1, ix_2, ix_3)$ and make the transformation

$$\mathbf{y}_i' = \sum_j a_{ij} y_j. \tag{19}$$

The result, expressed in terms of components of $\mathbf{x}$, is

$$\begin{aligned} x_1' &= \cosh\eta\, x_1 - \sinh\eta\, x_2 \\ x_2' &= -\sinh\eta\, x_1 + \cosh\eta\, x_2 \\ x_3' &= x_3; \end{aligned} \tag{20}$$

these are such that

$$x_1'^2 - x_2'^2 - x_3'^2 = x_1^2 - x_2^2 - x_3^2 \tag{21}$$

since $\cosh^2(\eta) - \sinh^2(\eta) = 1$, so we have succeeded in constructing a transformation that produces the right sort of invariant. All we have to do is generalize to four dimensions.

## 4.2 Derivation

Let's begin by introducing a vector with four components, $(x_0, x_1, x_2, x_3)$ where $x_0 = ct$ and the $x_i$ with $1 = 1, 2, 3$ are the usual Cartesian components of the position vector. Then introduce $\vec{y} \equiv (x_0, i\mathbf{x})$ which has the property that

$$\vec{y} \cdot \vec{y} = x_0^2 - \mathbf{x} \cdot \mathbf{x}. \tag{22}$$

This inner product is supposed to be an invariant under the transformation of $\mathbf{x}$ and $t$ that we seek. The transformation in question is a rotation through an imaginary angle $i\eta$ that mixes time and one spatial direction, which we pick to be the first ($y_1$ or $x_1$) without loss of generality. The matrix representing this rotation is

$$a = \begin{pmatrix} \cosh\eta & i\sinh\eta & 0 & 0 \\ -i\sinh\eta & \cosh\eta & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}. \tag{23}$$

Now operate with this matrix on $\vec{y}$ to produce $\vec{y}'$. If we write the components of the latter as $(x_0', i\mathbf{x}')$, we find the following:

$$\begin{aligned} x_0' &= \cosh\eta\, x_0 - \sinh\eta\, x_1 \\ x_1' &= -\sinh\eta\, x_0 + \cosh\eta\, x_1 \\ x_2' &= x_2 \\ x_3' &= x_3. \end{aligned} \tag{24}$$

It is a simple matter to show from these results that $x_0^2 - \mathbf{x} \cdot \mathbf{x}$ is an invariant, i.e.,

$$x_0^2 - \mathbf{x} \cdot \mathbf{x} = x_0'^2 - \mathbf{x}' \cdot \mathbf{x}' \tag{25}$$

which means we have devised an acceptable transformation in the sense that it preserves the separation between two events.

But what is the significance of $\eta$? Let us rewrite $\sinh\eta$ as $\cosh\eta \tanh\eta$. Then we have, in particular,

$$\begin{aligned} x_0' &= \cosh\eta\,(x_0 - \tanh\eta\, x_1) \\ x_1' &= \cosh\eta\,(x_1 - \tanh\eta\, x_0). \end{aligned} \tag{26}$$

The second of these is

$$x'_1 = \cosh\eta\,(x_1 - \tanh\eta\, ct). \tag{27}$$

Suppose that we are looking at an object at rest at the origin of $K'$, and the space-time point $(t, \mathbf{x})$ is this object's location. Then $x'_1 = 0$ for all $t'$. As seen from $K$, the object is at $x_1 - vt$ given that $\mathbf{v}$, the velocity of the object (and of $K'$) relative to $K$, is parallel to $\boldsymbol{\epsilon_1}$.

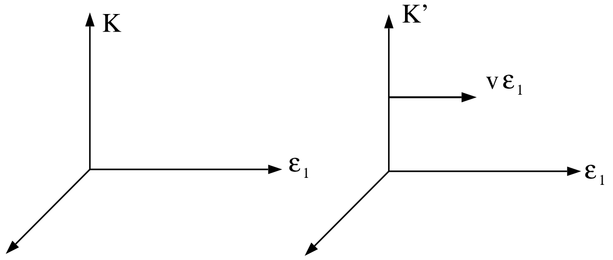


*Figure 8: Coordinates for our Lorentz transform.*

This is consistent with Eq. (27) provided

$$\tanh\eta \equiv \frac{v}{c} \equiv \beta \tag{28}$$

where $\beta$ is defined as the speed $v$ relative to the speed of light. From this relation, we find further that

$$\cosh\eta = 1/\sqrt{1-\beta^2} \equiv \gamma; \tag{29}$$

this expression defines $\gamma$, a parameter that comes up repeatedly in the special theory of relativity.

Our determination of the transformation, called the *Lorentz transformation*[4], is now complete. We find that, given a frame $K'$ moving at velocity $\mathbf{v} = v\boldsymbol{\epsilon_1}$ relative to

[4]H. A. Lorentz devised these transformations prior to Einstein's development of the special theory of relativity; they had in fact been used even earlier by Larmor and perhaps others. Furthermore, it was known that Maxwell's equations were invariant under these transformations, meaning that if these are the right transformations (as opposed to the Galilean transformations), Maxwell's equations are eligible for "law of nature" status.

$K$, a space-time point $(t, \mathbf{x})$ in $K$ becomes, in $K'$, the space-time point $(t', \mathbf{x}')$ with

$$x_0' = \gamma(x_0 - \beta x_1) \qquad x_1' = \gamma(x_1 - \beta x_0) \qquad x_2' = x_2 \qquad x_3' = x_3. \tag{30}$$

The inverse transformation can be extracted from these equations in a straightforward manner; it may also be inferred from the fact that $K$ is moving at velocity $-\mathbf{v}$ relative to $K'$ which tells us immediately that

$$x_0 = \gamma(x_0' + \beta x_1') \qquad x_1 = \gamma(x_1' + \beta x_0') \qquad x_2 = x_2' \qquad x_3 = x_3'. \tag{31}$$

## 4.3 Elapsed Proper Time Revisited

Let us try to use this transformation to calculate something. First, we revisit the proper time. For an object at rest in $K'$, $\mathbf{x}'$ does not change with time. Also, from our transformation,

$$ct = \gamma(ct' + \beta x_1'), \tag{32}$$

The differential of this transformation, making use of the fact that the object is instantaneously at rest in $K'$, gives, $dt = \gamma dt'$ since $dx_1'$ is second-order in powers of $dt'$. Stated in another fashion, we are considering the transformation of two events or space-time points. They are the locations of the object at times $t'$ and $t' + dt'$. Because the object is at rest in $K'$ at time $t'$, its displacement $d\mathbf{x}'$ during the time increment $dt'$ is of order $(dt')^2$ and so may be discarded. The corresponding elapsed time $dt$ in $K$ is thus found to be $dt = \gamma dt'$, using the Lorentz transformations of the two space-time points. This equation may also be written as

$$dt' = dt/\gamma = \sqrt{1 - v^2/c^2}dt. \tag{33}$$

The left-hand side of this equation is the elapsed proper time of the object while $dt$ is the elapsed time measured by observers at rest relative to $K$. If we introduce $\mathbf{u}$, the velocity of the object relative to $K$, and notice that $\mathbf{u} = \mathbf{v}$ at time t, then we can write $dt'$ in terms of $\mathbf{u}(t)$ as

$$dt' = \sqrt{1 - |\mathbf{u}(t)|^2/c^2}dt, \tag{34}$$

where now $dt'$ in the elapsed proper time of the object which moves at velocity $\mathbf{u}$ relative to frame $K$. We can integrate this relation to find the finite elapsed proper time during an arbitrary time interval (in $K$),

$$\tau_2 - \tau_1 = \int_{t_1}^{t_2} dt\sqrt{1 - |\mathbf{u}(t)|^2/c^2}. \tag{35}$$

## 4.4 Proper Length and Length Contraction

Next, we shall examine the *Fitzgerald-Lorentz contraction.* Define the *proper length* of an object as its length, measured in the frame where it is at rest. Let this be $L_0$, and let the rest frame be $K'$, moving at the usual velocity ($\mathbf{v} = v\boldsymbol{\epsilon}_\mathbf{1}$) relative to $K$.

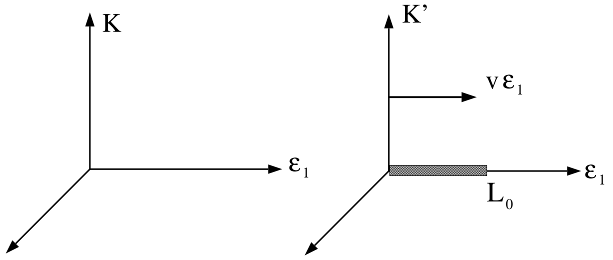


*Figure 9: Length contraction occurs along the axis parallel to the velocity.*

The relative geometry is shown in the figure. The rod, or object, is positioned in $K'$ so that its ends are at $x_1' = 0$, $L_0$. They are there for all $t'$. In order to find the length of the rod in $K$, we have to measure the positions of both ends at the **same** time $t$ as measured in $K$. We can find the results of these measurements from the Lorentz transformation

$$x_1' = \gamma(x_1 - \beta x_0). \tag{36}$$

Use this relation first with $x_1'$ equal to 0 and then with $x_1' = L_0$, using the same time $x_0$ in both cases, and take the difference of the two equations so obtained. The result is

$$L_0 = \gamma(x_{1R} - x_{1L}) \equiv \gamma L \tag{37}$$

where $x_{1R}$ and $x_{1L}$ are the positions of the right and left ends of the rod at some particular time, or $x_0$. The difference of these is $L$, the length of the rod as measured in frame $K$.

Our result for $L$ can be written as

$$L = L_0/\gamma = \sqrt{1-\beta^2}L_0. \tag{38}$$

This length is smaller than $L_0$ which means that the object is found (is measured) in $K$ to be shorter than its proper length or its length in the frame where it is at rest. Notice, however, that if we did the same calculation for its length in a direction perpendicular to the direction of $\mathbf{v}$, we would find that it is the same in $K$ as in $K'$. Consequently the transformation of the object's volume is

$$V = V_0/\gamma = \sqrt{1-\beta^2}\, V_0 \tag{39}$$

where $V_0$ is the *proper volume* or volume in the rest frame, and $V$ is the volume in a frame moving at speed $\beta c$ relative to the rest frame.

# 5 Transformation of Velocities

Because we know how $\mathbf{x}$ and $t$ transform, we can determine how anything that involves functions of these things transforms. For example, velocity. Let an object have velocity $\mathbf{u}$ in $K$ and velocity $\mathbf{u}'$ in $K'$ and let $K'$ move at velocity $\mathbf{v}$ relative to $K$. We wish to determine how $\mathbf{u}'$ is related to $\mathbf{u}$. In $K'$, the object moves a distance $d\mathbf{x}' = \mathbf{u}'dt'$ in time $dt'$. A similar statement, without any primed quantities, holds in $K$. The infinitesimal displacements in time and space are related by Lorentz transformations:

$$dt = \gamma(v)\left(dt' + (v/c^2)dx'\right) \qquad dx = \gamma(v)(dx' + vdt') \qquad dy = dy' \qquad dz = dz', \tag{40}$$

where we have let $\mathbf{v}$ be along the direction of $x_1$ as usual. Taking ratios of the displacements to the time increment, we have

$$u_x = \frac{dx}{dt} = \frac{dx' + vdt'}{dt' + (v/c^2)dx'} = \frac{dx'/dt' + v}{1 + (v/c^2)(dx'/t')} = \frac{u'_x + v}{1 + vu'_x/c^2}, \tag{41}$$

$$u_y = \frac{1}{\gamma(v)} \frac{u'_y}{1 + vu'_x/c^2}, \tag{42}$$

and

$$u_z = \frac{1}{\gamma(v)} \frac{u'_z}{1 + vu'_x/c^2}. \tag{43}$$

These results may be summarized in vectorial form:

$$\mathbf{u}_{||} = \frac{u'_{||} + \mathbf{v}}{1 + \mathbf{v} \cdot \mathbf{u}'/c^2} \qquad u_{\perp} = \frac{u'_{\perp}}{\gamma(v)(1 + \mathbf{v} \cdot \mathbf{u}'/c^2)} \tag{44}$$

where the subscripts "$||$" and "$\perp$" refer respectively to the components of the velocities $\mathbf{u}$ and $\mathbf{u}'$ parallel and perpendicular to $\mathbf{v}$. Notice too that

$$\mathbf{u}_{||} = \left(\frac{\mathbf{u} \cdot \mathbf{v}}{v^2}\right) \mathbf{v} \quad \text{and} \quad \mathbf{u}_{\perp} = \mathbf{u} - \mathbf{u}_{||} = \mathbf{u} - \left(\frac{\mathbf{u} \cdot \mathbf{v}}{v^2}\right) \mathbf{v}. \tag{45}$$

It is sometimes useful to express the transformations for velocity in polar coordinates $(u, \theta)$ and $(u', \theta')$ such that

$$u_{||} = u \cos\theta \quad \text{and} \quad u_{\perp} = u \sin\theta, \tag{46}$$

etc.; the appropriate expressions are

$$\tan\theta = \frac{u' \sin\theta'}{\gamma(v)(u' \cos\theta' + v)} \quad \text{and} \quad u = \frac{[u'^2 + v^2 + 2u'v\cos\theta' - (vu'\sin\theta'/c)^2]^{1/2}}{1 + u'v\cos\theta'/c^2}. \tag{47}$$

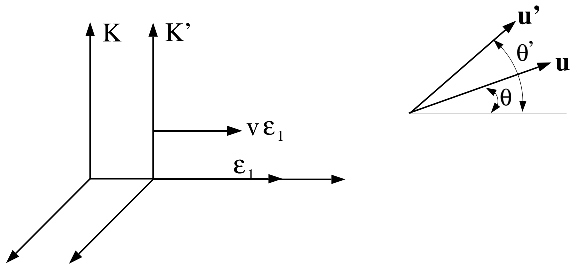


*Figure 10: Frames for velocity transform.*

The inverses of all of these velocity transformations are easily found by appropriate symmetry arguments based on the fact that the velocity of $K$ relative to $K'$ is just $-\mathbf{v}$.

In it interesting that the velocity transformations are, in contrast to the ones for $\mathbf{x}$ and $t$, nonlinear. They must be nonlinear because there is a maximum velocity which is the speed of light; combining two velocities, both of which are close to, or equal to, $c$, cannot give a velocity greater than $c$. A linear transformation would necessarily allow this to happen, so a nonlinear transformation is required. To see how the transformations rule out finding a frame where an object moves faster than $c$, let us consider the transformation of a velocity $|\mathbf{u}'| = c$. From the second of Eqs. (47), we see that

$$u = \frac{c^2 + v^2 + 2cv\cos\theta' - v^2(1-\cos^2\theta')]^{1/2}}{1 + v\cos\theta'/c} = c\frac{[(1 + v\cos\theta'/c)^2]^{1/2}}{1 + v\cos\theta'/c} = c. \tag{48}$$

Thus do we find what we already knew: If something moves at speed $c$ in one frame, then it moves at the same speed in any other frame. More generally, if we had used any $u' \le c$ and $v \le c$, we would have recovered a $u \le c$.

## 5.1 Aberration of Starlight

An interesting example of the application of the velocity transformation is the observed aberration of starlight. Suppose that an observer is moving with speed $v$ at right angles to the direction of a star that he is watching. If a Galilean transformation is applied to the determination of the apparent direction of the star, one finds that it is seen at an angle $\phi$ away from its true direction where $\tan\phi = v/c$.

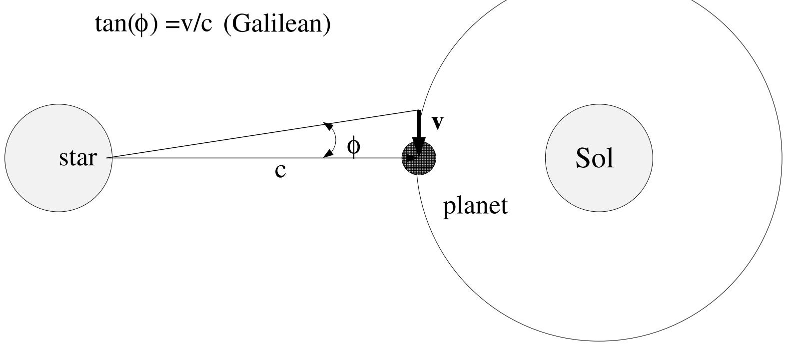

*Figure 11: Due to the finite velocity of light, a star is seen an angle $\phi$ away from its true direction.*

One can measure this angle by waiting six months. The velocity $\mathbf{v}$ is provided by the earth's orbital motion; six months later it is reversed and if the observer then looks for the same star, it position will have shifted by $2\phi$, at least according to the Galilean transformation.

But that prediction is not correct. Consider what happens if the Lorentz transformation is used to compute the angle $\phi$. Using Eq. (47), we see that the angle $\theta$ at which the light from the star appears to be headed in the frame $K$ of the observer is

$$\tan\theta = \frac{c\sin\theta'\sqrt{1-v^2/c^2}}{c(\cos\theta'+v/c)}. \tag{49}$$

where $\theta'$ is its direction in the frame $K'$ which is the rest frame of the sun.

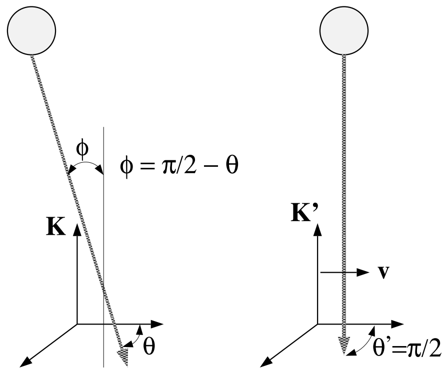


*Figure 12: Coordinates for stargazing.*

Now suppose that $\mathbf{v}$ is at $\pi/2$ radians to the direction of the light's motion in frame $K'$ so that $\theta'=\pi/2$. Then we find $\tan\theta = c/\gamma(v)v$. To compare with the prediction of the Galilean transformation, we need to find the angle $\phi$, which is to say, $\pi/2-\theta$. From a trigonometric identity, we have

$$\tan\theta = \frac{\tan(\pi/2)-\tan\phi}{1+\tan(\pi/2)\tan\phi} = \frac{1}{\tan\phi}, \tag{50}$$

and so

$$\tan\phi = v\gamma(v)/c \qquad \text{or} \qquad \sin\phi = v/c. \tag{51}$$

This is the correct answer; the tangent of the angle $\phi$ differs from the prediction of the Galilean transformation by a factor of $\gamma$ which is second-order in powers of $v/c$.

# 6 Doppler Shift

The Doppler shift of sound is a well-known and easy to understand phenomenon. It depends on the velocities of the source and observer relative to the medium in which the waves propagate. For electromagnetic waves, this medium does not exist and so the Doppler shift for light takes on its own special - and relatively simple! - form.

Suppose that in frame $K$ there is a plane wave with wave vector $\mathbf{k}$ and frequency $\omega$. Put an observer at some point $\mathbf{x}$ and set him to work counting wave crests as they go past him.

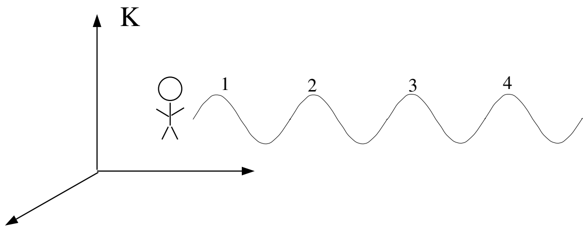


*Figure 13: An observer counts wave crests.*

Let him begin with the crest which passes the origin at $t = 0$ and continue counting until some later time $t$. How many crests does he count? We can decide by first determining when he starts. The starting time is $t_0 = (\mathbf{k} \cdot \mathbf{x})/kv_w$ where $v_w$ is the velocity of the wave in frame $K$. The observer counts from $t_0$ to $t$ and so counts $n$ crests where

$$n = (t - t_0)/T; \tag{52}$$

$T$ is the period of the wave, $T = 2\pi/\omega$. Hence,

$$n = \frac{1}{2\pi}\left(\omega t - \frac{\omega}{kv_w}\mathbf{k} \cdot \mathbf{x}\right) = \frac{1}{2\pi}(\omega t - \mathbf{k} \cdot \mathbf{x}), \tag{53}$$

since $\omega = v_w k$.

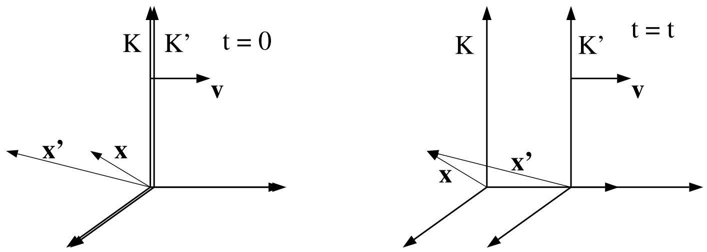


*Figure 14: Frame alignments for the Doppler problem.*

Now let the same measurement be performed by an observer at rest at a point $\mathbf{x}'$ in frame $K'$ which moves at $\mathbf{v}$ relative to $K$. We choose $\mathbf{x}'$ in a special way; it must be coincident with $\mathbf{x}$ at time $t$ (measured in $K$). This observer also counts crests, starting with the one that passed his origin at time $t' = 0$ and stopping with the one that arrives when he (the second observer) is coincident with the first observer. Given the usual transformations, the four-dimensional coordinate origins coincide, and so both observers count the same number of crests. Repeating the argument given for the number counted by the first observer, we find that the number counted by the second observer can be written as

$$n = \frac{1}{2\pi}(\omega' t' - \mathbf{k}' \cdot \mathbf{x}') \tag{54}$$

where $\omega'$ and $\mathbf{k}'$ are the frequency and wave vector of the wave in $K'$ and $(t', \mathbf{x}')$ is the spacetime point that transforms into $(t, \mathbf{x})$. Thus we find

$$\omega t - \mathbf{k} \cdot \mathbf{x} = \omega' t' - \mathbf{k}' \cdot \mathbf{x}'. \tag{55}$$

The significance of this relation is that the phase of the wave is an invariant. Further it appears to be the inner product of $(ct, i\mathbf{x})$ and $(\omega/c, i\mathbf{k})$. Because we know how $(t, \mathbf{x})$ transforms to $(t', \mathbf{x}')$, we can figure out how $(\omega/c, i\mathbf{k})$ transforms to $(\omega'/c, i\mathbf{k}')$. Let $\omega/c \equiv k_0$ and $\omega'/c \equiv k_0'$ and consider Eq. (55) with the transformations Eq. (30) used for $t'$ and $\mathbf{x}'$:

$$\omega t - k_1 x_1 - k_2 x_2 - k_3 x_3 = \omega' \gamma(t - \beta x_1/c) - k_1' \gamma(x_1 - \beta ct) - k_2' x_2 - k_3' x_3. \tag{56}$$

Because $t$ and $\mathbf{x}$ are completely arbitrary, we may conclude that

$$\omega = \gamma(\omega' + \beta c k_1') \qquad k_1 = \gamma(k_1' + \beta\omega'/c) \qquad k_2 = k_2' \qquad k_3 = k_3', \tag{57}$$

or

$$k_0 = \gamma(k_0' + \beta k_1') \qquad k_1 = \gamma k_1' + \beta k_0') \qquad k_2 = k_2' \qquad k_3 = k_3'. \tag{58}$$

We recognize the form of these transformations; they tell us that $(k_0, \mathbf{k})$ transforms in the same way as $(x_0, i\mathbf{x})$, i.e., via the Lorentz transformation.

Let's spend a few minutes thinking about the conditions under which our result is valid. We assumed when making the argument that we have a plane wave in both $K$ and $K'$ which means, more or less, that we are giving Maxwell's equations Law of Nature status since we assumed that the relevant equation of motion produces plane wave solutions in both frames. In fact, our results are not correct for waves in general, because many types of waves will not have this property (plane waves remain plane waves relative to all reference frames if they are plane waves relative to one frame). But they are correct for electromagnetic waves in vacuum.

Finally, let us look at an alternative form for our transformations. Let

$$\mathbf{k}' = k'(\cos\theta'\,\boldsymbol{\epsilon}_1 + \sin\theta'\,\boldsymbol{\epsilon}_2); \tag{59}$$

the component of $\mathbf{k}'$ perpendicular to $\boldsymbol{\epsilon}_1$ is defined to be in the direction of $\boldsymbol{\epsilon}_2$. Further, $\mathbf{k}' = \omega'/c$. Then the transformation equations may be used to produce the relations

$$k_1 = \gamma k'(\cos\theta' + \beta) \qquad \mathbf{k}_2 = k'\sin\theta'\,\boldsymbol{\epsilon}_2 \qquad \text{and} \qquad \omega = \gamma\omega'(1 + \beta\cos\theta') \tag{60}$$

where $\mathbf{k}_2$ is the component of $\mathbf{k}$ which is perpendicular to $\boldsymbol{\epsilon}_1$. ¿From these results it is easy to show that $\omega - ck$, no surprise, and that

$$\cos\theta = \frac{\cos\theta' + \beta}{1 + \beta\cos\theta'}; \tag{61}$$

$\theta$ is the angle that $\mathbf{k}$ makes with the direction of $\mathbf{v}$, (or $\boldsymbol{\epsilon}_1$); that is,

$$\mathbf{k} = k(\cos\theta\,\boldsymbol{\epsilon}_1 + \sin\theta\,\boldsymbol{\epsilon}_2). \tag{62}$$

### 6.1 Stellar Red Shift

The last of Eqs. (60) in particular may be used to describe the Doppler shift of the frequency of electromagnetic waves in vacuum. A well-known case in point is the "redshift" of light from distant galaxies.

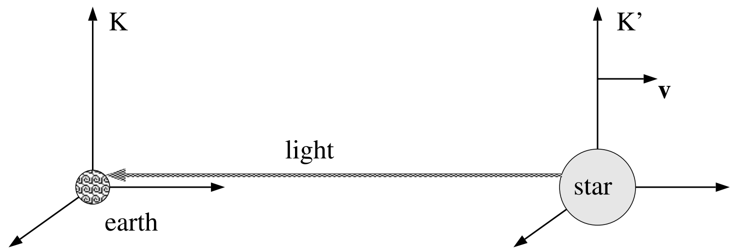


*Figure 15: Light from receding stars in $K'$ is redshifted when seen in $K$.*

Given an object receding from the observer in $K$ and emitting light of frequency $\omega'$ in its own rest frame, $K'$, we have $\cos\theta' \approx -1$ and

$$\omega = \gamma\omega'(1-\beta) = \omega'\sqrt{\frac{1-\beta}{1+\beta}}. \tag{63}$$

For, e.g., $\beta = 1/2$, $\omega = \omega'/\sqrt{3}$. The observer sees the light as having much lower frequency than that with which it is emitted; it is "red-shifted."

## 7 Four-tensors and all that

[5] It is no accident that $(x_0, \mathbf{x})$ and $(k_0, \mathbf{k})$ transform from $K$ to $K'$ in the same way. They are but two of many sets of four objects or elements that have this property. They are called *four-vectors*. More generally, there are sets of $4^p$ elements, with $p = 0, 1, 2...$, which have very similar transformation properties and which are called *four-tensors of rank $p$*. The better to manipulate them when the time comes, let us spend a little time now learning some of the basics of tensor calculus.

[5]The introduction to tensor calculus given in this section is largely drawn from J. L. Synge and A. Schild, *Tensor Calculus*, (University of Toronto Press, Toronto, 1949).

Consider the usual frames $K$ and $K'$ with coordinates $\bar{x}$ and $\bar{x}'$, respectively; $\bar{x}$ stands for $(x_0, \mathbf{x})$ and similarly for $x'$. Let there be some transformation from one frame to the other which gives

$$\bar{x}' = x'(\bar{x}), \tag{64}$$

with an inverse,

$$\bar{x} = x(\bar{x}'). \tag{65}$$

These transformations need not in general be linear.

A **contravariant vector** or *rank-one tensor* is defined to be a set of four quantities or elements $a^\alpha$, $\alpha = 0, 1, 2, 3$, which transform from $K$ to $K'$ according to the rule

$$a'^\alpha = \sum_{\beta=0}^{3} \frac{\partial x'^\alpha}{\partial x^\beta} a^\beta \equiv A^\alpha{}_\beta a^\beta. \tag{66}$$

This equation serves to define $A^\alpha{}_\beta$,

$$A^\alpha{}_\beta \equiv \frac{\partial x'^\alpha}{\partial x^\beta}; \tag{67}$$

we have also introduced in the last step the summation convention that a Greek index, which appears in a term as both an upper and a lower index, is summed from zero to three.

For any contravariant vector or tensor, we are going to introduce also a *covariant* vector or tensor whose components will be designated by subscripts. Define a **covariant vector** or *rank-one tensor* as a set of four objects $b_\alpha$, $\alpha = 0, 1, 2, 3$, which transform according to the rule

$$b'_\alpha = \sum_{\beta=0}^{3} \frac{\partial x^\beta}{\partial x'^\alpha} b_\beta \equiv A_\alpha{}^\beta b_\beta \tag{68}$$

where we have defined

$$A_\alpha{}^\beta \equiv \frac{\partial x^\beta}{\partial x'^\alpha}. \tag{69}$$

The generalization to tensors of ranks other than one is straightforward. For example, a rank-two contravariant tensor comprises a set of sixteen objects $T^{\alpha\beta}$ which transform according to the rule

$$T'^{\alpha\beta} = A^{\alpha}{}_{\gamma} A^{\beta}{}_{\delta} T^{\gamma\delta} \tag{70}$$

and a rank-two covariant tensor has sixteen elements $T_{\alpha\beta}$ which transform according to the rule

$$T'_{\alpha\beta} = A_{\alpha}{}^{\gamma} A_{\beta}{}^{\delta} T_{\gamma\delta}. \tag{71}$$

Mixed tensors can also be of interest. The rank-two mixed tensor $\bar{T}$ is a set of sixteen elements $T^{\alpha}{}_{\beta}$ which transform according to

$$T'^{\alpha}{}_{\beta} = A^{\alpha}{}_{\gamma} A_{\beta}{}^{\delta} T^{\gamma}{}_{\delta}. \tag{72}$$

Generalizations follow as you would expect.

The *inner product* of $\bar{a}$ and $\bar{b}$ can be[6] defined as

$$\bar{a} \cdot \bar{b} \equiv b_{\alpha} a^{\alpha}. \tag{73}$$

Consider the transformation properties of the inner product:

$$\bar{a}' \cdot \bar{b}' = a'^{\alpha} b'_{\alpha} = A_{\alpha}{}^{\gamma} A^{\alpha}{}_{\delta} b_{\gamma} a^{\delta}; \tag{74}$$

however,

$$A_{\alpha}{}^{\gamma} A^{\alpha}{}_{\delta} = \frac{\partial x^{\gamma}}{\partial x'^{\alpha}} \frac{\partial x'^{\alpha}}{\partial x^{\delta}} = \frac{\partial x^{\gamma}}{\partial x^{\delta}} = \delta^{\gamma}_{\delta} \tag{75}$$

where

$$\delta^{\gamma}_{\delta} \equiv \begin{cases} 1 & \gamma = \delta \\ 0 & \gamma \neq \delta \end{cases} \tag{76}$$

Hence

$$\bar{a}' \cdot \bar{b}' = \delta^{\gamma}_{\delta} a^{\delta} b_{\gamma} = a^{\gamma} b_{\gamma} = \bar{a} \cdot \bar{b}. \tag{77}$$

The inner product is an invariant, also known as a *scalar* or *rank-zero tensor*.

Notice that when we wrote the Kronecker delta function, we gave it a superscript and subscript as though it were a rank-two mixed tensor. It in fact is one as we can

[6] We will present a different but equivalent definition later.

show by transforming it from one frame to another. Let $\delta_\alpha^\beta$ be defined as above in the frame $K$ and let it be defined to be a mixed tensor. Then we know how it transforms and so can find it in a different frame $K'$ (where we hope it will turn out to be the same as in frame $K$:

$$\delta'^{\alpha}_{\beta} = A^{\alpha}{}_{\gamma} A_{\beta}{}^{\delta} \delta^{\gamma}_{\delta} = A^{\alpha}{}_{\gamma} A_{\beta}{}^{\gamma} = \frac{\partial x'^{\alpha}}{\partial x^{\gamma}} \frac{\partial x^{\gamma}}{\partial x'^{\beta}} = \frac{\partial x'^{\alpha}}{\partial x'^{\beta}} = \delta^{\alpha}_{\beta} \tag{78}$$

which means that the thing we defined to be a rank-two mixed tensor is remains the same as the Kronecker delta function in all frames.

The operation which enters the definition of the inner product is to set a contravariant and a covariant index equal to each other and then to sum them. This operation is called a *contraction* with respect to the pair of indices in question. It reduces the rank of something by two. That is, the sixteen objects $b_\alpha a^\beta$ form a rank-two tensor, as may be shown easily by checking how it transforms (given the transformation properties of $b_\alpha$ and $a^\beta$). After we perform the contraction, we are left with a rank-zero tensor.

## 7.1 The Metric Tensor

Now think about how we can use these things in relativity. We have a fundamental invariant which is the separation between two events; specifically,

$$(ds)^2 = (dx^0)^2 - (dx^1)^2 - (dx^2)^2 - (dx^3)^2 \tag{79}$$

is an invariant,

$$(ds)^2 = (ds')^2. \tag{80}$$

We would like to write this as an inner product $d\bar{x} \cdot d\bar{x}$, where

$$d\bar{x} \cdot d\bar{x} = dx_\alpha dx^\alpha. \tag{81}$$

However, in order that we can do so, it must be the case that the covariant four-vector $d\bar{x}$ have the components

$$dx_0 = dx^0 \quad \text{and} \quad dx_i = -dx^i \text{ for } i = 1, 2, 3. \tag{82}$$

In general, the components of the contravariant and covariant versions of a four-vector are related by the *metric tensor* $\bar{g}$ which is a rank-two tensor that can be expressed in covariant, contravariant, or mixed form (just like any other tensor of rank two or more). In particular, the *covariant metric tensor* is defined for any system by the statement that the separation can be written as

$$(ds)^2 \equiv g_{\alpha,\beta} dx^\alpha dx^\beta, \tag{83}$$

plus the statement that it is a symmetric tensor.

How do we know that this is a tensor? From the fact that its double contraction with the contravariant vector $\bar{x}$ is an invariant and from the fact that it is symmetric, one can prove that it is a rank-two covariant tensor.[7]

In three-dimensional Cartesian coordinates in a Euclidean space such as we are accustomed to thinking about, the covariant metric tensor is just the unit tensor. In curvilinear coordinates (for example, spherical coordinates) it is some other (still simple) thing. For the flat four-dimensional space that one deals with in the special theory of relativity, we can see from Eqs. (79) and (83), and from the condition that $\bar{g}$ is symmetric, that it must be

$$g_{00} = 1, \qquad g_{ii} = -1, \ i = 1, 2, 3, \qquad \text{and} \qquad g_{\alpha\beta} = 0, \ \alpha \neq \beta. \tag{84}$$

Next, we introduce the *contravariant metric tensor*. First, we take the determinant of the matrix formed by the covariant metric tensor,

$$g \equiv \det[g_{\alpha\beta}] = -1 \tag{85}$$

Then one introduces the cofactor, written as $\Delta^{\alpha\beta}$, of each element $g_{\alpha\beta}$ in the matrix. The elements of the contravariant metric tensor are defined as

$$g^{\alpha\beta} \equiv \frac{\Delta^{\alpha\beta}}{g}. \tag{86}$$

[7] See, e.g., Synge and Schild.

We need to demonstrate that this thing is indeed a contravariant tensor. From the standard definitions of the determinant and cofactor, we can write

$$g_{\alpha\beta}\Delta^{\alpha\gamma} = g_{\beta\alpha}\Delta^{\gamma\alpha} = \delta_\beta^\gamma g \tag{87}$$

from which it follows that

$$g_{\alpha\beta}g^{\alpha\gamma} = \delta_\beta^\gamma = g_{\beta\alpha}g^{\gamma\alpha}. \tag{88}$$

When contracted (as above) with a covariant tensor, the thing we call a contravariant tensor produces a mixed tensor. In addition, it is symmetric which follows from the symmetry of the covariant metric tensor. This is sufficient to prove that the elements $g^{\alpha\beta}$ do form a contravariant tensor.

It is easy to work out the elements of the contravariant metric tensor if one knows the covariant one; for our particular metric tensor they are the same as the elements of the covariant one.

The metric tensor is used to convert contravariant tensors or indices to covariant ones and conversely. Consider for example the elements $x_\alpha$ defined by

$$x_\alpha = g_{\alpha\beta}x^\beta. \tag{89}$$

It is clear that the result is a covariant tensor of rank one. It is the covariant version of the position four-vector $\bar{x}$ and has elements $(x^0, -\mathbf{x})$. Similarly, we may recover the contravariant version of a four-vector or tensor from the covariant version of the same tensor by using the contravariant metric tensor:

$$x^\alpha = g^{\alpha\beta}x_\beta = g^{\alpha\beta}g_{\beta\gamma}x^\gamma = \delta_\gamma^\alpha x^\gamma = x^\alpha. \tag{90}$$

More generally, one may raise or lower as many indices as one wishes by using the appropriate metric tensor as many times as needed. Among other things, we can thereby construct a mixed metric tensor,

$$g_\alpha^\beta = g_{\alpha\gamma}g^{\gamma\beta}; \tag{91}$$

Using the explicit components of the covariant and contravariant metric tensors, one finds that this is precisely the unit mixed tensor, i.e., the Kronecker delta,

$$g_{\alpha}^{\beta} = \delta_{\alpha}^{\beta} \tag{92}$$

Finally, we earlier defined the inner product of two vectors by contracting the covariant version of one with the contravariant version of the other; we can now see that there are numerous other ways to express the inner product:

$$\bar{a} \cdot \bar{b} = a^{\alpha} b_{\alpha} = g_{\alpha\gamma} a^{\alpha} b^{\gamma} = g^{\alpha\gamma} a_{\gamma} b_{\alpha}; \tag{93}$$

In particular, the separation is now seen to be the same as $\bar{x} \cdot \bar{x}$,

$$(s)^2 = g_{\alpha\beta} x^{\alpha} x^{\beta} = x^{\alpha} x_{\alpha}. \tag{94}$$

There is one piece of unfinished business in all of this. We have defined a metric tensor; it was defined so that the separation is an invariant. We still do not know (if we assume we haven't as yet learned about Lorentz transformations) the components $A^{\alpha}{}_{\beta}$ and $A_{\beta}{}^{\alpha}$ of the transformation matrices. Just any old transformations won't do; it has to be consistent with our metric tensor, i.e., with the condition that the separation is invariant. This implies some conditions on the transformations. We shall return to this point later.

## 7.2 Differential Operators

Differential operators also have simple transformation properties. Consider the basic example of the four operators $\partial/\partial x^{\alpha}$. The transformation of this from one frame to another is found from the relation

$$\frac{\partial}{\partial x'^{\alpha}} = \frac{\partial x^{\beta}}{\partial x'^{\alpha}} \frac{\partial}{\partial x^{\beta}} \equiv A_{\alpha}{}^{\beta} \frac{\partial}{\partial x^{\beta}}. \tag{95}$$

The components of this operator transform in the same way as the components of a covariant vector which means that the four differential operators $\partial/\partial x^{\alpha}$ form a covariant four-vector. That being the case the elements

$$A_\alpha{}^\beta = \frac{\partial x^\beta}{\partial x'^\alpha} \tag{96}$$

are the elements of a rank-two mixed tensor, and that is why we have all along used for them notation which suggests that they are components of such a tensor.

It is equally true that $\partial/\partial x_\alpha$ is a contravariant four-vector operator. Consider

$$\begin{aligned}\frac{\partial}{\partial x'_\alpha} = \frac{\partial x_\beta}{\partial x'_\alpha}\frac{\partial}{\partial x_\beta} &= g_{\beta\gamma}\frac{\partial x^\gamma}{\partial x'^\delta}\frac{\partial x'^\delta}{\partial x'_\alpha}\frac{\partial}{\partial x_\beta} \\ &= g_{\beta\gamma}g^{\delta\alpha}A_\delta{}^\gamma\frac{\partial}{\partial x_\beta} = A^\alpha{}_\beta\frac{\partial}{\partial x_\beta}.\end{aligned} \tag{97}$$

Since the operator transforms in the same way as a contravariant four-vector, it is a contravariant four-vector!

Either of these four-vectors is called the *four-divergence.* Let's introduce some new notation for them:

$$\partial_\alpha \equiv \frac{\partial}{\partial x^\alpha} \tag{98}$$

is the way we shall write a component of the covariant four-divergence, and

$$\partial^\alpha \equiv \frac{\partial}{\partial x_\alpha} \tag{99}$$

is the way we write a component of the contravariant four-divergence.

We can construct some interesting invariants using the four-divergence. For example, the inner product of one of them with an four-vector produces an invariant,

$$\partial^\alpha A_\alpha = \partial_\alpha A^\alpha = \frac{\partial A^0}{\partial x^0} + \nabla\cdot\mathbf{A}. \tag{100}$$

Also, the four-dimensional Laplacian

$$\partial^\alpha\partial_\alpha = \frac{\partial^2}{\partial x^{0^2}} - \nabla\cdot\nabla \equiv \Box \tag{101}$$

is an invariant, or scalar, operator.

### 7.3 Notation

It is natural to present four-vectors using column vectors and rank-two tensors using matrices. Thus a four-vector such as $\bar{x}$ becomes

$$\bar{x} = \begin{pmatrix} x^0 \\ x_1 \\ x^2 \\ x^3 \end{pmatrix}, \tag{102}$$

and its transpose is

$$\tilde{x} = (x^0 \; x^1 \; x^2 \; x^3). \tag{103}$$

Using this notation, we can write, e.g., the inner product of two vectors as

$$\bar{a} \cdot \bar{b} = a_\alpha b^\alpha = g_{\alpha\beta} a^\alpha b^\beta = \tilde{a}\bar{g}\bar{b}. \tag{104}$$

Notice however, that if $\tilde{a}$ were the transpose of the covariant vector, we would write the inner product as $\tilde{a}\bar{b}$. The notation leaves something to be desired. Be that as it may, we can write a transformation as

$$x'^\alpha = A^\alpha{}_\beta x^\beta \qquad \text{or } \bar{x}' = \bar{A}\bar{x}. \tag{105}$$

We will only make use of this abbreviated notation when it is necessary to cause lots of confusion.

# 8 Representation of the Lorentz transformation

Our next task is to find a general transformation matrix[8] $\bar{A}$. As pointed out earlier, the basic fact we have to work with is that the separation is invariant,

$$g_{\alpha\beta} dx^\alpha dx^\beta = g_{\alpha\beta} dx'^\alpha dx'^\beta. \tag{106}$$

[8] This could be the fully contravariant version, the fully covariant version or one of the two mixed versions. If we know one of them, we know the others because we can lower and raise indices with the metric tensor.

Knowing this, and knowing also that

$$x'^{\alpha} = A^{\alpha}{}_{\beta} x^{\beta}, \tag{107}$$

it is a standard and straightforward exercise in linear algebra to show that

$$\det|A| = \pm 1. \tag{108}$$

Just as in three dimensions, there are proper and improper transformations which satisfy our requirements. The proper ones may be arrived at via a sequence of infinitesimal transformations starting from the identity, $A^{\alpha}{}_{\beta} = g^{\alpha}_{\beta}$. All transformations generated in this manner have determinant $+1$. The improper ones cannot be constructed in this way, even though some of them can have determinant $+1$. An example is $A^{\alpha}{}_{\beta} = -g^{\alpha}_{\beta}$; it has determinant $+1$ but is an improper transformation and cannot be arrived at by a sequence of infinitesimal transformations.

In this investigation we shall construct proper Lorentz transformations and shall build them from infinitesimal ones. Let's start by writing

$$A^{\alpha}{}_{\beta} = \delta^{\alpha}_{\beta} + \Delta\omega^{\alpha}{}_{\beta}, \tag{109}$$

where $\Delta\omega^{\alpha}{}_{\beta}$ is an infinitesimal. From the invariance of the interval, one can easily show that of the sixteen components $\Delta\omega^{\alpha}{}_{\beta}$, the diagonal ones must be zero and the off-diagonal ones must be such that

$$\Delta\omega^{\alpha\beta} = -\Delta\omega^{\beta\alpha}; \tag{110}$$

notice that both indices are now contravariant, in contrast to the previous equation. If we write the preceding relation with one contravariant and one covariant index, we will find the same $-$ sign if the two indices are 1, 2, or 3, and there will be no $-$ if one index is 0 and the other is one of 1, 2, or 3. Evidently, it is simpler to use a completely contravariant form. [9]

[9] The point is, we can use any form for the tensor that we like because all forms can be found from any single one. Therefore, it makes sense to use that form in which the relations are simplest, if there is one.

These results demonstrate that we have just six independent infinitesimals. We may take them to be a set of six numbers without indices if we introduce suitable basis matrices. One such set of matrices is given by

$$(K_1)^\alpha{}_\beta = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}; \qquad (K_2)^\alpha{}_\beta = \begin{pmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \tag{111}$$

$$(K_3)^\alpha{}_\beta = \begin{pmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{pmatrix}; \qquad (S_1)^\alpha{}_\beta = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 \\ 0 & 0 & 1 & 0 \end{pmatrix} \tag{112}$$

$$(S_2)^\alpha{}_\beta = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \end{pmatrix}; \qquad (S_3)^\alpha{}_\beta = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}. \tag{113}$$

The most general infinitesimal transformation can now be written as

$$\bar{A} = \bar{g} - \Delta\boldsymbol{\omega} \cdot \bar{\mathbf{S}} - \Delta\boldsymbol{\zeta} \cdot \bar{\mathbf{K}}; \tag{114}$$

where $\Delta\boldsymbol{\omega}$ contains three independent infinitesimal components as does $\Delta\boldsymbol{\zeta}$; these are, respectively, just infinitesimal coordinate rotations and infinitesimal relative velocities.

Powers of the matrices $\bar{K}_i$ and $\bar{S}_i$ have some very special properties. For example,

$$(\bar{K}_1)^2 = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \quad \text{and} \quad (\bar{S}_1)^2 = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix}; \tag{115}$$

consequently, powers of the matrices tend to repeat; the periods of these cycles are two and four for the $\bar{K}$'s and the $\bar{S}$'s, respectively so that, for any $m$ and integral $n$,

$$(\bar{K}_i)^{m+2n} = (\bar{K}_i)^m \quad \text{and} \quad (\bar{S}_i)^{m+4n} = (\bar{S}_i)^m. \tag{116}$$

We have seen above what is the second power in each case; for the $\bar{K}$'s, the third power is the same as the first and for the $\bar{S}$'s, one finds the negative of the first,

$$(\bar{S}_i)^3 = -\bar{S}_i; \tag{117}$$

and finally, the fourth power of one of the $\bar{S}$'s has two 1's on the diagonal, much like the even powers of the $\bar{K}$'s.

We can construct the matrix for a finite transformation by making a sequence of many infinitesimal transformations. To this end consider some finite $\boldsymbol{\omega}$ and $\boldsymbol{\zeta}$ and relate them to the infinitesimals by

$$\Delta\boldsymbol{\omega} = \boldsymbol{\omega}/n \quad \text{and} \quad \Delta\boldsymbol{\zeta} = \boldsymbol{\zeta}/n, \tag{118}$$

where $n$ is a very large number. Now apply $\bar{A}$ (given by Eq. (114)) to $\bar{x}$ $n$ times, thereby producing some $\bar{x}'$:

$$x'^\alpha = \left(g - \frac{\boldsymbol{\omega}\cdot\mathbf{S}}{n} - \frac{\boldsymbol{\zeta}\cdot\mathbf{K}}{n}\right)^\alpha_{\ \alpha_1} \left(g - \frac{\boldsymbol{\omega}\cdot\mathbf{S}}{n} - \frac{\boldsymbol{\zeta}\cdot\mathbf{K}}{n}\right)^{\alpha_1}_{\ \alpha_2} \cdots \left(g - \frac{\boldsymbol{\omega}\cdot\mathbf{S}}{n} - \frac{\boldsymbol{\zeta}\cdot\mathbf{K}}{n}\right)^{\alpha_{n-1}}_{\ \alpha_n} x^{\alpha_n} \tag{119}$$

We want to take the $n \to \infty$ limit of this expression. In general,

$$\lim_{n\to\infty}\left(1 + \frac{a}{n}\right)^n = e^a, \tag{120}$$

as one can show by, e.g., considering the logarithm. Applying this fact, we find that

$$x'^\alpha = A^\alpha_{\ \beta} x^\beta \tag{121}$$

where

$$A^\alpha_{\ \beta} = \left(e^{-\boldsymbol{\omega}\cdot\mathbf{S} - \boldsymbol{\zeta}\cdot\mathbf{K}}\right)^\alpha_{\ \beta}. \tag{122}$$

We can get a little understanding of what this equation is telling us by considering some special cases which are also familiar. For example, let $\boldsymbol{\omega} = 0$ and $\boldsymbol{\zeta} = \zeta \boldsymbol{\epsilon_1}$. Then

$$\begin{aligned} \bar{A} = e^{-\zeta \bar{K}_1} &= 1 - \zeta \bar{K}_1 + \frac{\zeta^2}{2}\bar{K}_1^2 - \frac{\zeta^3}{6}\bar{K}_1^3 + ... \\ = 1 - \bar{K}_1 \left(\zeta + \frac{\zeta^3}{6} + ...\right) &+ \bar{K}_1^2 \left(1 + \frac{\zeta^2}{2} + ...\right) - \bar{K}_1^2 \\ &= 1 - \bar{K}_1^2 + (\cosh\zeta)\bar{K}_1^2 - (\sinh\zeta)\bar{K}_1 \end{aligned} \tag{123}$$

or

$$A^\alpha{}_\beta = \begin{pmatrix} \cosh\zeta & -\sinh\zeta & 0 & 0 \\ -\sinh\zeta & \cosh\zeta & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \tag{124}$$

which should be familiar. Similarly, if $\boldsymbol{\omega} = \omega \boldsymbol{\epsilon_1}$ with $\boldsymbol{\zeta} = 0$, one finds

$$A^\alpha{}_\beta = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & \cos\omega & \sin\omega \\ 0 & 0 & -\sin\omega & \cos\omega \end{pmatrix} \tag{125}$$

which we recognize as a simple rotation around the $x$-axis.

Our general result for $\bar{A}$ allows us to find the transformation matrix for any combination of $\boldsymbol{\omega}$ and $\boldsymbol{\zeta}$. In particular, one can show that for $\boldsymbol{\omega} = 0$ and general $\boldsymbol{\zeta}$ such that $\boldsymbol{\beta}$ has magnitude $\tanh\zeta$ and is in the direction of $\boldsymbol{\zeta}$. Writing the components of $\boldsymbol{\beta}$ as $\beta_i$, $i = 1, 2, 3$, we find that $\bar{A}$ is

$$A^\alpha{}_\beta = \begin{pmatrix} \gamma & -\gamma\beta_1 & -\gamma\beta_2 & -\gamma\beta_3 \\ -\gamma\beta_1 & 1 + \frac{(\gamma-1)\beta_1^2}{\beta^2} & \frac{(\gamma-1)\beta_1\beta_2}{\beta^2} & \frac{(\gamma-1)\beta_1\beta_3}{\beta^2} \\ -\beta_2 & \frac{(\gamma-1)\beta_1\beta_2}{\beta^2} & 1 + \frac{(\gamma-1)\beta_2^2}{\beta^2} & \frac{(\gamma-1)\beta_2\beta_3}{\beta^2} \\ -\gamma\beta_3 & \frac{(\gamma-1)\beta_1\beta_3}{\beta^2} & \frac{(\gamma-1)\beta_2\beta_3}{\beta^2} & 1 + \frac{(\gamma-1)\beta_3^2}{\beta^2} \end{pmatrix}, \tag{126}$$

in case anybody wanted to know.

# 9 Covariance of Electrodynamics

In this section we are going to demonstrate the consistency of the Maxwell equations with Einstein's first postulate. But first we must decide more precisely what it means for a "law of nature" to be "the same" in all inertial frames. The relevant statement is this: an equation expressing a law of nature must be **invariant in form** under Lorentz transformations. When this is the case, the equation is said to be *Lorentz covariant* or simply *covariant*, which has nothing to do with the definition of covariant as opposed to contravariant tensors. And what is meant by the phrase "invariant in form" which appears above? It means that the quantities in the equation must transform in well-defined ways (as particular components of some four-tensors, for example) and that when terms are grouped in an appropriate manner, each group transforms in the same way as each of the other groups. In order to determine whether the Maxwell equations can have this property, we must first figure out how each of the physical objects in those equations, that is, $\mathbf{E}$, $\mathbf{B}$, $\rho$, and $\mathbf{J}$, transforms.

## 9.1 Transformations of Source and Fields

### 9.1.1 $\rho$ and $\mathbf{J}$

Let's start with the electric charge. It is an experimental observation that charge is an invariant. If a system has a particular charge $q$ as measured in one frame, then it has the same charge $q$ when the measurements are made in a different frame. From this (experimental) fact and things we already know, we can determine how charge density and current density transform.

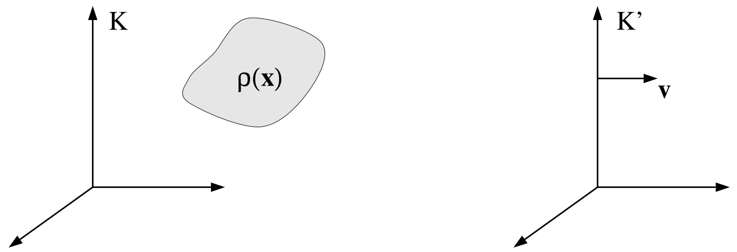


*Figure 16: The charge density transforms like time.*

Suppose that we have a system with charge density $\rho$, as measured in $K$, and $\rho'$ as measured in $K'$. Then in volume $d^3x$ in $K$, there is charge $dq$ where

$$dq = \rho d^3x = \rho d^3x \, dt/dt \tag{127}$$

where we have introduced an infinitesimal time element $dt$ as well. Similarly, in $K'$, the charge $dq'$ in the volume element $d^3x'$ can be written as

$$dq' = \rho' d^3x' dt'/dt'. \tag{128}$$

Now, if $d^3x'$ is what $d^3x$ transforms into (that it, if it is the same volume element as $d^3x$), then charge invariance implies that $dq = dq'$. Further, if $dt'$ is what $dt$ transforms into, then we can say that

$$c\, d^3x' dt' \equiv d^4x' = \left| \frac{\partial(x'^0, x'^1, x'^2, x'^3)}{\partial x^0, x^1, x^2, x^3)} \right| d^4x \equiv |\det[\bar{A}]| d^4x. \tag{129}$$

But the determinant of $\bar{A}$ is unity, so we have shown that a spacetime volume element is an invariant,

$$d^4x = d^4x'. \tag{130}$$

As applied to the present inquiry, we use this statement along with the equality of $dq$ and $dq'$ (and the invariance of $c$) to conclude that

$$\rho/dt = \rho'/dt'. \tag{131}$$

This relation can be true only if the charge density transforms in the same way as the time; that is, it must be the $0^{th}$ component of a four-vector.

Where are the other three components of this four-vector? They are the current density. Since $\mathbf{J}$ is $\rho$ times a velocity, which is in turn the ratio $d\mathbf{x}/dt$, we can write

$$\mathbf{J} = \rho\mathbf{u} = \rho\frac{d\mathbf{x}}{dt}; \tag{132}$$

in view of the fact that $\rho/dt$ is an invariant, $\mathbf{J}$ must transform in the same way as $d\mathbf{x}$, which is to say, as the 1,2,3 components of a (contravariant) four-vector. Hence we have the *contravariant current four-vector*

$$J^\alpha = (c\rho, \mathbf{J}); \tag{133}$$

the covariant current four-vector is

$$J_\alpha = (c\rho, -\mathbf{J}). \tag{134}$$

Knowing this, we are not surprised to find that the charge conservation equation is a four-divergence equation,

$$\frac{\partial\rho}{\partial t} + \nabla\cdot\mathbf{J} = 0 \qquad \text{or } \partial_\alpha J^\alpha = 0. \tag{135}$$

Notice that this equation is "covariant" in the sense introduced earlier; both sides are scalars.

### 9.1.2 Potentials

Now we shall proceed by demanding that all the relevant equations be Lorentz covariant. We shall apply this requirement to equations that we already have and see what are the implications for the fields $\mathbf{E}$ and $\mathbf{B}$ and also see that no contradictions arise. Let's start with the equations for the potentials in the Lorentz gauge. The equations of motion are

$$\Box\mathbf{A}(\mathbf{x},t) = \frac{4\pi}{c}\mathbf{J}(\mathbf{x},t) \qquad \text{and} \qquad \Box\Phi(\mathbf{x},t) = 4\pi\rho(\mathbf{x},t); \tag{136}$$

these can all be written in the very brief notation

$$\Box A^\alpha(\mathbf{x},t) = \frac{4\pi}{c}J^\alpha(\mathbf{x},t) \tag{137}$$

where we have introduced

$$A^\alpha \equiv (\Phi, \mathbf{A}) \tag{138}$$

which must be a contravariant four-vector if the equations of motion above are the correct equations of motion for the potential in the Lorentz gauge in every inertial frame. Notice that the potentials in gauges other than the Lorentz gauge will not form a four-vector.

The Lorentz condition, which is satisfied by potentials in the Lorentz gauge, is

$$\nabla \cdot \mathbf{A}(\mathbf{x}, t) + \frac{1}{c}\frac{\partial \Phi}{\partial t} = 0; \tag{139}$$

this equation may also be written as a four-divergence of a four-vector,

$$\partial_\alpha A^\alpha = 0. \tag{140}$$

### 9.1.3 Fields, Field-Strength Tensor

Let's look next at $\mathbf{E}$ and $\mathbf{B}$; these are given by

$$\mathbf{B}(\mathbf{x}, t) = \nabla \times \mathbf{A}(\mathbf{x}, t) \quad \text{and} \quad \mathbf{E}(\mathbf{x}, t) = -\nabla\Phi(\mathbf{x}, t) - \frac{1}{c}\frac{\partial \mathbf{A}(\mathbf{x}, t)}{\partial t}. \tag{141}$$

Look at just the $x$-components:

$$E_x = -\frac{1}{c}\frac{\partial A_x}{\partial t} - \frac{\partial \Phi}{\partial x} = -\frac{\partial A^1}{\partial x^0} - \frac{\partial \Phi}{\partial x^1} = -\frac{\partial A^1}{\partial x_0} + \frac{\partial A^0}{\partial x_1} = -\partial^0 A^1 + \partial^1 A^0. \tag{142}$$

Similarly, a component of the magnetic induction turns out to be, e.g.,

$$B_x = -\partial^2 A^3 + \partial^3 A^2. \tag{143}$$

Given the four-vector character of the differential operators and of the potentials, we can see that these particular components of the electric field and magnetic induction are elements of a rank-two tensor which we have expressed here in contravariant form. Let us define the *field-strength tensor* $\bar{F}$ by

$$F^{\alpha\beta} \equiv \partial^\alpha A^\beta - \partial^\beta A^\alpha. \tag{144}$$

This turns out to be

$$F^{\alpha\beta} = \begin{pmatrix} 0 & -E_x & -E_y & -E_z \\ E_x & 0 & -B_z & B_y \\ E_y & B_z & 0 & -B_x \\ E_z & -B_y & B_x & 0 \end{pmatrix}. \tag{145}$$

Because the tensor is antisymmetric, it has just six independent entries, these being the six components of $\mathbf{E}$ and $\mathbf{B}$.

The corresponding covariant tensor is easily worked out. It is the same as the contravariant one except that the signs of the entries in the first column and the first row are reversed. A somewhat different object which contains the same information is the *dual field-strength tensor* $\bar{\mathcal{F}}$ which is defined by

$$\mathcal{F}^{\alpha\beta} \equiv \epsilon^{\alpha\beta\gamma\delta}\frac{1}{2}F_{\gamma\delta} \tag{146}$$

where the *fully antisymmetric rank-four unit pseudotensor* with components $\epsilon^{\alpha\beta\gamma\delta}$ is in turn defined by specifying (1) that in frame $K$

$$\epsilon^{\alpha\beta\gamma\delta} \equiv \begin{cases} 1 & \text{if } \alpha\beta\gamma\delta \text{ is an even permutation of 1234} \\ -1 & \text{if } \alpha\beta\gamma\delta \text{ is an odd permutation of 1234} \\ 0 & \text{otherwise,} \end{cases} \tag{147}$$

and (2) that it transforms to other frames as a rank-four pseudotensor must,

$$(\epsilon')^{\alpha\beta\gamma\delta} \equiv \det[\bar{A}]A^{\alpha}{}_{\phi}A^{\beta}{}_{\chi}A^{\gamma}{}_{\psi}A^{\delta}{}_{\omega}\epsilon^{\phi\chi\psi\omega}. \tag{148}$$

Applying this definition, one can show that the components of this pseudotensor are given by Eq. (147) not only in frame $K$ but in all inertial frames.

Although $\bar{\epsilon}$ is a pseudotensor as opposed to a true tensor, the distinction will not be important for us so long as we stick to proper Lorentz transformations or to improper ones that have determinant +1. In what follows, we will refer to it as a tensor even though we know better; similarly we will refer to the dual tensor as a "tensor" (as was done in the definition) even though it is in fact a pseudotensor.

Returning now to the original point, $\mathcal{F}^{\alpha\beta}$ is, explicitly,

$$\mathcal{F}^{\alpha\beta} = \begin{pmatrix} 0 & -B_x & -B_y & -B_z \\ B_x & 0 & E_z & -E_y \\ B_y & -E_z & 0 & E_x \\ B_z & E_y & -E_x & 0 \end{pmatrix}. \tag{149}$$

## 9.2 Invariance of Maxwell Equations

Now we know how everything transforms; it remains to be seen whether the Maxwell equations are Lorentz covariant. The inhomogeneous equations are

$$\nabla \cdot \mathbf{E}(\mathbf{x},t) = 4\pi\rho(\mathbf{x},t) \quad \text{and} \quad \nabla \times \mathbf{B}(\mathbf{x},t) - \frac{1}{c}\frac{\partial \mathbf{E}(\mathbf{x},t)}{\partial t} = \frac{4\pi}{c}\mathbf{J}(\mathbf{x},t). \tag{150}$$

The first of these is

$$\frac{\partial F^{10}}{\partial x^1} + \frac{\partial F^{20}}{\partial x^2} + \frac{\partial F^{30}}{\partial x^3} = \frac{4\pi}{c}J^0. \tag{151}$$

Because $F^{00} \equiv 0$, we may add a term $\partial F^{00}/\partial x^0$ to the left-hand side of this equation and then find that it reads

$$\partial_\alpha F^{\alpha 0} = \frac{4\pi}{c}J^0. \tag{152}$$

This equation is clearly the $0^{th}$ component of a four-vector equation in which the left-hand side is obtained by taking the divergence of a rank-two tensor. The other three inhomogeneous Maxwell equations may be analyzed in similar fashion and the four may be concisely written as

$$\partial_\alpha F^{\alpha\beta} = \frac{4\pi}{c}J^\beta \tag{153}$$

where $\beta$ =0,1,2, and 3. These are manifestly Lorentz covariant.

The homogeneous Maxwell equations are

$$\nabla \cdot \mathbf{B}(\mathbf{x},t) = 0 \quad \text{and} \quad \nabla \times \mathbf{E}(\mathbf{x},t) = -\frac{1}{c}\frac{\partial \mathbf{B}(\mathbf{x},t)}{\partial t}. \tag{154}$$

The first one can be written as

$$\frac{\partial \mathcal{F}^{10}}{\partial x^1} + \frac{\partial \mathcal{F}^{20}}{\partial x^2} + \frac{\partial \mathcal{F}^{30}}{\partial x^3} = 0, \tag{155}$$

or, since $\mathcal{F}^{00} = 0$,

$$\partial_\alpha \mathcal{F}^{\alpha 0} = 0. \tag{156}$$

The others can be expressed in similar fashion, and all four are contained in the following equation:

$$\partial_\alpha \mathcal{F}^{\alpha\beta} = 0, \tag{157}$$

where $\beta$ =0, 1, 2, and 3. This form is clearly covariant, establishing the covariance of Maxwell's equations. These equations are components of a rank-one pseudotensor. They may also be written as components of a rank-three tensor. Notice that $\nabla \cdot \mathbf{B} = 0$ is, in tensor notation,

$$\frac{\partial F^{31}}{\partial x^2} + \frac{\partial F^{23}}{\partial x^1} + \frac{\partial F^{12}}{\partial x^3} = 0. \tag{158}$$

The remaining three homogeneous Maxwell equations can be expressed in similar fashion, and all four can be written as

$$\partial^\alpha F^{\beta\gamma} + \partial^\gamma F^{\alpha\beta} + \partial^\beta F^{\gamma\alpha} = 0 \tag{159}$$

where $\alpha$, $\beta$, and $\gamma$ are any three of 0,1,2,3, giving four equations. The other possible choices of the superscripts (involving repetition of two or more values) give nothing (They give 0=0). Hence we have succeeded in writing each of the homogeneous Maxwell equations in the form of an element of a rank-three tensor and the Lorentz covariant equation we have constructed simply says that this tensor is equal to zero.

# 10 Transformation of the electromagnetic field

The transformation properties of $\mathbf{E}$ and $\mathbf{B}$ are easily worked out by making use of our knowledge of how a rank-two tensor must transform:

$$(F')^{\alpha\beta} = A^\alpha{}_\gamma A^\beta{}_\delta F^{\gamma\delta}, \tag{160}$$

or, in matrix notation,

$$\bar{F}' = \bar{A}\bar{F}\tilde{A} \tag{161}$$

where $\tilde{A}$ is the transpose of the matrix representing $\bar{A}$. If we pick a frame $K'$ which is moving at velocity $\mathbf{v} = c\beta\boldsymbol{\epsilon}_1$, then

$$\bar{A} = \begin{pmatrix} \gamma & -\beta\gamma & 0 & 0 \\ -\beta\gamma & \gamma & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \equiv \tilde{A}. \tag{162}$$

Given the field tensor from Eq. (145), we have

$$\bar{F}\tilde{A} = \begin{pmatrix} \beta\gamma E_x & -\gamma E_x & -E_y & -E_z \\ \gamma E_x & -\beta\gamma E_x & -B_z & B_y \\ \gamma E_y - \beta\gamma B_z & -\beta\gamma E_y + \gamma B_z & 0 & -B_x \\ \gamma E_z + \beta\gamma B_y & -\beta\gamma E_z - \gamma B_y & B_x & 0 \end{pmatrix} \tag{163}$$

and

$$\bar{F}' = \begin{pmatrix} 0 & -E_x & -\gamma E_y + \beta\gamma B_y & -\gamma E_z - \beta\gamma B_y \\ E_x & 0 & \beta\gamma E_y - \gamma B_z & \beta\gamma E_z + \gamma B_y \\ \gamma E_y - \beta\gamma B_z & -\beta\gamma E_y + \gamma B_z & 0 & -B_x \\ \gamma E_z + \beta\gamma B_y & -\beta\gamma E_z - \gamma B_y & B_x & 0 \end{pmatrix}. \tag{164}$$

This is an antisymmetric tensor - as it should be - and we can equate individual elements to the appropriate components of $\mathbf{B}'$ and $\mathbf{E}'$. One finds

$$\begin{aligned} B'_x &= B_x & B'_y &= \gamma(B_y + \beta E_z) & B'_z &= \gamma(B_z - \beta E_y) \\ E'_x &= E_x & E'_y &= \gamma(E_y - \beta B_z) & E'_z &= \gamma(E_z + \beta B_y). \end{aligned} \tag{165}$$

By examining these relations for a bit, one can see that

$$\begin{aligned} \mathbf{E}'_{\parallel} &= \mathbf{E}_{\parallel} & \mathbf{E}'_{\perp} &= \gamma[\mathbf{E}_{\perp} + (\boldsymbol{\beta} \times \mathbf{B})] \\ \mathbf{B}'_{\parallel} &= \mathbf{B}_{\parallel} & \mathbf{B}'_{\perp} &= \gamma[\mathbf{B}_{\perp} - (\boldsymbol{\beta} \times \mathbf{E})] \end{aligned} \tag{166}$$

where the subscripts refer to components of the fields parallel or perpendicular to $\boldsymbol{\beta}$.

From the transformations one may see that when $\mathbf{E} \perp \mathbf{B}$, it is possible to find a frame where one of $\mathbf{E}'$ and $\mathbf{B}'$ (which one?) vanishes. This is achieved by picking $\boldsymbol{\beta} \perp \mathbf{E}$ and $\boldsymbol{\beta} \perp \mathbf{B}$ with an appropriate magnitude. For example, if $|\mathbf{B}| > |\mathbf{E}|$, we take

$$\boldsymbol{\beta} = \beta(\mathbf{E} \times \mathbf{B})/|\mathbf{E}||\mathbf{B}| \tag{167}$$

where $\beta$ is to be such that $\mathbf{E}'_\perp = 0$, or

$$\begin{aligned} 0 &= \mathbf{E} + \boldsymbol{\beta} \times \mathbf{B} = \mathbf{E} + \beta[(\mathbf{E} \times \mathbf{B}) \times \mathbf{B}]/|\mathbf{E}||\mathbf{B}| \\ &= \mathbf{E} - \beta\mathbf{E}|\mathbf{B}|/|\mathbf{E}| = \mathbf{E}(1 - \beta|\mathbf{B}|/|\mathbf{E}|) \end{aligned} \tag{168}$$

so that we find

$$\beta = |\mathbf{E}|/|\mathbf{B}| \tag{169}$$

which is possible if $B > E$.

## 10.1 Fields Due to a Point Charge

Another example of the use of the transformations is the determination of the fields of a charge moving at constant velocity. Suppose a charge $q$ has velocity $\mathbf{u} = \beta c \boldsymbol{\epsilon}_1$ relative to frame $K$. Let $K'$ move at this velocity relative to $K$ so that the charge in at rest in the primed frame. Further, choose the coordinates so that the charge is at $\mathbf{x}' = 0$. Then the fields in this frame are

$$\mathbf{B}'(\mathbf{x}', t') = 0 \qquad \text{and} \qquad \mathbf{E}'(\mathbf{x}', t') = \frac{q}{r'^3}\mathbf{x}'. \tag{170}$$

Let us restrict attention, without loss of generality, to the $z' = 0$ plane. There, the electric field is

$$\mathbf{E}' = \frac{q}{(\sqrt{x'^2 + y'^2})^3}[x'\boldsymbol{\epsilon}_1{}' + y'\boldsymbol{\epsilon}_2] \tag{171}$$

Using the transformations of the electromagnetic field, we find that the nonvanishing components of the fields in frame $K$ are

$$E_x = \frac{qx'}{(x'^2 + y'^2)^{3/2}} \qquad E_y = \frac{\gamma q y'}{(x'^2 + y'^2)^{3/2}} \tag{172}$$

and

$$B_z = \frac{\gamma\beta q y'}{(x'^2 + y'^2)^{3/2}}. \tag{173}$$

In order for these expressions to be of any use, we should express the fields in terms of $t$ and $\mathbf{x}$ rather than the primed spacetime variables.

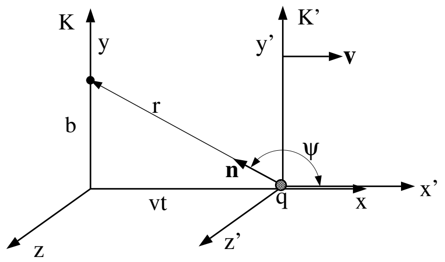


*Figure 17: Charge fixed in $K'$ detected by an observer at $P$ in $K$.*

We shall look in particular at the fields at the point $\mathbf{x} = b\boldsymbol{\epsilon}_2$. His position translates, via the Lorentz transformations, into $K'$ as

$$y' = y \qquad x' = -\gamma vt \quad \text{and} \quad z' = 0. \tag{174}$$

Using these in the expressions for $\mathbf{E}$ and $\mathbf{B}$, we find

$$\begin{aligned} E_x &= -\gamma qvt/[b^2 + (\gamma vt)^2]^{3/2} \\ E_y &= \gamma qb/[b^2 + (\gamma vt)^2]^{3/2} \\ B_z &= \gamma\beta qb/[b^2 + (\gamma vt)^2]^{3/2}. \end{aligned} \tag{175}$$

It is instructive to study these results. They tell us the field at a point (0,b,0) in $K$ when a charge $q$ goes along the $x$ axis with speed $v$, passing the origin at time $t = 0$. The fields are zero at large negative times, then $E_x$ rises and falls to zero at $t = 0$ and repeats this pattern with the opposite sign at positive times. The other two rise to a maximum value at $t = 0$ and then fall to zero at large positive time. The duration in time of the pulse if of order $b/\gamma v$ and becomes very short if $v \to c$ because then $\gamma$ becomes arbitrarily large. The maximum field strengths are, for $E_y$, $\gamma q/b^2$, and, for $B_z$, $\gamma\beta q/b^2$. Notice that for a highly relativistic particle, $\beta \to 1$, $E_y \approx B_z$

and also that the maximum pulse strength scales as $\gamma$ which means it becomes very large (but for a very short time) as the velocity approaches $c$.