Edwin F. Taylor
Massachusetts Institute of Technology

John Archibald Wheeler
Princeton University

Spacetime Physics

W. H. Freeman and Company
San Francisco

Printed in the United States of America

Library of Congress Catalogue Card Number: 65-13566

International Standard Book Number: 0-7167-0336-X

PAGES 21 THROUGH 40

A second edition of this book has been published:
W. H. Freeman and Company, New York, 1992, ISBN 0 7167 2327 1

A treatment of general relativity by the same authors:
Exploring Black Holes
Introduction to General Relativity
Addison Wesley, San Francisco, 2000, ISBN 0 201 38423 X

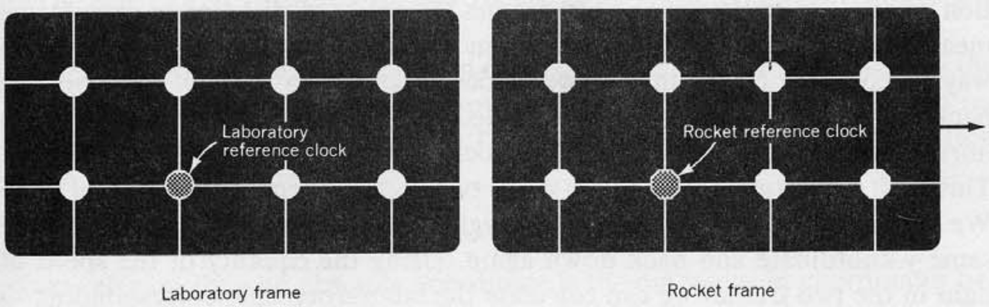


Fig. 11. Laboratory and rocket frames, further schematized from Fig. 10. The central reference clock of each frame is shaded.

Laboratory and rocket observers record single event

A firecracker explodes. The explosion is recorded by the clock in the laboratory lattice nearest to the explosion. It is recorded also by the clock in the rocket lattice nearest to the explosion. How do the coordinates of the recording laboratory clock compare with the coordinates of the recording rocket clock? One result can be derived immediately from the principle of relativity: the recording laboratory and rocket clocks will have the same y coordinates. To show this, let the recording rocket clock carry a wet paint brush that makes marks on the laboratory lattice as it moves past. Figure 12 shows this for the special case, $y = 1$ meter. The marks on the laboratory lattice serve to measure the laboratory y coordinate of the $y = 1$ rocket clock. These paint marks appear *on* the $y = 1$ laboratory clocks rather than *above* them or *below* them. For suppose that the paint marks appear on the lattice rods *below* the $y = 1$ laboratory clocks: Then *both* observers will agree that the $y = 1$ rocket clocks passed "inside" the $y = 1$ laboratory clocks. Permanent paint marks would verify this for all to see. Similarly, if the paint marks appear on the lattice rods *above* the $y = 1$ laboratory clocks, both observers will agree that the $y = 1$ rocket clocks passed "outside" the $y = 1$ laboratory clocks. In either case there would be a way to distinguish experimentally between the two frames. But no one has been able to distinguish between these two frames using any *other* experiment. The principle of relativity embodies the assumption that any such experimental distinction between inertial reference frames is impossible. Therefore we assume that no one could distinguish between the two frames using *this* experiment. It follows that the y coordinate of *any* event—such as the explosion that began this paragraph—will have the same y coordinate in the rocket frame as in the laboratory frame.

y coordinate of event is same in lab and rocket frames

By a similar argument the z coordinate of an event is the same in the rocket frame as in the laboratory frame. Notice that both the y coordinate and the z coordinate of an event are measured in a direction *perpendicular* to the direc-

z coordinate of event is same in lab and rocket frames

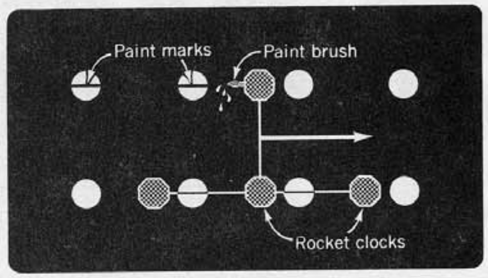


Fig. 12. Demonstration that y coordinate of an event is the same in laboratory and rocket frames.

tion of relative motion of the two frames. Equality of distance in each frame measured perpendicular to the direction of relative motion gives us a clean way to compare clocks in the two latticeworks. Let a flash of light bounce back and forth between a mirror mounted on the rocket reference clock and a mirror mounted on the $y = 1$ rocket clock directly above the reference clock. This flash will return to the origin every two meters of rocket light-travel time. We can trace the path of this flash of light in the laboratory frame up to the same y coordinate and back down again. Using the equality of the speed of light in the two frames we can calculate the laboratory time corresponding to the 2-meter round-trip time in the rocket frame. In the next section this study will lead to a demonstration of the invariance of the interval.

5. Invariance of the Interval

Looking for measure of separation AB that has same value in all inertial frames

Distance between two town gates is calculated from the difference between the x coordinates of the two gates and the difference between the y coordinates. How does one find the analogous physical quantity, the spacetime interval between two events? And between what two events shall this interval be evaluated?

Let *event A* be the *emission* of a flash of light. Let *event B* be the *reception* of this flash after its reflection from another object. All that matters in the end is the pair of events. Neither the light nor the object that reflects it is of any direct interest. Nevertheless, an analysis of the track of this pulse through spacetime reveals quickly and simply a quantity (the interval) that is associated with the two events and that has the same value in all inertial reference frames.

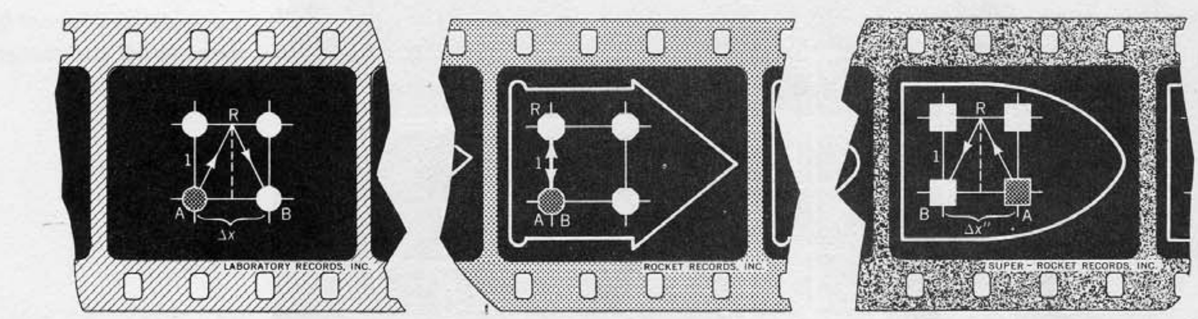
Event A: flash emission
Event B: flash reception

Event A: A spark plug fires. A flash of light flies up to reflector R in Fig. 13. Then the flash wings down. *Event B:* The flash is recorded. Now for the details (Figure 13).

The spark plug fires in the laboratory frame at the zero of time and at the origin of the x, y, z coordinate system (crosshatched). The rocket passes by with such timing that it records the spark as taking place also at *its* origin (likewise crosshatched) and at *its* zero of time. So much for the coordinates of the event of emission:

Details: lab and rocket coordinates of events A and B

$$x_{\text{emission}} = 0, \quad y_{\text{emission}} = 0, \quad t_{\text{emission}} = 0 \quad \text{(laboratory frame)}$$
$$x'_{\text{emission}} = 0, \quad y'_{\text{emission}} = 0, \quad t'_{\text{emission}} = 0 \quad \text{(rocket frame)}$$



A. Light path as observed in laboratory frame

B. Light path as observed in rocket frame

C. Light path as observed in super-rocket frame

Fig. 13. Emission and reflection of the reference flash, and its reception at origin of rocket frame.

The reflector is mounted on the rocket clock 1 meter directly above the origin.

The reception of the flash occurs in the rocket frame at the same place as the emission. The light flash travels a round-trip path of 2 meters. This trip requires 2 meters of light-travel time. The coordinates of the event of reception in the rocket frame are therefore

$$x'_{\text{reception}} = 0, \qquad y'_{\text{reception}} = 0, \qquad t'_{\text{reception}} = 2 \text{ meters}$$

More relevant than an absolute coordinate is the difference in coordinates between the event of reception and the event of emission

$$\begin{aligned} \Delta x' &= x'_{\text{reception}} - x'_{\text{emission}} = 0 \\ \Delta y' &= y'_{\text{reception}} - y'_{\text{emission}} = 0 \\ \Delta t' &= t'_{\text{reception}} - t'_{\text{emission}} = 2 \text{ meters} \end{aligned}$$

In the *laboratory* frame the light flash is received, not at the origin, but at the distance Δx to the right of the origin. High rocket speed implies a large Δx; low rocket speed, a small Δx. (The distance is shown as 1 meter in the figure, but the following analysis is correct for any distance.) In the laboratory frame the flash travels the hypotenuse of two right triangles. Each has a base of $(\Delta x/2)$ and an altitude of 1 meter. The total length of the light path is

$$2[1 + (\Delta x/2)^2]^{1/2}$$

Recall that the speed of light is the same in the laboratory frame as in the rocket frame—the preposterous-but-true character of nature! Therefore the time difference between emission and reception in the laboratory frame is given by the identical formula

$$\Delta t = t_{\text{reception}} - t_{\text{emission}} = 2[1 + (\Delta x/2)^2]^{1/2} \tag{4}$$

in meters of light-travel time.

Why is this time greater than 2 meters? Because the hypotenuse of a right triangle in Fig. 13,A, is greater than its altitude! There is no escape from the conclusion that *the time between emission and reception is not the same in the two reference frames.*

Time between A and B has different values for lab and rocket observers

Both time and space differences between the event of reception and the event of emission are summarized in Table 5.

Table 5. Difference in coordinates between the event of reception and the event of emission.

Laboratory frame	*Rocket frame*
$x_{\text{reception}} - x_{\text{emission}} = \Delta x$	$x'_{\text{reception}} - x'_{\text{emission}} = \Delta x' = 0$
$t_{\text{reception}} - t_{\text{emission}} = \Delta t = 2[1 + (\Delta x/2)^2]^{1/2}$	$t'_{\text{reception}} - t'_{\text{emission}} = \Delta t' = 2$ meters

The time lapse is different in the two reference frames; and so is the space separation—just as the coordinates Δx and Δy of the separation between two town gates are different for the Daytime and Nighttime surveyors! However, for the surveyors there was a combination of coordinates—the square of the *distance* between gates—that was the same for both of them

$$(\text{distance})^2 = (\Delta x)^2 + (\Delta y)^2 = (\Delta x')^2 + (\Delta y')^2$$

Is there any similar combination of coordinates for two events that will have the same value in the laboratory and rocket frames? Answer: Yes! the square of the *interval*

Interval between A and B has same value for lab and rocket observers

$$\text{(5)} \qquad (\text{interval})^2 = (\Delta t)^2 - (\Delta x)^2 = (\Delta t')^2 - (\Delta x')^2 = (2 \text{ meters})^2$$

as one checks directly by substituting in the quantities listed in Table 5.

The rocket frame chosen in which to analyze these two events is a rather special one, in that both emission and reception occur at the same place in it. Figure 13,C, shows the path of the reflected light in a second rocket frame ("super-rocket frame") that is moving even faster relative to the laboratory frame than is the first rocket. In this second rocket frame the difference between the x coordinates for emission and reception (double primes on symbols) $x''_{\text{reception}} - x''_{\text{emission}} = \Delta x''$ is a *negative* quantity because the reception occurs on the negative x axis in this frame. Nevertheless, $(-\Delta x'')^2 = (\Delta x'')^2$ and we can still use the right triangles in Fig. 13,C, to show that the total length of the light path in this second rocket frame is given by the expression $2[1 + (\Delta x''/2)^2]^{1/2}$—which is the same in form as that for the laboratory frame. The speed of light must have the same value in the second rocket frame as in the first rocket frame. Therefore the time between emission and reception is given by

$$t''_{\text{reception}} - t''_{\text{emission}} = \Delta t'' = 2[1 + (\Delta x''/2)^2]^{1/2}$$

Therefore

Interval AB has same value in all rocket frames!

$$(\Delta t'')^2 - (\Delta x'')^2 = (2 \text{ meters})^2$$

also, and in summary

$$\text{(6)} \qquad (\Delta t)^2 - (\Delta x)^2 = (\Delta t')^2 - (\Delta x')^2 = (\Delta t'')^2 - (\Delta x'')^2 = (2 \text{ meters})^2$$

Now forget the outgoing light flash, the reflector, and the returning light flash. They were only tools. They helped to identify the quantity that has the same value in different frames of reference. From now on focus on the quantity itself, the interval. Disregard the details of the derivation.

What is the same in two inertial frames; what is nearly the same; and what is different

What has been learned? Two events, A and B, occur at the same point in the rocket frame ($\Delta x' = 0$) but at different times ($\Delta t' = 2$ meters). Viewed in the laboratory frame, those same two events are separated in space by a distance Δx—the faster the rocket happens to be moving, the greater the distance. This result is hardly surprising. One is even entitled to say, "What could be more obvious!" The surprise comes elsewhere. First, *the time* Δt between the two events as recorded in the laboratory frame *does not have the same value as it has in the rocket frame*. Second, the time between A and B as punched out by the two relevant recording laboratory clocks is *greater* than the time between the same two events as recorded by the indentical reference clock of the rocket: $\Delta t \geqslant \Delta t'$. Third, the factor of increase of the time (see Table 5)

$$\Delta t/\Delta t' = [1 + (\Delta x/2)^2]^{1/2}$$

is close to unity (that is, the increase itself is very small) if the distance Δx covered by the rocket between events A and B is small. However, if the rocket moves very fast, Δx is a very great quantity, and the factor of discrepancy between the two times is enormous. Fourth, despite this newly found difference in *time* as recorded in the two reference frames, and despite the long-known difference between the *space* separation of the events in the two reference frames ($\Delta x \neq \Delta x' = 0$), there is nevertheless a quantity that *is* the same in the laboratory frame as the 2 meters of elapsed light-travel time between A and B in the rocket frame. This quantity is the interval,

$$(\text{interval}) = [(\Delta t)^2 - (\Delta x)^2]^{1/2}$$

The rocket speed may be very high. Then Δx will be very large. But then Δt is also very large. Moreover, the magnitude of Δt is perfectly tailored to the size of Δx. In consequence, the special quantity $(\Delta t)^2 - (\Delta x)^2$ has the value (2 meters)^2, no matter how great Δx and Δt individually may be.

One diagram illustrates the four great ideas of special relativity

All of these relationships can be seen at a glance in Fig. 13,A. The hypotenuse of the first right triangle is $\Delta t/2$. Its base is $\Delta x/2$. To say that $(\Delta t)^2 - (\Delta x)^2$ has a standard value, and thus to state that $(\Delta t/2)^2 - (\Delta x/2)^2$ has a standard value, is to say that the altitude of this right triangle has a fixed magnitude (1 meter in the diagram) no matter how fast the rocket is going. What then was the keystone of the argument establishing the fact that $(\Delta t)^2 - (\Delta x)^2$ has the value (2 meters)^2, no matter how fast the rocket is moving? The keystone was the principle of relativity, according to which there is no difference in the laws of physics between one inertial reference frame and another. This principle was put to use here in two very different ways. First, it was used to reason that distances at right angles to the direction of relative motion are recorded as of equal magnitude in the laboratory frame and in the rocket frame. Otherwise one frame could be distinguished from the other as the one with the shorter perpendicular distances. Second, the principle of relativity was employed to deduce that the speed of light must be the same in the laboratory frame as in the rocket frame—a deduction supported by the Kennedy-Thorndike experiment. The speed being the same, the fact that the light-travel path in the laboratory frame (the hypotenuses of two triangles) is longer than the simple round-trip path in the rocket frame (the altitudes of the two triangles: up 1 meter and down again) directly implies a longer *time* in the laboratory frame than in the rocket frame.

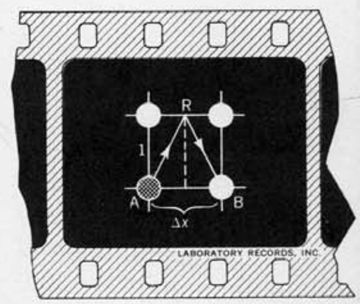


FIG. 13,A, PAGE 22

In brief, in the one elementary triangle of Fig. 13,A, are displayed the four great ideas that underlie all of special relativity: invariance of perpendicular distance, invariance of the speed of light, dependence of space *and time* coordinates upon the frame of reference, and invariance of the interval.

Is inequality of lab and rocket time lapses a paradox?

If Fig. 13,A, thus epitomizes all of special relativity in a form easy to remember, the foregoing analysis of the figure nevertheless leads to what at first sight seems to be a preposterous conclusion. How can it possibly make sense for the lapse of time between two events to be longer in the laboratory than in the rocket? Has it not already been argued that "distances at right angles to the direction of relative motion" are equal, "otherwise one frame could be distinguished from the other as the one with the shorter perpendicular distances"? What about the difference between *time* lapses in the two frames? Does not this difference give a way to differentiate physics in one frame from

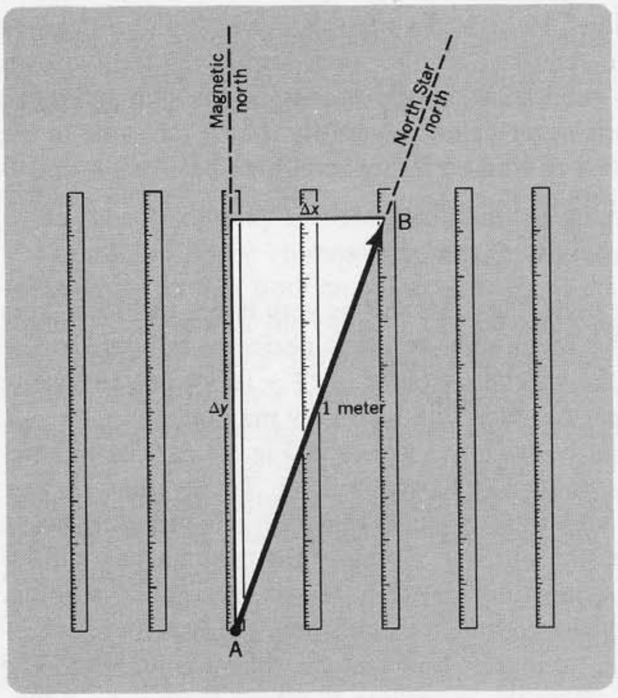


Fig. 14. The north-south separation of point B from point A (the "*northing* of B relative to A") depends upon the choice of the direction north.

physics in the other? Yet is not such a difference ruled out by the principle of relativity—the principle that one inertial reference frame is as good as another?

Relativity of time (Lorentz) compared with relativity of "northing" (Euclid)

For answers to these questions, turn back to the parable of the surveyors. Consider point B in Fig. 14. It is one meter straight north of another point, A, according to the reckoning of the Nighttime surveyor and his North Star north. Now consider the location of point B according to the Daytime surveyor and his magnetic north. Is the y separation Δy between A and B (surveying terminology: the *northing* of B relative to A) also one meter in the Daytime frame? No, Δy is *less* than one meter! How can this be? Because the altitude (Δy) of a right triangle is shorter than the hypotenuse (1 meter). Does this mean that the rules of surveying in the Daytime coordinate system are different from those in the Nighttime coordinate system? Evidently not! Similarly, there is no flaw in the construction or functioning of the laboratory clocks that makes them give longer readings for the time lapse AB. The "discrepancy" between the laboratory clocks and the rocket clock is caused instead by the character of spacetime geometry itself. That is the way the world is built! The analogy between the Lorentz geometry of spacetime and the Euclidean geometry of the surveyors' world is expanded in Table 6 (pages 28 and 29).

6. The Spacetime Diagram; World Lines

The spacetime diagram: simple way to display events

A simple way to look at the events of emission and reception of the last section is to plot the position of the event on the horizontal axis and the time of the event on the vertical axis of a *spacetime diagram* (Fig. 15). The light is *emitted* from a spark plug attached to the reference clock of the first rocket. This plug fires at the instant when this clock passes the laboratory reference clock. Both clocks then read zero. Therefore the event of emission is located at the origin of the spacetime diagram plotted by the rocket observer

$$x'_{\text{emission}} = 0, \qquad t'_{\text{emission}} = 0$$

It is also located at the origin of the spacetime diagram made by the laboratory observer

$$x_{\text{emission}} = 0, \qquad t_{\text{emission}} = 0$$

The further history of the relevant light ray looks different as plotted in the spacetime diagrams of the laboratory and the two rockets. In the first rocket the *reception* of the reflected flash occurs at $x' = 0$ and 2 meters of time later than the reference event

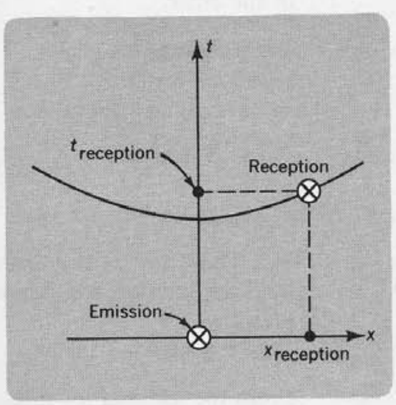
$$x'_{\text{reception}} = 0, \qquad t'_{\text{reception}} = 2 \text{ meters}$$

as already recorded in Table 5 and as seen more directly in Fig. 15,B. In the laboratory frame the event of reception is located to the right of the origin

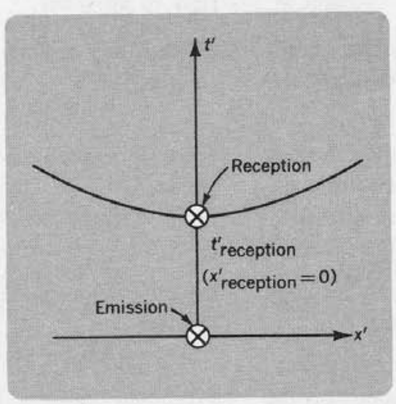
$$\begin{aligned} x_{\text{reception}} &= \text{a positive quantity} \\ t_{\text{reception}} &= [(2 \text{ meters})^2 + x^2_{\text{reception}}]^{1/2} \\ &= \text{a time } \textit{greater} \text{ than 2 meters} \end{aligned}$$

as shown in Fig. 15,A. In the second rocket frame (the second rocket is moving faster than the first!) the event of reception appears to the left of the origin (Fig. 15,C)

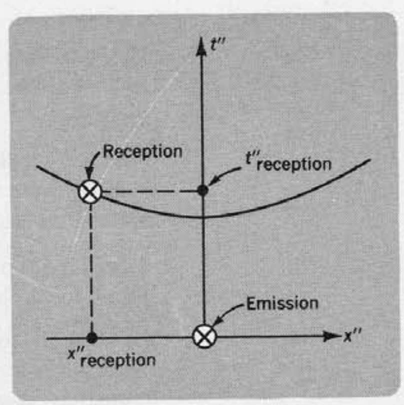
$$\begin{aligned} x''_{\text{reception}} &= \text{a negative quantity} \\ t''_{\text{reception}} &= [(2 \text{ meters})^2 + (x''_{\text{reception}})^2]^{1/2} \\ &= \text{a time } \textit{greater} \text{ than 2 meters} \quad \text{(again!)} \end{aligned}$$



A. Laboratory spacetime diagram



B. Rocket spacetime diagram



C. Super-rocket spacetime diagram

Fig. 15. Spacetime diagrams showing emission of the reference flash and its reception after reflection. The hyperbola drawn in each figure satisfies the equations $(\text{interval})^2 = t^2 - x^2 = t'^2 - x'^2 = t''^2 - x''^2$

Table 6. Comparison of the difference in northing between points A and B in Daytime and Nighttime coordinate systems and the difference in time between events A and B in laboratory and in rocket reference frames.

Questions	*Answers of a student of surveying concerning difference in northing between points A and B. (See Fig. 14.)*	*Answers of a student of spacetime physics concerning difference in time between events A and B. (See Fig. 13.)*
In which frame of reference does the separation of B from A appear simplest?	Coordinate system of Nighttime surveyor based on North-Star north.	Reference frame of rocket.
What is simplifying feature in this frame of reference?	Points have same x' coordinate; or $\Delta x' = 0$.	Events have same x' coordinate; or $\Delta x' = 0$.
Why does this feature simplify the measurement of the separation AB?	One meter stick oriented to North-Star north suffices (1) to verify that both points *do* have the same x' coordinate and (2) to measure directly the northing of B relative to A.	One recording clock attached to rocket frame suffices (1) to verify that both events *do* have the same x' coordinate and (2) to measure directly the time delay of B relative to A.
What is an alternative frame for analyzing the separation AB?	Coordinate system of Daytime surveyor based on magnetic north.	Reference frame of laboratory.
What complication is there in analyzing the separation in this alternative frame?	No single one of *his* meter sticks, oriented to magnetic north, can locate both A and B.	No single one of the recording laboratory clocks can register both A and B.
How is this difficulty met?	*Two* of these north-oriented meter sticks are needed, one located Δx meters to the right of the other.	*Two* of these laboratory clocks are needed, one located Δx meters to the right of the other.
What is the reading on the first of these measuring devices?	Point A at $y = 0$.	Event A at $t = 0$.
And the reading on the second of these measuring devices?	Point B located Δy meters north.	Event B delayed by Δt seconds.
Does the coordinate thus found for B directly measure its separation from A?	No! The northing Δy is less than the distance AB. More precisely, $\Delta y = [(\text{AB})^2 - (\Delta x)^2]^{1/2}$.	No! The delay Δt is greater than the interval AB. More precisely, $\Delta t = [(\text{AB})^2 + (\Delta x)^2]^{1/2}$.
Then how does one find the separation AB from measurements in this frame of reference?	From the formula for distance, $(\text{distance})^2 = (\Delta x)^2 + (\Delta y)^2$. (Test by substituting in the expression for Δy from the entry above!)	From the formula for interval, $(\text{interval})^2 = (\Delta t)^2 - (\Delta x)^2$ (Test by substituting in the expression for Δt from the entry above!)

What is the distinction in the present examples between measurements made in the primed and the unprimed frames?	Δy is *less* than $\Delta y'$ ($=$AB).	Δt is *greater* than $\Delta t'$($=$AB).
Isn't there something preposterous about this result?	Meaning that identical meter sticks give nonidentical northings?	Meaning that identical clocks record nonidentical times?
Yes! Does not this discrepancy prove that there is some inner contradiction in the reasoning?	No! A single Nighttime meter stick suffices to establish the distance AB. But there is no *single* Daytime meter stick with which one establishes the (lesser) magnetic northing of B relative to A. Therefore no Daytime meter stick can be said to disagree with the Nighttime meter stick.	No! A single rocket clock records the interval AB. But there is no *single* laboratory clock with which one establishes the (greater) laboratory time delay of B relative to A. Therefore no laboratory clock can be said to disagree with the rocket clock.
Is there some fundamental difference between the primed and the unprimed frame of reference that is responsible for the one-sided difference between coordinate values?	For $\Delta y < \Delta y'$? No!	For $\Delta t > \Delta t'$? No!
Then what *is* responsible for this one-sidedness?	The point B happens to lie on the same North-Star north line as A, but *not* on the same magnetic north line as A.	Event B by chance occurs at the same point in the rocket frame as A, but *not* at the same point as A in the laboratory frame.
How can the identical character of the physics in the two frames of reference be readily illustrated?	Pick a point C that has the same x coordinate as A (C in the line of magnetic north relative to A)	Pick an event C that has the same x coordinate as A (C at the same *place* as A in the laboratory frame, but later in *time*).
For such a choice of C, what is the distinction between measurements made in the primed and the unprimed frames?	Δy ($=$AC) is *greater* than $\Delta y'$.	Δt ($=$AC) is *less* than $\Delta t'$.
How can the discussion be summarized?	There is no paradox about northward component of AB having different values in two coordinate systems; the discrepancy is not a fault of the meter sticks; not even a fault at all; the "discrepancy" is caused by the inner workings of Euclidean geometry.	There is no paradox about time lapse from A to B having different values in two reference frames; the discrepancy is not a fault of the clocks; not even a fault at all; the "discrepancy" is caused by the very structure of the geometry of the spacetime in which all physics takes place.

Invariant interval corresponds to hyperbola on spacetime diagram

The different points marked "reception" in the different spacetime diagrams all refer to the *same event*. The event is the same but its coordinates in different frames are different. What do these different coordinates of the same event have in common? They all satisfy the equation

$$(\text{time separation})^2 - (\text{space separation})^2 = (\text{interval})^2 = \text{constant}$$

This is the equation of a *hyperbola*. Therefore, *an event which is plotted on the hyperbola $t^2 - x^2 = (constant)$ in the spacetime diagram of any laboratory or rocket frame will be plotted somewhere on a hyperbola with the same equation in the spacetime diagram of every other laboratory and rocket frame.*

Is there likewise a single curve that correlates the different coordinate values obtained by the Daytime and Nighttime surveyors for a single gate? The x and y coordinates of, say, gate A with respect to the town square depends on the choice of the north direction (Fig. 16). The Daytime and Nighttime plots of this gate are shown in Fig. 17, parts A and B. Think of a third and still different set of axes rotated even more than the Nighttime axes relative to the Daytime axes. For the surveyor who uses this third set of axes the x'' coordinate of gate A may be negative (Fig. 17,C).

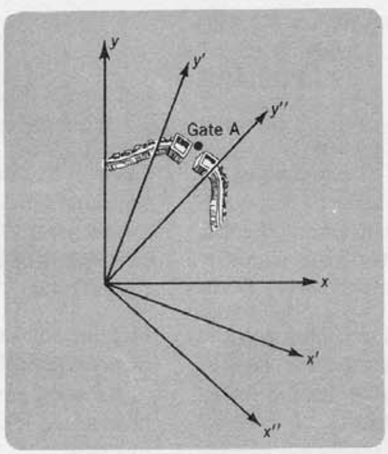
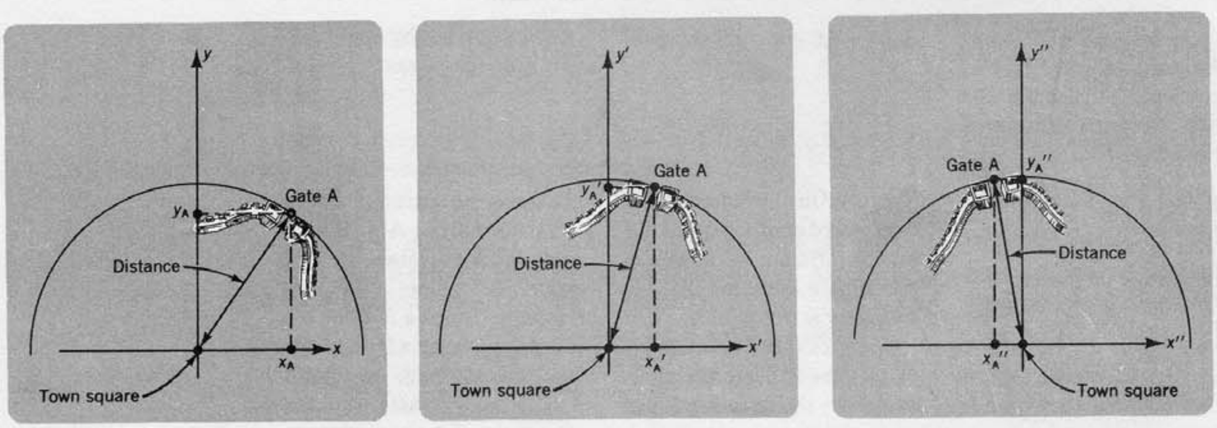


Fig. 16. Relative standards of north for Daytime, Nighttime, and a third surveyor respectively.



A. Daytime plot B. Nighttime plot C. Plot by third surveyor

Fig. 17. Coordinates of Gate A as observed by Daytime, Nighttime, and a third surveyor respectively. The circle drawn in each figure satisfies the equations $(\text{distance})^2 = x^2 + y^2 = x'^2 + y'^2 = x''^2 + y''^2$

Invariant distance corresponds to circle on xy diagram

The different points marked "gate A" in the different plots all refer to the *same gate*. The gate is the same but its coordinates in different plots are different. What do these different coordinates of the same gate have in common? They all satisfy the equation

$$(x \text{ separation})^2 + (y \text{ separation})^2 = (\text{distance})^2 = \text{constant}$$

This is the equation of a *circle*. Therefore, *a point that is plotted on the circle $x^2 + y^2 =$ (constant) in the coordinate system of any surveyor will be plotted somewhere on a circle with the same equation in the coordinate system of every other surveyor.*

Here is the fundamental difference between textbook Euclidean geometry and the real Lorentz geometry of spacetime. In Euclidean geometry the *distance* between two points is an invariant, and as a result, for all surveyors gate A will lie somewhere in the *xy* plane on a *circle* centered on the town square. In Lorentz geometry the *interval* between events is an invariant, and as a result, for all laboratory and rocket observers a given event will lie somewhere on a *hyperbola* in the spacetime diagram when referred to the reference event.

In Euclidean geometry the length—or its square—is always a *positive* quantity

$$(\Delta x)^2 + (\Delta y)^2 = (\Delta x')^2 + (\Delta y')^2 \geqslant 0$$

Three types of separation between two events: timelike lightlike spacelike

In contrast, the squared interval of Lorentz geometry

$$(\Delta t)^2 - (\Delta x)^2 = (\Delta t')^2 - (\Delta x')^2$$

may be *positive*, *negative*, or *zero*, depending on whether the time or the space component predominates. Moreover, whichever of these three descriptions characterizes the interval in one reference frame also characterizes the interval in *any other* reference frame, because the interval has the same value in all frames. Accordingly we find that nature provides a fundamental way to classify the relation between two events. An interval between two events is called *timelike*, *lightlike*, or *spacelike* depending on whether the squared interval is positive, zero, or negative, respectively, as shown in Table 7.

Table 7. Classification of the relation between two events.

Description	*Squared interval*	*Name*
Time part of interval dominates over space part	positive	timelike interval
Time part of interval equals space part	zero	lightlike interval
Space part of interval dominates over time part	negative	spacelike interval

The value of the interval between two events is represented by different symbols depending on whether it is timelike or spacelike. The value of a timelike interval is given the Greek letter tau (τ) and is called the *invariant timelike distance* between two events or the *proper time* (or sometimes the *local time*) between the two events

Proper time and proper distance

(7) $$\Delta\tau = [(\Delta t)^2 - (\Delta x)^2]^{1/2}$$

The value of a spacelike interval between two events is given the Greek letter sigma (σ) and is called the *invariant spacelike distance* or the *proper distance* between the two events

$$\Delta\sigma = [(\Delta x)^2 - (\Delta t)^2]^{1/2} \tag{8}$$

World line of particle

Figure 18 represents as a function of time the location of a particle which started along the x axis from the origin at $t = 0$. Such a plot of position versus time in a spacetime diagram is called the *world line* of the particle. Each lattice clock encountered by the particle punches out the time of coincidence. Thus the world line of the particle can be considered to be made up of these separate events of coincidence. No particle has ever been observed to travel faster than light. Therefore a particle will always travel *less* than one meter of distance in one meter of light-travel time. It follows that events along this world line will have a greater time separation than their separation in space: the world line of a particle will consist of events that are *timelike* with

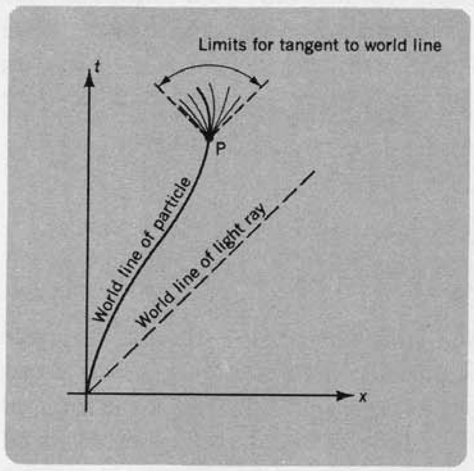


Fig. 18. Timelike world line of a particle.

respect to the initial event and to each other. In other words, *a particle must follow a timelike world line.* A timelike world line is characterized at every point P by a local tangent which lies between the world lines of light rays originating at that point. These light rays travel one meter of distance in one meter of light-travel time. Events along the world line of a light ray have equal space and time separations. Therefore the world line of a light ray consists of events that are *lightlike* with respect to the initial event and to each other. In other words *light rays follow lightlike world lines.*

Path in space has length

Distance is a central idea in all applications of Euclidean geometry. For instance, using a flexible tape measure it is easy to measure the distance s along a path that starts at the town square and winds out through gate A (Fig. 19,A). The distance Δs between any two nearby points on the path (for instance, those marked 3 and 4 in the figure) can also be *calculated* using the difference in coordinates Δx and Δy of the two points with respect to any coordinate system. Since distance is invariant, the distance between these two points will be the same when calculated in any coordinate system even though the separate coordinates Δx and Δy have different values in different coordinate systems. Elsewhere along the path the distance between another pair of nearby points will also be independent of the coordinate system used in evalu-

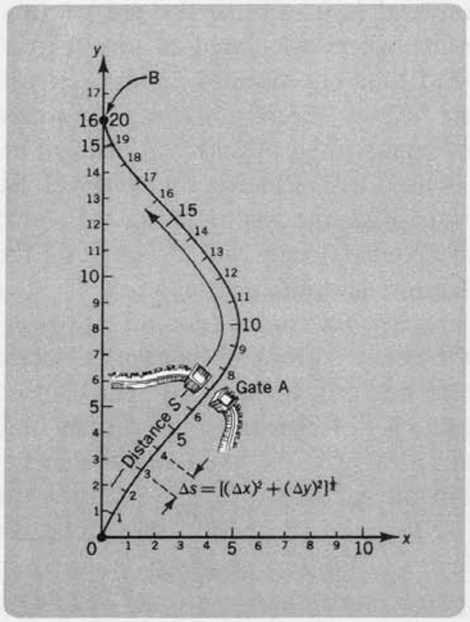


Fig. 19, A. Distance along a winding path which starts at the town square. Notice that the total *distance along the winding path* from point O to point B is *greater* than the *distance along the straight y axis* from point O to point B.

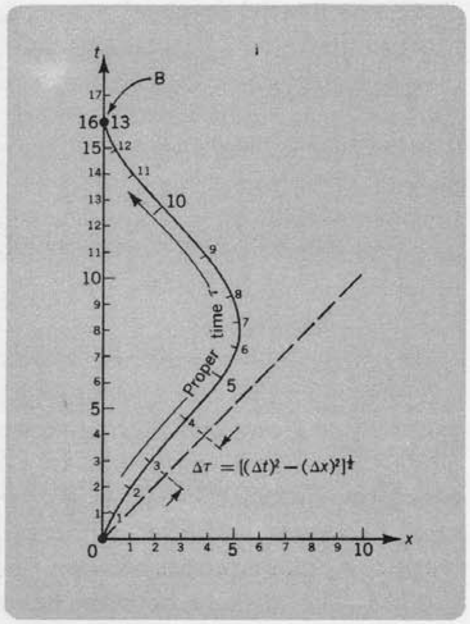


Fig. 19, B. Proper time along a curved world line in a spacetime diagram. Notice that the total *proper time along the curved world line* from event O to event B is *smaller* than the *proper time along the straight t axis* from event O to event B.

ating that distance. So too for the *sum* of the lengths of all the segments of the path! Thus different surveyors using different coordinate systems will all agree on the distance along a given path from a specified initial point O to a specified final point B.

It is possible to proceed from O to B along quite another path—for example, along the straight line OB in Fig. 19,A. The length of this alternative path is evidently different from that of the original path. This difference in length of different paths between O and B is a feature of Euclidean geometry so well known as to occasion hardly any comment and certainly no surprise. In Euclidean geometry a curved path between two specified points is *longer* than a straight path between the same two points. The existence of the difference of length between two paths violates no law. No one would claim that a tape measure fails to perform properly when laid along a curved path.

Length is shortest for direct path

Proper time is to a world line in Lorentz geometry what length is to a path in Euclidean geometry. The world line is started at an event O and ended at an event B. There are many different world lines that start at O and end at B. The lapse of proper time on each is well defined; but it differs between one world line and another. Is this surprising? Then it is appropriate to look more closely at how the proper time is defined and measured.

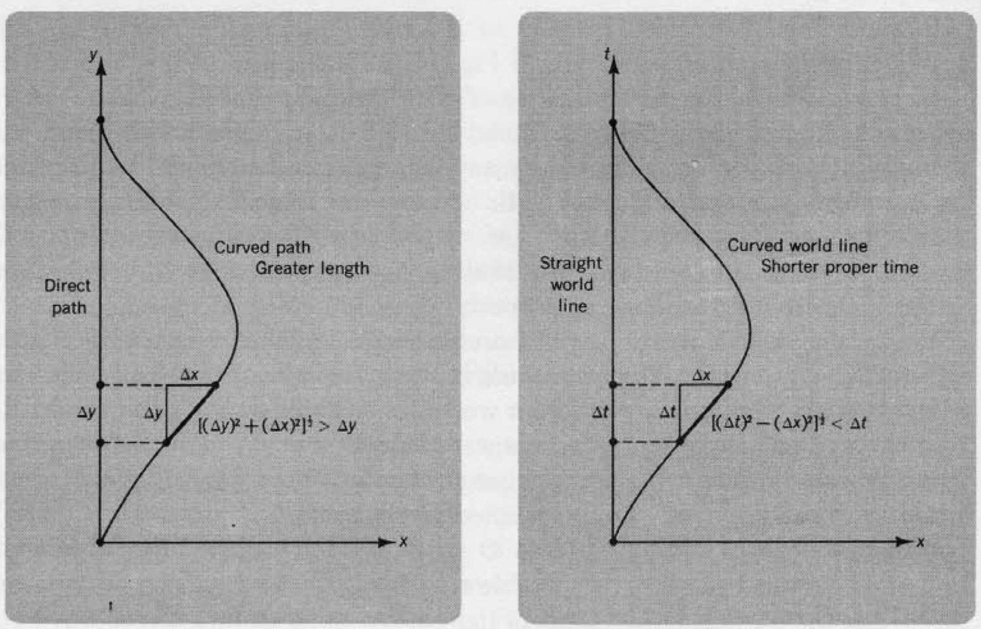
Consider a particle moving from O to B along the curved world line of Fig. 19,B. In this example, the particle travels along the x axis at a changing speed. Let the particle emit a flash of light every meter of time as recorded on a clock carried with it. The proper time $\Delta\tau$ between any two consecutive flashes (for instance, those marked 3 and 4 in the figure) can be *calculated* using the difference in coordinates Δx and Δt of the two events measured in a

Stretch of world line measured by proper time particular inertial frame. Because the interval is invariant, the proper time between these two events will be the same when calculated in any inertial frame, even though the separate space and time coordinates Δx and Δt will have different values in different reference frames. Elsewhere along the world line the interval between another pair of consecutive flashes will also be independent of the reference frame which is used in evaluating that interval. So too for the *sum* of the proper-time intervals of all the flashes along the world line! Thus observers in different inertial reference frames will all agree on the proper time along a given world line from a specified initial event O to a specified final event B.

Proper time is longest for direct world line

It is possible to proceed from event O to event B along quite another world line—for example, along the straight world line OB in Fig. 19,B. The elapsed proper time along this alternative world line is different from the proper time along the original world line. In Lorentz geometry a curved world line between two specified events is *shorter* than the direct world line between the same two events—shorter as measured by the elapsed proper time along the world line. This contrast between Euclidean and Lorentz geometry is shown in Fig. 20. The *distance* between nearby points along a *curved path* is always equal to or *greater* than the y displacement between those two points. In contrast, the *proper time* between nearby events along a curved *world line* is always equal to or *less* than the corresponding time along the direct world line. The determination of proper time is a fundamental method of comparing different world lines between two events.

The change of slope of the world line from point to point in Fig. 19,B, and Fig. 20,B, means that the clock being carried along the world line changes velocity: it is *accelerated*. Different clocks will behave differently when accelerated unless these clocks are sufficiently small. As a rule a clock can withstand a great acceleration only if it is small



A. Euclidean geometry B. Lorentz geometry

Fig. 20. Contrast between Euclidean and Lorentz geometry. In Lorentz geometry the *curved* world line is traversed in the *shorter* proper time.

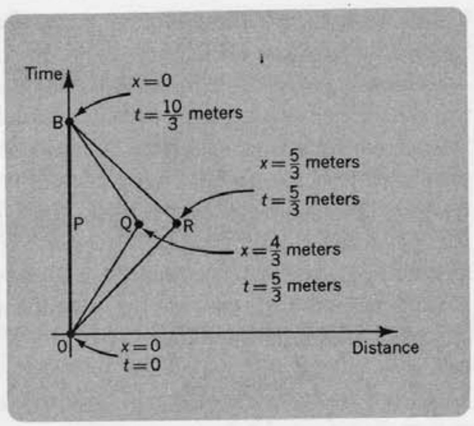


Fig. 21. Three alternative world lines connecting event O and event B. The sharp changes of speed at events Q and R have been drawn for the ideal limit of small (acceleration-proof) clocks.

and compact. The smaller the clock, the more acceleration it can withstand, and the sharper the curves on the world line can be. In all figures like Fig. 19,B, and Fig. 20,B, we assume the ideal limit of infinitesimally small clocks.

Proper time from O to B compared for three world lines

We are now free to analyze a motion in which the particle and the clock are subject to a great acceleration. In particular, consider a simple special case of the world line of Fig. 19,B. That world line gradually changed in slope as the particle speeded up and slowed down. Now make the period of speeding up shorter and shorter (great driving force!); also make the period of slowing down shorter and shorter. In this way the proportion of time spent in steady motion at high speed becomes greater and greater. Thus come eventually to the limiting case where the times of acceleration and deceleration are too short even to show up on the scale of the spacetime diagram (world line OQB in Fig. 21). In this simple limiting case the whole history of the motion is specified by (1) the initial event O, (2) the final event B, and (3) the coordinate x of the turnaround point Q, halfway in time between O and B. In this case it is particularly easy to see how the lapse of proper time between O and B depends upon the coordinate x of the halfway point—and thus to compare the three world lines OPB, OQB, and ORB.

Path OPB is the world line of a particle that does not move: $x = 0$ for all time. The proper time from O to B by way of P is evidently equal to the time as measured in the inertial reference system.

$$\tau_{\mathrm{OPB}} = 10/3 \text{ meters of light-travel time}$$

In contrast, on the way from O to B by way of R, each stretch is lightlike: for each segment the space and time components of the displacement are equal, and

$$\tau_{\mathrm{ORB}} = (\text{twice proper time on stretch OR}) = 2\,[(\text{time})^2 - (\text{distance})^2]^{1/2} = 0$$

Of course no clock can travel as fast as the speed of light. Therefore the world line ORB is not actually attainable. However, it is the ideal limit of world lines that actually *are* attainable. Or, in other words, one can find a speed sufficiently close to the speed of light, and yet *less* than the speed of light, so that a trip with this speed first one way then the other will bring an ideal clock back to $x = 0$ with a lapse of proper time as short as one pleases.

As distinguished from the limiting case ORB, the world line OQB demands an amount of proper time

$$\begin{aligned}\tau_{\mathrm{OQB}} &= (\text{twice proper time on stretch OQ}) \\ &= 2\,[(5/3)^2 - (4/3)^2]^{1/2} \\ &= 2\left[\frac{25-16}{9}\right]^{1/2} \\ &= 2 \text{ meters of light-travel time}\end{aligned}$$

This is *less* proper time than the proper time $\tau_{OPB} = 3\ 1/3$ meters that characterized the "direct" world line OPB!

Evidently *proper* time in the real physical world of spacetime differs remarkably from the *distance* of textbook Euclidean geometry. Distance is shortest for the direct route: "A straight line is the shortest distance between two points." In contrast, the lapse of proper time is less for the traveler who travels away, accelerating to high speed, then reverses his course and comes back, than for the man who stays home! (See Ex. 27 and Ex. 49 on the clock paradox). In brief, proper time is the appropriate measure of time as it will be observed by a particle that travels along a world line, just as the graduations along a flexible tape provide the appropriate measure of distance covered by a traveler along a winding path.

7. Regions of Spacetime

Interval including y and z coordinates

Thus far in dealing with the interval between two events, A and B, we have had occasion to consider only the situation in which they have the same y and z coordinates. In this situation the separation in space between the two events is measured by the single quantity

$$\text{distance} = \Delta x$$

The interval is given by the expression

$$[(\Delta t)^2 - (\Delta x)^2]^{1/2}$$

However, the orientation of the x, y, and z coordinate axes is evidently a matter of arbitrary choice. With another orientation of the axes the component Δx of the separation between the two events will ordinarily have quite a different value. Yet the separation in space between the two events is quite independent of any choice of orientation, and is given by the expression

$$(\text{distance})^2 = (\Delta x)^2 + (\Delta y)^2 + (\Delta z)^2$$

In other words, this is the quantity that must replace $(\Delta x)^2$ in the full formula for the interval. Thus we have the complete expression for the interval between two events

$$\text{A at } (t, x, y, z)$$

and

$$\text{B at } (t + \Delta t, x + \Delta x, y + \Delta y, z + \Delta z)$$

in the form

$$(9) \qquad \begin{aligned}(\text{interval of proper time})^2 &= (\text{time})^2 - (\text{distance})^2 \\ &= (\Delta t)^2 - (\Delta x)^2 - (\Delta y)^2 - (\Delta z)^2\end{aligned}$$

when the interval is timelike; and when it is spacelike,

$$(10) \qquad \begin{aligned}(\text{interval of proper distance})^2 &= (\text{distance})^2 - (\text{time})^2 \\ &= (\Delta x)^2 + (\Delta y)^2 + (\Delta z)^2 - (\Delta t)^2\end{aligned}$$

How is one to understand the new kind of geometry described by an expression for "interval of proper distance" that contains three plus signs—as in ordinary Euclidean geometry—but also one minus sign? One can follow

Minkowski on unity of spacetime

Minkowski (1908) and introduce a new quantity w to measure time, a quantity defined by

$$w = (-1)^{1/2}\, t$$

or

$$\Delta w = (-1)^{1/2}\, \Delta t \tag{11}$$

Then the expression for the interval of proper distance takes the form

$$(\text{interval of proper distance})^2 = (\Delta x)^2 + (\Delta y)^2 + (\Delta z)^2 + (\Delta w)^2$$

The signs are now all positive. The geometry superficially appears to be that of Euclid—in four dimensions, of course, instead of three. Impressed by this formula, Minkowski wrote his famous words, "Henceforth space by itself, and time by itself, are doomed to fade away into mere shadows, and only a kind of union of the two will preserve an independent reality."† Today this union of space and time is called spacetime. Spacetime is the arena in which stars, atoms, and people live and move and have their being. Space is different for different observers. Time is different for different observers. Spacetime is the same for everyone.

Minkowski's insight is central to the understanding of the physical world. It focuses attention on those quantities, such as interval, which are the same in all frames of reference. It brings out the relative character of quantities, such as velocity, energy, time, distance, which depend upon the choice of frame of reference.

Difference between time and space

Today we have learned not to overstate Minkowski's argument. It is right to say that time and space are inseparable parts of a larger unity. It is wrong to say that time is identical in quality with space. Why is it wrong? Is not time measured in meters, just as distance is? Are not the x and y coordinates of the surveyor quantities of identical physical character? By analogy, are not the x and t coordinates of the spacetime diagram of the same nature as one another? How else could it be legitimate to treat these quantities on an equal footing, as in the formula $[(\Delta x)^2 + (\Delta y)^2 + (\Delta z)^2 - (\Delta t)^2]^{1/2}$ for a spacelike interval? Equal footing, yes; same nature, no. There is a *minus* sign in this formula that no sleight of hand can ever conjure away. *This minus sign marks the difference in character between space and time.* It does not really remove this minus sign to introduce the imaginary number $\Delta w = (-1)^{1/2}\Delta t$. It would if w were a real quantity. But w is not real. No clock ever reads $(-1)^{1/2}$ seconds, or $(-1)^{1/2}$ meters. Real clocks show real time: $\Delta t = 7$ seconds, for example. Consequently the term $-(\Delta t)^2$ is always opposite in sign to the distance term $(\Delta x)^2 + (\Delta y)^2 + (\Delta z)^2$. No twisting or turning can ever make the two signs the same.

The difference in sign between the time term and the space terms in the expression for the interval gives Lorentz geometry a unique feature, which is new and quite different from anything in Euclidean geometry. In Euclidean geometry it is never possible for the distance AB between two points to be zero unless all three of the quantities Δx, Δy, and Δz are simultaneously zero. In contrast, the interval AB between two events can vanish even when the

†A. Einstein *et al.*, *The Principle of Relativity*, (Dover Publications, New York).

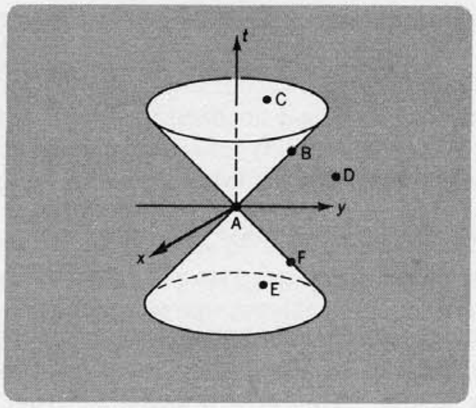


Fig. 22. Spacetime diagram showing x, y, and t coordinates of events for which $z = 0$.

separations Δx, Δy, Δz in space and Δt in time between B and A are individually quite large.

Case where interval vanishes

Under what condition does the interval AB vanish? *The interval vanishes when the time part of the separation between A and B is identical in magnitude to the space part of the separation*

$$\Delta t = \pm[(\Delta x)^2 + (\Delta y)^2 + (\Delta z)^2]^{1/2} \tag{12}$$

What is the physical interpretation of this condition? The expression on the right is the distance between the two points. But light travels one meter of distance in one meter of light-travel time. Therefore the expression on the right also represents the time that is *needed* by light to travel the distance between A and B. On the other hand, Δt represents the time that is *available* to travel this distance. In other words, condition (12) is satisfied—and the interval AB vanishes—when a light flash starting at event A can arrive precisely in time for event B (or when a flash starting at B can arrive at A). *The interval between two events is zero when they can be connected by one light ray.*

It is interesting to map out in an appropriate diagram the location of *all* events B that can be connected with one given event A by a light ray. For simplicity let event A occur at the origin of the spacetime diagram. Let the coordinates x, y, z of event B be taken to have any values. Then the time coordinate of event B has either the value

$$t_{\text{future}} = +(x^2 + y^2 + z^2)^{1/2} \tag{13}$$

or the value

$$t_{\text{past}} = -(x^2 + y^2 + z^2)^{1/2} \tag{14}$$

Light cones: partitions in spacetime

It simplifies the graphical presentation of this formula to limit attention to events B whose z coordinate is zero. Then it is appropriate to construct a spacetime diagram with two spatial coordinates x and y and the time coordinate t, as shown in Fig. 22. Every event B in this diagram that is separated from A by a zero interval ("lightlike interval") lies either on the "future light cone" of A (plus sign in Eq. 13) or on the "past light cone" of A (minus sign in Eq. 14).

In Fig. 22 consider all those events that have time coordinates 7 meters later than the zero time of flash A. These events lie on a plane 7 meters above the

xy plane, and parallel to the xy plane. Among these events those which lie on the future light cone of A are on a circle. This circle has a radius of 7 meters. This circle (circle in the present x, y, t diagram; a sphere in a full x, y, z, t diagram!) is the locus of the pulse of radiant energy which emerged from A. Observed at a later time, the pulse has expanded to a still larger radius. Thus the forward light cone tells the history of the expanding spherical pulse that started at A. Similarly the backward light cone tells the history of a converging pulse of radiation, so perfectly focused that it collapses at the origin at time zero.

Spacetime classified into five regions relative to event A

The light cone is a unique feature of Lorentz geometry; there is no such feature in Euclidean geometry. Moreover, associated with the light cone, Lorentz geometry has a characteristic of the greatest importance for the structure of the physical world. It provides the following ordering of all events with respect to their causal relationship to any chosen event A (Fig. 22).

1. Can a *particle* emitted at A affect what *is going* to happen at C? If so, C lies *in* the future light cone of A.
2. Can a *light ray* emitted at A affect what *is going* to happen at B? If so, B lies *on* the future light cone of A.
3. Can *no effect whatever* produced at A affect what happens at D? If so, D lies *outside* the light cone of A.
4. Can a *particle* emitted at E affect what *is happening* at A? If so, E lies *in* the past light cone of A.
5. Can a *light ray* emitted at F affect what *is happening* at A? If so, F lies *on* the past light cone of A.

Now, the light cone of event A—and the light cone of every other event—has an existence in spacetime quite apart from any coordinates that may be used to describe it. Therefore the possibilities mentioned in the five preceding questions, that one event will affect another event are independent of the reference frame in which this connection between events is observed. In this sense *the causal connection between two events is preserved in every reference frame.*

Figure 23 summarizes the relations between a selected event A and all other events of spacetime.

8. The Lorentz Transformation

Coordinates more convenient than interval for describing trip of π-meson

At heights of 10 to 30 kilometers above the earth, cosmic rays are continually striking the nuclei of oxygen and nitrogen atoms and producing π-mesons, both charged and neutral. Follow one of the π^+-mesons on its way down (Fig. 24). In the reference frame attached to this particle ("rocket frame") the average life of the π^+-meson is 2.55×10^{-8} seconds. In this rocket frame let the coordinates of the event of birth be $x' = 0$, $t' = 0$ (Fig. 25,B). Let the coordinates of the event of explosion of the π-meson (into muon plus neutrino) be written

$$x' = 0, \qquad t' = \tau_\pi$$

How do these events appear to the laboratory observer? As recorded by his clocks, how long does the π-meson live from birth to death? Or how much is

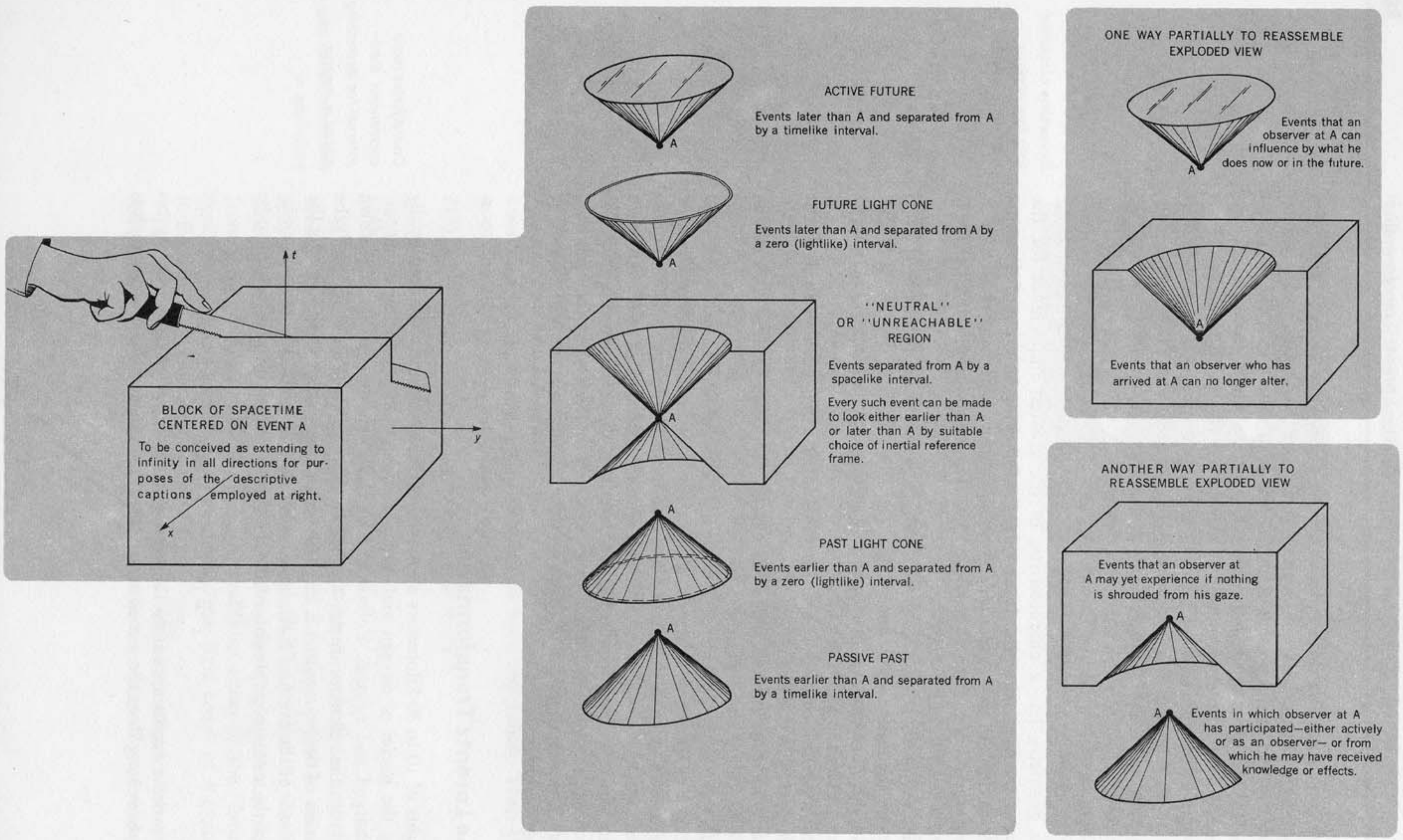


Fig. 23. Exploded view of the five regions into which the events of spacetime fall apart when classified with respect to a selected event A.