Chapter 2
The Lorentz-Transformation

Abstract The notion of the coordinate time is introduced. The interrelation between constancy of c and synchronization is analyzed. Lorentz-transformations are derived and using them the relativistic effects are reconsidered. The causality paradox is discussed and the twin paradox is described from the point of view of both siblings.

Keywords Time · Synchronization · Lorentz-transformation · Causality · Twin paradox

2.1 The Coordinate Time

Physical processes take place in space and time. Their ultimate constituents are the pointlike instantaneous *events* characterized by four spacetime coordinates, i.e. by three space coordinates and a moment of time. *Spacetime* itself is the four-dimensional continuum of these points.

Intuition is a much less reliable guide, concerning coordinate systems, in spacetime than in three-dimensional space alone. Therefore, even the simplest spacetime coordinate systems require careful description.

As it has already been stressed in Sect. 1.1, the importance of coordinate systems ($\mathcal{K}$) lies in the fundamental role they play in theoretical calculations. In sharp contrast to reference frames ($\mathcal{R}$) and inertial frames ($\mathcal{I}$), coordinate systems exist only in our minds: think of the description of planets' motion in Newtonian physics. When working in a given reference frame, the most convenient coordinate system is that in which our reference frame is at rest. Such a coordinate system is called *attached* to the reference frame: $\mathcal{R}\longleftrightarrow\mathcal{K}, \mathcal{R}'\longleftrightarrow\mathcal{K}'$ or $\mathcal{I}\longleftrightarrow\mathcal{K}, \mathcal{I}'\longleftrightarrow\mathcal{K}'$.

When, for example, a railcar is chosen to serve as a reference frame, one can imagine Cartesian coordinates attached to it whose origin is at the center of the car, x-axis is parallel to the rails and z-axis points upwards perpendicular to Earth's

P. Hraskó, *Basic Relativity*, SpringerBriefs in Physics,
DOI: 10.1007/978-3-642-17810-8_2,

surface. Then the location at the moment t of a point mass is determined by the triplet $x(t),\ y(t),\ z(t)$ of functions of time. Here t is the same time that in Sect. 2.1 has also been denoted by this letter and was advised to understand and use in the same way as learned in secondary school physics. Now ‘it is time’ to analyse it in some more detail.

Let us first give a name to it: it will be called *coordinate time*. This is a very appropriate term since on graphical representations of the function $s = f(t)$, which describes the motion, one of the coordinate axes is t-axis.

It was, perhaps, Galileo who, in his famous experiments with balls rolling down a slope, used for the first time the notion of time in the same sense we give now to coordinate time. Galileo measured the time it took to reach marked distances on the slope and established that this time is proportional to the square root of the corresponding distance. To measure the time, he made a hole in the bottom of a bucket, suspended a jar immediately below it and filled the bucket with water, keeping the hole closed. He let water to flow from the bucket into the jar only during the time the ball was rolling from its initial position on the top of the slope up to one of the marks. The weight of the water in the jar was then proportional to the time during which the ball travelled the distance chosen. The experiment was naturally a very inaccurate one but as we will show now its idealized version permits us to formulate precisely what coordinate time means.

Galileo had to roll the ball many times before he could draw the conclusion which is now written in the form $s = \frac{1}{2}gt^2$. In a modern version of the experiment a single rolling down would suffice. Imagine that the slope is bordered with photoelectric cells each of which is provided by appropriate electric circuitry to generate a sharp signal at the moment when the ball passes by. The signals are collected in a time delay analyzer controlled by an ideal clock. Then the output of the analyzer will be the collection of points of the path versus time curve $s = f(t)$. Since the number of the photoelectric cells may in principle be arbitrarily large the curve can be determined very accurately.

This accuracy, however, cannot be made *arbitrarily high* since the method has an inbuilt drawback: the necessity to transmit signals from one place to another. The time interval required for the signals to travel from the photoelectric cells to the analyzer must obviously be the same for all of the cells since otherwise the curve at the output of the analyzer would be distorted. Therefore, all the cables must be of equal length. But this is not enough because the speed of the signals, though largely determined by the construction of the cables, is slightly influenced also on how they are mounted, i.e. on their shape, environment and mutual arrangement. Since it is certainly impossible to ensure completely identical conditions for all the cables, methods, requiring signal transmittion, must be abandoned.

This difficulty, however, can be easily circumvented if every photoelectric cell is provided by its own clock and camera. In this case, the signal of any particular cell may be used to take a snapshot of the face of its clock at the moment when the ball rolls by. These data can also be used to reconstruct the law of motion $s = f(t)$, provided the clocks are *synchronized correctly*.

The most natural method of synchronisation consists in collecting all the clocks together in a suitable place where they can be easily surveyed and synchronized by bringing their hands simultaneously into the same position. Since they are of ideal construction by assumption, the rhythm of all of them is exactly the same and they will retain their synchronism forever. Therefore, they can be transported safely to their original places without the risk of being desynchronized (*Newtonian synchronization*).

The persistence of syncronization can be verified by means of the following procedure. After synchronization having been performed but clocks still being together, one of them is transported to a distant place and back again. If in the course of transportation its synchronism with the rest of clocks is not spoiled, Newtonian method of synchronization may be accepted as a correct one.

Now we are at a position to formulate the general definition of the coordinate time: *coordinate time would be measured by a set $\mathcal{S}$of uniformly scattered correctly synchronized ideal clocks at rest if they were indeed there*. As we could observe in Chap. 1, the *meaning* of basic physical quantities, such as velocity, momentum, energy, etc., is the same in both Newtonian and relativistic physics; it is their *properties* which are different, sometimes dramatically, in the two theories. The same is true for coordinate time. Its definition as given above in the context of Newtonian physics remains true in relativity theory as well. It is only the *method of synchronization* which is to be modified since, owing to twin paradox (or time dilation), the Newtonian synchronization would not, as a matter of fact, pass the test procedure described above.[1]

But relativity theory offers an alternative based on the constancy of light speed in inertial frames (*Einstein synchronization*). Consider a pair of clocks A and B, resting in an inertial frame at some distance from each other. A light signal sent from clock A at the moment of time t_{A1} is immediately reflected by a mirror from clock B and arrives back to clock A at the moment t_{A2}. Both t_{A1} and t_{A2} are, of course, read off from the clock A. Let the moment of reflection be t_B as seen on clock B. Now, if the two clocks were correctly synchronized, the equality of light velocity in both direction could be expressed by the equation $t_B - t_{A1} = t_{A2} - t_B$ from which $t_B = \frac{1}{2}(t_{A1} + t_{A2})$. If at the moment of reflexion the actual time on clock B was t'_B then it should be set ahead by an amount $t_B - t'_B = \frac{1}{2}(t_{A1} + t_{A2}) - t'_B$.

A and B may be either two particular real clocks or two arbitrary members of the virtual multitude $\mathcal{S}$ of clocks which define coordinate time in the inertial frame $\mathcal{I}$ chosen. Then constancy of light speed in $\mathcal{I}$ ensures that the distance Δl between two arbitrary points, travelled by light in time Δt, is given by the equation

$$\Delta l^2 = \Delta x^2 + \Delta y^2 + \Delta z^2 = c^2 \Delta t^2. \tag{2.1.1}$$

Since inertial frames are moving with respect to each other, all of them are provided by its own set of virtual clocks at rest, which show coordinate time in it.

[1] Twin paradox vanishes only if transportation velocity is equal to zero. But then the procedure would last infinitely long.

Individual clocks in each particular set are assumed to be correctly synchronized but, owing to time dilation, no synchronism can exist between a pair of clocks, belonging to different inertial frames. The only possibility to relate coordinate times of two different inertial frames consists in selecting a particular clock in both of them which are found at the same point of space and to assign on them the same time to their moment of encounter.

The standard procedure is this: let the inertial frames $\mathcal{I}$ and $\mathcal{I}'$ be equipped with Cartesian coordinates $\mathcal{K}$ and $\mathcal{K}'$, the corresponding axes of which are parallel to each other and, moreover, their x-axes are in common. Their relative motion is along this common axis. Both $\mathcal{I}$ and $\mathcal{I}'$ have their own sets $\mathcal{S}$ and $\mathcal{S}'$ of virtual clocks which define coordinate time in them. Let us denote the particular members of $\mathcal{S}$ and $\mathcal{S}'$ at the origins of $\mathcal{K}$ and $\mathcal{K}'$ by $\mathcal{O}$ and $\mathcal{O}'$. The coordinate times in the two frames are then related by the assumption that $\mathcal{O}$ and $\mathcal{O}'$ show zero time at the moment of their encounter (*standard setting*).

Remember now the definition of the proper time of a body: it would be measured by an ideal clock if it was attached to the body. Evidently, proper time and coordinate time are logically independent notions and, therefore, they need not coincide numerically. In Newtonian physics they actually do, in complete agreement with the Newtonian sychronization procedure. The splitting of the unique notion of time into two was recognized only in relativity theory.

2.2 Independence of the Constancy of *c* from Synchronization

It is sometimes claimed that Einstein synchronization of distant clocks A and B is circular. The argument is very simple: Einstein synchronization is based on the equality of light velocity on the path from A to B and back from B to A, but measurement of light velocity in one direction between two distant points is impossible unless the clocks at these points have already been synchronized.

This argument is, however, fallacious. It is true that one-way measurement of light velocity can be performed only if clocks at the endpoints are synchronized correctly. But since they need not show the correct coordinate time, they can be synchronized without light signals by transporting them from a common site in a symmetrical manner. The procedure consists of the following steps:

1. Let our laboratory be an inertial frame. To ascertain this no measurement of time is required: isolated bodies at rest must remain so and axes of gyroscopes must be pointing continuously to the same piece of the wall.
2. Two ideal clocks are synchronized at a common site and transported in a symmetrical way into the points A and B. The symmetry of transportation is ensured by mounting them on two completely identical 'land rovers' which start to move at the same moment of time in exactly opposite direction. Their motion is controlled by the clocks themselves by means of identical computers, executing identical programs.

3. At the points A and B the clocks are either gaining or losing with respect to the coordinate time (they do not belong to the set S) but since their synchronism is not destroyed they make it possible to compare light velocities in the direction $A \longrightarrow B$ and $B \longrightarrow A$.
4. The procedure is suitable to measure even *the value* of light velocity, since the distance between the points can be calculated from the number of turns and perimeter of the 'land rover's wheels. Repeating it many times, the isotropy of light velocity can be verified.

As we see, the thought experiment described is capable to prove constancy of light speed if it is true, or to disprove it if it is false. It provides, therefore, solid logical foundation for Einstein's synchronization prescription.

2.3 The Minkowski Coordinates

If an inertial frame is endowed with both Cartesian coordinates and a set S of correctly synchronized virtual clocks then it is called that a *spacetime coordinate system* $(\mathcal{K}, \mathcal{S})$ is attached to it (*Minkowski coordinates*). Instead of these pair, Minkowski coordinate systems will mostly be denoted by the single symbol $\mathcal{M}$. As coordinate systems generally do, Minkowski coordinate system too exisits only in our imagination. It serves to assign four well-defined spacetime coordinates to any pointlike instantaneous event: three space coordinates x, y, z in $\mathcal{K}$ and a moment of time shown on the virtual clock at this point.

Consider two inertial frames $\mathcal{I}$ and $\mathcal{I}'$ with Minkowski coordinates $(\mathcal{K}, \mathcal{S}) \equiv \mathcal{M}$ and $(\mathcal{K}', \mathcal{S}') \equiv \mathcal{M}'$ attached to them. Assume that a flash of light is given off at some point at a given moment of time (event A) and observed at a distant point at a later moment of time (event B). Either of these events have both primed and unprimed Minkowski coordinates whose differences are

$$\begin{array}{ll} \Delta x = x_B - x_A & \Delta x' = x'_B - x'_A \\ \Delta y = y_B - y_A & \Delta y' = y'_B - y'_A \\ \Delta z = z_B - z_A & \Delta z' = z'_B - z'_A \\ \Delta t = t_B - t_A & \Delta t' = t'_B - t'_A \end{array}$$

The equality of light speed in both $\mathcal{I}$ and $\mathcal{I}'$ is then expressed by the relation

$$\frac{\Delta l}{\Delta t} \equiv \frac{\sqrt{\Delta x^2 + \Delta y^2 + \Delta z^2}}{\Delta t} = c \Longleftrightarrow \frac{\Delta l'}{\Delta t'} \equiv \frac{\sqrt{\Delta x'^2 + \Delta y'^2 + \Delta z'^2}}{\Delta t'} = c,$$

which can also be expressed in the form

$$\Delta x^2 + \Delta y^2 + \Delta z^2 = c^2 \Delta t^2 \Longleftrightarrow \Delta x'^2 + \Delta y'^2 + \Delta z'^2 = c^2 \Delta t'^2, \qquad (2.3.1)$$

where $\Longleftrightarrow$ means that either part of the relation follows from the other.

Let us now suppose that the two Minkowski coordinate systems constitute a standard setting as defined at the end of Sect. 2.1. Then the origin of $\mathcal{K}$ and $\mathcal{K}'$ are moving on their common x-axis and the virtual clocks at the origin of both of them show zero time at the moment of their encounter. Assume that the event A (the light flash) takes place at just this moment of time at the common origin of $\mathcal{K}$ and $\mathcal{K}'$. Then all the eight A-coordinates are equal to zero and (2.3.1) is reduced to

$$x^2 + y^2 + z^2 = c^2t^2 \Longleftrightarrow x'^2 + y'^2 + z'^2 = c^2t'^2 \tag{2.3.2}$$

(subscript B has been omitted). The meaning of this relation is this: if E is any event with spacetime coordinates $x,\ y,\ z,\ t$ in $\mathcal{M}$ and $x',\ y',\ z',\ t'$ in $\mathcal{M}'$ and $x^2 + y^2 + z^2 = c^2t^2$ then $x'^2 + y'^2 + z'^2 = c^2t'^2$ is also true and vice versa.

In Sect. 1.1 it has been stated that in their original form Maxwell equations are valid in the inertial frames. But Maxwell equations can be written down only if the spacetime coordinate system has already been specified. The most natural assumption is that these equations are valid in Minkowski coordinates, the decisive argument being, that the constancy of light speed which is the consequence of Maxwell equations, is the basic property of Minkowski coordinates attached to any inertial frame.

The question now arises how to assign spacetime coordinates to an accelerating (non-inertial) reference frame $\mathcal{R}$. The problem is that in such frames light velocity is not isotropic and is, in general, even time dependent. Einstein synchronization is, therefore, out of the question. If Newtonian synchronization was a valid procedure it could be applied in any reference frame but Einstein synchronization is restricted to inertial frames.

But this is not a drawback since accelerating frames are always related in some definite way to the inertial ones. Newtonian equations of motion, for example, are originally given in the form suitable only in inertial frames. When they are applied in an accelerating frame $\mathcal{R}$, their mathematical form can be computed from the known relation of $\mathcal{R}$ to some inertial frame $\mathcal{I}$. Similarly, when in relativity theory a calculation is performed, for some reason or another, in a non-inertial frame $\mathcal{R}$, the coordinate time is always chosen in relation to the Minkowski coordinates of an inertial frame (see Sect. 2.12 for an example of such a procedure).

2.4 The Lorentz-Transformation

Let us take up the discussion of the previous section which led to (2.3.1) and (2.3.2) and ask: How to calculate in a standard setting the primed spacetime coordinates of an event E if its unprimed coordinates are given? Formulae, answering this question, are known as the $Lorentz-transformation$.

The analogous question in Newtonian physics can be answered immediately: it is the $Galilei-transformation$

$$t' = t, \quad x' = x - Vt, \quad y' = y, \quad z' = z \tag{2.4.1}$$

which connects the primed and unprimed spacetime coordinates of an event in a standard setting of $\mathcal{K}$ and $\mathcal{K}'$ attached to $\mathcal{I}$ and $\mathcal{I}'$. In (2.4.1) V is the constant velocity of $\mathcal{I}'$ with respect to $\mathcal{I}$ along the common x-axis in positive direction if $V > 0$ and in negative direction if $V < 0$. For $V = 0$ (2.4.1) reduces to the identity transformation as should be.

Galilei-transformations are *linear transformations*, hence they have the same form when applied to coordinate differences of a *pair* of events:

$$\Delta t' = \Delta t, \quad \Delta x' = \Delta x - V\Delta t, \quad \Delta y' = \Delta y, \quad \Delta z' = \Delta z.$$

This property is an essential one. It is the consequence of the fact that the origin of a coordinate system attached to a given reference frame can be freely chosen. Its choice, therefore, must not lead to observable consequences. The coordinates themselves depend on the position of the origin but the coordinate differences are independent of it. The form containing Δ-s is, therefore, the more fundamental one. Formulae (2.4.1) are meaningful only with respect to the standard setting in which the origin of the Cartesian coordinates and the zero moment of time (i.e. the origin of the spacetime coordinate systems attached to the inertial frames) have already been chosen.

The transformations (2.4.1) are, moreover, homogeneous since no constant term independent of the unprimed coordinates is found on their right hand side. This is the direct consequence of the fact that in a standard setting the primed coordinates of the event at $x = y = z = t = 0$ are equal also to zero: $x' = y' = z' = t' = 0$.

Considerations concerning homogeneous linearity evidently apply to Lorentz-transformations too. In order to write (2.3.1) as (2.3.2) we had to take event A of the light flash for the origin of the coordinate system. By this choice we tacitly assumed that the origin of an attached coordinate system may be anywhere in spacetime. As a matter of fact, (2.3.1) and (2.3.2) can be both valid only for linear transformations. As for homogeniety, it remains valid for the same reason as above.

The peculiarity of Lorentz-transformations lies in the requirement (2.3.2) which they have to obey. Galilei transformations evidently violate it. Our task is, therefore, to find those homogeneous linear transformations with V-dependent coefficients which obey (2.3.2) and for $V = 0$ reduce to the identity transformation.

As a first step suppose that we already know the formulae of Lorentz-transformations and substitute them into the expression $c^2t'^2 - x'^2 - y'^2 - z'^2$. We then obtain a quadratic expression of unprimed coordinates which, according to (2.3.2), must vanish when $c^2t^2 - x^2 - y^2 - z^2 = 0$. This can only happen if it is *proportional* to this expression:

$$c^2t'^2 - x'^2 - y'^2 - z'^2 = F(V) \cdot \left(c^2t^2 - x^2 - y^2 - z^2\right). \tag{2.4.2}$$

The constant of proportionality can, of course, depend on V. We show now that it is an even function of V.

We have already made use of the fact that the origin of the attached coordinates may be freely chosen. This playes the role of a constraint on the possible form of Lorentz-transformations, enforcing them to be linear. $F(V)$ is constrained to be an even function due to another freedom in the choice of the standard setting. Let us rotate the Cartesian coordinate systems $\mathcal{K}$ and $\mathcal{K}'$ in a standard setting by 180° around their respective z-axes. We arrive at another standard setting attached to the same bodies in which, however, the velocity of $\mathcal{I}'$ with respect to $\mathcal{I}$ has opposite sign. Since both settings are standard ones, neither V nor $-V$ is the 'true' relative velocity and if $F(V)$ and $F(-V)$ were different from each other it would be impossible to make a choice between them. We are, therefore, compelled to assume that $F(-V) = F(V)$.

The inverse of a Lorentz-transformation is also a Lorentz-transformation with V replaced by $-V$. Applying (2.4.2) to it we will have

$$c^2t^2 - x^2 - y^2 - z^2 = F(-V) \cdot \left(c^2t'^2 - x'^2 - y'^2 - z'^2\right).$$

Substituting this into (2.4.2), the relation

$$c^2t^2 - x^2 - y^2 - z^2 = F(V) \cdot F(-V) \cdot \left(c^2t^2 - x^2 - y^2 - z^2\right)$$

is obtained, from which $F(V) \cdot F(-V) = 1$ follows. Since $F(-V) = F(V)$ we find that $F(V) = \pm 1$. But Lorentz-transformation for $V = 0$ reduces to the identity transformation. Hence (2.4.2) leads to $F(0) = +1$ and so $F(V) = 1$. We have, therefore,

$$c^2t'^2 - x'^2 - y'^2 - z'^2 = c^2t^2 - x^2 - y^2 - z^2 \tag{2.4.3}$$

which is a constraint much stronger than (2.3.2). Since the transformation is linear (2.4.3) is applicable also to coordinate differences.

In order to constraint further the possible form of the Lorentz-transformations, we will consider a version of Einstein's train thought experiment in which the explosives are above and below the source of light at equal distances from it. In this setting constancy of light velocity leads to no unexpected consequences: The explosions are simultaneous with respect to both the car and the platform.

In the corresponding standard setting x-axis is parallel to the rails. If the vertical direction is taken for the z-axis, we will have the following relations between the coordinate differences of explosions:

$$\begin{aligned} \Delta t' &= \Delta t = 0, \quad \Delta x' = \Delta x = 0 \\ \Delta y' &= \Delta y = 0, \quad |\Delta z'| = |\Delta z| \neq 0. \end{aligned}$$

Analogous relations (with the role of y and z directions exchanged) are obtained when the explosives are installed in the y direction with respect to the light source. The corresponding relations for the *original* version of the experiment are

$$\begin{aligned} &\Delta t' \neq 0, \quad \Delta t = 0, \quad \Delta x' \neq 0, \quad \Delta x \neq 0 \\ &\Delta y' = \Delta y = 0 = \Delta z' = \Delta z = 0, \end{aligned}$$

the unprimed frame being the rest frame of the train.

To meet these requirements Lorentz-transformations should be of the form

$$\left.\begin{aligned} ct' &= A(V) \cdot ct + B(V) \cdot x \\ x' &= C(V) \cdot ct + D(V) \cdot x \\ y' &= y; \quad z' = z. \end{aligned}\right\} \tag{2.4.4}$$

The use of ct instead of t is expedient because the former has the same dimension of length as space coordinates do and all the coefficients will be dimensionless. Due to the special form of (2.4.4) relation (2.4.3) reduces to

$$c^2t'^2 - x'^2 = c^2t^2 - x^2. \tag{2.4.5}$$

The coefficients $A(V)$ and $B(V)$ can be determined by their physical meaning. Consider, for example, a clock at the origin of $\mathcal{K}$ and the events there when the readings on the clock are t and $t + \Delta t$. For this pair of events we have $\Delta x = 0$ and, according to the first of the equations (2.4.4), the time $\Delta t'$ which elapsed betwen them in $\mathcal{K}'$ is equal to $A(V)\Delta t$. But Δt is evidenly equal to the proper time $\Delta\tau$ between the events and, since the clock moves with the velocity -V in $\mathcal{K}'$, $\Delta t'$ must be equal to $\Delta\tau/\sqrt{1-(-V)^2/c^2} = \Delta\tau/\sqrt{1-V^2/c^2}$. Hence

$$A(V) = \frac{1}{\sqrt{1-V^2/c^2}} \equiv \gamma V.$$

Return now to the original setting of Einstein's train thought experiment again as discussed in Sect. 1.3. There V denoted the velocity of the train and we retain this meaning of it. If $\mathcal{I}$ is the rest frame of the train then for the explosions we have $\Delta t = 0$ and $\Delta x = l_0$ where l_0 is the proper length of the traincar. In the rest frame $\mathcal{I}'$ of the platform, which moves with the velocity -V with respect to the train, the time $\Delta t'$ elapsed between the explosions is equal to $B(-V)l_0$. But we know from Problem 2 that $\Delta t' = V \cdot l_0/(c^2\sqrt{1-V^2/c^2})$, therefore $B(-V) = V/(c\sqrt{1-V^2/c^2})$ and so

$$B(V) = -\frac{V/c}{\sqrt{1-V^2/c^2}}.$$

The rest of the coefficients follows from (2.4.5). Substituting (2.4.4) into this relation and equating the coefficient of c^2t^2, x^2 and ctx, we have

$$A^2 - C^2 = 1, \quad B^2 - D^2 = -1, \quad AB - CD = 0. \tag{2.4.6}$$

Substituting A and B, we obtain $C(V) = +B(V)$ and $D(V) = +A(V)$ or $C(V) = -B(V)$ and $D(V) = -A(V)$. The correct solution is the first one, since at $V = 0$ (2.4.4) must reduce to the identity transformation, and so we must have $D(0) = +A(0) = 1$ [and $C(0) = B(0) = 0$]. The Lorentz-transformations are, therefore, given by the formulae

$$\left.\begin{aligned} ct' &= \frac{ct - \frac{V}{c}x}{\sqrt{1-V^2/c^2}} \\ x' &= \frac{x - Vt}{\sqrt{1-V^2/c^2}} \\ y' &= y; \quad z' = z \end{aligned}\right\} \tag{2.4.7}$$

In the $c \longrightarrow \infty$ limit they become identical to Galilei-transformations (2.4.1).

Lorentz-transformations are often called *pseudorotations* of the Minkowski coordinate system around the origin. The term is based on the similarity of (2.4.3) to the equation

$$x'^2 + y'^2 + z'^2 = x^2 + y^2 + z^2,$$

which holds true for rotations of Cartesian coordinates in 3D (three dimensional) geometric space. The transformations (2.4.7) are, therefore, pseudorotations in the (ct, x) plane of spacetime. To stress that they are not the most general case of Lorentz-transformations, the name *boost* is sometimes used to denote them, but in what follows we will continue to call them Lorentz-transformation in conformity with general usage.

In the last step of derivation of (2.4.7) we made use of our prior knowledge of time dilation and train thought experiment. Either of them would suffice. If no such prior information is available at all we may start from the equation $x = Vt$ of the origin of $\mathcal{K}'$ in $\mathcal{K}$ which in $\mathcal{K}'$ is simply $x' = 0$. Using (2.4.4), this last equation can be expressed through unprimed coordinates as $C.(ct) + Dx = 0$ or $x = -\frac{C}{D}(ct)$. Since this is just another form of the equation $x = Vt$, we have $C/D = -V/c$. This formula together with (2.4.6) is sufficient to derive (2.4.7).

The material of Chap. 1 is not, therefore, a necessary prerequisite to derive Lorentz-transformations. On the contrary, most of phenomena discussed in Chap. 1 can be obtained in a systematic way from Lorentz-transformations. Nevertheless, Chap. 1 does not seem at all superfluous. First, Lorentz-transformation rests on Einstein synchronization whose proper understanding is greatly facilitated by prior knowledge of time dilation. Second, mass–energy relation is a piece of knowledge independent of Lorentz-transformation. Relativistic dynamics requires both of them for its foundation (see Sect. 2.19).

2.5 Classification of Spacetime Intervals

Both three dimensional geometric space and spacetime of special relativity are *manifolds* equipped with *metric*.

Points of geometric space and spacetime are in one-to-one correspondence with triples or quadruples of real numbers respectively; the former is, therefore, a three dimensional manifold, the latter is a four dimensional one. The correspondence can be accomplished in infinitely many different ways, all of them being a possible *coordinate system* in the corresponding manifold.

The possession of a metric is the expression of the fact that any two points of the manifold are at a well-defined 'squared distance' from each other. The squared distance is, in principle, a measurable quantity and, therefore, independent of the coordinate system chosen: it is *invariant*. In 3D geometric space the formula for the squared distace has its simplest mathematical form in Cartesian coordinates while in spacetime it is most simple in Minkowski coordinates:

$$\Delta l^2 = \Delta x^2 + \Delta y^2 + \Delta z^2, \tag{2.5.1}$$

$$\Delta s^2 = c^2\Delta t^2 - \Delta x^2 - \Delta y^2 - \Delta z^2 \tag{2.5.2}$$

(for infinitesimal intervals $\Delta \longrightarrow d$). Formula (2.5.1) refers to 3D *euclidean* space. The meaning of the adjective 'euclidean' is that Δl^2 is positive definite, since it contains only positive terms. On the other hand, spacetime is a *pseudo−euclidean* manifold since the expression of Δs^2 contains both positive and negative terms. As a consequence, it may be equal to zero or may even take on negative values. It is, therefore, quite misleading to write it as square of a quantity but, since it is the spacetime counterpart of Δl^2, the notation Δs^2 is an appropriate one.

An important aspect of (2.5.1) and (2.5.2) is expressed in the form of negation. The meaning of (2.5.1), for example, is not fully exhausted by the statement that the sum on the right hand side has the same value in all of the coordinate systems which are rotated with respect to each other; its full content consists in the assertion that it is *the sum only* which remains invariant. Its separate terms depend on the coordinate system chosen and so they are of no intrinsic geometric significance.

The same is true for (2.5.2). A pair of events defines a unique Δs^2 between them, but their time difference and spatial distance depend on the coordinate system chosen. Since different inertial frames involve different Minkowski coordinate systems, this property of spacetime is an expression of the relativity of simultaneity and of the Lorentz contraction on geometric language.

Depending on the sign of Δs^2, pairs of events (or spacetime intervals between them) are of three fundamentally different types. Since Δs^2 is invariant, the type of a given pair is the same with respect to all the inertial frames. Let the events be denoted by E_a and E_b and assume that in the inertial frame $\mathcal{I}$ chosen E_a precedes E_b: $\Delta t = t_b - t_a > 0$ (and $\Delta x = x_b - x_a$). For the sake of simplicity we assume that $\Delta y = \Delta z = 0$.

1. When $\Delta s^2 = 0$ the pair is called *lightlike*, because E_a and E_b can be mediated by light pulse. Indeed, since $\Delta s^2 = c^2\Delta t^2 - \Delta l^2$, for $\Delta s^2 = 0$ we have $\Delta t = \Delta l/c$.
2. If $\Delta s^2 > 0$ the events are *timelike*. For timelike pairs (and only for them) an inertial frame $\mathcal{I}_0$ is always found in which the events happen at the same place. The time $\Delta t_0 = t_{0b} - t_{0a}$ elapsed between the events in $\mathcal{I}_0$ is proportional to Δs since, owing to $\Delta x_0 = x_{0b} - x_{0a} = 0, \Delta t_0^2 = \Delta s^2/c^2$.
   Two consecutive events on a pointlike body always constitute a timelike pair. If in $\mathcal{I}$ the body moves with constant velocity then $\mathcal{I}_0$ is its rest frame and Δt_0 is identical to the proper time interval between the events on it: $\Delta\tau^2 = \Delta s^2/c^2$.
3. When, finally, $\Delta s^2 < 0$ the pair is *spacelike*. According to the first equation of (2.4.7), in $\mathcal{I}'$ the time elapsed between them is equal to

$$\Delta t' = \frac{\Delta t - (V/c^2)\Delta x}{\sqrt{1 - V^2/c^2}} = \Delta t \times \frac{1 - V/V_\Delta}{\sqrt{1 - V^2/c^2}}, \tag{2.5.3}$$

where $V_\Delta = c^2(\Delta t/\Delta x)$ is a quantity of the dimension of velocity. For a spacelike pair $V_\Delta < c$. The proof is simple: for $\Delta s^2 < 0$ we have

$$V_\Delta^2 = c^2 \frac{c^2 \Delta t^2}{\Delta x^2} = c^2 \frac{\Delta x^2 + \Delta s^2}{\Delta x^2} < c^2.$$

Therefore, an inertial system $\mathcal{I}_\Delta$ exists which in $\mathcal{I}$ moves with velocity V_Δ. As it can be seen from (2.5.3), in this frame the events E_a and E_b are simultaneous. In the frames $\mathcal{I}'$ for which $1 < V/V_\Delta$, E_b precedes $E_a(t'_b < t'_a)$ while in those with $1 > V/V_\Delta$ the event E_a takes place earlier. The conclusion is that *time order within a pair of spacelike events depends on the inertial frame chosen.*[2]

2.6 Spacetime Diagrams

Spacetime diagrams (or Minkowski diagrams) help in visualisation of the content of Lorentz-transformations. Representation of Galilei-transformations by analogous diagrams is also possible but, owing to the simplicity of these transformations, they are never used. Yet we are going to start with this latter type of diagrams because they may be instrumental in proper understanding of how and why diagrams of this kind may be useful.

Both (2.4.1) and (2.4.7) refer to standard setting in which coordinates y and z remain intact. Let us denote the residual (t, x) part of the coordinate system, attached to $\mathcal{I}$, by $\mathcal{M}_2$. On a sheet of paper it can be represented as a rectilinear coordinate system (t, x) whose vertical axis is, by tradition, the t-axis.

The points of the (t, x) coordinate plane represent events. For any point, the position and the moment of time of the corresponding event is obtained by parallel projection along the axes. The motion of a point mass is a continuous array of events which on the diagram is represented by a continuous curve called *world line*. In Newtonian physics world lines can be of any form with proviso that to any t there corresponds a unique value of x since no body can be found in several different places at the same moment of time. To a given x, on the contrary, there belong as many different moments of time as many times the body returns to the same place x.

The world line of a uniformly moving body is a straight line $x = vt + x_0$. The angle β made by this line with the t-axis is connected to the velocity trough the formula $v = \tan \beta$. For a point mass at rest $\beta = 0$: the world line of a body at rest is parallel to the t-axis. In particular, the world line of a point mass, resting at the origin $x = 0$, is the t-axis itself, and it is often convenient to look at the t-axis as the $x = 0$ axis.

---

[2] In the train thought experiment, for example, explosions are simultaneous in the train's rest frame, the explosion at the rear end is the first one in the platform's rest frame and this time order is reversed as seen from another train, moving faster in the same direction

Straight lines parallel to the x-axis cannot be world lines of point masses since they belong to infinite velocity: $\beta = 90^\circ$ and so $v = \tan 90^\circ = \infty$. The events they contain are in $\mathcal{M}_2$ simultaneous to each other (the same t belongs to all of them). In particular, all the events on the x-axis take place at the moment $t = 0$ and the x-axis is, therefore, often called the $t = 0$ axis.

What has so far been said refers to both Newtonian and relativistic physics (in the latter case the velocity on the world lines must be smaller than c). They are summarized on Fig. 2.1. Let us now confine ourselves to Galilean-transformations. When, using (2.4.1), we change our unprimed coordinates $\mathcal{M}_2$ to primed ones, attached to another inertial frame, nothing happens to the event E: it remains *the same event* though referred to two different coordinate systems. The correct representation of this process is, therefore, to leave the point which represents E at its original place, but to replace the unprimed rectilinear coordinate system $\mathcal{M}_2$ with an oblique $\mathcal{M}'_2$ one in such a way that the coordinates (t', x') of E, obtained by parallel projection along the new axes, be equal to those given by (2.4.1).

This is very easy to do. According to Galilei-transformations (2.4.1), equation $x' = 0$ of the origin of $\mathcal{M}'_2$ is described in $\mathcal{M}_2$ by the equation $x' = x - Vt = 0$. The straight line $x = Vt$ passes through the origin and makes an angle $\beta = \arctan V$ with the t-axis. But we have seen that the straight line $x' = 0$ is identical to the t'-axis of $\mathcal{M}'_2$ (see Fig. 2.2). The t- and t'-axes are, therefore, different from each other. The x'-axis, however, must be the same as the x-axis. This last axis contains the events which in $\mathcal{M}$ take place at the moment $t = 0$. Since time in Newtonian physics is absolute $(t' = t)$, x'-axis consists of these same events (points), and the $t' =$ constant coordinate lines are all of horizontal direction. As a conseqence, the scales on the t- and t'-axes are different from each are, since the

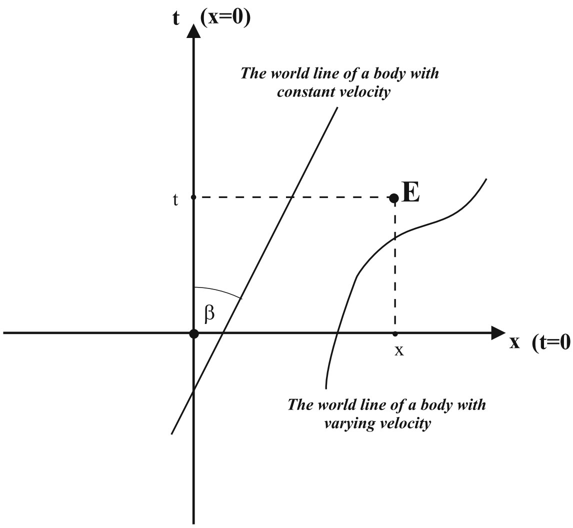

Fig. 2.1 World lines on the Minkowski diagram

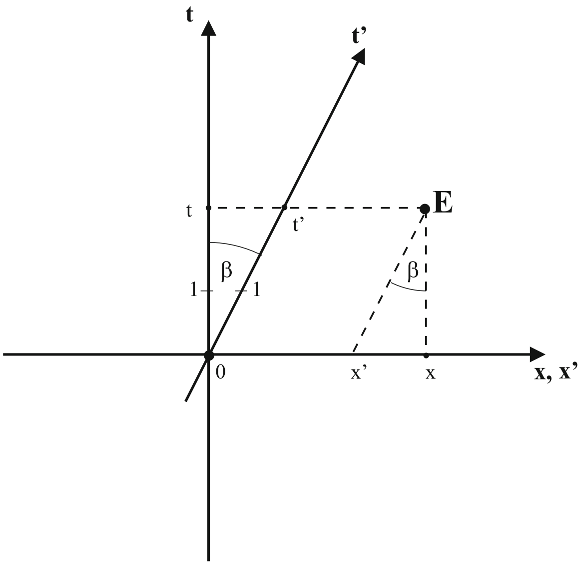

Fig. 2.2 Representation of Galilei-transformation on the Minkowski diagram

geometric distance between the origin and the points, belonging to 1 s on the t- and t'-axes is in the former case smaller than in the latter.

As seen on Fig. 2.2, the coordinates (t, x) and (t', x') of the point E are indeed related to each other by the transformation (2.4.1). The equality $t' = t$ is obviously fulfilled and

$$x' \equiv \overline{Ox'} = \overline{Ox} - \overline{x'x} = x - \tan\beta \cdot t = x - Vt$$

is also true.

When $\mathcal{I}'$ is moving in the negative direction ($V < 0, \beta < 0$), the t'-axis lies to the left of the t-axis. On spacetime diagrams rectilinear coordinates are by no means distinguished with respect to the oblique ones: it is the *relative direction* of the time axes which matters. When, for example, the t-axis of $\mathcal{M}$ is taken so as to make the angle $\beta = -\tan V$ with respect to the vertical direction (and the former coordinates t and x of E are measured along these oblique axes) then the t'-axis will be the vertical one.

While Galilean spacetime diagrams have in fact no useful application, visualisation of the more complicated Lorentz-transformations by means of spacetime diagrams may often be helpful. The basic principle is the same in this more difficult case too: the equation of the t'-axis with respect to $\mathcal{M}_2$ is obtained from the condition $x' = 0$, and that of the x'-axis is determined by the equation $t' = 0$, provided t' and x' are expressed in them, with the help of the Lorentz-transformation (2.4.7), through the unprimed coordinates.

For the equation of the t'-axis we obtain again $x = Vt$ so this axis makes on relativistic spacetime diagrams the same angle $\beta = \arctan V$ with the t-axis as it does on Galilean diagrams. From now on, however, it will be more convenient to work with the variable ct instead of t. It is often denoted by x^0 which is very convenient from a mathematical point of view, but we will continue to use the explicite form ct. Since $x = Vt = (V/c) \cdot ct$, the ct' axis makes an angle $\beta = \arctan(V/c)$ with the upward ct-axis in clockwise direction (for $\beta > 0$).

The essential novelty of relativistic diagrams with respect to the Galilean ones consists in the direction of the x'-axis. According to relativity theory events, which are simultaneous with each other in different inertial frames, belong to different domains of spacetime. Since x-axis (the line $t = 0$) and the x'-axis ($t' = 0$) are built up of simultaneous events in $\mathcal{M}_2$ and $\mathcal{M}_2'$ respectively, they cannot coincide with each other. If we substitute $t' = 0$ into the first line of (2.4.7) we obtain that the x'-axis is indeed different from the x-axis, its equation in unprimed coordinates being $ct = (V/c) \cdot x$. This line passes through the origin and, as it can be easily shown, makes the same angle β with the x-axis (in counterclockwise direction) which is made by the two time axes with each other.

Let us discuss now the scales on the primed and unprimed axes. As we have already observed, on Galilean diagrams the scales on the t- and the t'-axes are different: pieces of equal lengths on them correspont to different intervals of time. On relativistic diagrams this distortion, analogous to the distortions on geographical maps, is amplified and extended to the x- and x'-axes too. The measure of distortion can be assessed by means of the hyperbolas $c^2t^2 - x^2 = \text{constant} \neq 0$ which, according to (2.4.5), have the same mathematical form in all Minkowski coordinates. Take the points $ct' = \pm\kappa$ on the t'-axis for which $c^2t'^2 - x'^2 = \kappa^2$. The content of (2.4.5) is that the unprimed coordinates of this point satisfy the equation $c^2t^2 - x^2 = \kappa^2$. Hence, the points on all the possible time axes, possessing the same coordinate, lie on hyperbolas $c^2t^2 - x^2 = \text{constant} > 0$. The analogous statement for points on the x-axes contains hyperbolas $c^2t^2 - x^2 = \text{constant} < 0$.

The common asymptotes of the two hyperbolas are the bisectrices $x = \pm ct$ of the angle between the coordinate axes. As it can be established by inspection of the relativistic diagrams (or proved with the help of Lorentz-transformation) the equation of them is of the same form in all Minkowski coordinates. Since their points satisfy equation $c^2t^2 - x^2 = 0$ they do not indeed have common points with the hyperbolas $c^2t^2 - x^2 = \text{constant} \neq 0$ though are approaching them infinitely close.

The points on relativistic diagrams can be classified according to their types with respect to the event E_0 at the origin. Since any point of spacetime may be chosen for the origin this classification refers in fact to two arbitrary events (or spacetime intervals between them). The interval between the points of the asymptotes $x = \pm ct$ and the origin is lightlike. They constitute the *light cone* of E_0. The upper half ($t > 0$) of the light cone is the path a flash of light at the origin takes through spacetime. It is the *lightlike future* of E_0. The lower part ($t < 0$) consists of those events the light signals from which can reach the origin. They are, therefore, the *lightlike past* of the event at the origin.

The light cone splits spacetime into several domains. Points *outside* the light cone are those which are spacelike with respect to $E_0(c^2t^2 - x^2 < 0)$. As we have seen in the previous section, for any event E in this domain an inertial frame can be found in which E and E_0 are simultaneous. The points of the outer domain constitute the ‘interval of simultaneity’ with respect to E_0 which was discussed in Sect. 1.3 in connection with the Mars rover. On the spacetime diagram of $\mathcal{M}_2$ the outside domain appears divided into two disjuct parts but this is true only for the two dimensional sections of it. If Minkowski coordinates are imagined in three dimensional space where y-direction can also be visualized the outside domain is seen to be a connected part of spacetime.

The *inside* domain of light cone contains events which are timelike with respect to E_0. Like the light cone itself, this domain splits also into two disconnected part according to the sign of t. Events with $t > 0$ constitute the *timelike future* of the origin while points with $t < 0$ constitute its *timelike past*. To any event E in these domains an inertial frame can be found in which E and E_0 happen at the same place. E precedes or follows E_0, depending on whether it belongs to the timelike past or future of it. This order of events cannot be altered by any choice of the coordinate system.

The properties of the relativistic spacetime diagrams discussed so far are summarized on Fig. 2.3. Remember, how Lorentz-transformations are represented on them. Primed and unprimed coordinates of an event E are obtained by means of parallel projection along the corresponding axis, if only the unity is taken on them properly (Fig. 2.4). Primed axes on Fig. 2.4 correspond to motion of $\mathcal{I}'$ with positive velocity ($V > 0,\ \beta > 0$). In the opposite case ($V < 0,\ \beta < 0$) the time axis is rotated in counterclockwise, the x-axis in clockwise direction. The primed axes

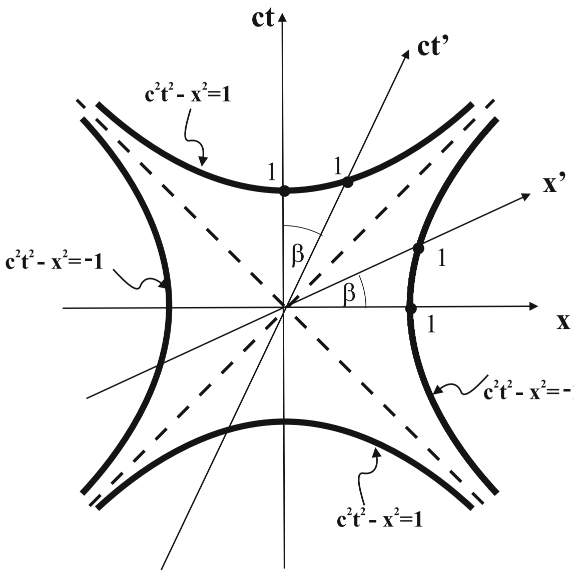

Fig. 2.3 Structure of the relativistic spacetime diagram

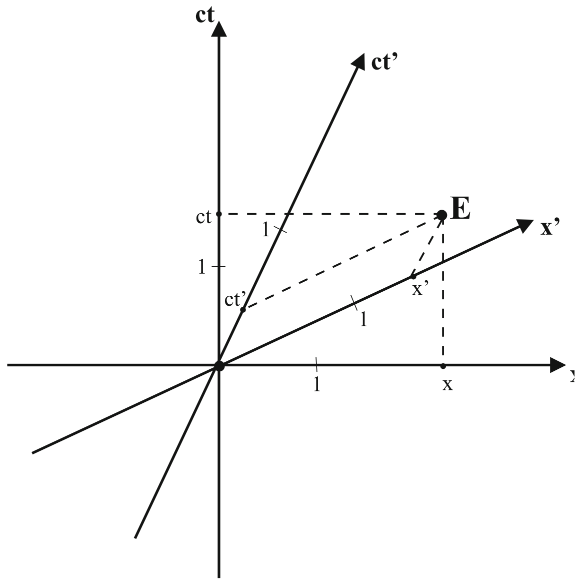

Fig. 2.4 Representation of Lorentz-transformation on the Minkowski diagram

will then make an obtuse angle with each other. Since primed and unprimed frames are equivalent, rectilinear coordinates are by no means distinguished with respect to the oblique ones. If, for example, the primed coordinates are taken in the vertical and horizontal direction then the unprimed coordinates will make an obtuse angle between them.

2.7 The Causality Paradox

If superluminal signals existed, sending messages into past would be possible due to the relativity of simultaneity. In a fit of despair, for example, one could hire a killer in the past to put one's father to death when he is still in his infancy. Suppose the murder did his job. On who's behalf was he acting?

Influences annulling their own cause would beset with perplexing consequences if subjects were able to choose and act as their own will dictated (as in the story above). Let us illustrate with an example how superluminal velocity could in relativity theory lead to influence of this kind.

Consider two inertial frames $\mathcal{I}$ and $\mathcal{I}'$ in standard setting equipped at their origins O and O' with identical cannons C and C', aiming at each other. Examine the following scenario, described in terms of the coordinate time of $\mathcal{I}$: at the moment t_1 cannon C fires a projectile toward $\mathcal{I}'$ which hits it at the moment t_2. The shot, however, misses cannon C' which retaliates immediately. The projectile fired by C' hits cannon C at the moment t_3 and completely destroys it. Since we are investigating consequences of superluminal signals, the muzzle velocity v of the cannons will be allowed to be either smaller or greater than light velocity (but it must be greater than the relative velocity V of the inertial frames in order to make the scenario possible). The problem is to calculate t_3 as a function of t_1, V and v.

In a standard setting the trajectory of O' in $\mathcal{K}$ is $x = Vt$. The trajectory of the projectile in $\mathcal{K}$ shot by C is $x = v(t - t_1)$. The moment t_2 of its arrival to O' is determined by the equation

$$Vt_2 = v(t_2 - t_1),$$

from which we obtain

$$t_2 = \frac{v}{v - V}\, t_1. \tag{2.7.1}$$

At this moment O' is found at the point

$$x_2 = Vt_2 = \frac{Vv}{v - V}\, t_1 > 0 \tag{2.7.2}$$

of $\mathcal{K}$. Let us denote by u the velocity with respect to $\mathcal{K}$ of the projectile fired by C'. According to the rule of relativistic velocity addition we find that

$$u = \frac{V - v}{1 - \frac{Vv}{c^2}}, \tag{2.7.3}$$

since we have now in (1.6.1) $V' = u$, $V = -v$ and $U = -V$. The trajectory of this second projectile in $\mathcal{K}$ is

$$x = x_2 + u(t - t_2).$$

At the moment t_3 of its hit it is found at $x = 0$, hence

$$t_3 = t_2 - \frac{x_2}{u}. \tag{2.7.4}$$

Substituting (2.7.1), (2.7.2) and (2.7.3) into (2.7.4) we obtain, after some rearrangements, the final formula

$$t_3 = \frac{1 - \frac{V^2}{c^2}}{\left(1 - \frac{V}{v}\right)^2} \cdot t_1. \tag{2.7.5}$$

Remember that v need not be smaller than c but it cannot be less than V. For $v > V$ the derivative of t_3 with respect to v is negative, therefore, t_3 is a decreasing function of v as it should be. It is very large when v is only slightly greater than V: as $v \longrightarrow V, t_3 \longrightarrow \infty$. In the opposite limit, when v tends to infinity, the denominator of (2.7.5) is practically equal to unity and $t_3 \approx (1 - V^2/c^2) \cdot t_1$. This value is *smaller* than t_1! When, therefore, v is sufficiently great, cannon C is destroyed *before* it could begin the duel. But then no retaliation by C' is followed, cannon C remains ready to shoote at t_1 which it indeed does, its projectile hits O', etc., etc.

We arrived at a *logical contradiction* because we are both affirming and negating the same facts.

In order to analyze the sequence of the events in some more detail, denote the consecutive events of firing by C, C′ and the destruction of the cannon C at O by E_1, E_2 and E_3 whose coordinates in $\mathcal{M}$ are $(0,\, t_1), (x_2,\, t_2)$ and $(0,\, t_3)$, respectively. In $\mathcal{I}$ E_2 happens later than $E_1 (t_2 > t_1)$ and in $\mathcal{I}'$ E_3 follows E_2 $(t'_3 > t'_2)$.

When $v < c$ nothing paradoxical happens. The pairs $E_1,\, E_2$ and $E_2,\, E_3$ are timelike and their time order is the same in any inertial frame (it is invariant). E_3 is, therefore, subsequent to E_2 in the frame $\mathcal{I}$ too. As a consequence, in this frame E_3 follows $E_1 : t_3 > t_1$. But E_1 and E_3 are timelike with respect to each other since they both happen at the same place and so their time order is also invariant. The time sequence of the events is, therefore, $E_1 \rightarrow E_2 \rightarrow E_3$ as expected, independently of the coordinate system chosen.

This order, however, can be changed dramatically when the projectiles move with faster-than-light velocity. In this case the pairs $E_1,\, E_2$ and $E_2,\, E_3$ are spacelike and E_3 may precede E_2 in $\mathcal{I}$ in spite of the fact that in $\mathcal{I}'$ the former is subsequent to the latter. When t_3 becomes sufficiently smaller than t_2 the causality paradox arises. If, for example, $t_1 = 75\ s$, $V = c/2$ and $v = 8c$ then $t_3 = 64\, s < t_1$.

Let us denote by v_k the critical value of v, above which the paradox $t_3 > t_1$ occurs. When v is equal to v_k, t_3 is equal to t_1. From this condition, using (2.7.5), we obtain for the critical velocity the following expression:

$$v_k = \frac{V}{1 - \sqrt{1 - V^2/c^2}} > c. \tag{2.7.6}$$

When $v > v_k$ formula (2.7.3) leads to *positive* velocity u of the response projectile[3] shot by C′; in the numerical example above $u = +5c/2$. Does this mean the projectile flies *away from O* instead of approaching it? No, it flies toward O since the sign of u coincides with the direction of motion (i.e. with the sign of dx) only if $dt > 0$. But when the return projectile travels in $\mathcal{I}$ backward in time $(dt < 0)$ then a positive u is needed to ensure that $dx = u \cdot dt$ be negative.

On Fig. 2.5 the spacetime diagram of the scenario is shown for two special values of v: on Fig. 2.5a $v = c$ and on Fig. 2.5b $v = \infty$. When $v = c$ the trajectories of the projectiles make an angle 45° with the ct-axis and the time sequence of the events is the natural one. It is the more so the smaller the velocity v is. On Fig. 2.5b both the pairs $E_1,\, E_2$ and $E_2,\, E_3$ are simultaneous but with respect to different inertial frames: the trajectory $E_1 \longrightarrow E_2$ is parallel to the x-axis, while the trajectory $E_2 \longrightarrow E_3$ is parallel to the x'-axis. This is the reason why t_3 turns out smaller than t_1. The coordinate time t is obviously identical to the proper time of the cannon C, resting in the origin of $\mathcal{K}$, and so the moments t_1 and t_3 are independent of the coordinate system chosen.

[3] The velocity u is positive already for $v > c^2/V$ According to (2.7.4), in this domain $t_3 < t_2$. Since $v_k > c^2/V$ it is true *a fortiori* in the domain $v > v_k$.

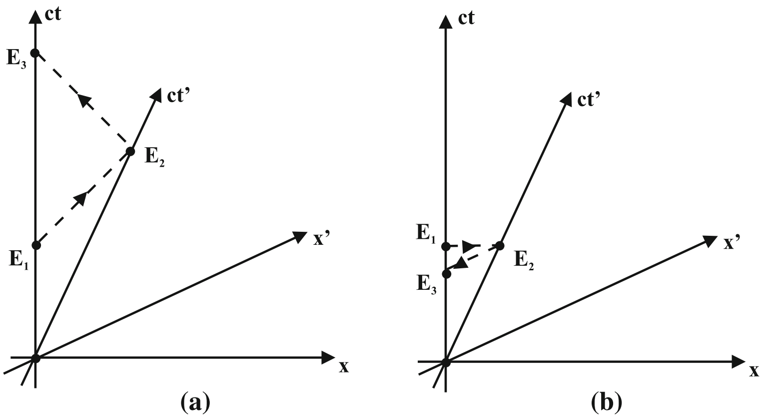

Fig. 2.5 Illustration of the causality paradox

As we see, superluminal signals are a potential source of the causality paradox. The paradox itself consists in reversal of the time order within a causally related timelike pair of events (E_1 and E_3 in the example). This reversal is made possible by the noninvariance of time order within spacelike pairs. The problem cannot be disposed of by saying that the cause is always the event wich took place first because cause and effect (the shot and the explosion) are events intrinsically different from each other.

It has been shown in Sect. 1.7 that bodies cannot be accelerated up to the velocity of light. Why then bother about superluminal velocities and causal paradox? The answer is that signals are not necessarily moving bodies, the obvious example being a light pulse. There may also exist elementary particles which are born in a radioactive process as faster-than-light objects. No such exotic particle has so far been observed but a name has already been given to them: They are called *tachyons*.

The mathematical formalism of relativity theory does not forbid tachyons but their real existence would certainly undermine causality. Since causaliy plays an outstanding role in physics, we close this section by stating its content. In physics causality means that (1) cause and its effect are different in nature and (2) the effect always follows in time its cause.

2.8 Demonstration of Time Dilation on Spacetime Diagram

In Sect. 1.3 we outlined an argument to the effect that the symmetry of time dilation is incompatible with the Newtonian notion of time. At the same time it has been argued that contradictions can be removed if simultaneity of distant events is allowed to depend on the inertial frame in which they are observed. Spacetime diagram provides an ideal tool for making this argument transparent.

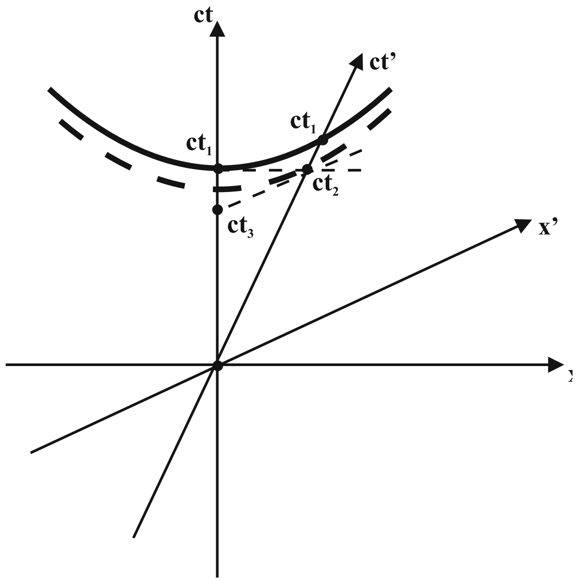

Fig. 2.6 Explanation of the time dilation

Let the clocks $\mathcal{O}$ and $\mathcal{O}'$ are resting at the origins O and O' of $\mathcal{K}$ and $\mathcal{K}'$ in a standard setting. The ct- and ct'-axes on Fig. 2.6 are the world lines of them. The origin is the event of their encounter when both clocks show zero second. In the rest frame of $\mathcal{O}$ clock $\mathcal{O}'$ is moving. When $\mathcal{O}$ shows t_1 s (event ct_1 on the ct-axis) the hand of $\mathcal{O}'$ points to t_2 (event ct_2 on the ct'-axis). These two points are found on the same (horizontal) x-coordinate line. Owing to the different scales on these axes, t_2 is smaller than t_1. This is time dilation from the 'unprimed point of view'.

Analogously, in the rest frame of $\mathcal{O}'$ the clock $\mathcal{O}$ is moving. When the former shows t_2 s the reading on the latter is t_3 s. Again, the difference of the scales leads to $t_3 < t_2$ which is time dilation from the 'primed point of view'.

In this latter case simultaneity is determined by the (oblique) x'- coordinate lines. Notice that the x-coordinate line which connects the points ct_1 and ct_2 on the corresponding time axes is tangent to the hyperbola $c^2t^2 - x^2 = c^2t_1^2$. Similarly, the x'-coordinate line which passes through the points ct_2 and ct_3 on them is tangent to the hyperbola $c^2t^2 - x^2 = c^2t_2^2$. The first of these statement is obvious but the second one needs verification.

As a matter of fact, the second statement would be evident too if the primed coordinate system had been drawn as the rectilinear one. This would be a permitted choice since primed and unprimed frames are equivalent to each other. The property to be proved is expressed by the following theorem: the tangent to the hyperbolas $c^2t^2 - x^2 = \text{constant} > 0$ at the point of their intersection with the ct'-axis is parallel to the x'-axis, and the tangent to the hyperbolas $c^2t^2 - x^2 = \text{constant} < 0$ at the point of their intersection with the x'-axis is parallel to the ct'-axis.

The proof is very simple. The differential of the equation $c^2t^2 - x^2 = \text{constant}$ of the hyperbola is $2ct \cdot d(ct) = 2x \cdot dx$. Its slope is, therefore, given by the equation

$$\frac{d(ct)}{dx} = \frac{x}{ct}.$$

But on the ct'-axis (i.e. on the $x' = 0$ axis) $x = Vt$, therefore, the slope of the hyperbola at its point of intersection with this axis is

$$\frac{d(ct)}{dx} = \frac{V}{c}.$$

On the other hand, the equation of the x'-axis in the unprimed coordinates is $ct - \frac{V}{c}x = 0$ and its slope $d(ct)/dx$ is also equal to V/c. This proves the first part of the theorem and the second part can be proved in the same way.

Notice that Fig. 2.6 is identical to Fig. 2.5b, because the moments of the shoot and the hit are simultaneous events in the inertial frame where muzzle velocity is infinitely large.

2.9 Doppler-Effect Revisited

Consider a monochromatic electromagnetic plane wave whose phase in $\mathcal{I}$ is given by the expression

$$2\pi\left(\frac{x}{\lambda} - \nu t\right) = \frac{2\pi}{c}\nu(x - ct).$$

The second form has been obtained using the relation $\nu\lambda = c$. Let us express in this formula the unprimed coordinates $(ct,\ x)$ through the primed ones. To this end we have to reverse in (2.4.7) the role of primed and unprimed coordinates and, since $\mathcal{K}$ is moving with respect to $\mathcal{K}'$ with the velocity $-V$, the sign of V must also be changed to the opposite. Then

$$\begin{aligned}\frac{2\pi}{c}\nu(x - ct) &= \frac{2\pi\nu}{c\sqrt{1 - V^2/c^2}}\left[(x' + Vt) - \left(ct' + \frac{V}{c}x'\right)\right] = \\ &= \frac{2\pi\nu}{c\sqrt{1 - V^2/c^2}}\left[\left(1 - \frac{V}{c}\right)(x' - ct')\right].\end{aligned}$$

Comparing this expression with the phase $\frac{2\pi}{c}\nu'(x' - ct')$ of the same light wave in $\mathcal{I}'$ we obtain for the frequency ν' the expression

$$\nu' = \nu\sqrt{\frac{1 - V/c}{1 + V/c}}$$

in agreement with the formula (1.2.7). We see that the same formula is applied to both the rate of light signals and the frequency of light waves.

In Sect. 1.2 we have emphasized that ν and ν' are frequencies measured in the proper time of the apparatuses which are at rest in $\mathcal{I}$ and $\mathcal{I}'$, respectively, while the present derivation concerns frequencies in the coordinate times t and t'. But frequencies are measured by clocks at rest for which $dt = d\tau$ so from the point of view of frequencies the two different kinds of time are equivalent to each other.

2.10 The Connection of the Proper Time and Coordinate Time in Inertial Frames

A clock 'laid down' in an inertial frame goes at the rhythm of the coordinate time. Therefore, the coordinate time and the proper time of a clock at rest is related to each other by the equation $d\tau = dt$ independently of their synchronisation. But clocks moving in $\mathcal{I}$ go slower. How many times slower?

The trajectory of a pointlike clock in $\mathcal{K}$ attached to $\mathcal{I}$ is described by equations of the form

$$x = f(t), \quad y = g(t), \quad z = h(t)$$

in which t is the coordinate time. The square of the instantaneous velocity $\mathbf{v}(t)$ is then

$$v^2 = \left(\frac{df}{dt}\right)^2 + \left(\frac{dg}{dt}\right)^2 + \left(\frac{dh}{dt}\right)^2 \equiv \dot{x}^2 + \dot{y}^2 + \dot{z}^2. \tag{2.10.1}$$

The functions $f(t), g(t)$ and $h(t)$ are such that $v^2 < c^2$.

At the moment t the clock is at rest with respect to the instantaneous rest frame $\mathcal{I}'$, moving with *uniform* velocity $\mathbf{v}(t)$. At this moment its proper time is flowing with the same speed as the coordinate time t' of $\mathcal{I}'$: $d\tau = dt'$. Since at this same moment the velocity of the clock in $\mathcal{I}'$ is zero, this equation can also be written as

$$c^2 d\tau^2 = c^2 dt'^2 = c^2 dt'^2 - dx'^2 - dy'^2 - dz'^2. \tag{2.10.2}$$

But the expression on the right hand side is the relativistic squared distance ds^2 whose value can be calculated by means of this same formula in any inertial frame:

$$c^2 dt'^2 - dx'^2 - dy'^2 - dz'^2 = c^2 dt^2 - dx^2 - dy^2 - dz^2. \tag{2.10.3}$$

Substituting this into the right hand side of (2.10.2), factoring out $c^2 dt^2$ and using (2.10.1), we obtain

$$c^2 d\tau^2 = c^2 dt^2 (1 - v^2/c^2).$$

Therefore, at the point of the trajectory where the velocity is equal to v the proper time flows

$$\frac{d\tau}{dt} = \sqrt{1 - v^2/c^2} \equiv 1/\gamma(v) \tag{2.10.4}$$

times slower than the coordinate time.

This formula has already been obtained in Sect. 1.4 when descriptions of the Doppler-effect from the point of view of two different inerial frames were compared with each other. That was a convincing *heuristic* argument in which the time t within an inertial frame was taken identical to the familiar Newtonian time. That was a natural simplifying assumption but not a misleading one since in the intuitive notion of the Newtonian time subtlities of synchronization play no role.

2.11 The Magnitude of the Twin Paradox

The twin paradox had already been discussed in Sect. 1.4 but the calculation of its magnitude requires Minkowski coordinates introduced only in the present section. The most straightforward formulation of it uses two ideal clocks. The paradox consists in the statement that if they move on different trajectories then they in general measure two different values of the proper time between subsequent encounters.

Let the trajectories of the clocks in $\mathcal{M}$ attached to $\mathcal{I}$ is given by the functions

$$x = f_i(t), \quad y = g_i(t), \quad z = h_i(t) \qquad (i = 1, 2).$$

The proper time measured by them between encounters at the moments t_a and t_b is equal to $\Delta\tau_i = \int_{t_a}^{t_b} d\tau_i$. Since $d\tau_i = dt\sqrt{1 - v_i{}^2/c^2}$, we have

$$\Delta\tau_i = \int_{t_a}^{t_b} d\tau_i = \int_{t_a}^{t_b} \sqrt{1 - \frac{1}{c^2}\left[\left(\frac{df_i}{dt}\right)^2 + \left(\frac{dg_i}{dt}\right)^2 + \left(\frac{dh_i}{dt}\right)^2\right]}\, dt. \tag{2.11.1}$$

The magnitude of the paradox is the difference between $\Delta\tau_1$ and $\Delta\tau_2$. In the special case when clock 1 remains at rest we obtain

$$\Delta\tau_1 = t_b - t_a \quad \Delta\tau_2 < (t_b - t_a) = \Delta\tau_1,$$

which indicates that the shorter elapsed time is shown by that clock which was subjected to larger acceleration.

The integral $\int d\tau$ by which the elapsed time is calculated is mathematically akin to the path integral $\int dl$ in analytic geometry which is used to calculate the arc length of a curve. The distinctive property of this type of integrals is that they are independent of the coordinate system chosen in conformity with the fact that both proper time and arc length are invariant quantities.

The magnitude of the twin paradox was experimentally verified in the Gravity Probe A (GP-A) experiment of NASA in 1976. The probe was launched nearly vertically upward, reaching a height of 10,000 km. It housed a hydrogen maser, a highly accurate frequency standard, to compare the time elapsed on the spacecraft and on the Earth. The experiment confirmed the prediction of relativity theory to an accuracy of about 70 ppm.

In this experiment, however, gravitation played an important role and, according to the general relativity theory, *gravitation influences the proper time of bodies*. It can be shown that in the vicinity of the Earth (2.10.4) is modified to

$$d\tau = dt\sqrt{1 - v^2/c^2 + 2\Phi/c^2}, \tag{2.11.2}$$

in which $\Phi = -GM/r$ is the Earth's gravitational potential which tends to zero at infinity. In the GP-A experiment the validity of this formula was verified.

We notice finally that twin paradox (time dilation in general) has found already its way to the engineering application since the GPS navigation system would not function correctly if the internal clocks were not adjusted for relativity.

2.12 The Coordinate Time in Accelerating Frames: the Twin Paradox

In this section our task will be to compare the detailed description of the twin paradox as experienced by either of the twins, say Alice and Bob, when Alice remains at rest in an inertial frame, and her twin-brother Bob makes a trip. Let the place they live in call Blackwood and assume that Bob decides to travel by train from Blackwood to Whitewood, which is at some distance from it, and then back to Blackwood.

The area, surrounding Blackwood and extending beyond Whitewood, constitutes the rest frame of Alice who will be awaiting Bob's return patiently at the railway station at Blackwood. Since we will forget Earth's rotation, her rest frame can be taken for an inertial one. The clocks at the railway stations within this area will all be assumed to show the correct Minkowski coordinate time.

Early morning of the day of departure Bob and Alice walk to the railway station at Blackwood. They shoot a quick look at the station's clock and ascertain that their ideal wrest-watches are keeping good time. On the way to Whitewood Bob compares the reading τ_B on his watch (proper time!) with the time t on the clocks at the stations the train passes by (coordinate time!) and observes that it is the more behind the more time has passed since the departure. This tendency continues to hold on the way back too. When he finally gets off the train at Blackwood and compares his watch with that of Alice, they find that while he has spent a time T_B on the train she was waiting for him for the longer time T_A: the difference $T_A - T_B$ is 'the magnitude of the twin paradox'. It could have been approximately calculated in advance by means of (2.11.1), using the time table of Bob's train. When the train moves with the same constant velocity in both directions and spends no time at Whitewood then $\tau_B = \frac{T_B}{T_A}t$ with $T_B = T_A\sqrt{1 - V^2/c^2}$.

Is it possible to give a similarly unambiguous account of how Alice's watch gets more and more *ahead* with respect to the coordinate time of the train's rest frame? As we have already stressed, the *magnitude* of the effect is an invariant quantity which can be unambigouosly calculated in any reference frame. This follows from the invariance of the integral (2.11.1) whose value is independent of whether the paths of Bob and Alice are referred to the coordinates t, x in which Alice rests at $x = 0$ or to θ , ξ attached to the train where Bob is resting at, say, $\xi = 0$ (the notation of the coordinates attached to the train by greek characters serves to remind us that they do not constitute a Minkowski coordinate system). But this is the global effect only which in itself tells nothing on how the time lag is building up gradually, station by station.

When the train is declared to be at rest it is Alice who, together with the whole area of Blackwood and Whitewood, sets to depart in the direction opposite to the original motion of the train. In this alternative description she is expected to compare her watch with the clocks which are at rest on the train but this is clearly impossible unless the train is at least as long as the distance between Blackwood and Whitewood and appropriate clocks are indeed installed on it at more or less regular distances from each other.

Though a train of such an enormous length would be a rather monstrous construction, it does not seem to contradict any physical principle and so we may be, perhaps, reconciled to its existence. But the proper synchronization of the clocks mentioned above does constitute a principal problem. Accelerations, decelerations and changes in the direction of motion of the *area* with respect to the train at rest are felt as inertial forces *in the train* and make the speed of light to depend in an irregular way on both of its direction of propagation and time. As a consequence, *no universal method* of synchronization of clocks on the train exists which would correspond to the synchronization of clocks at the stations, resting in an inertial frame. As a result, no universal scenario exists of how Alice's watch is gaining gradually more and more.

The behaviour of any particular member of the train's clocks can be followed up unambiguously, *provided* its reading is prescribed at some point of its trajectory. As noted above, this is ensured by the possibility to relate the integral in (2.11.1) to Minkowski coordinates in an inertial frame. Since synchronization consists precisely of such kind of prescriptions detailed analysis of the twin paradox is possible also in coordinate systems attached to accelerating reference frames like our train once a coordinate time is fixed in them in some particular way. But the claim to give a universal account of it prior to chosing a particular kind of synchronization is unfounded since any prediction of this kind could be verified only using real clocks synchronized in some definite manner. No sensible physical theory should give answer to such an ill posed problem. In this respect accelerating reference frames differ significantly from the inertial ones in which Einstein synchronization and Minkowski coordinates are always a possible preferred choice.

What has been said above will now be illustrated by two particular examples of synchronization in the train's rest frame. Let us assume first that, while staying at Blackwood, the clocks on the train are adjusted to show the coordinate time t in the Blackwood–Whitewood area (remember that the train is now longer than the distance between them). From the point of view of the train (and of Bob on it) at the moment of departure Alice (and the whole area) begins to move toward the rear end of the train. Passing by the 'public clocks' on the train Alice compares the reading on her watch τ_A (proper time!) with the time θ (coordinate time!) shown on these clocks.

Assume that the point $\xi = 0$ of the train where Bob is sitting reaches Whitewood at the moment $t = 0$ and at this moment the velocity of the whole train suddenly changes sign. Such a sudden synchronous velocity reversal can be accomplished only if all the cars are motor-cars and the motormen on them are aware of the time t due to e.g. the public clocks along the railroad.

Our object of study is the dependence of τ_A on θ i.e. the function $f_1(\theta)$ in the relation $\tau_A = f_1(\theta)$ (the index 1 refers to the first method of synchronization). If the train's motion is to a good approximation uniform with the same velocity V both to Whitewood and backward from it we have $f_1(\theta) = \frac{T_A}{T_B}\theta$ where $T_B = T_A\sqrt{1 - V^2/c^2}$. Indeed, looking at the series of encounters of Alice with the clocks on the train from the point of view of her rest frame, we observe that the later a train clock passes by Alice the longer path it had already travelled and, therefore, the more it is losing compared to her watch. This result is the strict analogue of Bob's experience which has been summed up in the formula $\tau_B = \frac{T_B}{T_A}t$

Consider now the second method of synchronization. Assume that the coordinate time θ on the train has been chosen according to Einstein's synchronization procedure when the train was moving with the constant velocity V toward Blackwood. Suppose that at that time Bob has already been sitting on the train and compares his clock with that of Alice later while in motion through Blackwood. The velocity reversal takes place at the moment $\theta = 0$, when Bob arrives at Whitewood. The motormen now perform this operation according to their own correctly synchronized clocks.

Since the rest frame of Alice is an inertial one, formula (2.11.1) is applicable in it, and leads immediately to the relation $T_B = T_A\sqrt{1 - V^2/c^2}$. But there is a problem here. Now the rest frame of Bob is also an inertial one both on the way toward Whitewood and back from it. Therefore, (2.11.1) is applicable in them also and leads obviously to the opposite result $T_A = T_B\sqrt{1 - V^2/c^2}$. This last formula is certainly wrong but it may not be immediately clear how to rectify it.

To solve the puzzle let us attach a coordinate system to the train. Consider the auxiliary inertial frame $\mathcal{I}$ which moves with velocity $+V$ with respect to the ground and attach Minkowski coordinates t, x to it. The trajectory of any given point P of the train is given in it by the equation

$$x = \begin{cases} \xi & \text{if } t \leq 0, \\ \xi + Ut & \text{if } t \geq 0, \end{cases} \tag{2.12.1}$$

where ξ is a constant and

$$U = \frac{-2V}{1 + V^2/c^2} \tag{2.12.2}$$

is equal to the velocity of the train on its way back from Whitewood *with respect to* $\mathcal{I}$. It has been assumed that the direction of motion of the train has changed at $t = 0$.

Each piece of the train has its own constant ξ which can, therefore, be taken as space coordinate of the coordinate system attached to the train. Bob will be assumed to be sitting at $\xi = 0$. The most natural choice for the coordinate time θ of the attached system is to assume that θ is equal to the proper time of the correctly synchronized clocks (virtual or real, including Bob's watch) which are fixed on the train and show zero at the moment of the velocity reversal:

$$\theta = \begin{cases} t & \text{if } t \leq 0, \\ t\sqrt{1 - U^2/c^2} & \text{if } t \geq 0. \end{cases} \tag{2.12.3}$$

Then the relation between the coordinates t, x and θ, ξ can be summarized as

$$x = \begin{cases} \xi & \text{if } \theta \leq 0 \\ \xi + \frac{U\theta}{\sqrt{1-U^2/c^2}} & \text{if } \theta \geq 0 \end{cases} \quad t = \begin{cases} \theta & \text{if } \theta \leq 0 \\ \frac{\theta}{\sqrt{1-U^2/c^2}} & \text{if } \theta \geq 0 \end{cases} \tag{2.12.4}$$

and

$$\xi = \begin{cases} x & \text{if } t \leq 0 \\ x - Ut & \text{if } t \geq 0 \end{cases} \quad \theta = \begin{cases} t & \text{if } t \leq 0 \\ t\sqrt{1 - U^2/c^2} & \text{if } t \geq 0. \end{cases} \tag{2.12.5}$$

In these formulae U must be expressed through V using (2.12.2). If in $ds^2 = c^2dt^2 - dx^2$ we express the differentials by means of (2.12.4) through $d\theta$ and $d\xi$ we obtain the expression

$$ds^2 = \begin{cases} c^2d\theta^2 - d\xi^2 & \text{if } \theta < 0 \\ c^2d\theta^2 + \frac{4V/c}{1-V^2/c^2}cd\theta \cdot d\xi - d\xi^2 & \text{if } \theta > 0 \end{cases} \tag{2.12.6}$$

where U has been expressed through V.

The construction of the attached coordinate system has now been completed. The next task is to write down the trajectory of Alice in it. It is obvious that at $t < 0$ (which is the same as $\theta < 0$) this trajectory is $\xi = -V(\theta + \theta_e)$ where θ_e is a constant. Since Bob is sitting at the point $\xi = 0$ their first encounter takes place at the moment $\theta = -\theta_e < 0$. At that moment Bob's watch shows $-\theta_e$ s. Since the train at $\theta > 0$ moves with the same speed as before, the second encounter will take place when the reading on Bob's watch is $+\theta_e$ (i.e. $T_B = 2\theta_e$). The trajectory of Alice is, therefore, determined by the equations[4]

[4] The velocity of Alice at $\theta > 0$ can also be obtained directly, using (2.12.5) and (2.12.2):

$$\frac{d\xi}{d\theta} = \frac{dx - U \cdot dt}{dt\sqrt{1 - U^2/c^2}} = \frac{-V - U}{\sqrt{1 - U^2/c^2}} = +V,$$

since the velocity $\frac{dx}{dt}$ of Alice with respect to the auxiliary frame $\mathcal{I}$ is equal to $-V$ all the time.

$$\xi = \begin{cases} -V(\theta + \theta_e) & \text{if } \theta \leq 0 \\ +V(\theta - \theta_e) & \text{if } \theta \geq 0. \end{cases} \tag{2.12.7}$$

Now we can find the functional form of the relation $\tau_A = f_2(\theta)$ by integration:

$$\tau_A = \int_{-\theta_e}^{\theta} d\tau = \frac{1}{c} \int_{-\theta_e}^{\theta} \sqrt{ds^2} = \frac{1}{c} \int_{-\theta_e}^{\theta} \sqrt{\left(\frac{ds}{d\theta}\right)^2} \cdot d\theta.$$

Using (2.12.6) we obtain for the integrand the expression

$$\frac{1}{c}\sqrt{\left(\frac{ds}{d\theta}\right)^2} = \begin{cases} \sqrt{1 - \frac{1}{c^2}\left(\frac{d\xi}{d\theta}\right)^2} & \text{if } \theta < 0 \\ \sqrt{1 + \frac{4V/c}{1-V^2/c^2} \cdot \frac{1}{c}\frac{d\xi}{d\theta} - \frac{1}{c^2}\left(\frac{d\xi}{d\theta}\right)^2} & \text{if } \theta > 0. \end{cases} \tag{2.12.8}$$

According to (2.12.7) $\frac{d\xi}{d\theta} = \mp V$ in the first and second line, respectively. The integrand is, therefore, constant in both domains of θ which makes the integration trivial. After a little algebra we obtain

$$\tau_A = f_2(\theta) = \begin{cases} \sqrt{1 - V^2/c^2} \cdot (\theta + \theta_e) & \text{if } \theta \leq 0 \\ \frac{1+V^2/c^2}{\sqrt{1-V^2/c^2}}\,\theta + \sqrt{1 - V^2/c^2} \cdot \theta_e & \text{if } \theta \geq 0. \end{cases} \tag{2.12.9}$$

This is our final formula for the second method of synchronization which replaces $\tau_A = f_1(\theta) = (T_A/T_B)\theta$ of the first method in which $T_A/T_B = 1/\sqrt{1 - V^2/c^2}$. For *this ratio* (2.12.9) leads to this same expression since $T_A = f_2(\theta_e) = 2\theta_e/\sqrt{1 - V^2/c^2}$ and, as we have already seen, $2\theta_e = T_B$. But the functions $f_1(\theta)$ and $f_2(\theta)$ differ from each other (see Fig. 2.7).

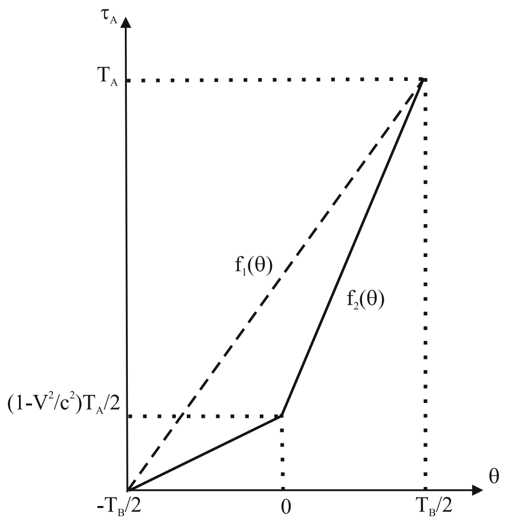

Fig. 2.7 Time dependence of the twin paradox

The puzzle, found in the naiv consideration of the second method, has been thereby solved. The mistake was committed in the treatment of the second half of the motion. It is true that the train is an inertial frame in this phase too but the coordinates attached to it, which are in conformity with the continuous operation of clocks fixed on the train, are not Minkowski coordinates since for $\theta > 0$ $ds^2 \neq c^2 d\theta^2 - d\xi^2$. Therefore, (2.11.1) is not valid in these coordinates.

The change in the form of ds^2 is the lasting effect of the instantaneous acceleration at $\theta = 0$. One of its consequences is that a free body which is originally at rest acquires at the moment $\theta = 0$ the velocity $-U = 2V/(1 + V^2/c^2)$ which, in the absence of friction, is a lasting effect. A less obvious persistent effect is *desynchronization of correctly synchronized pairs of ideal clocks* attached to the train. Assume that the light velocity measured by a pair of clocks fixed on the train is originally the same in both directions along the coordinate ξ. This will no longer be the case at $\theta > 0$. The light velocity $\frac{d\xi}{d\theta}$ can be obtaind from the condition $\frac{ds^2}{d\theta^2} = 0$, using (2.12.6) divided by $d\theta^2$. For $\theta < 0$ we obtain $\left(\frac{d\xi}{d\theta}\right)_{\pm} = \pm c$, but for $\theta > 0$ the two solutions of the equation

$$\left(\frac{d\xi}{d\theta}\right)^2 - \frac{4V}{1 - V^2/c^2}\left(\frac{d\xi}{d\theta}\right) - c^2 = 0$$

are of different magnitude in the positive and negative direction:

$$\left(\frac{d\xi}{d\theta}\right)_{\pm} = c\,\frac{V/c \pm (1 + V^2/c^2)}{1 - V^2/c^2}.$$

This difference can be ascribed only to clocks' desynchronization since, in spite of the fact that in $\theta > 0$ the form of ds^2 is different from the Minkowskian form $c^2 d\theta^2 - d\xi^2$, the train is, none the less, an inertial frame in this domain of time too. This can be formally proven by demonstrating that θ, ξ can be replaced by a new attached coordinate system t', x' in which $ds^2 = c^2 dt'^2 - dx'^2$ holds true.

The first step is to replace θ by t':

$$c\theta = ct' - 2\frac{V/c}{1 - V^2/c^2}\;\xi.$$

Then we have after some algebra

$$ds^2 = c^2 dt'^2 - \left(\frac{1 + V^2/c^2}{1 - V^2/c^2}\right)^2 d\xi^2.$$

Rescaling ξ according to the relation

$$\xi = \frac{1 - V^2/c^2}{1 + V^2/c^2}\;x' \tag{2.12.10}$$

leads to the desired Minkowski form of ds^2. The crucial moment here is that objects with fixed ξ will have fixed x' coordinate too. The primed coordinates are, therefore, indeed attached to the train.

Consider two fixed points at ξ and $\xi + \Delta\xi$. Since at negative times θ, ξ are Minkowski coordinates, the distance Δl between them is equal to $|\Delta\xi|$. For the same reason, at positive times their distance is changed to $\Delta l' = |\Delta x'|$. Using (2.12.10) we then obtain

$$\Delta l' = \frac{1 + V^2/c^2}{1 - V^2/c^2}\,\Delta l.$$

This is another persistent consequence of the velocity reversal: increase of the distance between given points of the train. Therefore, the material composition of our 'model train' which serves to modellize the coordinate transformation (2.12.4), (2.12.5) must be such as to make the necessary deformation possible.

Closing this section the main message of it is emphazised again: *detailed behaviour of the twin paradox from the point of view of the accelerating twin depends on how the coordinate time is specified in his/her rest frame*.

2.13 The Coordinate Time in Accelerating Frames: the Rotating Earth

Rotating Earth is an accelerating reference frame considerably more relevant than the train commuting between Blackwood and Whitewood. But consider first the case of the rotating disc. The coordinate time on it may be defined by the readings of (virtual) clocks rotating together with the disc. It may be assumed that, before the disc starts rotating, these clocks keep in synchronism with those (virtual) clocks that are at rest on the ground and show the Minkowski time in the inertial frame $\mathcal{I}$ attached to the latter.

As the disc sets rotating time dilation causes the clocks attached to it to lose. This slowering of them can in principle be easily established since in the course of their rotation either of these clocks periodically passes by the clock on the ground in the neighbourhood of which it had rested before the rotation began. The time lag is the larger the farther the clock is located from the axis of rotation but this fact by no means forbids to base coordinate time on their readings. Spacetime coordinate system based on this definition of coordinate time and on a cylindrical coordinate system whose z-axis coincides with the axis of rotation will be denoted by $\mathcal{K}_1'$.

As a matter of fact this coordinate system is never used in practice because calculations usually become much simpler if coordinate time is chosen instead in the following manner: at the point P of the rotating disc the coordinate time is defined by the reading on that clock *fixed to the ground* which P is just passing by. Coordinate system based on this notion of coordinate time will be called $\mathcal{K}_2'$.

The coordinate system attached to the ground is the unprimed $\mathcal{K}$. Its spatial part is a cylindrical system of the same orientation as described above. Then transformation from $\mathcal{K}$ to $\mathcal{K}_2'$ is given by the formulas

$$t' = t, \quad r' = r, \quad \varphi' = \varphi - \omega t, \quad z' = z. \tag{2.13.1}$$

Clocks attached to the disc do not show the coordinate time, they are losing more and more with respect to it. Much as unnatural this choice may seem it is usually the most convenient one.

The coordinate time on the Earth should conform with the practice that the duration of the solar day must be the same at every point of the Earth i.e. it must be independent of the latitude. This requirement excludes the choice $\mathcal{K}_1'$ since, owing to time dilation, the time elapsed between two successive returns of the Sun to the local meridian would be on smaller latitudes shorter than on higher ones. Since the speed of the Earth's rotation is very slow with respect to light velocity this problem is not a practical one. Let us now suppose, however, that Earth rotates much faster than it actually does and, as a consequence, the value of $\sqrt{1 - V^2/c^2}$ on the Equator is considerably smaller than unity but for the sake of simplificity forget the revolution of the Earth around the Sun. Then the solar day which determines the rhythm of everyday life will coincide with the sidereal day equal to the period of rotation of the Earth with respect to the fixed stars. By this assumption the case of the Earth becomes indistinguishable from that of the rotating disc provided $\mathcal{I}$ is identified with the inertial frame of the fixed stars. The coordinate system $\mathcal{K}_2'$ adapted to the Sun's path as seen from the Earth is then defined by the correctly synchronized virtual clocks, resting in that frame.

But on the hypothetical Earth, rotating with considerable angular velocity, this natural choice of coordinate time would require certain developments in the construction of time keeping devices since on different latitudes they should be calibrated to different speeds. Clocks, therefore, should provide an option to set the precise value of the latitude at which they are to work. We are fortunately not obliged to deal with problems of this kind till time dilation due to the rotation of our real Earth can be considered negligibly small.

2.14 Lorentz Contraction Revisited

Lorentz contraction is illustrated on the spacetime diagram of Fig. 2.8. Consider two points fixed in the inertial frame $\mathcal{I}_0$ whose distance is equal to l_0 (proper distance). In Minkowski coordinates their world lines are a pair of vertical lines at, say, $x_0 = 0$ and $x_0 = l_0$. Lorentz contraction is the consequence of the geometrical fact that segments of different length are cut out by these world lines from the axes x_0 and x. The position of the hyperbola $c^2t_0^2 - x_0^2 = -l_0^2$ on the figure shows that the distance l cut out from the x axis is smaller than l_0.

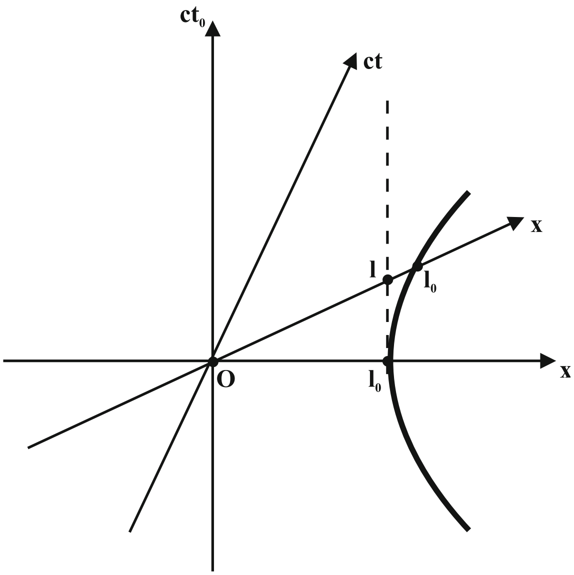

Fig. 2.8 To the derivation of the Lorentz contaction formula

Let us express the magnitude of the Lorentz contracted length between the points by means of Lorentz transformation as

$$l = x = \frac{x_0 - Vt_0}{\sqrt{1 - V^2/c^2}},$$

where (x_0, ct_0) are the coordinates in $\mathcal{K}_0$ of the point of intersection of the world line $x_0 = l_0$ with the x-axis. According to Sect. 2.6, the equation of the x-axis in $\mathcal{K}_0$ is $ct_0 = \frac{V}{c}x_0$. Hence $(x_0, ct_0) = (l_0, Vl_0/c)$. Using these values in the above formula we obtain

$$l = x = \frac{x_0 - Vt_0}{\sqrt{1 - V^2/c^2}} = l_0 \frac{1 - V^2/c^2}{\sqrt{1 - V^2/c^2}} = l_0\sqrt{1 - V^2/c^2}.$$

Earlier, in Sect. 1.5, this formula has already been derived from time dilation which is the consequence of the relativity of simultaneity. As Fig. 2.8 shows clearly, contraction owes its existence to the nonzero angle between the x-axes, belonging to different inertial frames. As we have explained in Sect. 2.6, this property of Minkowski diagrams is the manifestation of the absence of absolute simultaneity in relativity theory. The derivation of Lorentz contraction in the present section leads, therefore, to the same conclusion as the derivation in Sect. 1.5: contraction is the consequence of the relativity of simultaneity.

When, instead of $\mathcal{K}_0$, the points are at rest in $\mathcal{K}$ then their distance becomes shorter with respect to $\mathcal{K}_0$. This conclusion can also be easily drawn from Fig. 2.8, since the world lines of the points are now the ct axis and the tangent to the hyperbola $c^2t_0^2 - x_0^2 = -l_0^2$ at its point of intersection with the x-axis (see the theorem proved in Sect. 2.8). The point of intersection of this tangent with the x_0-axis is obviously nearer to the origin than the proper distance l_0 of the points.

2.15 Is the Perimeter of a Spinning Disc Contracted?

Before tackling this problem let us return to the Bell's thought experiment with the rod transported by a pair of railcars (Sect. 1.5). There we arrived at the conclusion that Lorentz contraction may cause the rod to slip between the cars. But what would happen if the rod was tightly bolted to the cars? If the cars' motion were still kept in rigorous synchronicity the length of the rod would obviously remain unchanged but only at the cost of being streched. As a result, the work done by the engines of the motor-cars would be increased by the amount of the elastic energy of the rod and their fuel consumption might, therefore, be raised considerably.

We are now ready to examine the case of the disc. Let us first assume that it rests in the inertial frame $\mathcal{I}$ in a plane perpendicular to the vertical z-axis. Divide its perimeter into n arcs of equal length by means of the set of points $A_1, A_2, \ldots, A_n$, mark the corresponding points on the ground as $B_1, B_2, \ldots, B_n$ and then make the disc to rotate around the z-axis with a uniform angular velocity. The question is whether its dimensions will show up modifications correponding to Lorentz contraction.

Centrifugal force will tempt the radius to increase. But this is a 'case specific' phenomenon because the magnitude of the radial deformation is contingent on the peculiar elastic properties of the disc: under identical circumstances the deformation is the less the more rigid material the disc is composed of. Lorentz contraction, on the contrary, is independent of material composition and even ideally rigid discs are subjected to it. Moreover, since radial direction is perpendicular to the velocity of rotation, no Lorentz contraction is expected to occur along it. We can, therefore, safely assume that for an ideally rigid disc no deformations take place along the radius.

Along the circumference, however, Lorentz contraction may be in operation. It should manifest itself in the incongruence of the point set $A_1, A_2, \ldots, A_n$ on the disc with the corresponding set $B_1, B_2, \ldots, B_n$ on the ground. But an incongruousness of this kind is obviously impossible unless axial symmetry of the disc is destroyed by contraction which would be equivalent to its destruction. If this does not happen, all points A_i must obviously pass by the corresponding B_is simultaneously which is the same as to say that neither segment of the disc is contracted as compared to its counterpart on the ground.

Does not this conclusion plainly contradict Lorentz contraction of a single rod? The example of the rod transported by traincars reveals that contraction takes place only when the rod can change its size freely with respect to the cars. The rotating disc is, however, analogous to the case when the rod is fixed to the cars since adjecent segments of the disc mutually prohibit each other's deformation in tangential direction. Stresses may arise but as long as they do not lead to instability and destruction of axial symmetry they prevent Lorentz contraction to manifest itself.

Let us place now measuring rods (rulers) along the rim of the disc at rest without fastening them to each other or to the disc. For a sufficiently large disc the length of its circumference is equal to the number of such rods provided they fill it without gaps. When, however, the disc is made to rotate the rods suffer contraction, they will cease to fill the perimeter densely and gaps between them arise. What is the message to be drawn from this fact?

When a ruler is employed to measure the distance between a pair of points of a body it must be at rest with respect to the body. Therefore, the contraction of the measuring rods on the rotating disc does not prompt us to make any conclusion concerning its perimeter as seen from the frame in which it is rotating. As elucidated above, the circumference understood in this sense is independent of whether the disc is spinning or not. But the gaps between the rulers along its perimeter are an incontestable fact which proves unambiguously that with respect to the rest frame of the rotating disc, in which the rulers are resting, the length of the perimeter did increase. The gaps cannot be attributed to the change in the length of the rulers because no length etalon except ideally rigid measuring rods are assumed to exist. Since rulers along the radius do not contract, the radius of the rotating disc is of the same length as that of the disc at rest. Therefore, the ratio of the cirmumference to the radius of a rotating disc measured in its rest frame is larger than 2π.

2.16 Do Moving Bodies seem Shorter?

George Gamow, the famous russian-american physicist, told in his book *Mr. Tompkins in Wonderland* the story of a certain bank clerk Mr. Tompkins who had once dropped in a popular lecture on relativity theory but in the middle of it had fallen asleep and had found himself in a wonderland where light velocity had had the value of only 15 km/h.

> The hands of the big clock on the tower down the street were pointing to five o'clock and the streets were nearly empty. A single cyclist was coming slowly down the street and, as he approached, Mr Tompkins's eyes opened wide with astonishment. For the bicycle and the young man on it were unbelievably shortened in the direction of the motion, as if seen through a cylindrical lens. The clock on the tower struck five, and the cyclist, evidently in a hurry, stepped harder on the pedals. Mr Tompkins did not notice that he gained much in speed, but, as the result of his effort, he shortened still more and went down the street looking exactly like a picture cut out of cardboard. Then Mr Tompkins felt very proud because he could understand what was happening to the cyclist—it was simply the contraction of moving bodies, about which he had just heard.

This vivid description is based on the conviction that Lorentz contraction may be observed by seeing it. In books on relativity theory the statement is often found that a moving sphere is seen as an oblate ellipsoid flattened along the direction of its motion. But it is by no means obvious that when a moving body is *seen* the same property of it is observed which determines its length, i.e. the distance between its endpoints at a given moment of time in the inertial frame of the observer. By seeing we perceive the angle of sight rather than a length and the light fronts which reach our eyes at a given moment had generally emerged from the endpoints of the body at different moments of time.

The distortion that a passing sphere would appear to undergo if it were travelling a significant fraction of the speed of light, was described independently by James Terrell and Roger Penrose in 1959. They found that the sphere does not appear oblate but seems to undergo a peculiar rotation (*Penrose–Terrell effect*).

2.17 Velocity Addition Revisited

In Sect. 1.6 the law of velocity addition has been derived for the special case when the velocity V of the moving body is parallel to the relative velocity U of the inertial frames. Lorentz-transformation permits us to obtain the corresponding formula for the more general case in a single step.

In standard setting the x-axis can always be chosen parallel to the relative velocity: $(U_x, U_y, U_z) = (U, 0, 0)$, but the velocity $\mathbf{V}$ must be allowed to be of any direction. Then Eqs. (2.4.7) written for coordinate and time differentials lead at once to the required formulae:

$$\begin{aligned} V'_x &= \frac{dx'}{dt'} = \frac{dx - Udt}{dt - \frac{U}{c^2}dx} = \frac{V_x - U}{1 - V_x U/c^2}, \\ V'_y &= \frac{dy'}{dt'} = \frac{dy\sqrt{1 - U^2/c^2}}{dt - \frac{U}{c^2}dx} = \frac{V_y\sqrt{1 - U^2/c^2}}{1 - V_x U/c^2}, \\ V'_z &= \frac{dz'}{dt'} = \frac{dz\sqrt{1 - U^2/c^2}}{dt - \frac{U}{c^2}dx} = \frac{V_z\sqrt{1 - U^2/c^2}}{1 - V_x U/c^2}. \end{aligned} \tag{2.17.1}$$

As we have seen in Sect. 1.6, in relativity theory the relative velocity of two bodies differs from the rate of change of the distance between them. This is the direct consequence of the relativistic notion of the coordinate time. The rate of change of the distance refers to the coordinate time of a common $\mathcal{K}$ for both bodies, while, in calculating relative velocity, the velocity U of one of the bodies is understood in terms of the coordinate time in some $\mathcal{K}$, while the velocity V of the other is expressed in the coordinate time *of the rest frame* of the former.

2.18 Equation of Motion Revisited

In Sect. 1.7 the basic principles, underlying equations of motion of point particles in relativity theory, have been clarified, but the equations themselves have been derived only for the special case when the force is parallel to the velocity. Below the equations in the general case will be derived with the help of the Lorentz-transformations.

The key point of the derivation is the transformation rule of the acceleration which in $\mathcal{I}$ and $\mathcal{I}'$ is defined by the formulae

$$\mathbf{a} = (a_x,\ a_y,\ a_z) = \left(\frac{dv_x}{dt},\ \frac{dv_y}{dt},\ \frac{dv_z}{dt}\right)$$
$$\mathbf{a}' = (a'_x,\ a'_y,\ a'_z) = \left(\frac{dv'_x}{dt'},\ \frac{dv'_y}{dt'},\ \frac{dv'_z}{dt'}\right),$$

where $\mathbf{v}$ and $\mathbf{v}'$ are the particle's velocity in $\mathcal{I}$ and $\mathcal{I}'$ respectively. Since $dx = v_x dt$, the differentials dt and dt' are related to each other by Lorentz-transformation as

$$dt' = \frac{dt - (V/c^2)\,dx}{\sqrt{1 - V^2/c^2}} = \frac{1 - v_x V/c^2}{\sqrt{1 - V^2/c^2}}\,dt \tag{2.18.1}$$

(as usual, V is the relative velocity of $\mathcal{I}'$ with respect to $\mathcal{I}$ in a standard setting). According to the first of the equations (2.17.1) $v'_x = (v_x - V)/(1 - v_x V/c^2)$, hence

$$dv'_x = \frac{d}{dv_x}\left(\frac{v_x - V}{1 - v_x V/c^2}\right) dv_x = \frac{1 - V^2/c^2}{(1 - v_x V/c^2)^2}\,dv_x.$$

We then have

$$\frac{dv'_x}{dt'} = \frac{(1 - V^2/c^2)^{3/2}}{(1 - v_x V/c^2)^3}\,\frac{dv_x}{dt},$$

and finally

$$a'_x = \frac{(1 - V^2/c^2)^{3/2}}{(1 - v_x V/c^2)^3}\,a_x.$$

The differential of the second equation of (2.17.1) is

$$dv'_y = \frac{\partial}{\partial v_x}\left(\frac{v_y\sqrt{1 - V^2/c^2}}{1 - v_x V/c^2}\right) dv_x + \frac{\partial}{\partial v_y}\left(\frac{v_y\sqrt{1 - V^2/c^2}}{1 - v_x V/c^2}\right) dv_y.$$

Having performed the differentiations we obtain

$$dv'_y = v_y\sqrt{1 - V^2/c^2}\cdot\frac{(-V/c^2)}{(1 - v_x V/c^2)^2}\,dv_x + \frac{\sqrt{1 - V^2/c^2}}{1 - v_x V/c^2}\,dv_y.$$

Again, dividing by (2.18.1), we have for the primed y-component of the acceleration the expression

$$a'_y = -\frac{v_y V}{c^2}\cdot\frac{1 - V^2/c^2}{(1 - v_x V/c^2)^3}\,a_x + \frac{1 - V^2/c^2}{(1 - v_x V/c^2)^2}\,a_y.$$

Assume that the primed frame is just the instantaneous rest frame of the particle and the common x direction of $\mathcal{I}$ and $\mathcal{I}'$ is parallel to $\mathbf{v}$ at the moment chosen. Then

$$V = v_x \equiv v, \quad \text{and} \quad v_y = v_z = 0 \tag{2.18.2}$$

and we have

$$a'_x = \frac{a_x}{(1 - v^2/c^2)^{3/2}}.$$
$$a'_y = \frac{a_y}{1 - v^2/c^2}.$$

For the z-component we obtain similarly

$$a'_z = \frac{a_z}{1 - v^2/c^2}.$$

In Sect. 1.7 we pointed out that in the rest frame which is now identical to the primed one equations of motion must have their original Newtonian form $m\mathbf{a}' = \mathbf{F}'$. If in this equation the primed components are expressed through the unprimed ones we obtain the equation of motion valid in the unprimed frame where the point mass at the given moment of time is moving in x-direction[5]:

$$\begin{aligned} \frac{m}{(1 - v^2/c^2)^{3/2}}\, a_x &= F'_x, \\ \frac{m}{1 - v^2/c^2}\, a_y &= F'_y, \\ \frac{m}{1 - v^2/c^2}\, a_z &= F'_z. \end{aligned} \tag{2.18.3}$$

The primed components of the force must also be expressed through the corresponding unprimed components but this can only be done after having specified its mathematical form (see Sect. 2.21).

2.19 The Energy–Momentum Four Vector

Four-component quantities whose primed components are expressed through the unprimed ones via Lorentz-transformations are called *four-vectors*. The coordinates (ct, x, y, z) and the coordinate differentials in Minkowski coordinates constitute a four-vector.

In the expressions (1.7.9) and (1.10.1) of the momentum and energy the coordinate time can be replaced, using (2.10.4), with the proper time:

[5] In Sect. 2.21 we will dispose of this limitation.

$$\begin{aligned}\mathbf{p} &= (p_x,\ p_y,\ p_z) = \frac{m}{\sqrt{1-v^2/c^2}}\left(\frac{dx}{dt},\frac{dy}{dt},\frac{dz}{dt}\right) = m\left(\frac{dx}{d\tau},\frac{dy}{d\tau},\frac{dz}{d\tau}\right)\\ E &= \frac{mc^2}{\sqrt{1-v^2/c^2}} = mc\frac{d(ct)}{d\tau}.\end{aligned} \tag{2.19.1}$$

Under Lorentz-transformations $(d(ct),\ dx,\ dy,\ dz)$ behave as components of a four-vector while m and $d\tau$ remain unchanged. Therefore, $(E/c,\ p_x,\ p_y,\ p_z)$ is also a four-vector, called *the energy–momentum four vector* or *four-momentum*, whose primed and unprimed components are related to each other by Lorentz-transformation:

$$\left.\begin{aligned} E' &= \frac{E - Vp_x}{\sqrt{1-V^2/c^2}}\\ p'_x &= \frac{p_x - \frac{V}{c^2}E}{\sqrt{1-V^2/c^2}}\\ p'_y &= p_y; \quad p'_z = p_z. \end{aligned}\right\} \tag{2.19.2}$$

It has to be stressed that the kinetic energy K of a point mass *is not* a component of any four-vector. The knowledge of the mass–energy relation is, therefore, crucial for being able to form a four-vector which contains the momentum of the particle as three of its components. To *define* this four-vector without the preliminary clarification of the relation between the rest energy and the mass of the particle would be a mere hindsight without physical motivation.

For the momentum four-vector the invariant, corresponding to (2.5.2), is the quadratic expression $E^2/c^2 - p_x^2 - p_y^2 - p_z^2$. Its value, which is the same in any inertial frame, is easily found in the rest frame where $\mathbf{p} = 0$ and $E = E_0 = mc^2$:

$$\frac{E^2}{c^2} - p_x^2 - p_y^2 - p_z^2 \equiv \frac{E^2}{c^2} - p^2 = m^2c^2. \tag{2.19.3}$$

The validity of this relation can be verified also by substituting (2.19.1) into its left hand side and making use of (2.5.2). The mass is, therefore, invariant under Lorentz-transformation.

Solving this equation with respect to the energy we obtain the formula

$$E = c\sqrt{p^2 + m^2c^2} \tag{2.19.4}$$

which corresponds to the Newtonian expression $K = p^2/2m$ but includes the rest energy too. The relativistic equivalent of the Newtonian relation $v = p/m$ is arrived at by dividing (1.7.9) by (1.10.1):

$$\mathbf{v} = \frac{c^2}{E}\mathbf{p}. \tag{2.19.5}$$

The most significant property of the energy and the momentum is their conservation in isolated systems and their mathematical expression should have been derived from this requirement. This could be done within the Lagrangian approach which would confirm the correctness of (1.7.9) and (1.10.1). The use of them will now be illustrated on the example of the decay $Po^{210} \longrightarrow Pb^{206} + \alpha$ (cf. Sect. 1.11). The problem is to calculate the velocity of the α-particle from the known masses of the three particles.

Let the masses of Po^{210}, Pb^{206} and the α-particle be equal to M, m and μ, respectively. When the decaying polonium nucleus is at rest, the momenta of the lead nucleus and the α-particle will be of the same magnitude p and opposite direction. Then the conservation of energy is expressed by the equation

$$Mc^2 = c\sqrt{p^2 + m^2c^2} + c\sqrt{p^2 + \mu^2c^2},$$

from which p can be determined. The velocity of the α-particle can then be calculated by means of the formula $v = c^2 p/c\sqrt{p^2 + \mu^2c^2}$. Since in beta-decay the number of the final particles is three, energy and momentum conservation are not sufficient to fix their velocity unambiguously. Hence beta-electrons have, unlike α-particles in the above decay, continuous velocity distribution within a range determined by the conservation laws.

2.20 Massless Particles

If in (2.19.3) we put the mass equal to zero and use (2.19.5) too we arrive at the equations

$$E = cp, \quad \mathbf{v} = c\mathbf{n}, \tag{2.20.1}$$

where $\mathbf{n} = \mathbf{p}/p$. Particles whose energy and momentum obey these relations are permanently in state of motion with the speed of light with respect to any inertial frame. In this respect they differ significantly from particles with mass however small but different from zero. This contrast is closely related, through (2.19.3), to the sharp distinction between timelike and lightlike spacetime intervals.

Since particles of this kind are never at rest it is meaningless to speak of their rest frames. But, according to Sect. 1.7, mass is the measure of inertia of a particle at rest. Therefore, $m = 0$ particles are *genuinely massless* in the sense that the notion of mass is inapplicable to them (while the term 'zero mass particles' would suggest that they possess mass equal to zero). As an obvious corrolary: *mass–energy relation is inapplicable to massless particles.*

How the energy of a massless particle is altered when it is observed from an inertial frame, moving toward it?

Consider the standard setting of inertial frames $\mathcal{I}$ and $\mathcal{I}'$, and assume that a massless particle of energy E moves in the negative direction of the x-axis of $\mathcal{I}$. Its momentum is then equal to $\mathbf{p} = (-\,E/c,\ 0,\ 0)$. According to Lorentz-transformations, in $\mathcal{I}'$ its energy E' is larger than E:

$$E' = \frac{E - Vp_x}{\sqrt{1 - V^2/c^2}} = \frac{E + \frac{V}{c}E}{\sqrt{1 - V^2/c^2}} = \sqrt{\frac{1 + V/c}{1 - V/c}}\,E. \tag{2.20.2}$$

This increase in energy cannot be attributed to the change in the velocity since the latter is equal to the same c in both $\mathcal{I}$ and $\mathcal{I}'$. A massless particle must, therefore, posses a property, other than velocity, which determines its energy and *does* change when going from one inertial frame to the other.

For an explanation return to the discussion of the Doppler-effect in Sect. 1.2. When (1.2.7) is applied to the receiver which is approaching the emitter [in this case V has to be replaced by $(-V)$], the ratio ν'/ν is found *identical* to E'/E obtained from (2.20.2)! The conclusion from this coincidence is that the ratio of the energy of a massless particle to the frequency of a light wave is independent of the inertial frame chosen: $E/\nu = konstans$.

If, therefore, frequency was among the attributes of massless particles then it might be the required property we are looking for. This demand is indeed fulfilled in quantum theory but in the reverse direction, since light waves are endowed with particle properties: *photons* are massless particles associated with light waves. The value of the energy-to-frequency ratio is *the Planck-constant h*. It follows from the first equation of (2.20.1) that, beside energy, photons posses momentum too: $p = E/c = h\nu/c = h/\lambda$.

2.21 The Transformation of the Electromagnetic Field

According to the widespread belief Einstein discovered relativity theory in the course of his analysis of the negative result of the Michelson–Morley experiment. But this is a mere urban legend which is refuted by both the text of Einstein's original paper *On The Electrodynamics of Moving Bodies* and his later recollections.

In the first paragraph of this paper Einstein stated the motivation of his study with his usual clarity:

> It is known that Maxwell's electrodynamics—as usually understood at the present time—when applied to moving bodies, leads to asymmetries which do not appear to be inherent in the phenomena. Take, for example, the reciprocal electrodynamic action of a magnet and a conductor. The observable phenomenon here depends only on the relative motion of the conductor and the magnet, whereas the customary view draws a sharp distinction between the two cases in which either the one or the other of these bodies is in motion. For if the magnet is in motion and the conductor at rest, there arises in the neighbourhood of the magnet an electric field with a certain definite energy, producing a current at the places where parts of the conductor are situated. But if the magnet is stationary and the conductor in motion, no electric field arises in the neighbourhood of the magnet. In the conductor, however, we find an electromotive force, to which in itself there is no corresponding energy, but which gives rise—assuming equality of relative motion in the two cases discussed—to electric currents of the same path and intensity as those produced by the electric forces in the former case.

The embarassing fact here was that, in spite of the sharp difference between the underlying physical pictures, the calculation based on the Maxwell equations lead precisely to the same force in both cases, proving that the asymmetry is indeed not ‘inherent in the phenomena’. But this seemed only an accidental coincidence rather than a necessary consequence of the equivalence of the rest frames of the magnet and the conductor since Maxwell-equations are not invariant with respect to the Galilei-transformations. Einstein’s primary aim in his paper was, therefore, to replace Galilei-transformations with a generalization of them which ensure the same mathematical form of the Maxwell-equations in all inertial frames and reduce to Galilei-transformations in those domains of experience where these latter transformations have been proved to work well.

Since the value of the light speed may be deduced from the Maxwell-equations, Einstein perhaps started from the reasonable expectation that, if such transformations do indeed exist, they should lead automatically to constancy of the light speed and, moreover, they could possibly be found as the consequence of this constancy. This is at least as good an argument in favour of the constancy of light velocity as the negative result of the Michelson–Morley experiment.

Maxwell-equations and their invariance will not be discussed here. But it has to be noted that in his attempts to find the correct transformations of these equations one of Einstein’s difficult task was to find the transformation of the electric and magnetic field from one inertial frame to the other. For a standard setting of the coordinate systems these transformations written in the International System of Units (SI) are as follows:

$$E'_x = E_x, \quad E'_y = \frac{E_y - VB_z}{\sqrt{1 - V^2/c^2}}, \quad E'_z = \frac{E_z + VB_y}{\sqrt{1 - V^2/c^2}} \tag{2.21.1}$$

$$B'_x = B_x, \quad B'_y = \frac{B_y + \frac{V}{c^2} E_z}{\sqrt{1 - V^2/c^2}}, \quad B'_z = \frac{B_z - \frac{V}{c^2} E_y}{\sqrt{1 - V^2/c^2}}. \tag{2.21.2}$$

In Sect. 2.18 we have left open the problem of how to express the primed components of the force in terms of the unprimed ones, valid in the inertial frame in which the particle is moving. For a point charge the answer can now be given. Since a charge at rest is acted upon solely by the Coulomb-force, in (2.18.3) we can specify the force as $\mathbf{F}' = Q\mathbf{E}'$. Then, using (2.21.1), we obtain the following equation of motion for a point charge of magnitude Q and mass m:

$$\begin{aligned} \frac{m}{(1 - v^2/c^2)^{3/2}}\, a_x &= QE_x, \\ \frac{m}{1 - v^2/c^2}\, a_y &= Q\,\frac{E_y - vB_z}{\sqrt{1 - v^2/c^2}} \\ \frac{m}{1 - v^2/c^2}\, a_z &= Q\,\frac{E_z + vB_y}{\sqrt{1 - v^2/c^2}}. \end{aligned} \tag{2.21.3}$$

These equations are often cast into the form

$$
\begin{aligned}
m_{\|}a_x &= QE_x, \\
m_{\perp}a_y &= Q(E_y - vB_z) \\
m_{\perp}a_z &= Q(E_z + vB_y),
\end{aligned}
$$

where

$$m_{\|} = \frac{m}{(1 - v^2/c^2)^{3/2}} \quad \text{and} \quad m_{\perp} = \frac{m}{\sqrt{1 - v^2/c^2}}$$

are the longitudinal and transverse masses, respectively.

Equations (2.21.3) can be written as a single vector equation which is applicable when the force is of arbitrary direction:

$$\frac{m}{\sqrt{1 - v^2/c^2}}\,\mathbf{a} = Q\left[\mathbf{E} - \frac{1}{c^2}(\mathbf{E}\cdot\mathbf{v})\mathbf{v} + (\mathbf{v}\times\mathbf{B})\right] \tag{2.21.4}$$

(the $\cdot$ and the $\times$ denote scalar and vector product). The validity of this equation follows from the fact that for $\mathbf{v} = (v, 0, 0)$ it reduces to (2.21.3). When $v^2/c^2 \ll 1$ the force on the right hand side becomes equal to the sum of the Coulomb and Lorentz forces.

2.22 The Thomas-Precession

The amazing property of the gyroscope is that if no external torque is acted upon it the orientation of its spin axis remains fixed, regardless of the motion of the platform on which it is mounted. If, for example, the gyroscope frame is pinned on a disc at the distance r from its axis and its spinning axis is directed toward the point P at the wall of the laboratory then it will continue to point at P even after the disc is made to rotate. Thanks to this property gyroscopes became indispensable part of navigation systems. When e.g. the axis of the gyroscope of an airplane navigation system is seen to decline 5° to the right around the vertical axis than the cockpit crew knows that the aircraft has made a 5° turn to the left.

But this behaviour of the gyroscope is true only in Newtonian physics. According to relativity theory gyroscopes, moving on curvilinear trajectories, do not preserve the orientation of their spin axis. For example if mounted on a rotating disc their spin direction will *precess* with respect to the laboratory around the rotation axis of the disc. The angular velocity of this motion known as *Thomas-precession* is given by the formula

$$\omega_T = \omega[1 - \gamma(v)] = \omega\left[1 - \frac{1}{\sqrt{1 - v^2/c^2}}\right]. \tag{2.22.1}$$

Since ω_T is obviously negative the sense of Thomas-precession is opposite to that of the rotation of the support.

For the gyroscope mounted on the rotating disc we have $v = r\omega$ but (2.22.1) remains valid in the general case too. If the acceleration of the carrier of the gyroscope is not collinear to its velocity, i.e. when it rotates with some instantaneous angular velocity ω in the plane S which contains this pair of vectors then Thomas-precession will take place in S with the angular velocity given by (2.22.1).[6] If aircrafts could fly with speed comparable to light velocity then Thomas-precession would render navigation by means of gyroscopes less straightforward since the true deviation in the course of the aircraft would be less then indicated by the deflection of the gyroscope with respect to the cockpit.

2.23 The Sagnac Effect

In this section another phenomenon on a rotating disc will be discussed. Imagine a source of light placed at the point A_1 on the disc's circumference wich is capable to emit simultaneously a pair of light pulses along the tangent both in the direction of the rotation velocity and opposite to it. A series of mirrors reflects the signals to keep their track along the perimeter.

The point on the disc opposite to A_1 will be denoted by A_2 and a pair of diametrically opposite points B_1 and B_2 will also be marked on the ground. Suppose that the light flash at A_1 takes place at the moment when A_1 passes by B_1. The two signals emitted in diametrically opposite directions will meet again at B_2 after having travelled a distance πR along the circumference (R is the radius of the disc). In the laboratory frame, wich is an inertial one by assumption, the time required to cover this distance is equal to $\Delta t = \pi R/c$.

The point of encounter *on the disc* will obviously have an angular displacement with respect to A_2 equal to $\omega\Delta t$ where ω is the angular velocity of rotation. Therefore, the pulse wich propagated in the counter rotation direction travelled only a distance $(\pi - \omega\Delta t)R$ with respect to the disc. Accordingly, the distance covered by the pulse of opposite direction is equal to $(\pi + \omega\Delta t)R$. The path difference is, therefore, equal to $2\Delta t \times \omega R = 2S\omega/c$ where $S = \pi R^2$ is the area enclosed by the paths of the signals. As we see, the path difference is proportional to the instantaneous angular velocity of the device with respect to the inertial frames. In *optical gyroscopes* this path difference is monitored continuously in order to keep track of the orientation of the equipment in space. Instruments of this kind are as useful for purposes of navigation as their mechanical predecessors.

[6] In the general case the angular velocity *vector* is equal to $\omega = \frac{1}{v^2}\left(\mathbf{v} \times \frac{d\mathbf{v}}{dt}\right)$. When this is substituted into (2.22.1) we obtain for the Thomas-precession angular velocity *vector* the expression $\omega_T = (1-\gamma)\frac{1}{v^2}\left(\mathbf{v} \times \frac{d\mathbf{v}}{dt}\right) = -\frac{\gamma}{1+\gamma}\frac{1}{c^2}\left(\mathbf{v} \times \frac{d\mathbf{v}}{d\tau}\right)$.

The use of moving mirrors in the above description may raise doubt on the validity of the formula $\Delta t = \pi R/c$ since the velocity of the light pulses might perhaps be influenced by the motion of the mirrors which reflect it. Such an influence is, however, excluded by the basic principle of relativity theory, according to which the velocity of light in inertial frames is equal to c irrespective of its source. As a matter of fact, in gyroscopes available in commerce light pulses are replaced by monochromatic laser beams of wavelength λ and mirrors by fiber optics. The path difference can then be converted into a phase shift

$$\Delta\varphi = (2\pi/\lambda) \times (2S\omega/c) \tag{2.23.1}$$

which can be measured by interferometric methods. The phase shift itself is known as the *Sagnac effect*. Both mechanical and optical gyroscopes are suitable for the laboratory demonstration of the Earth's rotation. This fact alone is sufficient to assert that on the surface of the Earth light velocity is not isotropic.

Springer
http://www.springer.com/978-3-642-17809-2

Basic Relativity
An Introductory Essay
Hraskó, P.
2011, VIII, 104 p. 16 illus., Softcover
ISBN: 978-3-642-17809-2