1

Special relativity

1.1 Introduction

This chapter introduces the special theory of relativity from a perspective that is appropriate for proceeding to the general theory of relativity later on, from Chapter 4 onwards.[1] Several topics such as the manipulation of tensorial quantities, description of physical systems using action principles, the use of distribution function to describe a collection of particles, etc., are introduced in this chapter in order to develop familiarity with these concepts within the context of special relativity itself. Virtually all the topics developed in this chapter will be generalized to curved spacetime in Chapter 4. The discussion of Lorentz group in Section 1.3.3 and in Section 1.10 is somewhat outside the main theme; the rest of the topics will be used extensively in the future chapters.[2]

1.2 The principles of special relativity

To describe any physical process we need to specify the spatial and temporal coordinates of the relevant *event*. It is convenient to combine these four real numbers – one to denote the time of occurrence of the event and the other three to denote the location in space – into a single entity denoted by the four-component object $x^i = (x^0, x^1, x^2, x^3) \equiv (t, \boldsymbol{x}) \equiv (t, x^\alpha)$. More usefully, we can think of an event $\mathcal{P}$ as a point in a four-dimensional space with coordinates x^i. We will call the collection of all events as *spacetime*.

Though the actual numerical values of x^i, attributed to any given event, will depend on the specific coordinate system which is used, the event $\mathcal{P}$ itself is a geometrical quantity that is independent of the coordinates used for its description. This is clear even from the consideration of the spatial coordinates of an event. A spatial location can be specified, for example, in the Cartesian coordinates giving the coordinates (x, y, z) or, say, in terms of the spherical polar coordinates by providing (r, θ, ϕ). While the numerical values (and even the dimensions) of these coordinates are different, they both signify the same geometrical point in three-dimensional space. Similarly, one can describe an event in terms of any suitable set of four independent numbers and one can transform from any system of coordinates to another by well-defined coordinate transformations.

Among all possible coordinate systems which can be used to describe an event, a subset of coordinate systems, called the *inertial coordinate systems* (or inertial frames), deserve special attention. Such coordinate systems are defined by the property that a material particle, far removed from all external influences, will move with uniform velocity in such frames of reference. This definition is convenient and practical but is inherently flawed, since one can never operationally verify the criterion that no external influence is present. In fact, there is no fundamental reason why any one class of coordinate system should be preferred over others, except for mathematical convenience. Later on, in the development of general relativity in Chapter 4, we shall drop this restrictive assumption and develop the physical principles treating all coordinate systems as physically equivalent. For the purpose of this chapter and the next, however, we shall *postulate* the existence of inertial coordinate systems which enjoy a special status. (Even in the context of general relativity, it will turn out that one can introduce inertial frames in a sufficiently small region around any event. Therefore, the description we develop in the first two chapters will be of importance even in a more general context.) It is obvious from the definition that any coordinate frame moving with uniform velocity with respect to an inertial frame will also constitute an inertial frame.

To proceed further, we shall introduce two empirical facts which are demonstrated by experiments. (i) It turns out that all laws of nature remain identical in all inertial frames of reference; that is, the equations expressing the laws of nature retain the same form under the coordinate transformation connecting any two inertial frames. (ii) The interactions between material particles do not take place instantaneously and there exists a maximum possible speed of propagation for interactions. We will denote this speed by the letter c. Later on, we will show in Chapter 2 that ordinary light waves, described by Maxwell's equations, propagate at this speed. Anticipating this result we may talk of light rays propagating in straight lines with the speed c. From (i) above, it follows that the maximum velocity of propagation c should be the same in all inertial frames.

Of these two empirically determined facts, the first one is valid even in non-relativistic physics. So the key new results of special relativity actually originate from the second fact. Further, the existence of a uniquely defined speed c allows one to express time in units of length by working with ct rather than t. We shall accordingly specify an event by giving the coordinates $x^i = (ct, x^\alpha)$ rather than in terms of t and x^α. This has the advantage that all components of x^i have the same dimension when we use Cartesian spatial coordinates.

The two facts, (i) and (ii), when combined together, lead to a profound consequence: they rule out the absolute nature of the notion of simultaneity; two events which appear to occur at the same time in one inertial frame will not, in general, appear to occur at the same time in another inertial frame. For example, consider two inertial frames K and K' with K' moving relative to K along the x-axis with the speed V. Let B, A and C (in that order) be three points along the common x-axis with $AB = AC$ in the *primed* frame, K'. Two light signals that start from a point A and go towards B and C will reach B and C at the same instant of time as observed in K'. But the two events, namely arrival of signals at B and C, cannot be simultaneous for an observer in K. This is because, in the frame K, point B moves towards the signal while C moves away from the signal; but the speed of the signal is postulated to be the same in both frames. Obviously, when viewed in the frame K, the signal will reach B before it reaches C.

In non-relativistic physics, one would have expected the two light beams to inherit the velocity of the source at the time of emission so that the two light signals travel with *different* speeds $(c \pm V)$ towards C and B and hence will reach them simultaneously in both frames. It is the constancy of the speed of light, independent of the speed of the source, which makes the notion of simultaneity frame dependent.

The concept of associating a time coordinate to an event is based entirely on the notion of simultaneity. In the simplest sense, we will attribute a time coordinate t to an event – say, the collision of two particles – if the reading of a clock indicating the time t is simultaneous with the occurrence of the collision. Since the notion of simultaneity depends on the frame of reference, it follows that two different observers will, in general, assign different time coordinates to the same event. This is an important conceptual departure from non-relativistic physics in which simultaneity is an absolute concept and all observers use the same clock time.

The second consequence of the constancy of speed of light is the following. Consider two infinitesimally separated events $\mathcal{P}$ and $\mathcal{Q}$ with coordinates x^i and $(x^i + dx^i)$. We define a quantity ds – called the *spacetime interval* – between these two events by the relation

$$ds^2 = -c^2dt^2 + dx^2 + dy^2 + dz^2. \tag{1.1}$$

If $ds = 0$ in one frame, it follows that these two infinitesimally separated events $\mathcal{P}$ and $\mathcal{Q}$ can be connected by a light signal. Since light travels with the same speed c in all inertial frames, $ds' = 0$ in any other inertial frame. In fact, one can prove the stronger result that $ds' = ds$ for any two infinitesimally separated events, not just those connected by a light signal. To do this, let us treat ds^2 as a function of ds'^2 we can expand ds^2 in a Taylor series in ds'^2, as $ds^2 = k + a\,ds'^2 + \cdots$. The fact that $ds = 0$ when $ds' = 0$ implies $k = 0$; the coefficient a can only be a function of the relative velocity $\boldsymbol{V}$ between the frames. Further, homogeneity and isotropy of space requires that only the magnitude $|\boldsymbol{V}| = V$ enters into this function. Thus we conclude that $ds^2 = a(V)ds'^2$, where the coefficient $a(V)$ can depend only on the absolute value of the relative velocity between the inertial frames. Now consider three inertial frames K, K_1, K_2, where K_1 and K_2 have relative velocities $\boldsymbol{V}_1$ and $\boldsymbol{V}_2$ with respect to K. From $ds_1^2 = a(V_1)ds^2, ds_2^2 = a(V_2)ds^2$ and $ds_2^2 = a(V_{12})ds_1^2$, where $\boldsymbol{V}_{12}$ is the relative velocity of K_1 with respect to K_2, we see that $a(V_2)/a(V_1) = a(V_{12})$. But the magnitude of the relative velocity V_{12} must depend not only on the magnitudes of $\boldsymbol{V}_1$ and $\boldsymbol{V}_2$ but also on the angle between the velocity vectors. So, it is impossible to satisfy this relation unless the function $a(V)$ is a constant; further, this constant should be equal to unity to satisfy this relation. It follows that the quantity ds has the same value in all inertial frames; $ds^2 = ds'^2$, i.e. the infinitesimal spacetime interval is an invariant quantity. Events for which ds^2 is less than, equal to or greater than zero are said to be separated by *timelike, null* or *spacelike* intervals, respectively.

With future applications in mind, we shall write the line interval in Eq. (1.1) using the notation

$$ds^2 = \eta_{ab}dx^a dx^b; \quad \eta_{ab} = \text{diag}\,(-1, +1, +1, +1) \tag{1.2}$$

in which we have introduced the *summation convention*, which states that any index which is repeated in an expression – like a, b here – is summed over the range of values taken by the index. (It can be directly verified that this convention is a consistent one and leads to expressions which are unambiguous.) In defining ds^2 in Eq. (1.1) and Eq. (1.2) we have used a negative sign for c^2dt^2 and a positive sign for the spatial terms dx^2, etc. The sequence of signs in η_{ab} is called *signature* and it is usual to say that the signature of spacetime is $(-+++)$. One can, equivalently, use the signature $(+---)$ which will require a change of sign in several expressions. This point should be kept in mind while comparing formulas in different textbooks.

A continuous sequence of events in the spacetime can be specified by giving the coordinates $x^a(\lambda)$ of the events along a parametrized curve defined in terms of a suitable parameter λ. Using the fact that ds defined in Eq. (1.1) is invariant, we can define the analogue of an (invariant) arc length along the curve, connecting two events $\mathcal{P}$ and $\mathcal{Q}$, by:

$$s(\mathcal{P}, \mathcal{Q}) = \int_{\mathcal{P}}^{\mathcal{Q}} |ds| = \int_{\lambda_1}^{\lambda_2} \frac{|ds|}{d\lambda} d\lambda \equiv \int_{\lambda_1}^{\lambda_2} d\lambda \left| \left(\frac{d\boldsymbol{x}}{d\lambda}\right)^2 - c^2 \left(\frac{dt}{d\lambda}\right)^2 \right|^{1/2}. \tag{1.3}$$

The modulus sign is introduced here because the sign of the squared arc length ds^2 is indefinite in the spacetime. For curves which have a definite sign for the arclength – i.e. for curves which are everywhere spacelike or everywhere timelike – one can define the arclength with appropriate sign. That, is, for a curve with $ds^2 < 0$ everywhere, we will define the arc length with a flip of sign, as $(-ds^2)^{1/2}$. (For curves along the path of a light ray the arc length will be zero.) This arc length will have the same numerical value in all inertial frames and will be independent of the parametrization used to describe the curve; a transformation $\lambda \to \lambda' = f(\lambda)$ leaves the value of the arc length unchanged.

Of special significance, among all possible curves in the spacetime, is the one that describes the trajectory of a material particle moving along some specified path, called the *worldline*. In three-dimensional space, we can describe such a trajectory by giving the position as a function of time, $\boldsymbol{x}(t)$, with the corresponding velocity $\boldsymbol{v}(t) = (d\boldsymbol{x}/dt)$. We can consider this as a curve in *spacetime* with $\lambda = ct$ acting as the parameter so that $x^i = x^i(t) = (ct, \boldsymbol{x}(t))$. Further, given the existence of a maximum velocity, we must have $|\boldsymbol{v}| < c$ everywhere making the curve everywhere timelike with $ds^2 < 0$. In this case, one can provide a direct physical interpretation for the arc length along the curve. Let us consider a clock (attached to the particle) which is moving relative to an inertial frame K on an *arbitrary* trajectory. During a time interval between t and $(t + dt)$, the clock moves through a distance $|d\boldsymbol{x}|$ as measured in K. Consider now another inertial coordinate system K', which – at that instant of time t – is moving with respect to K with the same velocity as the clock. In this frame, the clock is momentarily at rest, giving $d\boldsymbol{x}' = 0$. If the clock indicates a lapse of time $dt' \equiv d\tau$, when the time interval measured in K is dt, the invariance of spacetime intervals implies that

$$ds^2 = -c^2dt^2 + dx^2 + dy^2 + dz^2 = ds'^2 = -c^2d\tau^2. \tag{1.4}$$

Or,

$$d\tau = \frac{[-ds^2]^{1/2}}{c} = dt\sqrt{1 - \frac{v^2}{c^2}}. \tag{1.5}$$

Hence $(1/c)(-ds^2)^{1/2} \equiv |(ds/c)|$, defined with a flip of sign in ds^2, is the lapse of time in a moving clock; this is called the *proper time* along the trajectory of the clock. The arclength in Eq. (1.3), divided by c, viz.

$$\tau = \int d\tau = \int_{t_1}^{t_2} dt\sqrt{1 - \frac{v^2(t)}{c^2}} \tag{1.6}$$

now denotes the total time that has elapsed in a moving clock between two events. It is obvious that this time lapse is smaller than the corresponding coordinate time interval $(t_2 - t_1)$ showing that moving clocks slow down. We stress that these results hold for a particle moving in an *arbitrary* trajectory and not merely for one moving with uniform velocity. (Special relativity is adequate to describe the

Cambridge University Press
978-0-521-88223-1 - Gravitation: Foundations and Frontiers
T. Padmanabhan
Excerpt
More information

physics involving accelerated motion and one does not require general relativity for that purpose.)

1.3 Transformation of coordinates and velocities

The line interval in Eq. (1.1) is written in terms of a special set of coordinates which are natural to some inertial frame. An observer who is moving with respect to an inertial frame will use a different set of coordinates. Since the concept of simultaneity has no invariant significance, the coordinates of any two frames will be related by a transformation in which space *and* time coordinates will, in general, be different.

It turns out that the invariant speed of light signals allows us to set up a possible set of coordinates for any observer, moving along an arbitrary trajectory. In particular, if the observer is moving with a uniform velocity with respect to the original inertial frame, then the coordinates that we obtain by this procedure satisfy the condition $ds = ds'$ derived earlier. With future applications in mind, we will study the *general* question of determining the coordinates appropriate for an *arbitrary* observer moving along the x-axis and then specialize to the case of a uniformly moving observer.

Before discussing the procedure, we emphasize the following aspect of the derivation given below. In the specific case of an observer moving with a uniform velocity, the resulting transformation is called the *Lorentz transformation*. It is possible to obtain the Lorentz transformation by other procedures, such as, for example, demanding the invariance of the line interval. But once a transformation from a set of coordinates x^a to another set of coordinates x'^a is obtained, we would also like to understand the operational procedure by which a particular observer can set up the corresponding coordinate grid in the spacetime. Given the constancy of the speed of light, the most natural procedure will be to use light signals to set up the coordinates. Since special relativity is perfectly capable of handling accelerated observers we should be able to provide an operational procedure by which *any* observer – moving along an arbitrary trajectory – can set up a suitable coordinate system. To stress this fact – and with future applications in mind – we will first obtain the coordinate transformations for the general observer and then specialize to an observer moving with a uniform velocity.

Let (ct, x, y, z) be an inertial coordinate system. Consider an observer travelling along the x-axis in a trajectory $x = f_1(\tau), t = f_0(\tau)$, where f_1 and f_0 are specified functions and τ is the proper time in the clock carried by the observer. We can assign a suitable coordinate system to this observer along the following lines. Let $\mathcal{P}$ be some event with inertial coordinates (ct, x) to which this observer wants to assign the coordinates (ct', x'), say. The observer sends a light signal from the

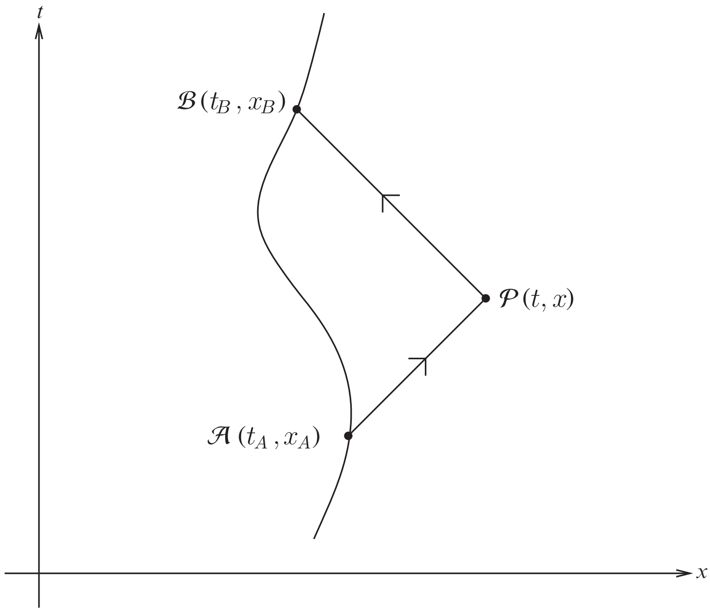

Fig. 1.1. The procedure to set up a natural coordinate system using light signals by an observer moving along an arbitrary trajectory.

event $\mathcal{A}$ (at $\tau = t_A$) to the event $\mathcal{P}$. The signal is reflected back at $\mathcal{P}$ and reaches the observer at event $\mathcal{B}$ (at $\tau = t_B$). Since the light has travelled for a time interval $(t_B - t_A)$, it is reasonable to attribute the coordinates

$$t' = \frac{1}{2}\,(t_B + t_A)\,; \quad x' = \frac{1}{2}\,(t_B - t_A)\,c \tag{1.7}$$

to the event $\mathcal{P}$. To relate (t', x') to (t, x) we proceed as follows. Since the events $\mathcal{P}(t, x)$, $\mathcal{A}(t_A, x_A)$ and $\mathcal{B}(t_B, x_B)$ are connected by light signals travelling in forward and backward directions, it follows that (see Fig. 1.1)

$$x - x_A = c(t - t_A); \quad x - x_B = -c\,(t - t_B)\,. \tag{1.8}$$

Or, equivalently,

$$x - ct = x_A - ct_A = f_1(t_A) - cf_0(t_A) = f_1\left[t' - (x'/c)\right] - cf_0\left[t' - (x'/c)\right], \tag{1.9}$$

$$x + ct = x_B + ct_B = f_1(t_B) + cf_0(t_B) = f_1\left[t' + (x'/c)\right] + cf_0\left[t' + (x'/c)\right]. \tag{1.10}$$

Given f_1 and f_0, these equations can be solved to find (x, t) in terms of (x', t'). This procedure is applicable to *any* observer and provides the necessary coordinate transformation between (t, x) and (t', x').

1.3.1 Lorentz transformation

We shall now specialize to an observer moving with uniform velocity V, which will provide the coordinate transformation between two inertial frames. The trajectory is now $x = Vt$ with the proper time given by $\tau = t[1-(V^2/c^2)]^{1/2}$ (see Eq. (1.6) which can be trivially integrated for constant V). So the trajectory, parameterized in terms of the proper time, can be written as:

$$f_1(\tau) = \frac{V\tau}{\sqrt{1-(V^2/c^2)}} \equiv \gamma V\tau; \qquad f_0(\tau) = \frac{\tau}{\sqrt{1-(V^2/c^2)}} \equiv \gamma\tau, \tag{1.11}$$

where $\gamma \equiv [1-(V^2/c^2)]^{-1/2}$. On substituting these expressions in Eqs. (1.9) and (1.10), we get

$$\begin{aligned} x \pm ct &= f_1\left[t' \pm (x'/c)\right] \pm c f_0\left[t' \pm (x'/c)\right] \\ &= \gamma\Big[\left[Vt' \pm (V/c)x'\right] \pm \left[ct' \pm x'\right]\Big] \\ &= \sqrt{\frac{1 \pm (V/c)}{1 \mp (V/c)}}(x' \pm ct'). \end{aligned} \tag{1.12}$$

On solving these two equations, we obtain

$$t = \gamma\left(t' + \frac{V}{c^2}x'\right); \quad x = \gamma\left(x' + Vt'\right). \tag{1.13}$$

Using Eq. (1.13), we can now express (t', x') in terms of (t, x). Consistency requires that it should have the same form as Eq. (1.13) with V replaced by $(-V)$. It can be easily verified that this is indeed the case. For two inertial frames K and K' with a relative velocity $\boldsymbol{V}$, we can always align the coordinates in such a way that the relative velocity vector is along the common (x, x') axis. Then, from symmetry, it follows that the transverse directions are not affected and hence $y' = y, z' = z$. These relations, along with Eq. (1.13), give the coordinate transformation between the two inertial frames, usually called the *Lorentz transformation*.

Since Eq. (1.13) is a linear transformation between the coordinates, the coordinate differentials (dt, dx^μ) transform in the same way as the coordinates themselves. Therefore, the invariance of the line interval in Eq. (1.1) translates to finite values of the coordinate separations. That is, the Lorentz transformation leaves the quantity

$$s^2(1,2) = |\boldsymbol{x}_1 - \boldsymbol{x}_2|^2 - c^2(t_1 - t_2)^2 \tag{1.14}$$

invariant. (This result, of course, can be verified directly from Eq. (1.12).) In particular, the Lorentz transformation leaves the quantity $s^2 \equiv (-c^2t^2 + |\boldsymbol{x}|^2)$ invariant since this is the spacetime interval between the origin and any event $(t, \boldsymbol{x})$. A quadratic expression of this form is similar to the length of a vector in three dimensions which – as is well known – is invariant under rotation of the coordinate axes. This suggests that the transformation between the inertial frames can be thought of as a 'rotation' in four-dimensional space. The 'rotation' must be in the tx-plane characterized by a parameter, say, χ. Indeed, the Lorentz transformation in Eq. (1.13) can be written as

$$x = x' \cosh \chi + ct' \sinh \chi, \qquad ct = x' \sinh \chi + ct' \cosh \chi, \tag{1.15}$$

with $\tanh \chi = (V/c)$, which determines the parameter χ in terms of the relative velocity between the two frames. These equations can be thought of as rotation by a complex angle. The quantity χ is called the *rapidity* corresponding to the speed V and will play a useful role in future discussions. Note that, in terms of rapidity, $\gamma = \cosh \chi$ and $(V/c)\gamma = \sinh \chi$. Equation (1.13) can be written as

$$(x' \pm ct') = e^{\mp\chi}(x \pm ct), \tag{1.16}$$

showing the Lorentz transformation compresses $(x + ct)$ by $e^{-\chi}$ and stretches $(x - ct)$ by e^{χ} leaving $(x^2 - c^2t^2)$ invariant.

Very often one uses the coordinates $u = ct - x$, $v = ct + x$, $u' = ct' - x'$, $v' = ct' + x'$, instead of the coordinates (ct, x), etc., because it simplifies the algebra. Note that, even in the general case of an observer moving along an arbitrary trajectory, the transformations given by Eq. (1.9) and Eq. (1.10) are simpler to state in terms of the (u, v) coordinates:

$$-u = f_1(u'/c) - cf_0(u'/c); \qquad v = f_1(v'/c) + cf_0(v'/c). \tag{1.17}$$

Thus, even in the general case, the coordinate transformations do not mix u and v though, of course, they will not keep the form of $(c^2t^2 - |x|^2)$ invariant. We will have occasion to use this result in later chapters.

The non-relativistic limit of Lorentz transformation is obtained by taking the limit of $c \to \infty$ when we get

$$t' = t, \quad x' = x - Vt, \quad y' = y, \quad z' = z. \tag{1.18}$$

This is called the *Galilean transformation* which uses the same absolute time coordinate in all inertial frames. When we take the same limit ($c \to \infty$) in different laws of physics, they should remain covariant in form under the Galilean transformation. This is why we mentioned earlier that the statement (i) on page 2 is not specific to relativity and holds even in the non-relativistic limit.

Exercise 1.1

Light clocks A simple model for a 'light clock' is made of two mirrors facing each other and separated by a distance L in the rest frame. A light pulse bouncing between them will provide a measure of time. Show that such a clock will slow down exactly as predicted by special relativity when it moves (a) in a direction transverse to the separation between the mirrors or (b) along the direction of the separation between the mirrors. For a more challenging task, work out the case in which the motion is in an arbitrary direction with constant velocity.

1.3.2 Transformation of velocities

Given the Lorentz transformation, we can compute the transformation law for any other physical quantity which depends on the coordinates. As an example, consider the transformation of the velocity of a particle, as measured in two inertial frames. Taking the differential form of the Lorentz transformation in Eq. (1.13), we obtain

$$dx = \gamma\left(dx' + V dt'\right), \quad dy = dy', \quad dz = dz', \quad dt = \gamma\left(dt' + \frac{V}{c^2}dx'\right), \tag{1.19}$$

and forming the ratios $\boldsymbol{v} = d\boldsymbol{x}/dt, \boldsymbol{v}' = d\boldsymbol{x}'/dt'$, we find the transformation law for the velocity to be

$$v_x = \frac{v_x' + V}{1 + (v_x' V/c^2)}, \quad v_y = \gamma^{-1}\frac{v_y'}{1 + (v_x' V/c^2)}, \quad v_z = \gamma^{-1}\frac{v_z'}{1 + (v_x' V/c^2)}. \tag{1.20}$$

The transformation of velocity of a particle moving along the x-axis is easy to understand in terms of the analogy with the rotation introduced earlier. Since this will involve two successive rotations in the t–x plane it follows that we must have additivity in the rapidity parameter $\chi = \tanh^{-1}(V/c)$ of the particle and the coordinate frame; that is we expect $\chi_{12} = \chi_1(v_x') + \chi_2(V)$, which correctly reproduces the first equation in Eq. (1.20). It is also obvious that the transformation law in Eq. (1.20) reduces to the familiar addition of velocities in the limit of $c \to \infty$. But in the relativistic case, none of the velocity components exceeds c, thereby respecting the existence of a maximum speed.

The transverse velocities transform in a non-trivial manner under Lorentz transformation – unlike the transverse coordinates, which remain unchanged under the Lorentz transformation. This is, of course, a direct consequence of the transformation of the time coordinate. An interesting consequence of this fact is that the *direction* of motion of a particle will appear to be different in different inertial frames. If $v_x = v\cos\theta$ and $v_y = v\sin\theta$ are the components of the velocity in the coordinate frame K (with primes denoting corresponding quantities in the frame