Special Relativity

Important Definitions

Event = something happening in fairly limited region of space and for short duration in time.

Mathematically, we idealize concepts -> **Event** = point in space and instant in time.

Universe -> event = 4 numbers require to “locate” an event…..

-> 3 numbers = spatial position and 1 number = time:

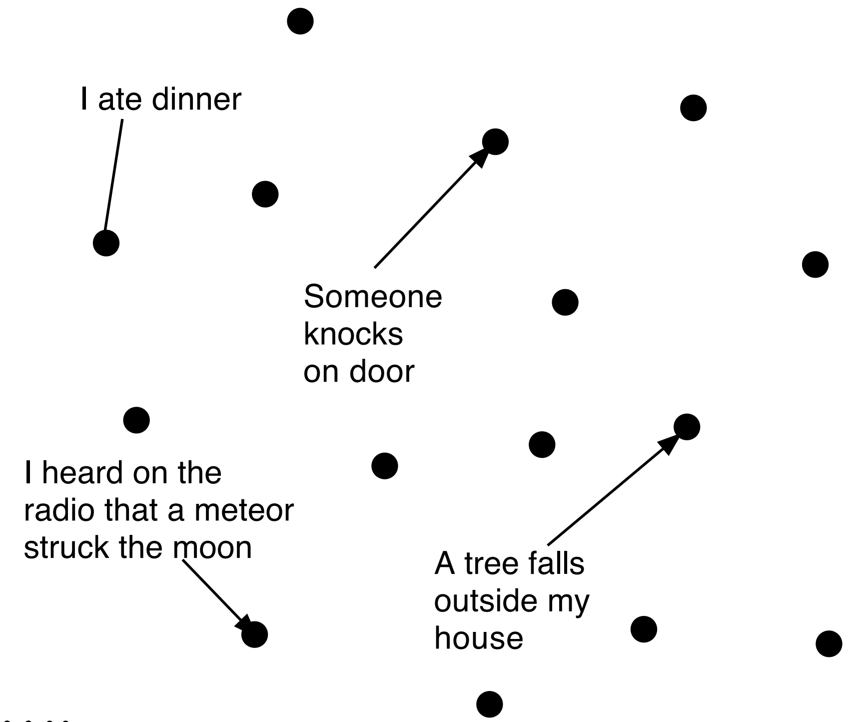
= 4-dimensional spacetime.

Everything that happens in universe = an event or a collection of events.

Events are **independent** of observers.

However, the 4 numbers describing event are **not independent** of observers, as we will see.

Spacetime is the collection of all possible events.



Events in Spacetime

Alternatively, we could say with extra details……..

Event = any physical occurrence that can be considered to happen at a definite place in space and at a definite instant in time.

The explosion of a small firecracker at a particular location in space and at a definite instant in time is a vivid example of an event.

The collision of two particles or the decay of a single particle at a certain place and time also defines an event.

The simple passage of a particle through a given mathematical point in space can also be treated as if it were an event (simply imagine that the particle sets of a firecracker at that point as it passes by).

Because an event occurs at a specific point in space and at a specific time, we can quantify when and where the event occurs by four numbers: three that specify the location of the event in some three-dimensional spatial coordinate system and one that specifies what time the event occurred.

These four numbers are called the spacetime coordinates of the event.

Note that the exact values of the spacetime coordinates of an event depend on certain arbitrary choices, such as the origin and orientation of the spatial coordinate axes and what time is considered to be $t = 0$.

Once these choices are made and consistently used, however, specifying the coordinates of physical events provides a useful method of mathematically describing motion.

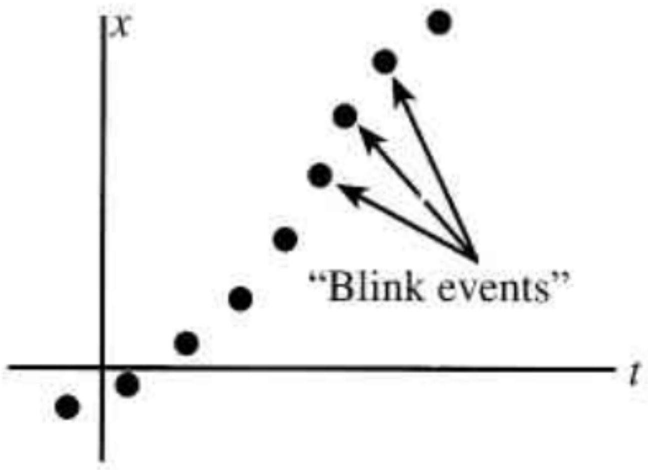
Specifically, we can quantify the motion of a particle by treating it as a series of events.

We can visualize this process in the following manner. Imagine an airplane moving along the x axis of some coordinate system.

The airplane carries a blinking strobe light.

Each blink of the strobe is an event in the sense that we are using the word here: it occurs at a definite place in space and at a definite instant of time.

We can describe the motion of the plane by plotting a graph of the position coordinate of each "blink event" vs. the time coordinate of the same, as illustrated in Figure.



Sketching out a graph of the motion (position vs. time) of an object by plotting the "blink events" that occur along its path.

If the time between blink events is reduced, one gets an even more detailed picture of the plane's motion.

Therefore, the plane's motion can be described in arbitrarily fine detail by listing the spacetime coordinates of a sufficient number of blink events distributed along its path.

The above example is a specific illustration of a general idea: the motion of any particle can be mathematically described to arbitrary accuracy by specifying the spacetime coordinates of a sufficient number of events suitably distributed along its path.

Studying the motion of particles is the most basic way to discover and test the laws of physics.

Therefore, the most fundamental task of a "laboratory" (as a place in which the laws of physics are to be tested) is to provide a means of measuring the spacetime coordinates of events.

How do we measure four coordinates or the “where/when” of an event?

One method is the so-called **many-observer model**.

It works as follows.

First the setup:

Measure and label grid locations ahead of experiment

Each observer has a clock

Synchronize all clocks ahead of experiment

Observers then move clocks to grid locations

They **assume** this process has no effect on synchronization

Now, an experiment — throw an eraser into the air

If eraser passes an observer's location(grid point) then observer records local time

—-> "where and when" information

The collection of such "where and when" information (when brought back together) gives the set of events **representing** the motion being observed

This **operational** definition of event —> 1 possible prescription for assigning numbers to the associated where and when information in a precise and reproducible way.

Now assume 2-dimensional spacetime = 1 spatial dimension and 1 time dimension(simplicity).

All physics that we derive in restricted universe easily extended to real 4-dimensional universe.

A particular set of coordinate axes and associated scales are chosen inside space-time.

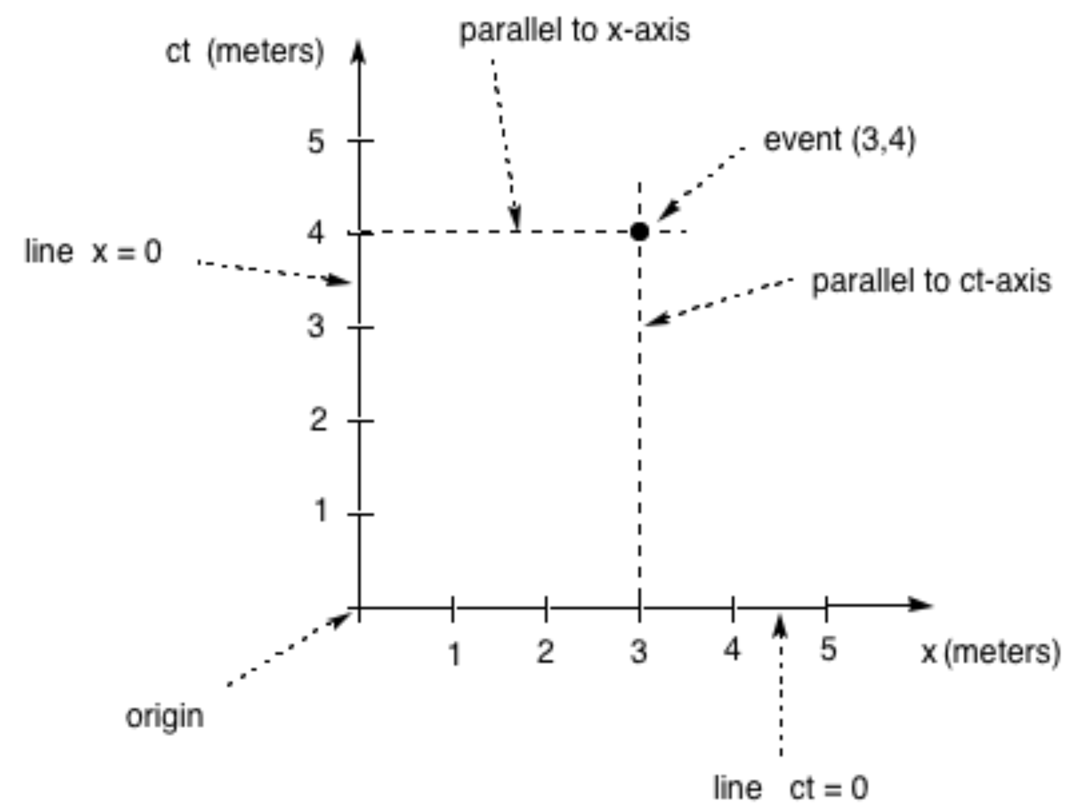
The choice is arbitrary!!

It is called a frame of reference.

—> allows us to relate events to measured quantities in experiments,

i.e., so that theorists can talk to experimentalists.

Represent events using spacetime diagram.



(2- dimensional case)

Definition of coordinate axes given by rules shown:

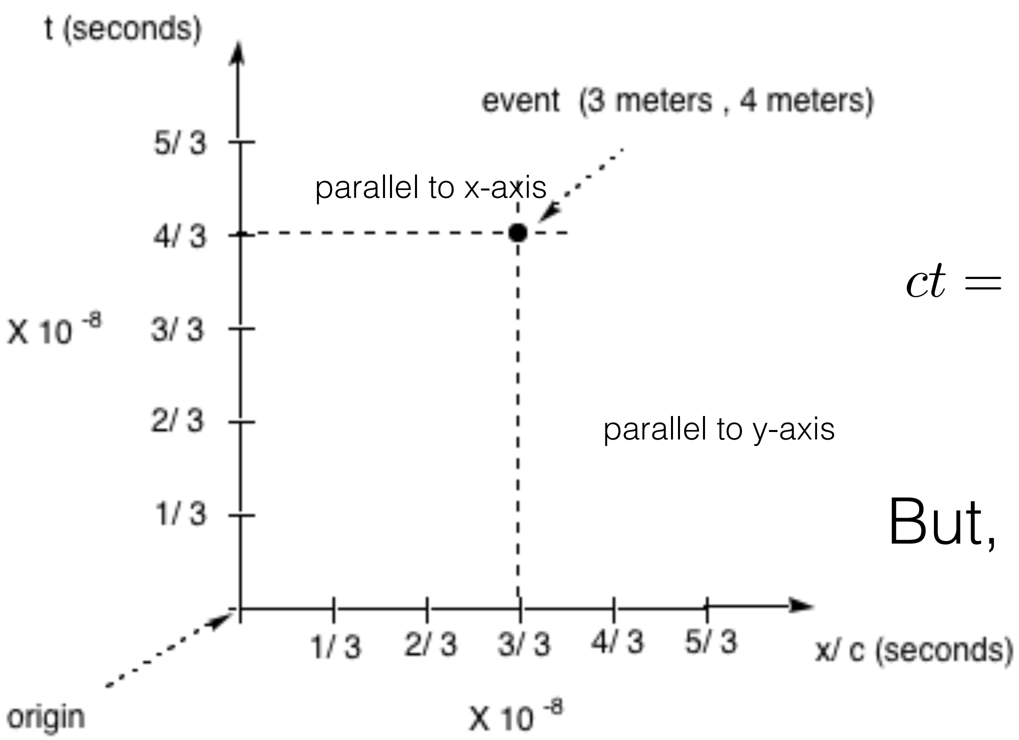
Note: use ct rather than t for vertical axis.
where c= speed of light ($c = 3.0\times10^8$ m/sec)

-> change in scale for vertical axis; reason will be clear later.

Conversion scales:

$$ct = 1\,meter \rightarrow t = \frac{1\,meter}{3.0 \times 10^8\,m/sec} = \frac{1}{3} \times 10^{-8}\,sec = 3.33\,nanosecond = 3.33\,ns$$

Many texts use a different scheme for labeling axes ->



Here -> t and x measured in sec instead of meters

$$ct = 1\ meter \rightarrow t = \frac{1}{3} \times 10^{-8}\,sec \text{ and } x = 1\ meter \rightarrow \frac{x}{c} = \frac{1}{3} \times 10^{-8}\,sec$$

But, carrying around all the powers of 10 is cumbersome.

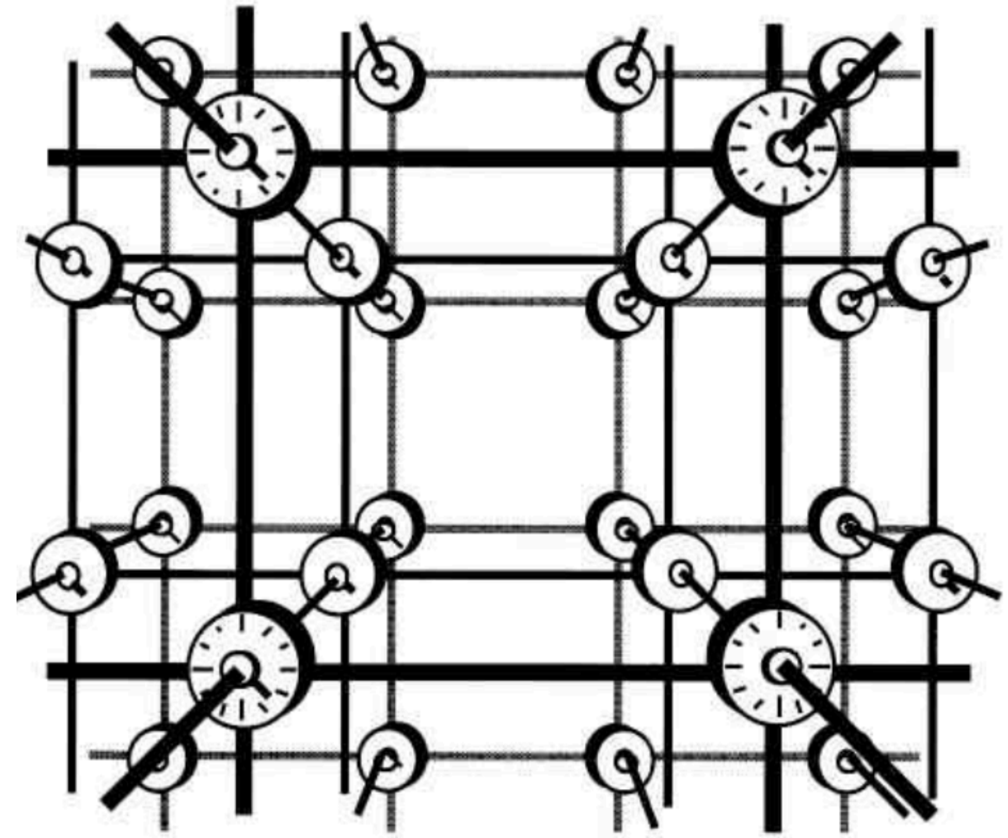
Note: parallel line definition of coordinatevalues rather than perpendicular definition(they are different as we shall see).

REFERENCE FRAMES IN GENERAL

How might this be done?

Imagine constructing a rigid cubical lattice (like a playground jungle gym) consisting of a large number of identical measuring sticks of some given length.

At each lattice intersection, you place one of a set of identical clocks, as shown in Figure



A reference frame visualized as a cubical lattice of clocks and measuring sticks.

We will now arbitrarily choose one clock in the framework to serve as the spatial origin of the reference frame: the origin defines a unique location in the frame against which all other locations are compared.

The measuring sticks radiating from the origin clock specify three mutually perpendicular directions that for the sake of argument we can call the *x direction*, the *y direction*, and the *z direction*.

We will take the direction opposite to the *x direction* to be the *negative x direction*, and so on.

Once the origin and these directions have been chosen, the location of any other clock in the reference frame can be uniquely described by stating the distance that one has to travel (as registered by the lattice measuring sticks) along the ±x direction, the ±y direction, and then the ±z direction to get from the origin clock to the second clock.

If we agree to always state these numbers in the order [x, y, z), then the location of any clock in the lattice can be specified as an *ordered set of three signed numbers*.

For example, the set [-3 m, 6 m, - 1 m] specifies the clock 3m in the x-direction, 6 m in the y direction, and 1 m in the -z direction from the origin.

These three numbers are called the **spatial coordinates** of the clock in the reference frame.

The next step is to select an event to represent the **origin of time** t = 0.

At the instant that this event occurs, all clocks in the lattice should be synchronized (in some manner) to read t = 0.

Now imagine that a different event (for example, the explosion of a firecracker) occurs somewhere in the lattice.

The time of the event is defined to be the time registered by the clock nearest to the event when the event occurs.

The location of this event in space is defined to be the lattice location of that clock, as specified by its spatial coordinates.

This might not seem to locate the event very precisely if the distance between clocks is large, but we can imagine building a very fine lattice if we need to make very fine measurements.

The measuring sticks in the lattice allow one to directly read the distance of this clock from an arbitrarily chosen lattice origin along the three mutually perpendicular directions defined by the cubical lattice.

The spacetime coordinates of an event are thus defined to be an ordered set of four signed numbers: one that specifies the reading of the clock nearest the event and three that specify the spatial coordinates of that clock.

One usually specifies the spacetime coordinates of an event with the time coordinate first.

We might say that the firecracker explosion occurs at [3 s, - 3 m, 6 m, - 1m]; that is, the event is registered by the clock that is 3 m, 6 m, and 1 m away from the origin in the -x, +y and - z directions, respectively, and that clock registered the event as occurring 3s after the event defining the origin of time.

Why is it important to have a clock at every lattice intersection?

The point is to make sure that there is a clock essentially at the location of any event to be measured.

If we attempt to read the time of an event using a clock located a substantial distance away, we need to make additional assumptions about how long it took the information that the event had occurred to reach that distant clock.

For example, if we read the time when the sound from an event reaches the distant clock, we should correct that reading by subtracting the time it takes sound to travel from the event to the clock.

But to do this, we must assume that we know the speed of sound in our lattice.

We can avoid making extra assumptions of this nature if we require that the time of each event be measured by a clock that is essentially *present* at the event.

Note that if we must have all these clocks, it is also essential that they all be synchronized in some meaningful and self-consistent manner.

It would be impossible to track the motion of a particle through the lattice if these clocks were not carefully synchronized, as adjacent clocks might differ wildly and thus give a totally incoherent picture of when the particle passes various lattice points.

The appropriate method for synchronizing lattice clocks is actually a subtle issue that we will explore in more detail later.

For now, it is sufficient to recognize that it must be done.

Once we have specified a synchronization method, the clock-lattice image just described represents a complete definition of a procedure that one can use, in principle, to determine the spacetime coordinates of an event.

This amounts to what is called an **operational definition** of these spacetime coordinates.

In general, an operational definition of a physical quantity defines the quantity by describing how the quantity may be measured.

Operational definitions provide a useful way of anchoring slippery human words to physical reality by linking the words to specific, repeatable procedures rather than to vague comparisons or analogies.

The procedure just described represents an admittedly idealized method for determining the spacetime coordinates of an event.

The actual methods employed by physicists may well differ from this description, but these methods should be equivalent to what is described above: the clock-lattice method defines a standard against which actual methods are to be compared.

It is such a simple and direct method that ti is inconceivable that any actual technique could yield different results and still be considered correct and meaningful.

But the real importance of the clock-lattice definition of spacetime coordinates is as a mental image that tersely and cleanly describes exactly what has to be done to determine the spacetime coordinates of an event (and thus quantify particle motions) without the obscuring complications that always arise in building real devices.

As a simple and vivid description of the bare necessities required to determine event coordinates, this mental image will make arguments that follow more straightforward.

With this in mind, we define the following technical words to aid us in future discussions:

Definition

A **reference frame** is defined to be a rigid cubical lattice of clocks as described above *or its equivalent,* with some self-consistent method of synchronizing clocks specified.

Definition

The **spacetime coordinates** of an event in a given reference frame are defined to be an ordered set of four numbers: one specifying the time of the event as registered by the nearest clock in the reference frame clock lattice and three specifying the spatial coordinates of that clock in the lattice.

Definition

An **observer** is defined to be a (possibly hypothetical) person who interprets the measurements made in a reference frame (i.e., the person who interprets the spacetime coordinates collected by a central computer from all the clocks).

A reference frame is often spoken of in connection with some object.

For example, one might refer to "the reference frame of the surface of the earth" or "the reference frame of the cabin of the plane" or "the reference frame of the particle."

In these cases, we are being asked to imagine a clock lattice (or equivalent) fixed to the object in question.

Sometimes the actual reference frame is referred to only obliquely, as in the phrase "an observer in the plane cabin finds . . . . "

Since the word observer in this text refers to someone who is using a reference frame to make measurements of the coordinates of events, the existence of a reference frame attached to the cabin of the plane is presumed.

A reference frame may be moving or at rest, accelerating, or even rotating about some axis.

The beauty of the definition of spacetime coordinates given above is that measurements of the coordinates of events (and thus measurements of the motion of objects) can be carried out in a reference frame no matter how it is moving, provided that the clocks in the frame can be reasonably synchronized in some manner.

Note that the act of "observing" in the context of the last definition is an act of interpretation of measurements generated by the reference frame apparatus and may have little or nothing to do with what that observer sees with his or her own eyes.

When we say that "an observer in such-and-such reference frame observes such-and- such," we are actually referring to the *conclusions* that the observer draws from the coordinate measurements generated by that reference frame lattice.

INERTIAL REFERENCE FRAMES

As we will see later , all available evidence will suggests that we can make the following general statement about the way the universe is constructed:

> *The laws of physics are the same inside a laboratory moving at a constant velocity as they are in a laboratory at rest.*

This is an unpolished statement of what we will call the **principle of relativity**.

This simple idea, based on common, everyday experience, is , as we will see, the foundation of Einstein's special theory of relativity.

All of that theory's exciting and mind-bending predictions about the nature of space and time follow as logical consequences of the principle of relativity!

Indeed, the remainder of this class is little more than a step-by-step unfolding of the rich implications of this statement.

Replacing the vaguely defined concept of a "laboratory" in the statement of the principle of relativity with the precisely defined concept of a "reference frame" would represent a substantial improvement in the clarity of that statement.

We now turn to the problem of clarifying the ambiguities in the concepts of laboratories "at rest" and laboratories moving at a "constant velocity."

How can we operationally define when a laboratory is physically at rest?

The principle of relativity is both a very new and a very old idea.

It was not first stated by Einstein (as one might expect) but by Galileo Galilei in a book published in 1632?

In the nearly three centuries that passed between Galileo's statement and Einstein's first paper on special relativity in 1905, the principle of relativity as it applied to the laws of mechanics was widely understood and used (in fact, it was generally considered to be a consequence of the particular characteristics of Newton's laws).

What Einstein did was to assert the applicability of the principle of relativity to all the laws of physics and most particularly to the laws of electromagnetism (which had just been developed and thus were completely unknown to Galileo).

Thus Einstein did not invent the principle of relativity; rather, his main contribution was to reinterpret it as being fundamental (more fundamental than Newton's laws or even than the ideas about time that up to that point had been considered obvious and inescapable) and to explore insightfully its implications regarding the nature of light, time, and space.

Our task in this text is to work out the rich and unexpected consequences of this principle. But it is important to make two cautionary statements before we proceed: (1) The principle of relativity is a postulate, and (2) it needs to be more precisely stated before we can extract any of its logical implications.

The principle of relativity is one of those core physical assumptions (like Newton's second law or the law of conservation of energy) that have to be accepted on faith: it cannot be proved experimentally or logically derived from more basic ideas (for example, it us not possible even in principle to test every physical law in every smoothly moving laboratory).

The value of such a postulate rests entirely on its ability to explain and illuminate experimental results.

The principle of relativity has weathered intense critical examination for more than 109 years.

No contradiction of the principle or its consequences has ever been conclusively demonstrated.

Moreover, the principle of relativity has a variety of unusual and unexpected implications that have been verified (to an extraordinary degree of accuracy) to occur exactly as predicted.

Therefore, while this principle cannot be proven, it has not yet been disproven, and physicists find it to be something that can be confidently believed.

The principle of relativity, simple as it is, is a very rich and powerful idea, one that the physics community has found to be not only helpful but crucial in the understanding of much of modern physics.

Turning to the other problem, the principle of relativity as stated earlier suffers from certain problems of both abstraction and ambiguity.

For example, what do we mean by "the laws of physics are the same"?

What exactly do we mean by "a laboratory at rest"?

How can we tell if a laboratory is "at rest" or not?

If we intend to explore the logical consequences of any idea, it is essential to state the idea in such a way that its meaning is clear and unambiguous.

Our task is to resolve these problems.

We will first replace the ambiguous phrases "laboratory," "at rest," and "constant velocity" with a single phrase involving more clearly defined terms.

Later, we will explore what we really mean by "the laws of physics are the same" in such laboratories.

In so doing, we will provide a firm foundation for exploring the implications of the principle of relativity.

Replacing the vaguely defined concept of a "laboratory" in the statement of the principle of relativity with the precisely defined concept of a "reference frame" would represent a substantial improvement in the clarity of that statement.

We now turn to the problem of clarifying the ambiguities in the concepts of laboratories "at rest" and laboratories moving at a "constant velocity."

How can we operationally define when a laboratory is physically at rest?

The answer is that we cannot!

The principle of relativity specifically states that laboratories moving with a constant velocity are physically equivalent to a laboratory at rest.

Therefore, there can be no physical basis for distinguishing a laboratory at rest from another moving at a constant velocity.

Imagine that you and I are in spaceships coasting at a constant velocity in deep space.

You will consider yourself to be at rest, while I am moving by you at a constant velocity.

I, on the other hand, will consider myself to be at rest, while you are moving by me at a constant velocity.

According to the principle of relativity, there is no physical experiment that can resolve our argument about who is at rest.

Which of us we consider to be at rest is arbitrary.

On the other hand, there are definitely laboratories where the laws of physics are clearly not the same as in other laboratories.

Imagine yourself to be in a jet accelerating for takeoff.

You seem to experience a magical force pressing you into the back of your chair.

Unlike real physical forces (such as electric or magnetic forces), this force does not appear to express an interaction between you and any other object in your vicinity; it seems simply to magically appear and disappear.

If you were to experience such a force while sitting at home reading a book, you would be deeply disturbed!

Yet you are not disturbed by this in the plane because you know from experience that the usual laws of physics are violated in an accelerating airplane.

The real issue is thus not whether a laboratory is moving or at rest, but how that laboratory is moving.

We will now develop an operational means of determining whether the standard laws of physics hold in a given laboratory or not without using the terms rest or constant velocity, indeed without referring to anything outside the laboratory at all.

Newton's first law of motion may be stated as follows:

> *An object that experiences zero external force will move in a straight line at a constant velocity.*

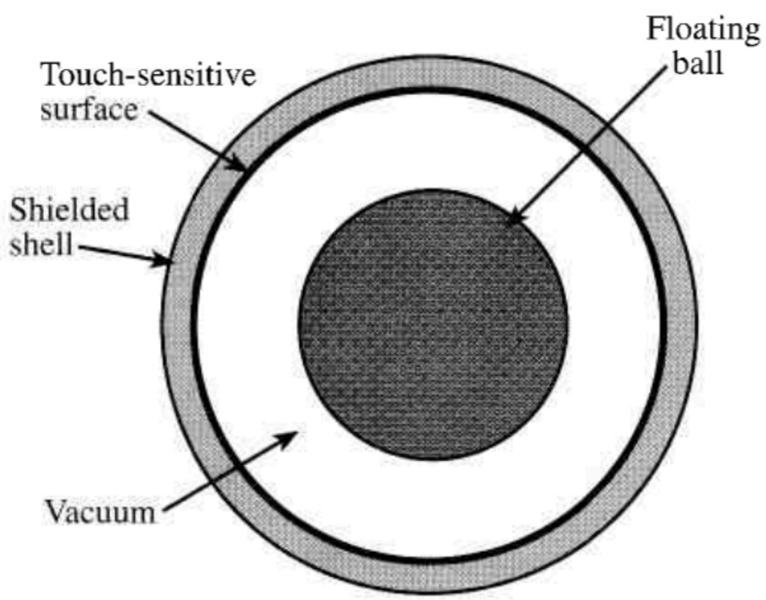
Specifically this law implies that if it is known that there are *no* physical forces on an object and it is *at rest* at a certain time, it will *remain at rest* for all time.

With this in mind, we can imagine constructing a device that can test the validity of Newton's first law.

Imagine a ball floating in a vacuum in the center of a spherical shell.

Imagine that the device has some mechanism that we can use to hold the ball at the center of the shell and then release it at rest so that it floats in the center of the shell.

If the ball drifts away from the center and contacts the touch-sensitive surface of the shell, the mechanism can be reactivated to reset the ball at rest in the center (Figure).



Cross section of a hypothetical first-law detector. A ball floats in the center of a touch-sensitive spherical shell which shields ti from outside forces. If the ball drifts away from the center and touches the outer shell, a mechanism (not shown) is activated that resets the ball at the center of the shell and then releases it at rest.

If the outer shell of this device is a good electrical conductor, the space inside will be shielded from any external electromagnetic fields.

If the ball floats in a vacuum and maintains a distance of more than a few micrometers from the outer shell, it will not be affected by forces due to air pressure, sound waves, contact forces, external nuclear forces (which have a very short range), and so on.

If we operate this device in deep space, far from any massive objects, it will also not be affected by gravitational forces.

If this is so, the ball should then be completely isolated from any external forces.

Newton's first law then predicts that if the ball is released at rest in the center of its spherical shell, it will remain exactly at rest at the center of the shell.

If Newton's first law is not true, then the ball will eventually drift away from the center and touch the sensitive inner surface of the detector shell and be reset to the center.

The reset action would thus signal a violation of Newton's first law.

The frequency of these reset actions will indicate the degree to which that law is violated.

This is just one example of how a first-law detector might (in principle) be constructed; you may be able to think of other approaches that could be used.

But if we assume that such a detector can be built, we can make the following definitions.

Definition

An **inertial clock** is a clock moving in such a manner that a first-law detector constructed as described above (or its equivalent) fixed to the clock detects no violation of Newton's first law.

Definition

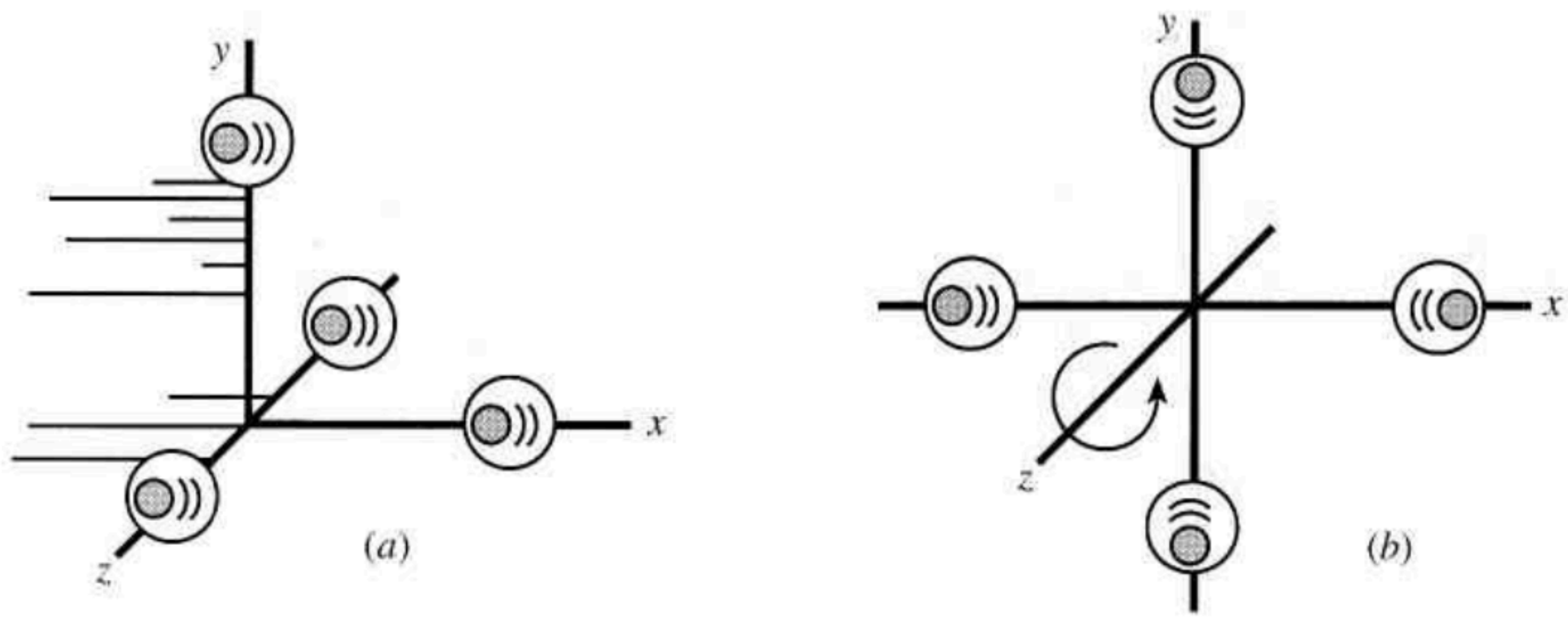
An **inertial reference frame** is a reference frame whose lattice clocks are all inertial clocks. Equivalently, an inertial reference frame is a reference frame in which Newton's first law is measured to be true at *every* point in the frame.

Again, note that the definition of an *inertial reference frame* amounts to being an *operational definition*, because we have defined *reference frame* by describing how to build one (see the definition of reference frame earlier) and *inertial* by describing explicitly *how* one might test a given frame to determine whether it qualifies as being inertial.

The definition of an inertial reference frame given above makes it easy to distinguish inertial and noninertial frames in realistic circumstances.

For example, a frame attached to an airplane accelerating for takeoff obviously cannot be an inertial frame; even without actually trying it, you know that the ball in the center of a first-law detector in such an accelerating plane will spontaneously drift toward the rear of the plane.

Similarly, a reference frame floating in deep space but rotating about some axis cannot be an inertial frame; you know that the ball in a first-law detector will drift away from the center of rotation in such a frame (see Figure).



First-law detectors in (a) a linearly accelerating reference frame and (b) a rotating reference frame. The frames are represented schematically by a set of coordinate axes. The drift directions of the floating balls inside the first-law detectors are also represented schematically.

A good example of an inertial reference frame would be a reference frame floating freely in deep space.

You can easily imagine that if you placed first-law detectors at various places in such a frame and reset each detector ball to rest in the center of its shell, each ball would continue to float at rest, fulfilling the terms of the definition of an inertial reference frame.

THE RELATIVE VELOCITY BETWEEN TWO INERTIAL FRAMES

The following statement is an immediate consequence of the definition of an inertial reference frame given earlier (as we will see in a moment):

> *Any inertial reference frame will be observed to move with a constant velocity by observers in any other inertial reference frame. Conversely, a rigid reference frame that moves at a constant velocity with respect to any inertial frame must itself be inertial.*

This statement is very important!

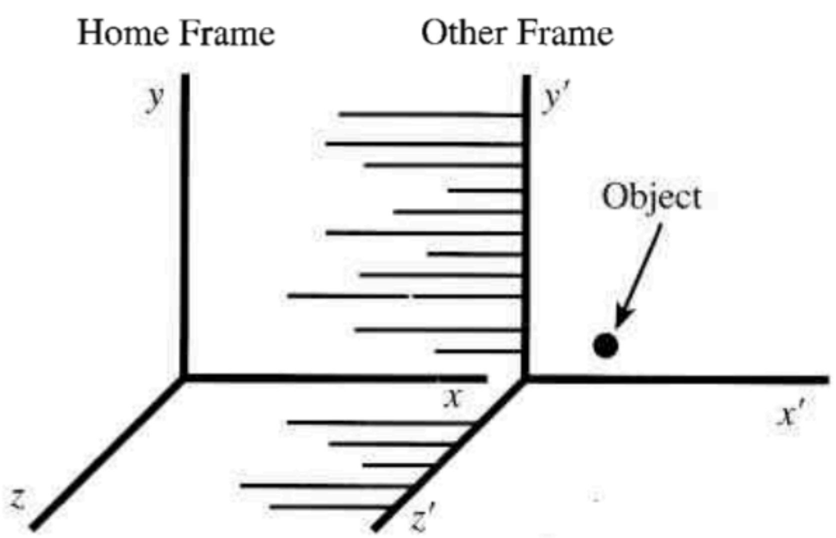
The fact that inertial frames move at constant velocities with respect to each other will be used over and over as we work out the consequences of the principle of relativity.

This statement also provides an easy way to distinguish between inertial and noninertial frames: the second part of the statement tells us that if we happen to know that any specific reference frame is inertial, then any frame that moves at a constant velocity with respect to that inertial frame must also be inertial.

For example, if a frame attached to the ground is inertial, then a frame attached to a plane accelerating for takeoff is not inertial because it is accelerating horizontally relative to the ground, while a frame attached to a plane flying horizontally at a constant velocity is inertial.

Here is an argument for the first part of the statement.

Imagine two inertial reference frames, which we will call the Home Frame and the Other Frame (see Figure).



A force-free object at rest in the Other Frame must move at a constant velocity with respect to the Home Frame by Newton's first law. But since the object is at rest with respect to the Other Frame, that frame itself will be observed to move with the same constant velocity with respect to the Home Frame.

"Home Frame" and "Other Frame" are capitalized in this text to emphasize that these phrases are actually names of inertial frames.

Since these frames are inertial, Newton's first law is true in both by definition.

Consider an object with no external forces acting on ti that also happens to be at rest in the Other Frame.

By Newton's first law, it will remain at rest in that frame.

Now observe the same object from the Home Frame.

Since the object has no forces acting on it, Newton's first law implies that ti must move at a constant velocity in that frame.

But if that object is at rest in the Other Frame, the whole Other Frame will be observed in the Home Frame to move with the same velocity as the object!

Therefore, the Other Frame will be observed to move with a constant velocity with respect to the Home Frame, as stated above.

THE FINAL VERSION OF THE PRINCIPLE OF RELATIVITY

Notice that we have *defined* inertial reference frames so that at least one physical law (Newton's first law) is true in every such frame.

The concept of an "inertial reference frame" also represents a precisely and operationally defined replacement for problematic "laboratories at rest" and "laboratories moving with a constant velocity."

The concept of "reference frame" precisely and efficiently captures the sense of a "laboratory" as a place where one does physical experiments.

This prompts us to rephrase the principle of relativity as follows:

> **The Principle of Relativity**
>
> The laws of physics are the same in every inertial reference frame.

This is our final polished statement of the principle of relativity.

It replaces the fuzzy, ambiguous concept of a "smoothly moving laboratory" with the well-defined concept of an "inertial reference frame."

What we are claiming is that if Newton's first law is obeyed in the set of reference frames that are "inertial," then all the laws of physics will be obeyed in that set of frames.

THE NEWTONIAN SOLUTION TO CLOCK SYNCHRONIZATION

An inertial frame floats in space, ready to use.

We would like to use it to measure coordinates of events happening within it.

But an important problem remains to be solved.

How do we synchronize all those clocks?

If the clocks are not synchronized, the time coordinate of an event will be meaningless: it will depend in an arbitrary, random way on which actual clock in the lattice records the event.

"The solution is easy," says a newtonian physicist.

“Everyone knows that time is absolute and flows equably without regard to anything external.

Any good clock will therefore measure the flow of this absolute time”.

Therefore, simply designate one clock to be a master clock, carry it around to each of the lattice clocks, and synchronize each lattice clock to the master.

Since the master clock and the lattice clocks all measure the flow of immutable absolute time, the motion of the master clock as it is carried from place in the lattice is irrelevant.

Once a lattice clock is set to agree with the master clock, it will certainly remain ni agreement with it, since both clocks measure the flow of absolute time.

Indeed, if the master clocks in two different reference frames are in agreement at any given event, then all the clocks in the two frames will always agree.

It does not matter whether the frames are ni motion with respect to each other; it does not even matter if they are inertial or not.

This follows from the self- evident absolute nature of time.

This picture of the nature of time is straightforward and believable.

It reflects the intuitive picture of time that most of us already hold.

But what are its consequences?

We will see shortly that if we assume that time is "universal and absolute," it is possible to link the spacetime coordinates of an event in one frame to those of the same event in a different frame using Galilean relativity.

The only problem with all of this is that the assumption that time is "universal and absolute" turns out to be *wrong*!

We will discuss the evidence for this incredible assertion and some of its implications later.

Finally, let me state what exactly we will mean when we say that "the laws of physics are the same" in all inertial frames?

Observers in different inertial frames may disagree about the values of various quantities (particularly velocities), but each observer will agree that if one takes the mathematical equation describing a physical law (like Newton's second law) and plugs in the values measured in that observer's frame, one will always find that the equation is satisfied.

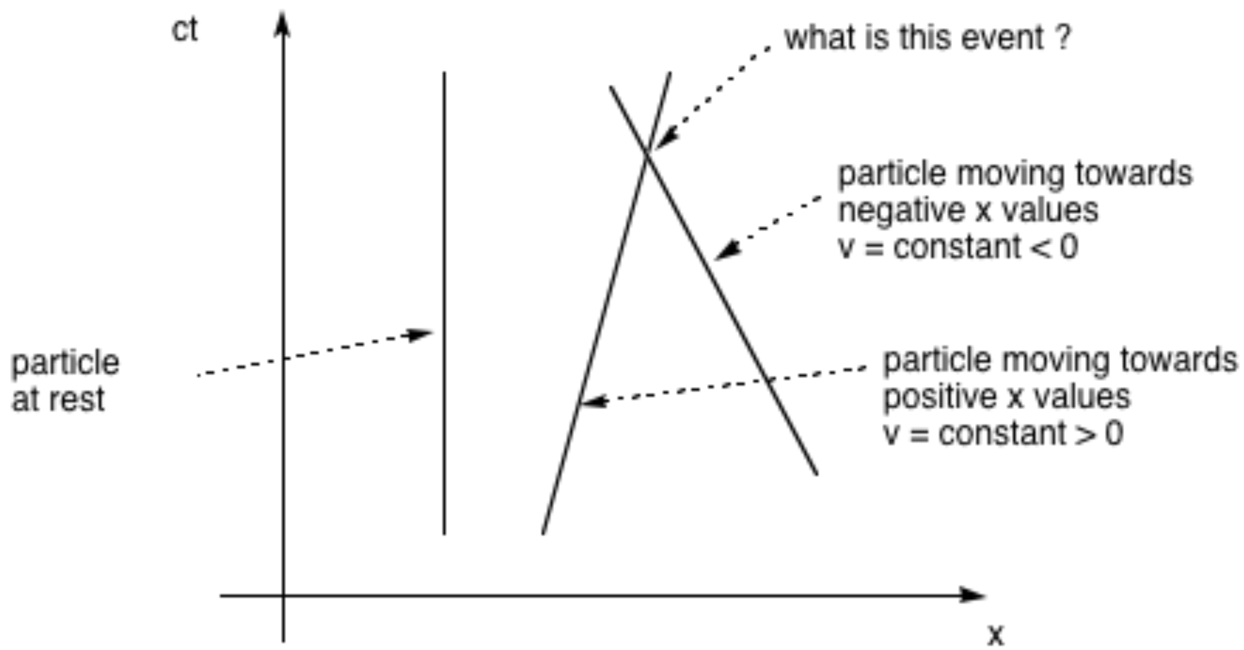
In other words, *the same basic equations will be found to describe the laws of physics in all inertial reference frames*.

Now back to the discussion of specifics...........

We will use x and ct axes in our frames of reference.

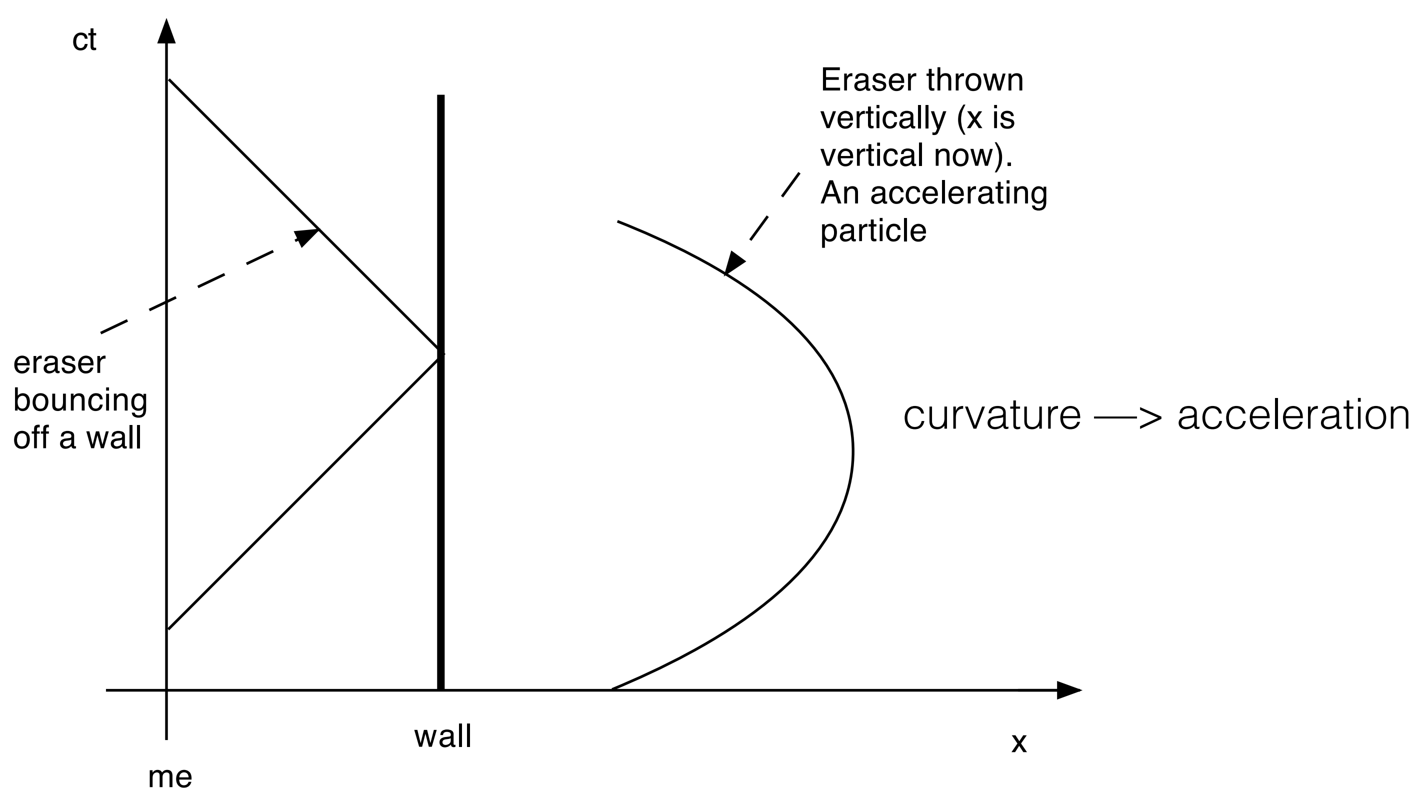
A collection of related events = **worldline**.

Examples of worldlines and other things shown

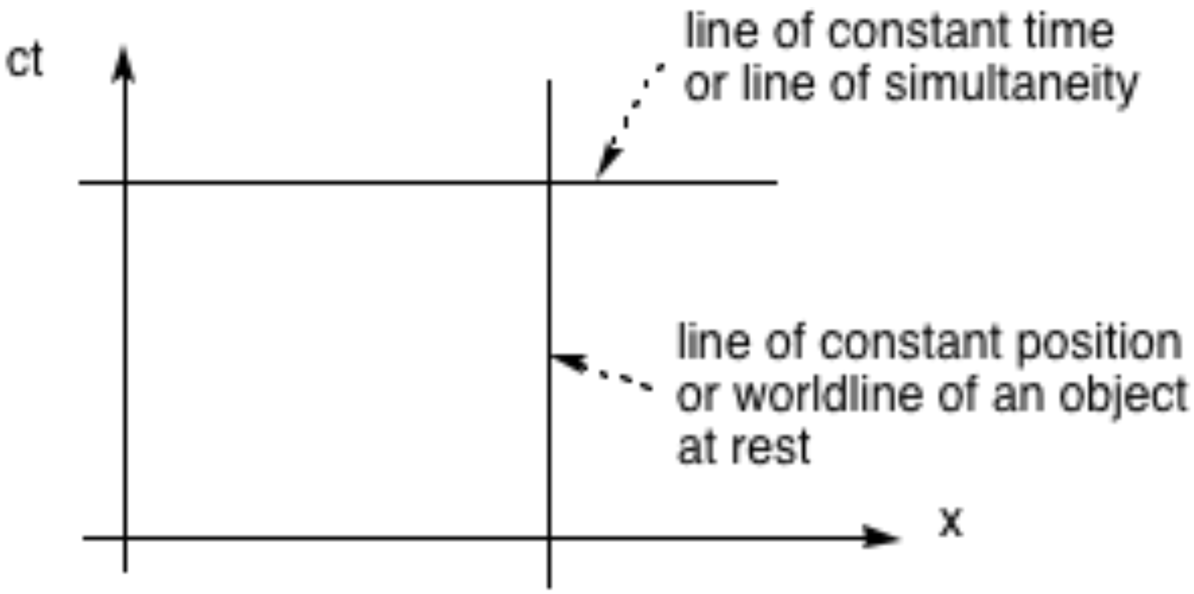


Your choice of axes

—> **your worldline** = ct-axis



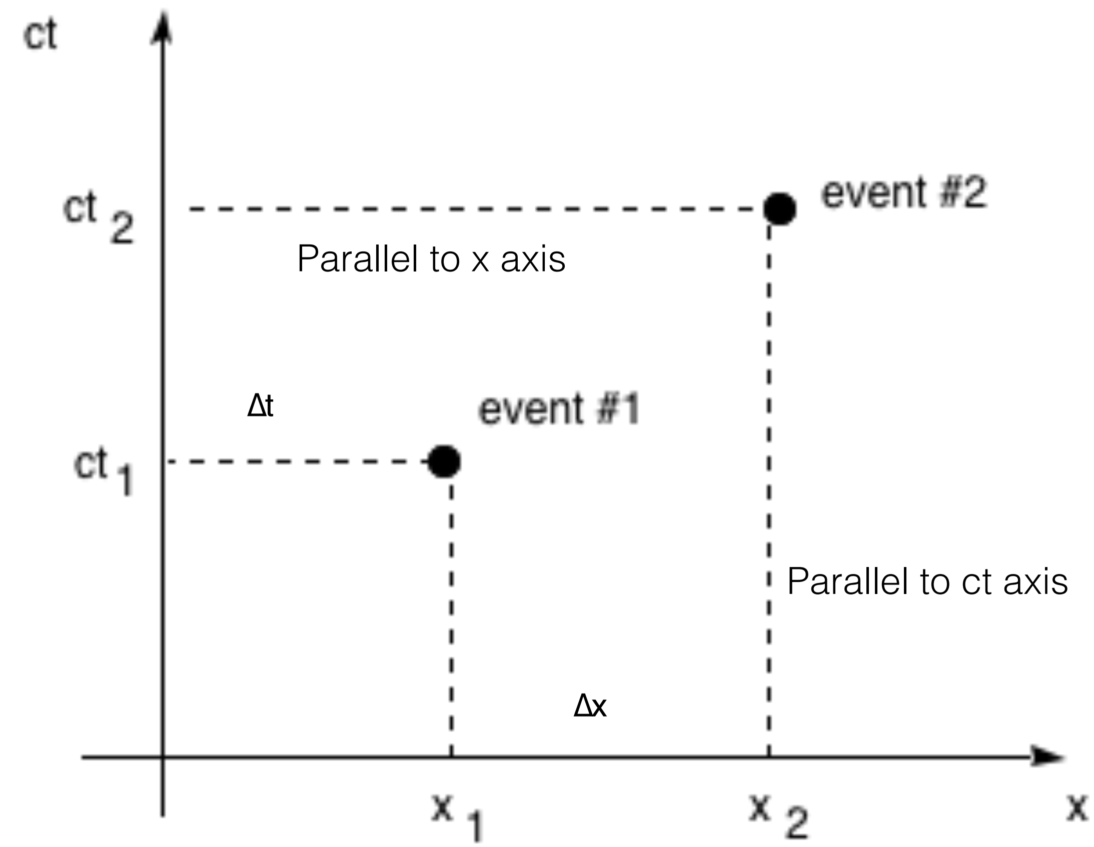
There are two very important lines on spacetime diagram are given below:



Now we discuss these concepts in detail.

We start with things from everyday experience.

Consider the diagram on **right**



Quantity $\Delta t = t_2 - t_1$ = **time-separation** or **coordinate time** between events

Quantity $\Delta x = x_2 - x_1$ = **spatial-separation** between events.

Can we also say at this point that Δt = time-interval between events

and

Δx = distance between events?

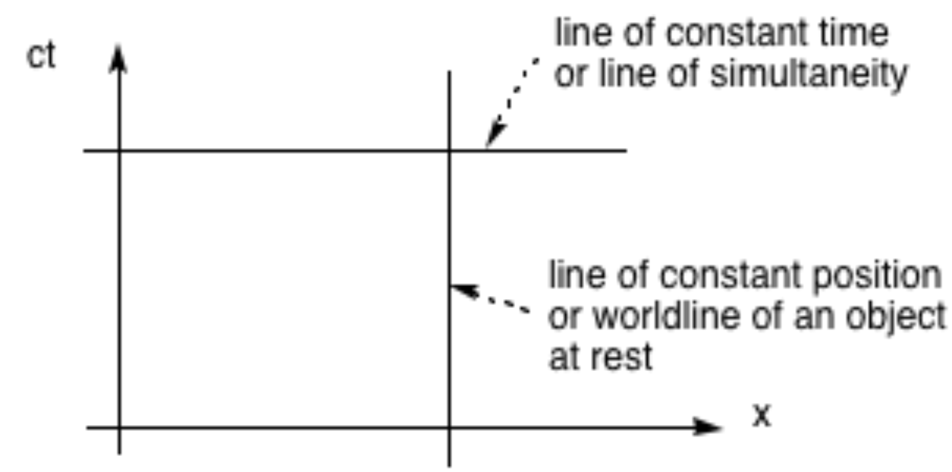
Answer is NO!

We must be very careful not to make assumptions when cannot prove statements;

A good rule = if do not know something is true, then should not assume it!!

It is clear, however,

that two events on same vertical line take place at same position

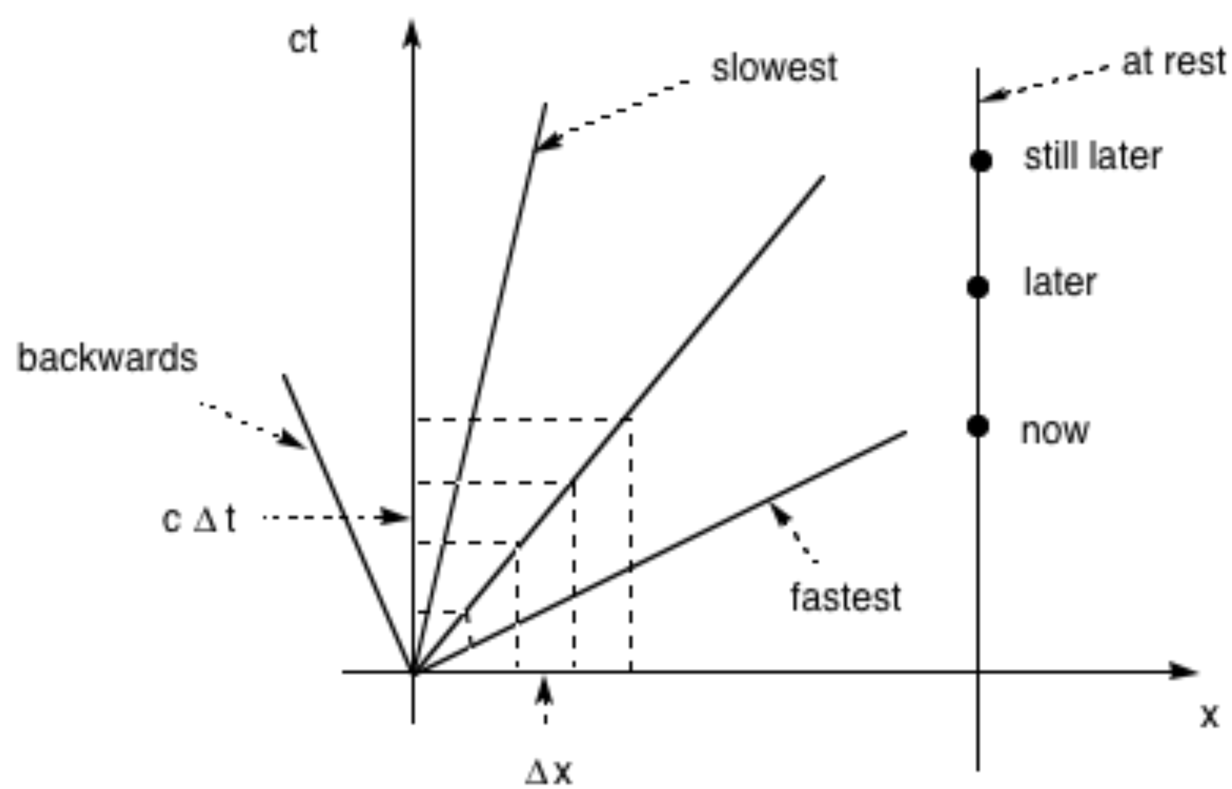


and

two events on same horizontal line take place at same time

(they are simultaneous) as earlier.

In diagram at right have several objects all moving with different speeds.



Before proceeding we must deal with the last roadblock in our ability to create these diagrams.

So let us step back for a bit…….

As I stated earlier, to Newton and his followers, time was self-evidently universal and absolute, flowing "equably without relation to anything external."

This makes time a frame-independent quantity: every observer in every reference frame (inertial or not!) should be able to agree on what time it is.

This assumption about time was considered "self-evident" because it is consistent with our immediate experience about the way that time works.

Galilean relativity was the basis of Newtonian physics until 1900!

You have deep understanding of Galilean relativity ingrained within your brain.

If you did not, then you would not have survived to be taking this class today.

Galilean relativity accurately describes the everyday world we live in.

The Galilean relative velocity formula, which we will derive later, is the **signature** of the old classical physics and everyday world.

Do an experiment:

Suppose we measure that I can throw an object with speed of 20m/s when at rest relative to you.

Then we can ask:

What speed will you observe me throwing it if I am running at speed of 10m/s relative to you?

The relative velocity rule we use everyday (the Galilean velocity relationship) is

$$\vec{V}_{\text{ball relative to you}} = \vec{V}_{\text{ball relative to me}} + \vec{V}_{\text{me relative you}}$$

Thus, the Galilean velocity relationship or velocity addition formula gives

30 m/s = 20 m/s + 10 m/s s.

Suppose instead that I am at rest and throw object at 20m/s and

you are running in direction opposite to that of moving object(towards me) at 10m/s.

What speed will you measure?

Galilean velocity addition formula gives

$$20 + 10 = 30 \text{ m/s}.$$

Finally, suppose instead that I am at rest and throw object at 20m/s and you are running in same direction as moving object(away from me) at 10m/s.

What speed will you measure?

Galilean velocity addition formula -> $20 - 10 = 10$ m/s.

Clear that in everyday experience with objects moving at everyday speeds, Galilean relativity works(classical theory valid).

Our discussion thus implies that if an object is measured to have a velocity $\vec{V}$ in one inertial reference frame (the Home Frame), its velocity $\vec{V}'$ measured in another inertial frame (the Other Frame) is given by the vector equation

$$\vec{V}' = \vec{V} - \vec{\beta}$$

where $\vec{\beta}$ is the velocity of the Other frame with respect to the Home Frame.

This equation is a direct consequence of the definition of an inertial reference frame and the assumption that time is frame-independent.

The problem is that light, or electromagnetic radiation or all of Maxwell's equations of Electromagnetism **do not satisfy** this velocity rule!!

In 1887, American physicists Albert Michelson and Edward Morley performed a sensitive experiment designed to prove the existence of this seemingly indiscernible ether.

If this ether filled all space, the earth must (as a result of its orbital motion around the sun) be moving through the ether at a speed comparable to its orbital speed of about 30 km/s.

This "ether wind" would have the effect of making the speed of light depend on the direction of its travel: a light wave moving against the ether wind would move more slowly than a wave moving across the wind.

Michelson and Morley therefore constructed a very sensitive experiment that compared the speed of two beams of light sent in perpendicular directions.

The presence of an ether wind would manifest itself in a slight difference between the speeds of light in these two directions.

To the surprise of everyone involved, it turned out that there was no discernible difference in the speeds of the two light waves.

The experiment was repeated with different orientations of the apparatus and at different times of the year (just in case the earth happened to be at rest with respect to the ether at the time of the first experiment), and by other physicists.

In all cases, the result was that the speed of light was apparently a fixed value, independent of the motion of the earth.

(In an experiment in the 1930s, R. J. Kennedy and E. M. Thorndike showed that although the relative speed of the earth's reference frame in July and its frame in January is roughly 60 km/s, the numerical value of the speed of light differed by less than 2 m/s during this time span.

This affirms the constant value of the speed of light with a fractional uncertainty of less than 10^8, that is, to better than eight decimal places.')

The cost?

The Galilean formula must be wrong.

But, as we will see shortly when we derive the formula, we will ask - what could possibly be wrong with their derivation?

At that point Einstein will intervene and say:

> "It is the idea of universal and absolute time that is wrong."

Special Relativity will follow as we will see.

For now, however, we will simply assume that the speed of light is a universal constant and use that fact to develop a procedure for synchronizing clocks.

A **light-second** is defined to be the distance that light travels in 1second of time.

Since 1983, the meter has in fact been officially defined by international agreement as the distance that light travels in 1/299,792,458 second.

Therefore there are exactly 299,792,458 meters in 1 light-second by definition.

I will use both regular units and these new units in different places so that you get used to both!

Definition of Synchronization

Two clocks in an inertial reference frame are defined to be synchronized if the time interval (in seconds) registered by the clocks for a light flash to travel between them is equal to their separation (in light-seconds).

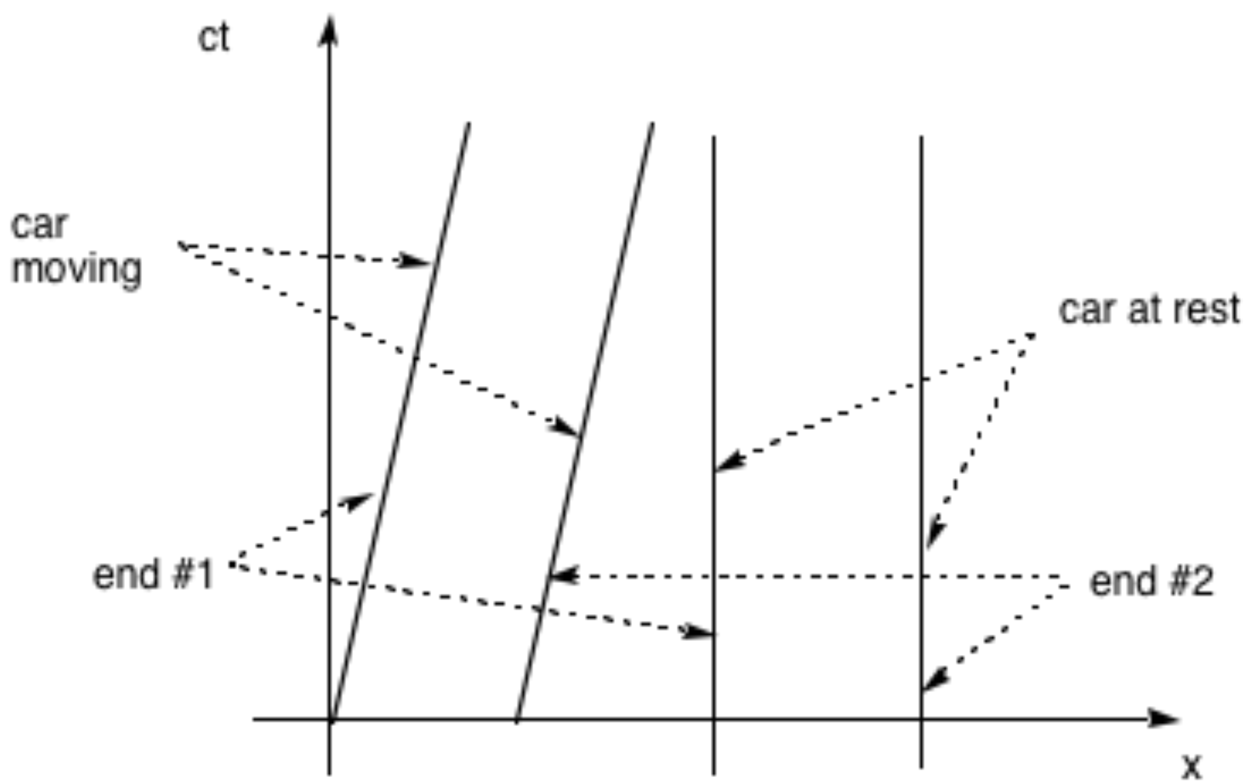
At this point each observer is able to calibrate their diagram's space and time axes.

So we can continue with our discussions……

Let now attempt to define how to measure the length of an object (the distance between it ends) and the time interval between two events.

Now imagine car on track —> must create a diagram to represent motion of car.

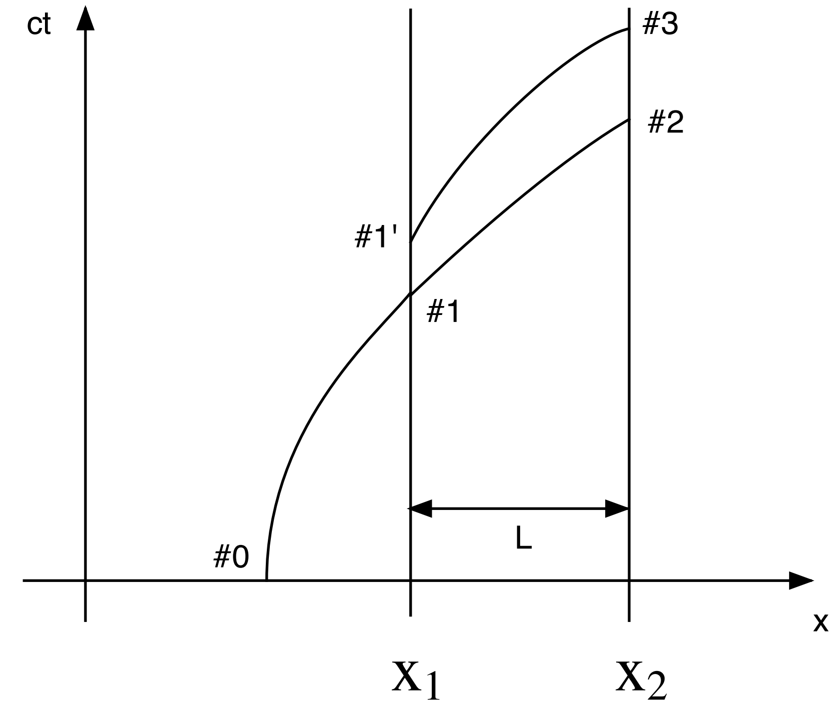
Diagram shows worldlines —> car at rest and car moving with constant speed in + x−direction.



We only show the worldlines of the ends of the car for simplicity.

In all cases, during any interval speed $v = \Delta x / \Delta t = 1/\text{slope(of worldline)}$.

Now let us attempt to measure the length of the car.



First consider a car at rest.

Diagram(right) represents me walking (in your frame of reference - these are your ct-x axes) first to one end of car and recording its position (x_1) and then walking to other end and recording its position (x_2).

At event #0, was previously(prior to event #0) standing at rest(talking) —> would draw vertical worldline.

Then walk over to one end of car (event #1 = (x_1, ct_1)).

Then walk over to other end of car (event #2 = (x_2,ct_2)).

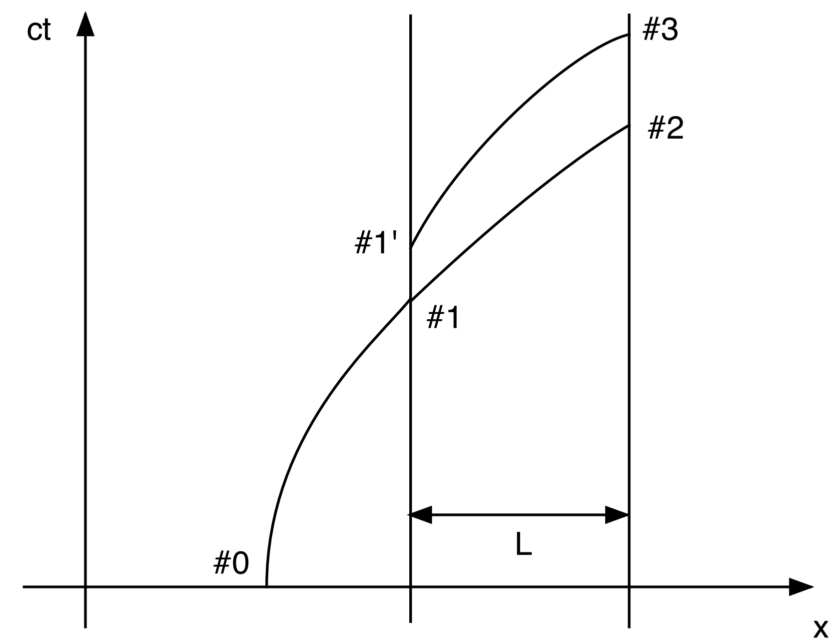
Remember you(your worldline) is the ct-axis on this diagram.

You are ct−axis(you are at rest at x=0) in your own frame of reference!!

You are the observer describing everything and drawing the diagram!!!

Alternatively, I could have delayed walking over to other end of car (until event $\#1' = (x_1, ct'_1)$)

and then gone over to other end (event $\#3 = (x_3, ct_3)$).



Length of car from my measurements

$\longrightarrow L = x_2 - x_1$ or $L = x_3 - x_1 = x_2 - x_1$.

For car at rest, length measurement is same no matter how long I delay getting to the other end (whether I use event #2 or #3).

Notice how car is just in spacetime.

We do not have to be there!!.

What is actually in spacetime for car?

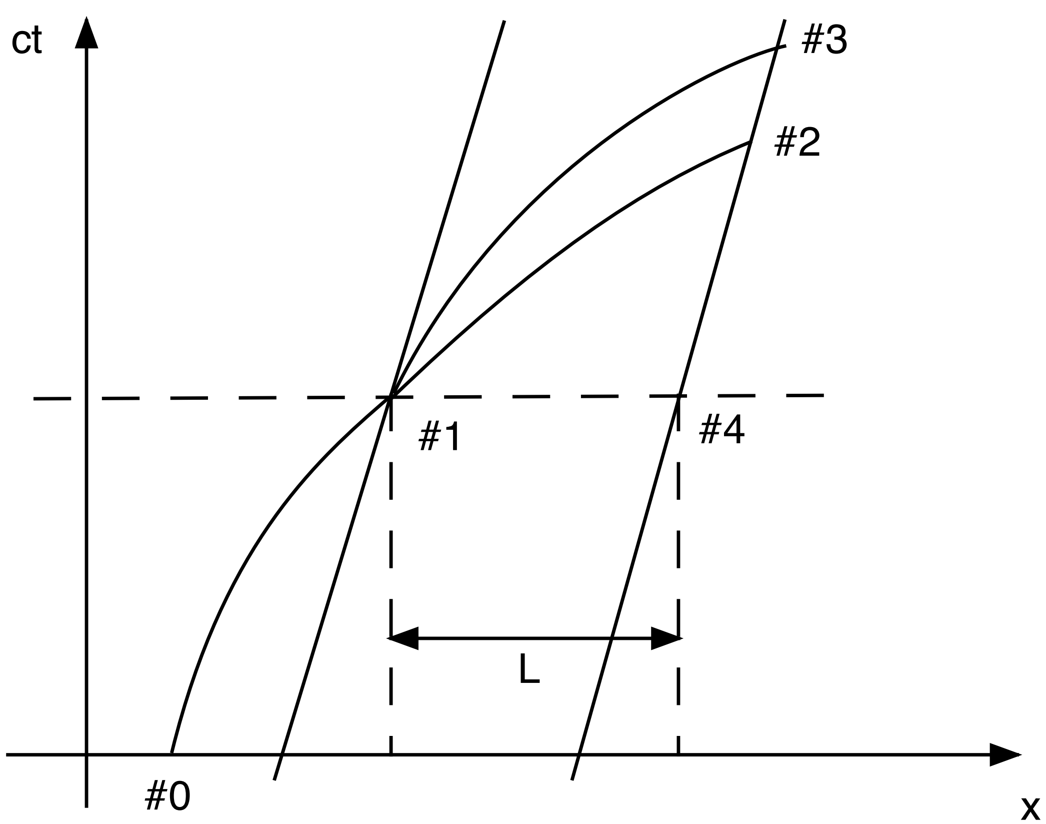
Look carefully.... **all** of its past, **all** of its future — everything about car is in spacetime!!!!

What is difference if one uses a many-observer model?

Two observers (of all observers) would happen to be located at the ends of the car - they record locations and then calculate length. No changes from earlier result for car at rest!!

At this point -> there are no problems with length measurements.

Now consider a **moving** car:



At event #0, standing at rest.

Then walk over to one end of car (event #1 = (x_1, ct_1)).

Then walk over to other end of car (event #2 = (x_2, ct_2)).

Alternatively, could have walked over more slowly to other end of car (event #3 = (x_3, ct_3)).

In this case, $x_3 \neq x_2$.

Is length of car $L_{12} = x_2 - x_1$ or $L_{13} = x_3 - x_1 > L_{12}$?

As can be seen from diagram, neither is correct result L(from previous measurement of car at rest).

Is there **operational** procedure that we can use to guarantee will always measure correct length (= length measured at rest)?

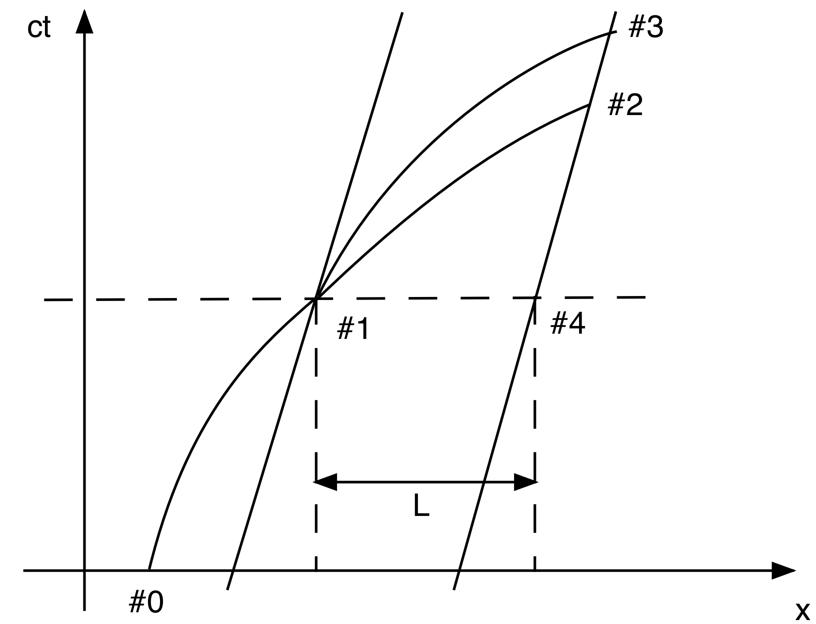


Diagram (dashed horizontal line)indicates answer.

If we measure location of ends of car at same time (simultaneously), namely, events #1 and #4, then we get correct length L (using any other line of simultaneity is OK also).

Of course -> impossible measurement for single observer, but not for many-observer model.

Have all observers close eyes and when their clock alarms go off (all set to go off simultaneously), then two observers will be located at ends of the car (even if moving) and length = spatial separation of their grid locations.

Thus we **define** length measurement as

spatial separation between endpoints of object measured simultaneously

Philosophers would not let me use word **define** for this **operational procedure;** luckily no philosophers here to challenge me!

A question: Have we just exchanged unknown meaning of length for new unknown, namely, simultaneity?

Answer is YES!

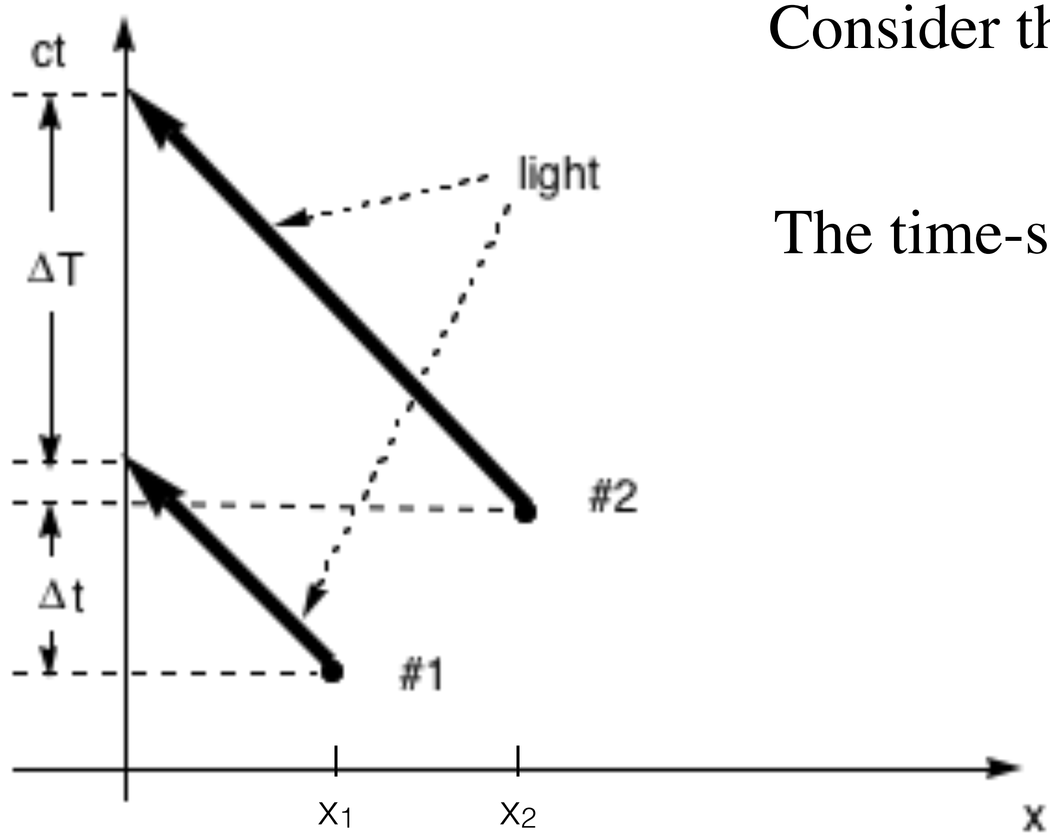
However, that is what **operational procedures** are all about and that is why they differ from definitions.

We proceed this way because we think, at this point,

we **are able** to define simultaneity unambiguously,

and a new procedure is **better than** not knowing how to measure length.

What about time intervals?



Consider the two events shown left:

The time-separation between events #1 and #2 is $\Delta t = t_2 - t_1$.

I am the observer. My worldline is ct−axis.

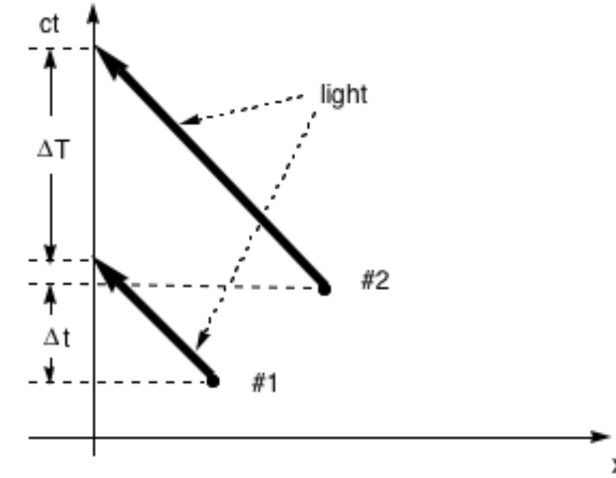
Can I measure this quantity?

I impose **restriction** that I can only have **confidence** in measuring instruments on my worldline (i.e., always with me)

in that case, then, the answer is NO!.

I can, however, measure quantity ΔT = time indicated on diagram, since that measurement **can** be made with clock I am carrying with me **on** my worldline in following manner.

Event #1 = (x_1,ct_1) takes place and sends out light beam towards me.



Light beam arrives at me (worldlines intersect) at time $t_1 + x_1/c$,

i.e., actual time of event t_1 + time takes light beam to reach me from distance x_1.

Similarly for event #2.

Thus, for $\Delta x = x_2 - x_1$.

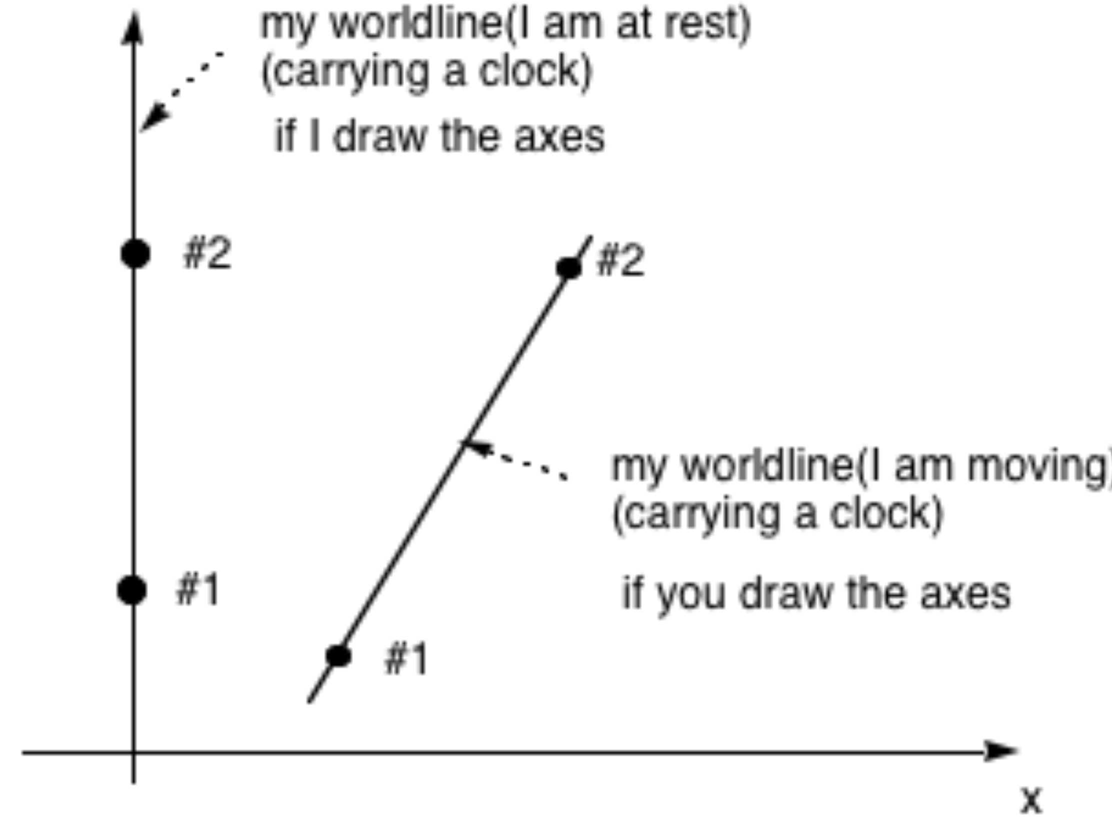
$$\Delta T = \left(t_2 + \frac{x_2}{c}\right) - \left(t_1 + \frac{x_1}{c}\right) = \Delta t + \frac{\Delta x}{c}$$

It is clear that $\Delta T > \Delta t$.

If I independently know spatial separation Δx between two events, then could infer(**calculate**) time-separation Δt, but this is **not** a measurement!

ΔT is time one **sees** between two events.

Is either of these values = time interval?



We just do not know!

Must create **operational** definition for time interval.

Done as follows. Looking at figure(right)

See that in both cases, time interval between two events is operationally defined as

difference in my clock readings,

i.e., **clock and thus me** must have a worldline that passes through **both** events in order to define the time-interval between the events in an **unambiguous** manner.

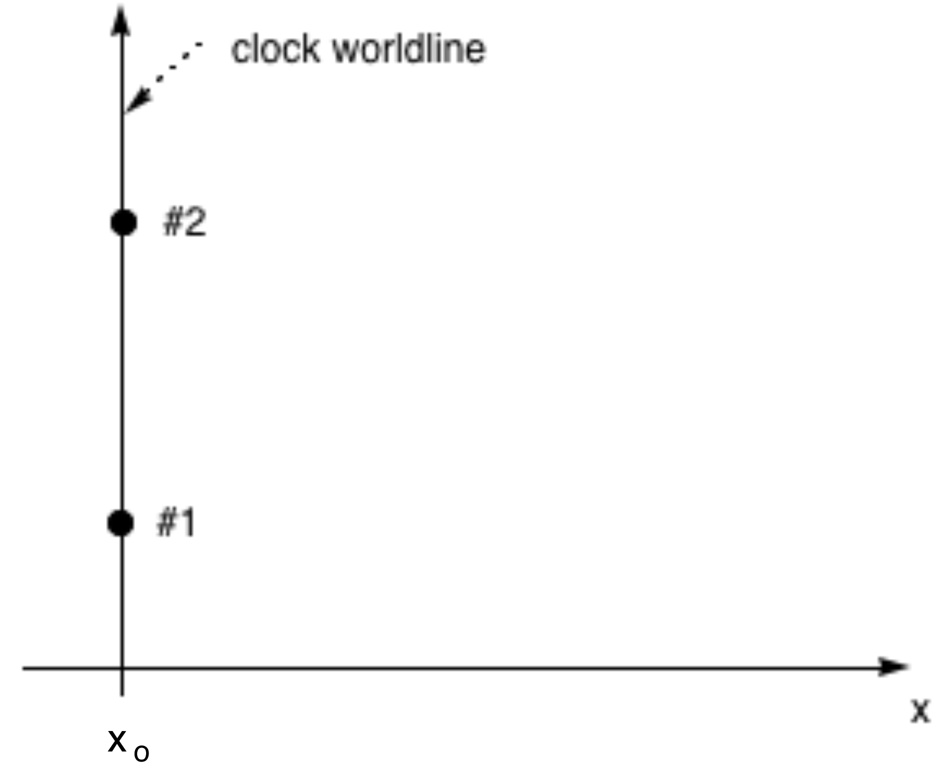
Prescription assumes that nothing happens to a clock when it moves that changes this result.

We do not know that this is true at this time.

Thus, to be safe, we define

> **time interval between two events = time-separation when clock is at rest and thus, the two events take place at same position according to observer carrying clock**

as shown(right):



2 events in this case are

$$\text{event } \#1 = (x_0, ct_1) \quad \text{and event } \#2 = (x_0, ct_2)$$

and time interval or elapsed time between events is given by $t_2 - t_1$.

Is this what you actually do? **NO**.

You move between events usually (changing your speed in process) and assume that has no effect on your clock or you stay still and infer time interval by measuring time that you see between two events.

As we shall see from the theory we are developing, this is OK for everyday world we live in but not in world where objects move with large speeds.

We are **assuming** that all of these measurement procedures are **objective**.

Suppose there is a rotten core in the apple of the scientific objectivity.

Physics works .... makes correct predictions.

Does it matter if we are really being subjective, i.e., that entire view of spacetime might be dependent on human observation or that measurements are relative to observer?

Philosophers spend a lot of time discussing this sort of stuff!

We do not!

So now we have operational definitions that allow a single observer looking at universe to describe events, measure distances and time intervals between events, and so on and report on what happened in some experiment.

Our problem arises, however, when **second** observer, **moving** relative to first observer, appears and **also** tries to describe experiment using same procedures.

Galilean Relativity

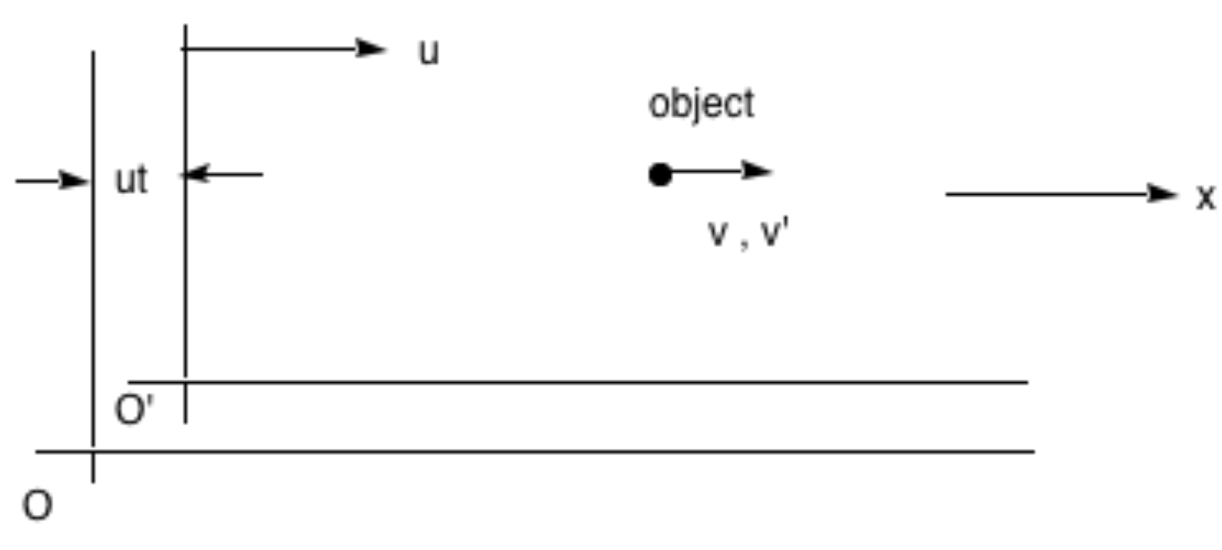
Central to any discussion of relativity that prevailed **alongside** Newtonian (pre-Einstein) physics is **concept of absolute time** as I mentioned earlier.

Newton and Galileo **assumed** that passage of time was **same** for all observers **no matter what they were doing**.

Thus if two observers **separately** measured time interval between two events, then they **assumed** that $\Delta t = t_2 - t_1 = t'_2 - t'_1 = \Delta t'$.

Suppose two observers are **moving** with respect to each other (along **common** x−direction) with **relative** speed u such that their respective origins coincide at $t = t' = 0$.

Then, at some time t later, we might have situation shown.



Know from **everyday experience(your survival)** that if observer O' measures a velocity v' and observer O measures a velocity v for some moving object, the relationship between these two measured velocities is given by $v' = v - u$.

Now to measure a velocity(assumed to be constant) of object, each observer must observe 2 events in its motion.

Suppose this has been done and have measured results for the 2 events as:

$$\#1 \rightarrow (x_1, ct_1) \text{ and } (x_1', ct_1')$$

$$\#2 \rightarrow (x_2, ct_2) \text{ and } (x_2', ct_2')$$

Then we have

$$v = \text{ velocity measured by } O(\text{frame S}) = \frac{x_2 - x_1}{t_2 - t_1} = \frac{\Delta x}{\Delta t}$$

$$v' = \text{ velocity measured by } O'(\text{frame S'}) = \frac{x_2' - x_1'}{t_2' - t_1'} = \frac{\Delta x'}{\Delta t'}$$

Now the absolute time concept says that $\Delta t = \Delta t'$ and this then implies that

$$v' = v - u$$

or

$$\frac{\Delta x'}{\Delta t'} = \frac{\Delta x}{\Delta t} - u$$

or

$$\Delta x' = \Delta x - u\Delta t$$

Now if we choose the events representing measurement of particle velocity to be

$$\text{event } \#1 \rightarrow (x = 0, ct = 0) \text{ and } (x' = 0, ct' = 0)$$

$$\text{event } \#2 \rightarrow (x, ct) \text{ and } (x', ct')$$

which -> a **choice of origin** for the space and time measurements

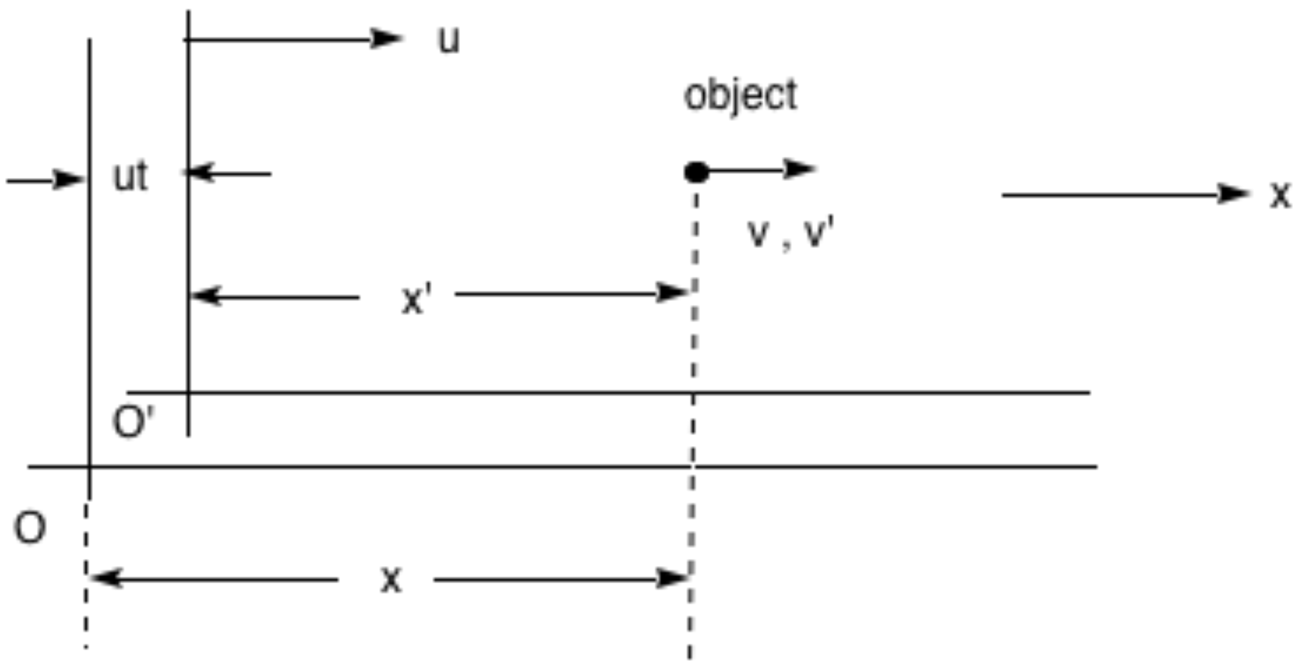
(always allowed because physical phenomena are not dependent on choice of origin - this is confirmed by many experiments).

We then obtain equations

$$ct' = ct \quad , \quad x' = x - ut = x - \frac{u}{c}ct \qquad \text{or} \qquad c\Delta t' = c\Delta t \quad , \quad \Delta x' = \Delta x - u\Delta t = \Delta x - \frac{u}{c}c\Delta t$$

as equations relating the two sets of observations —> equations of **Galilean Relativity** —> Galilean transformation equations or relations.

This relationship is as shown below:



They allow two observers in different frames of reference moving with constant speed relative to each other to compare their respective observations under the assumption that Newtonian/Galilean physics is valid.

Galilean relativity was the basis of Newtonian physics until 1900!

You have deep understanding of Galilean relativity ingrained within your brain.

If you did not, then you would not have survived to be taking this class today.

Galilean relativity accurately describes the everyday world we live in.

Relative velocity formula —> **signature** of old classical physics and everyday world.

This is **Old Original World View** of 19th century physics circa 1900 ....

a product of finest minds - **developed** over several centuries.

Everyone was **comfortable** with theory.

It was **internally consistent**.

It **worked** amazingly well(**agreed** with all experiments).

It is clear that in everyday experience with objects moving at everyday speeds, Galilean relativity works(classical theory valid).

First, what went wrong with theory as devised so far?

When measured by an observer at rest relative to experiment setup,

speed of light is $c = 3.0 \times 10^8$ m/sec = 186,000 mi/sec

-> very large compared to everyday speeds.

What is fastest humans have launched any object?

Apollo 10 Capsule – 24,791 mph = 11.07 km/s

Stardust – 28,856 mph = 12.90 km/s

Voyager 1 – 38,610 mph = 17.26 km/s

An iron manhole cover – 125,000 mph = 55.88 km/s

Helios Satellites – 157,078 mph = 70.22 km/s

What is fastest object ever seen in universe?

light	**299,792.4580000 km/s**
cosmic rays	**299,792.4579999 km/s**
protons in LHC	**299,792.4553019 km/s**

Observed speed of objects **depends** on motion of source and observer of object.

As I said earlier, Michelson and Morley, two American physicists, did experiment of this sort with light.

Found that speed of light always measured to be $c = 3.0 \times 10^8$ m/sec

no matter what source or observer of light was doing!

Experiments gave astonishing result:

speed of light = constant = $c = 3.0 \times 10^8$ m/sec
motion of source or observer

Leads to direct breakdown of Galilean relativity since Galilean relativity

-> 2 observers in relative motion both looking at light (moving object)

must have $c' = c - u \neq c$.

Clearly, a new theory is needed.

—-> Experiment forcing paradigm shift in theoretical understanding of world.

That is way physics works!

We will derive new theory assuming **one general principle and results of 2 experiments**:

Principle of Relativity

Laws of physics are identical for all observers in uniform relative motion.

Experiment #1:

Speed of light is universal constant c independent of motion of source or observer

Experiment #2:

Experimentally observed that when source of light and detector of light are moving relative to each other with speed v wavelength of observed light changes with relative speed.

Experimental result given by formula

$$\lambda = k(v)\lambda_0$$

where λ = wavelength observed by observer moving with speed v wrt light source,

λ_0 = wavelength observed by observer at rest wrt light source (v = 0) and

$$k(v) = \sqrt{\frac{c+v}{c-v}}$$

where c = speed of light and $v > 0 \rightarrow$ source and observer are moving away from each other

$v < 0 \rightarrow$ source and observer moving towards each other.

—-> Famous galactic red shift observed by astronomers for light received on earth from distant galaxies moving away from earth

Wavelength of light wave related to frequency and period of light wave by formula

$$\lambda f = c = \frac{\lambda}{T}$$

where f = frequency (hertz = Hz = oscillation/sec), T = period (sec) and f = 1/T .

Other physicists(classes) might derive this theory with a smaller number of assumptions.

For clarity, however, at level we are working here, the derivations will be clearer if we use an extra experimental result to start.

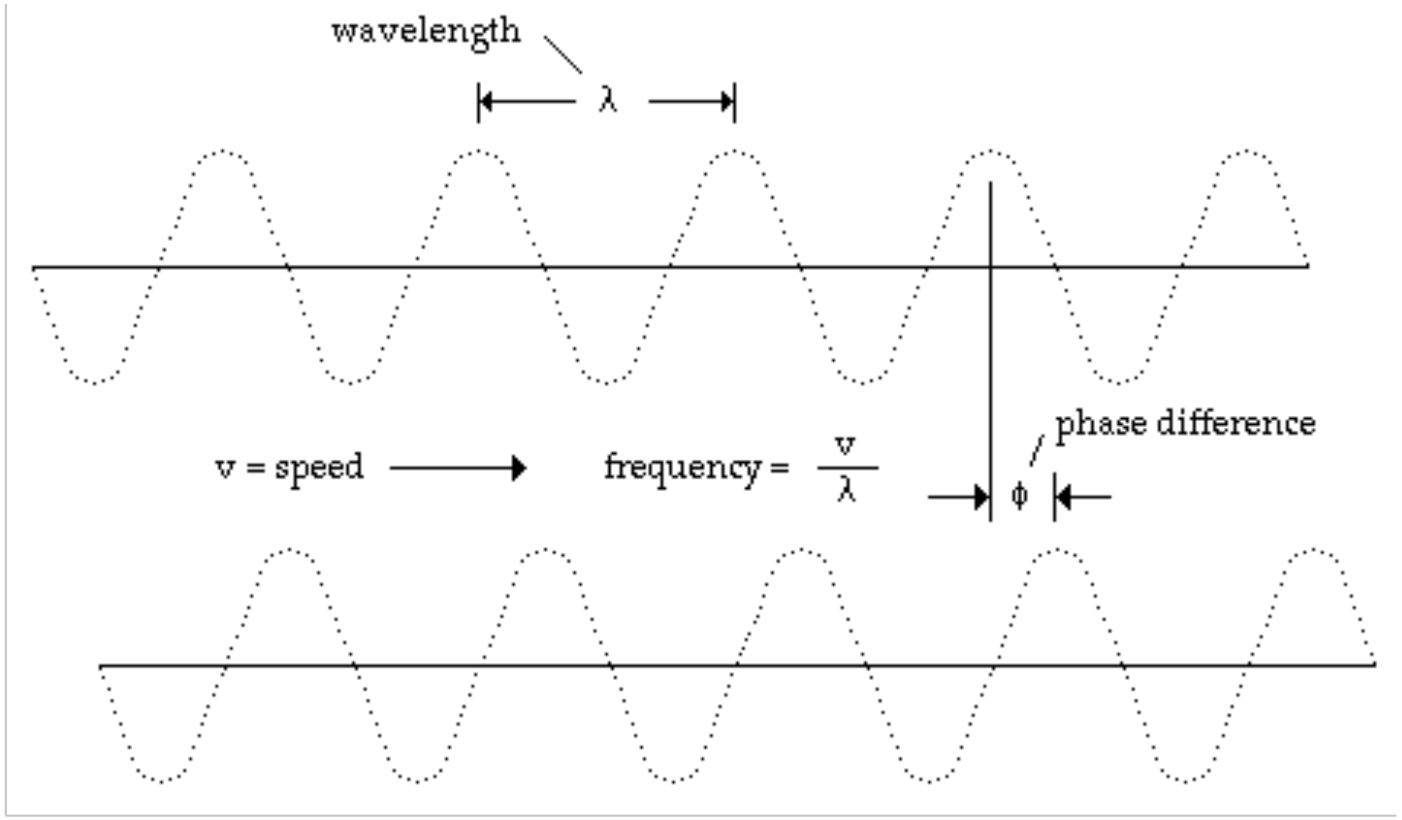
With a lot more work could do the same derivation leaving it out asyou will see from other approaches I present during the lectures.

Short Digression: Review of Wave Properties

Waves are periodic phenomena in space and time.

Sinusoidal wave illustrates typical wave ...

but only need periodicity property.



Wavelength = distance between "like" points

Frequency = f = 1/(time for point to repeat) = 1/period = 1 / T

Amplitude = maximum displacement = A.

Energy of classical wave proportional to (A = Amplitude)^2 and independent of frequency.

Example:

$$y = A\cos(kx - \omega t)$$

$$k = \frac{2\pi}{\lambda} \quad , \quad \omega = 2\pi f = \frac{2\pi}{T}$$

$$y = A\cos\left(\frac{2\pi}{\lambda}x - \frac{2\pi}{T}t\right) = A\cos 2\pi\left(\frac{x}{\lambda} - \frac{t}{T}\right)$$

$$v = \text{ wave speed } = \lambda f$$

Fix t → photograph of waveform in space → wavelength, i.e., t = 0 —> $y = A\cos 2\pi\left(\frac{x}{\lambda}\right)$

Fix x → oscillation in time → frequency or period, i.e., x = 0 —> $y = A\cos 2\pi\left(\frac{t}{T}\right)$

When add waves

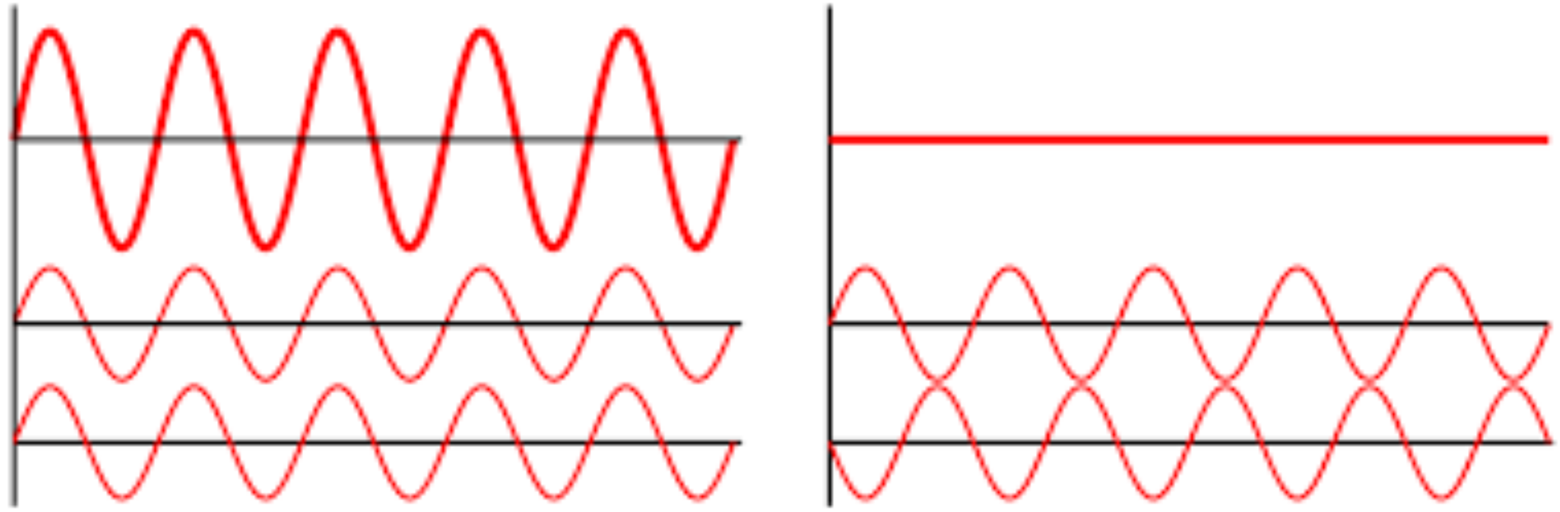
-> simple algebraic sum of their amplitudes at each space-time point

-> principle of superposition.

Interference Types:

constructive → phase difference = 0 or peaks line up with peaks

destructive → phase difference = 1/2 wavelength or peaks line up with valleys



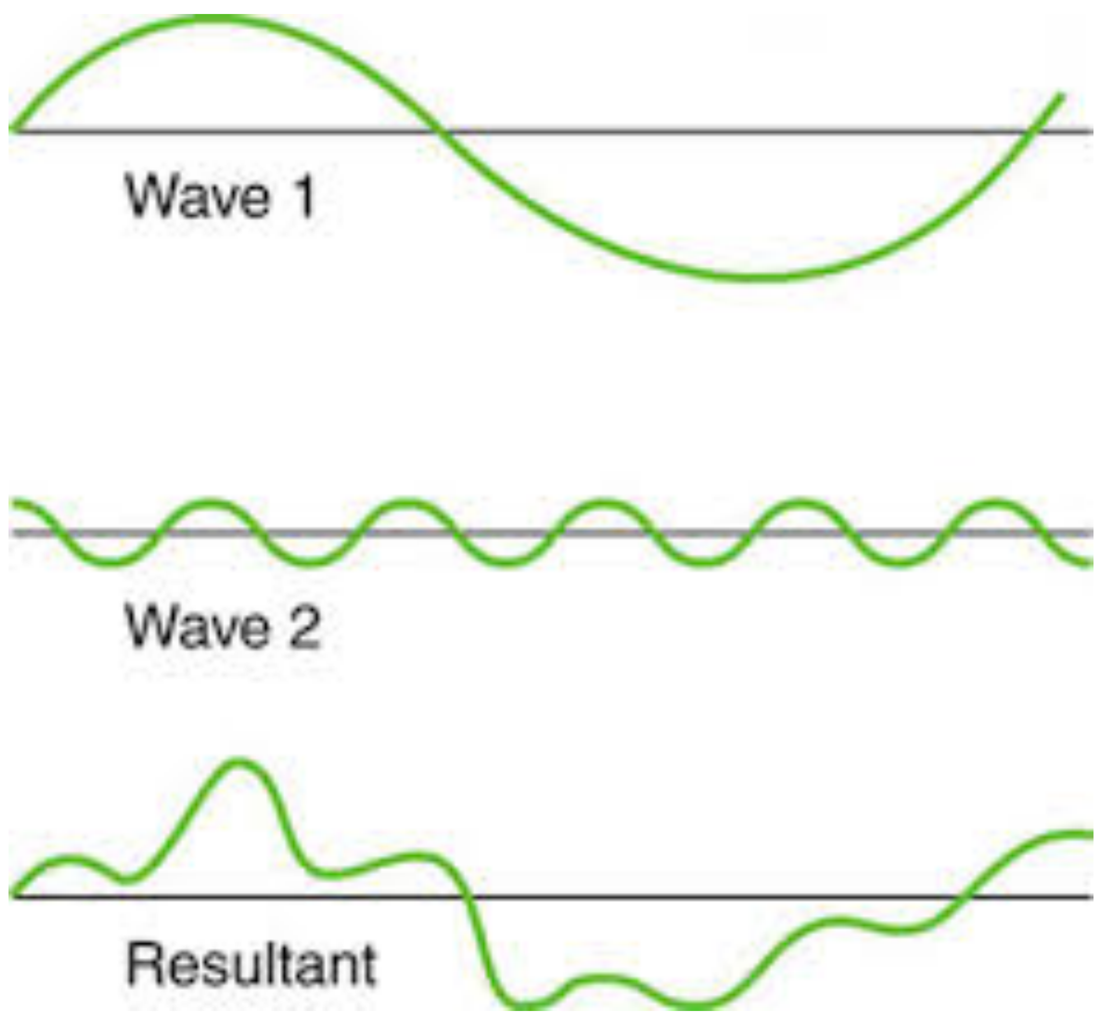
Wave 1 + Wave 2

Wave 1

Wave 2

Constructive

Destructive



Now, continuing with our the earlier discussion of our assumption, we write

$$c = \frac{\lambda}{T} \text{ (for an observer at rest wrt source)}$$

$$c' = c = \frac{\lambda'}{T'} \text{ (for an observer moving wrt source)}$$

$$T' = T_{\text{observer moving wrt source}} = k(v)T = k(v)T_{\text{observer at rest wrt source}}$$

We have explicitly assumed the result of two experiments in writing this formula, namely,

(1) c = speed of light = constant for all observers

and

(2) red-shift relation between time intervals.

Last equation follows as below.

$$c' = \frac{\lambda'}{T'} = c = \frac{\lambda}{T} \rightarrow \frac{T'}{T} = \frac{\lambda'}{\lambda} = \frac{k(v)\lambda}{\lambda} = k(v)$$

Using these results as our theoretical assumptions we can now derive special relativity.

Spacetime Diagrams

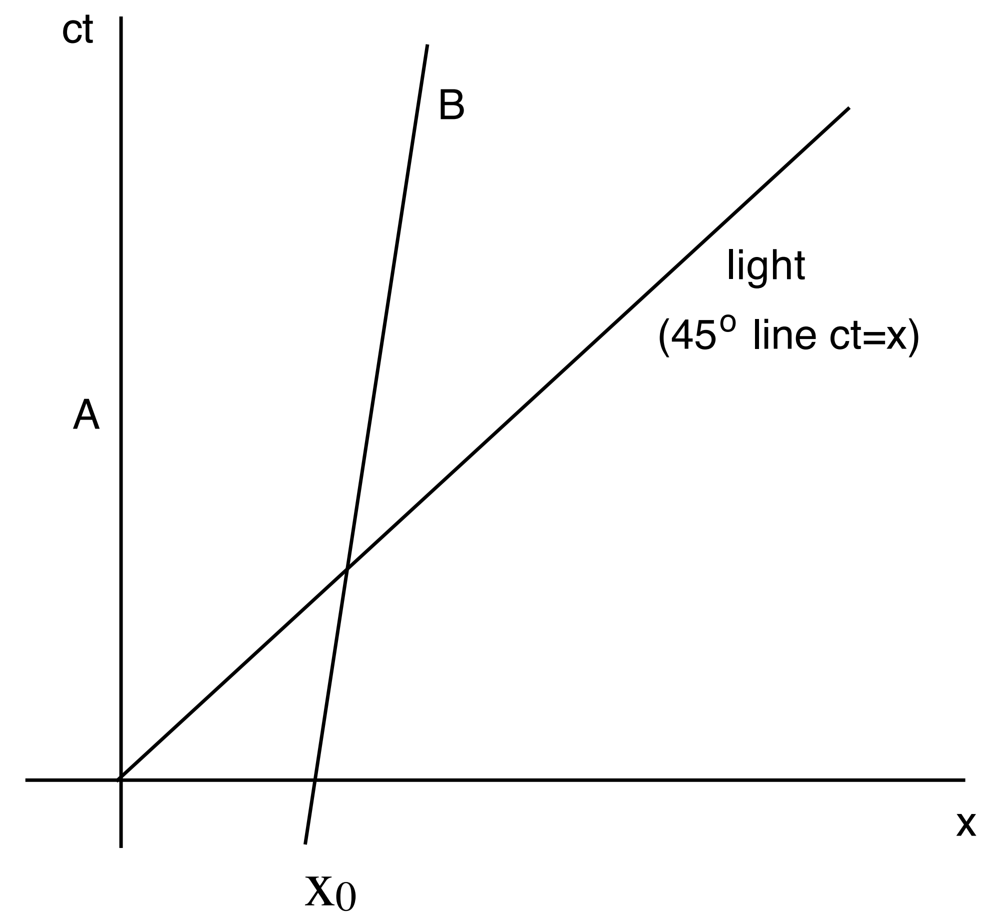
Consider 2 observers A and B.

Observer B moving away from observer A with constant speed $v < c$

(1–dimensional motion for simplicity)

and separated (in x-direction) at $t = 0$ by distance x_0.

—-> Represented by spacetime diagram shown:



Included worldline of light beam that started at (0,0).

Now clear why we choose vertical axis to be ct rather than just t —> world line of light then always a 45° line!

We assume that each observer carries their own clock.

Radar Method

Each observer will need to determine events on worldline of other observer

using only measurements available on their own worldline.

We can only trust information recorded by instruments moving with us (on same worldline)

—> must figure out how A (or B) can measure (x, ct) values

for an event not on their own worldline.

Remember, on my own worldline,

it is true that time interval between events that I experience

(my worldline must pass through both of them)

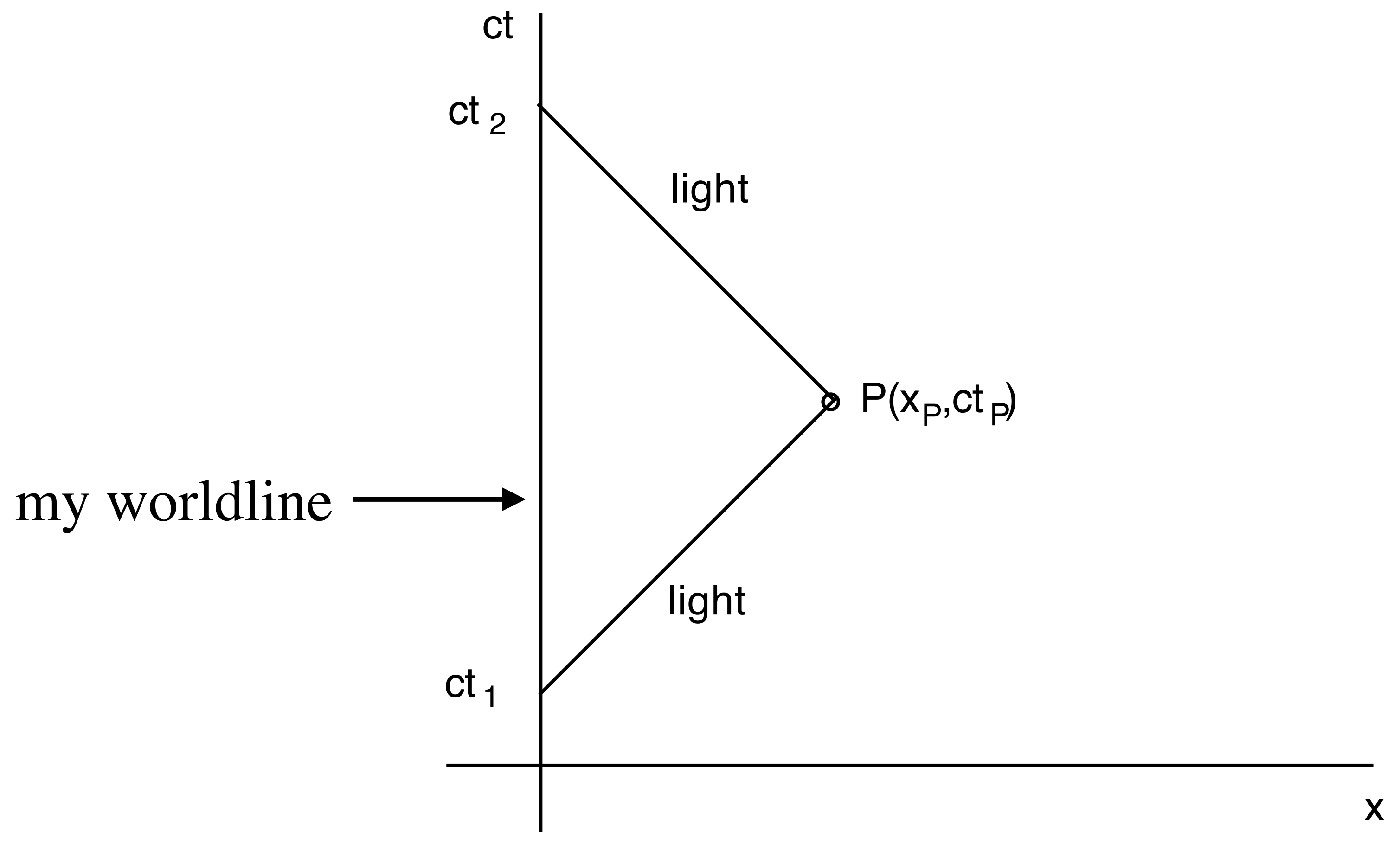
is directly measured by the clock that I carry

and

my position is constant (usually assumed to be zero).

Method we now develop => **radar method**.

Consider the diagram at right:



Observer A assigns coordinates to event P by bouncing light signal off of whatever occurring at P.

Light signal is sent out at event $(0, ct_1)$ and is received back at event $(0, ct_2)$.

Note, it is very important that both events are on A's worldline.

We then have (only using $\Delta x = c\Delta t$ for light)

$$(x_p - x_{me}) = (x_P - 0) = x_P = c(t_P - t_1) = (ct_P - ct_1) \qquad \text{going out}$$

$$(x_p - x_{me}) = (x_P - 0) = x_P = c(t_2 - t_P) = (ct_2 - ct_P) \qquad \text{coming back}$$

or

$$ct_P - ct_1 = ct_2 - ct_P \to ct_P = \frac{c(t_2 + t_1)}{2}$$

which is the average of sending and receiving times(that makes sense).

Then, substituting, we obtain

$$x_P = (ct_2 - ct_P) = \frac{c(t_2 - t_1)}{2}$$

Thus, any observer (a particular worldline)

can determine the coordinates of an event off their worldline by only using light,

which has constant speed for all observers

and

only measuring time values on their own clock (the clock on the same worldline).

Emphasize - that we must only use information about events we actually experience

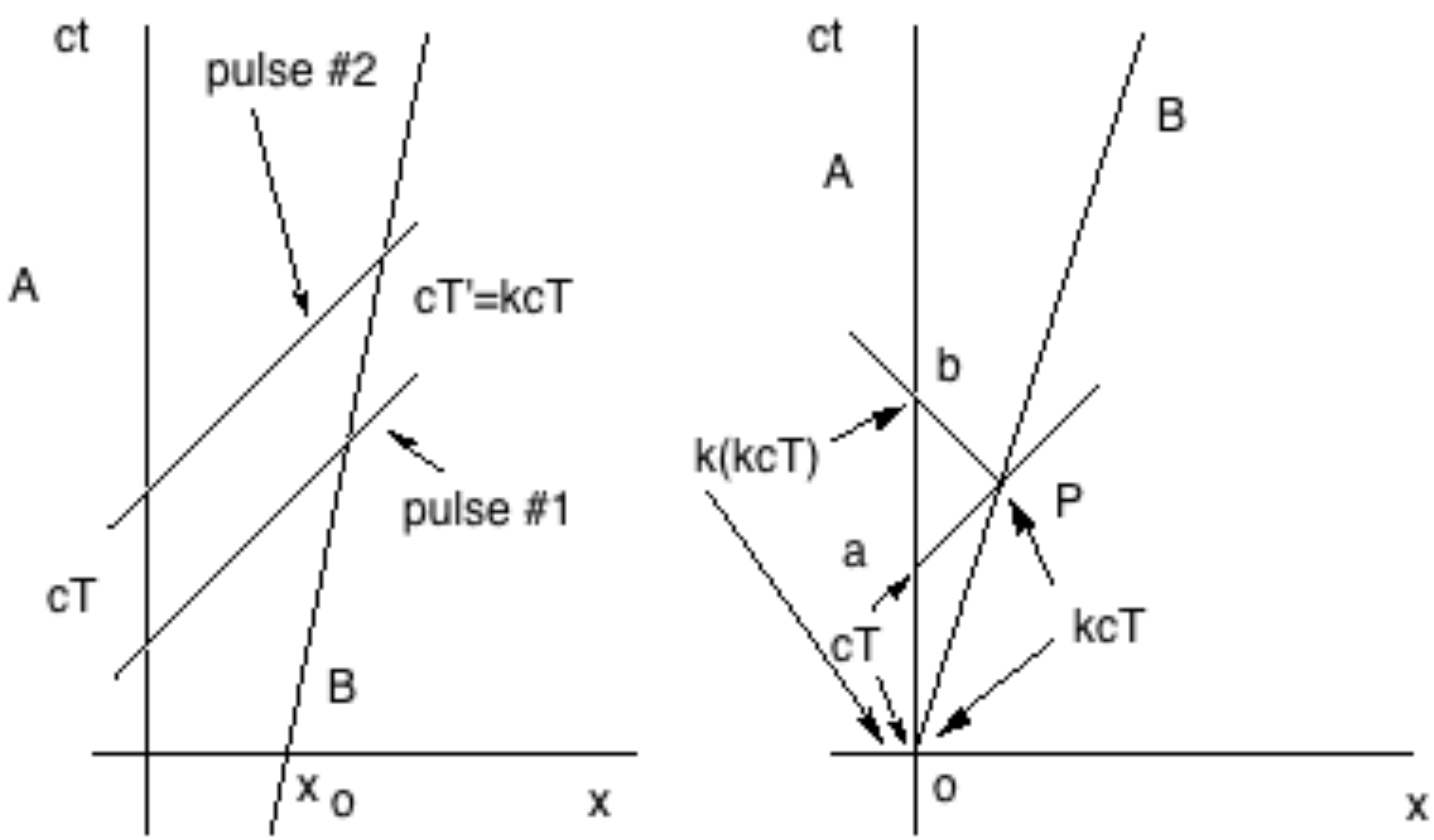
(are on our worldline), otherwise we **cannot** be certain of their validity.

Special Relativity now follows ....

Radar procedure + assumptions -> derive new theory called Special Relativity (Einstein 1905).

We consider experiments represented by worldlines in spacetime diagrams below.

In each case, observers A and B are assumed to be moving away from each other with speed v.



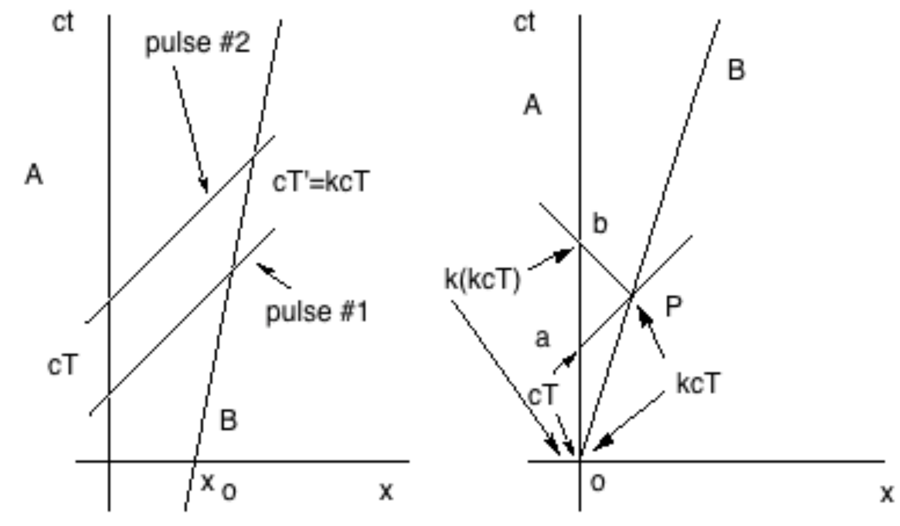
(a) (b)

Part(a), 2 pulses separated in time by T sent from A to B.

Part(b), 2 pulses(one at t = 0) sent from A to B and B then sends each back to A.

Part(a), B's worldline given by equation

$$x = x_0 + vt = x_0 + \frac{v}{c}ct \quad \text{(B is at } x_0 \text{ at } t = 0\text{)}$$



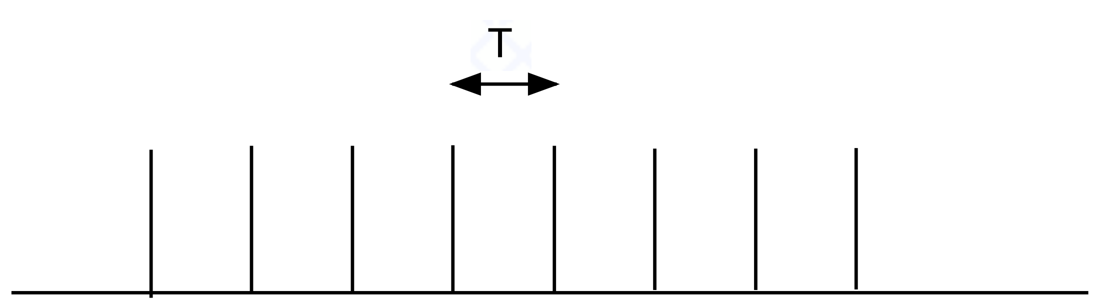
(a) (b)

Part(b), B's worldline given by equation

$$x = \frac{v}{c}ct \quad \text{(B is at } x = 0 \text{ at } t = 0\text{)}$$

In both cases, assuming light being sent out

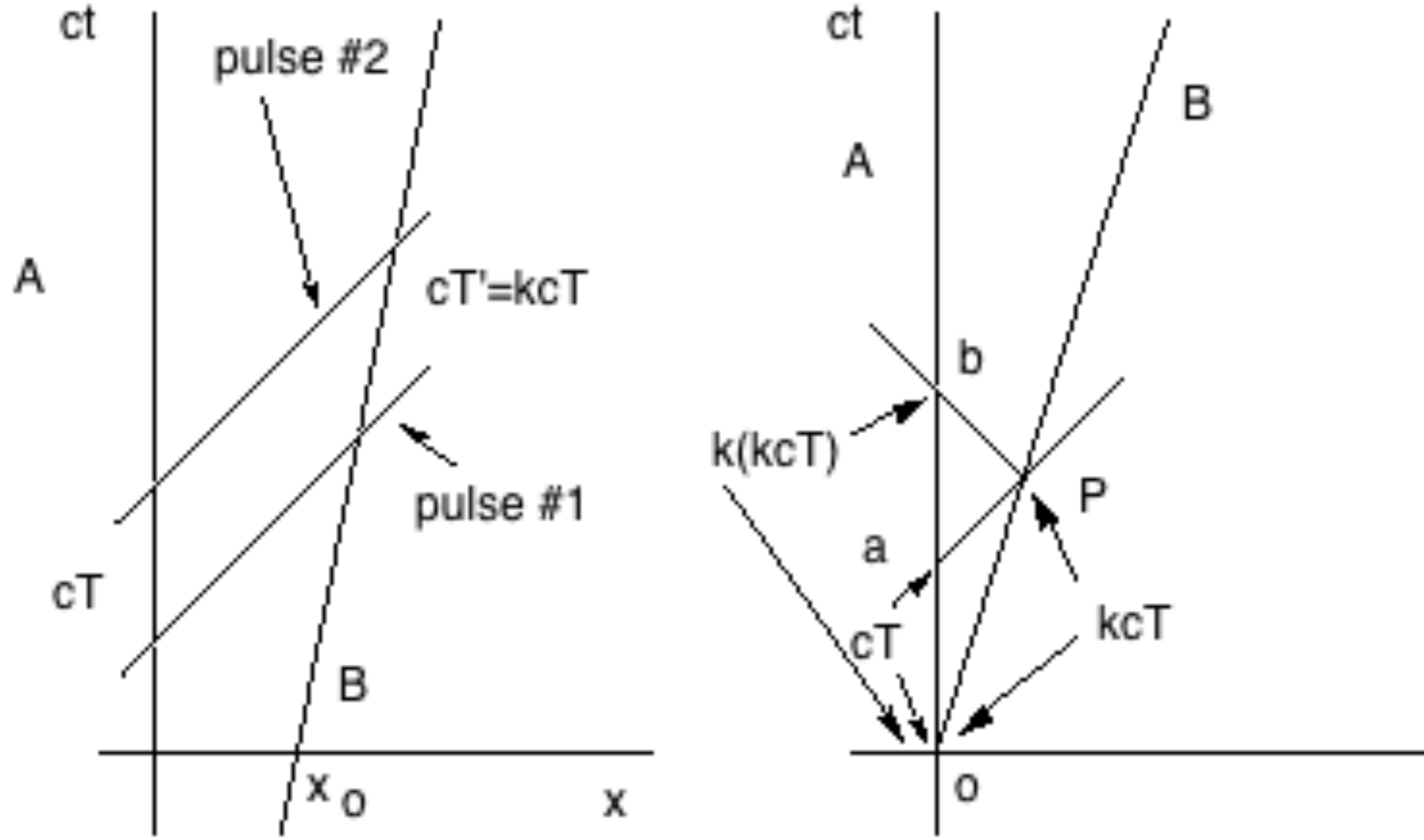
= series of pulses separated by time T in frame of source (A) as shown.



For 1st experiment(a), assumptions —> interval between reception of 2 signals seen by B

(according to clock traveling with B), is cT′

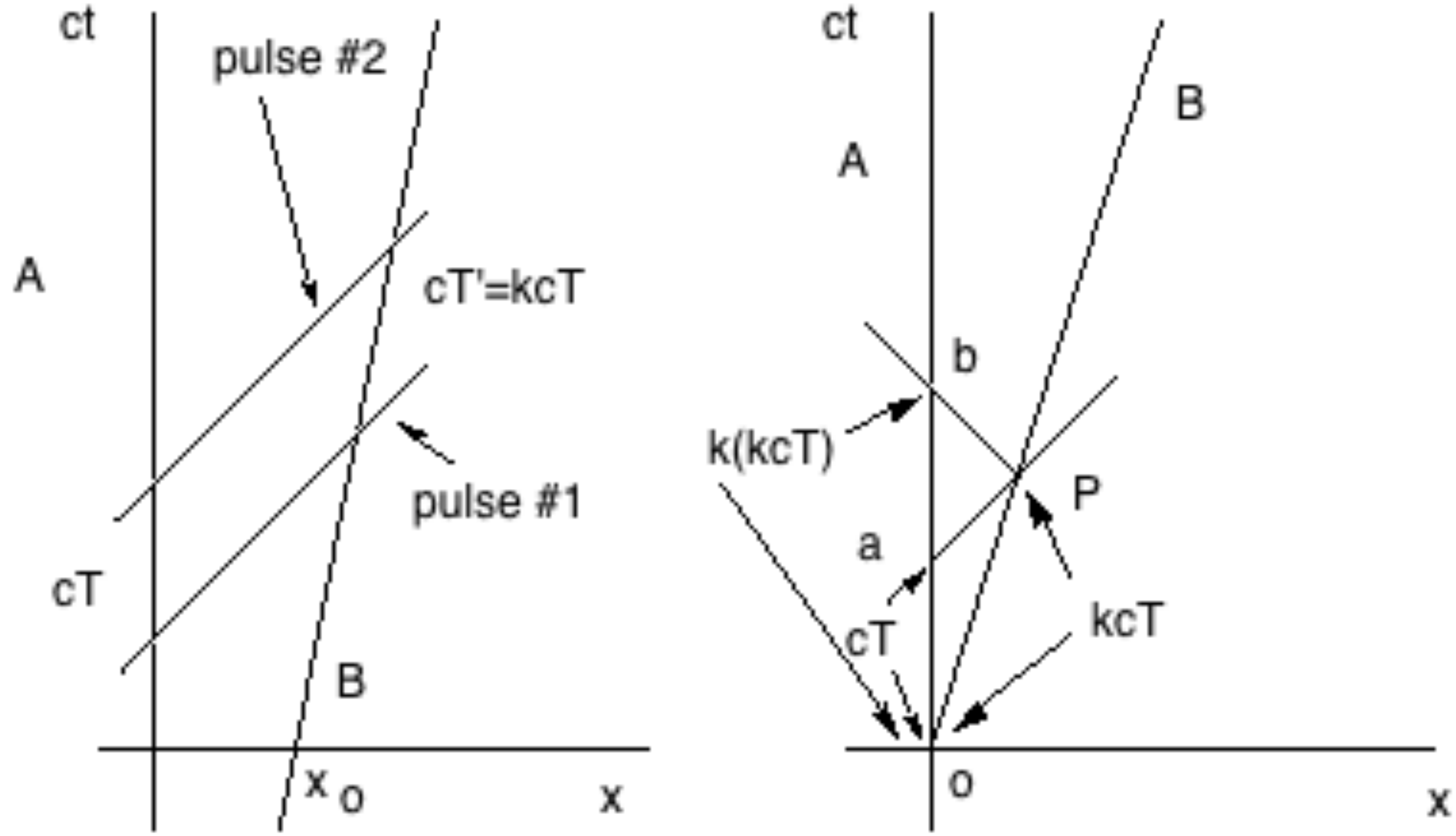
and this interval proportional to cT (see diagram) with proportionality factor k(v) that depends only on relative velocity between A and B, that is



(a) (b)

$$cT' = k(v)cT \quad , \quad v = \text{ velocity of B wrt A}$$

That is one of our assumptions!



(a) (b)

In 2nd experiment(b), see 2 pulses separated by time T sent out by A

(1st when at **same** spacetime point (origin));

received by B separated by cT and then sent back to A and received separated by k(cT).

Used fact that physical laws independent of relative motion(assumption (1))

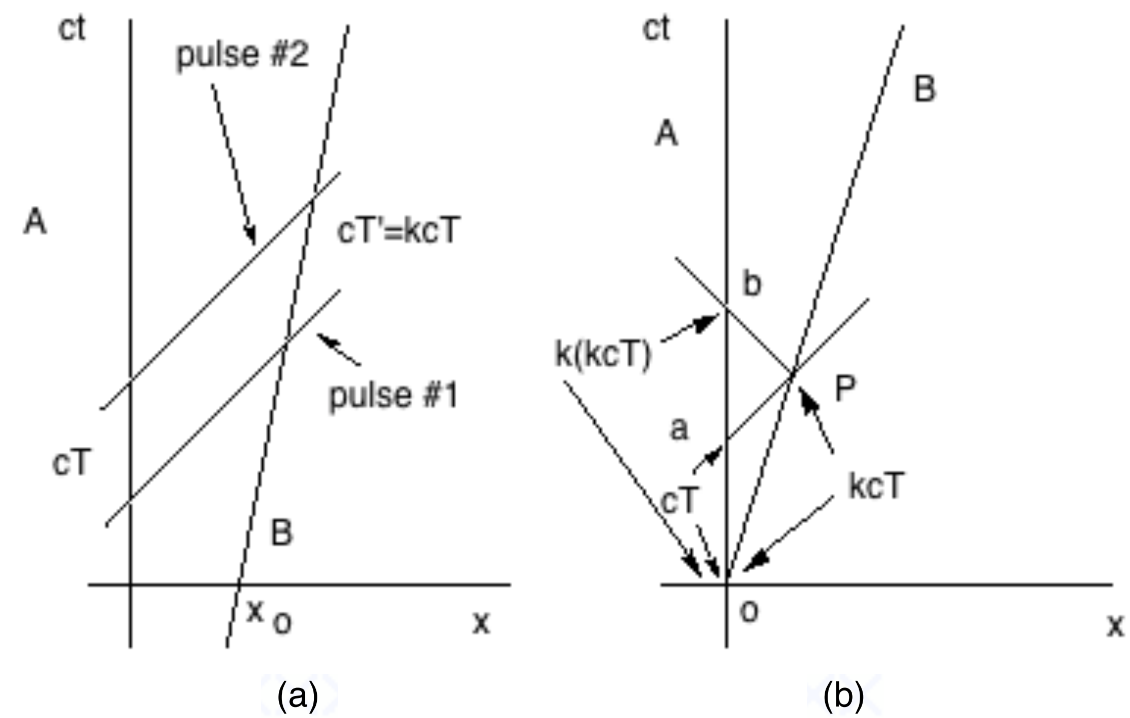
-> relationship between A and B is reciprocal

-> if B emits 2 signals separated by interval cT (according to B's clock),

then A must receive them with interval kcT (according to A's clock).

The last result emphasizes that

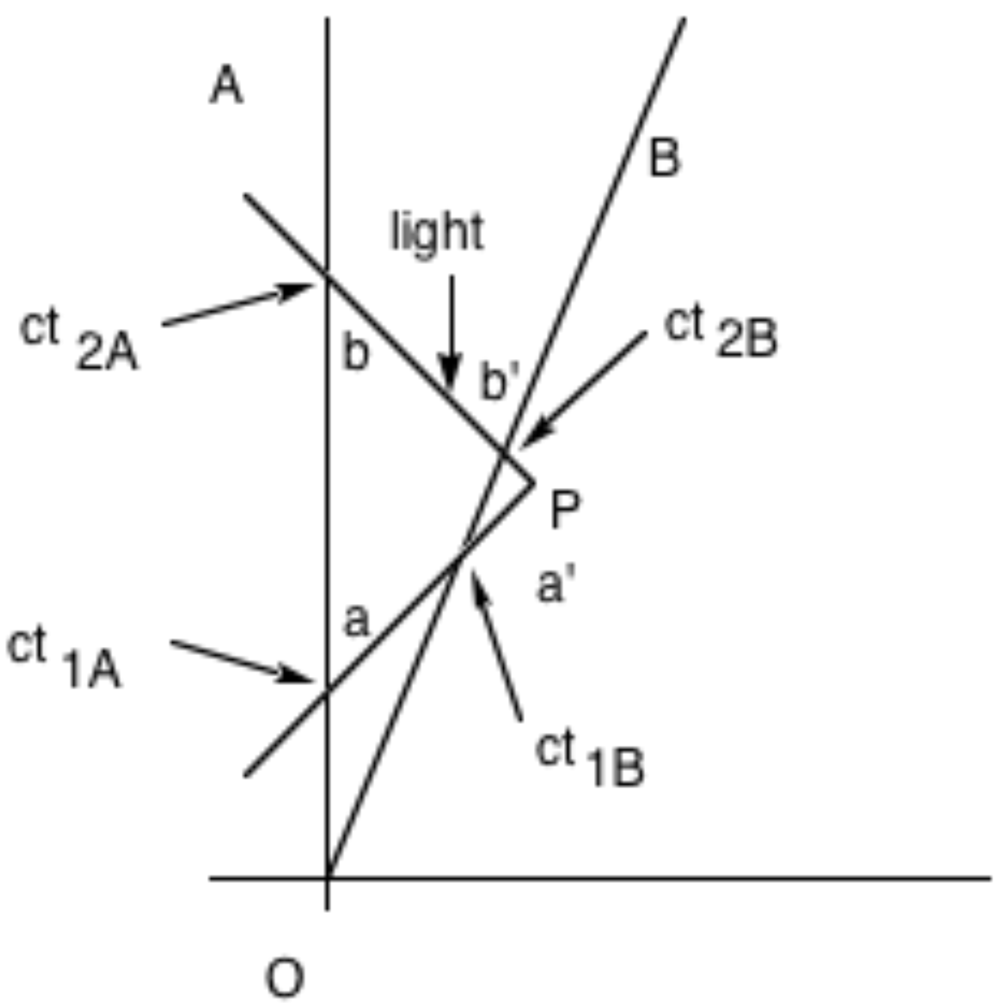
Relativity is truly relative!!



Therefore intervals go like

$$cT \to kcT \to k(kcT)$$

as in previous diagram

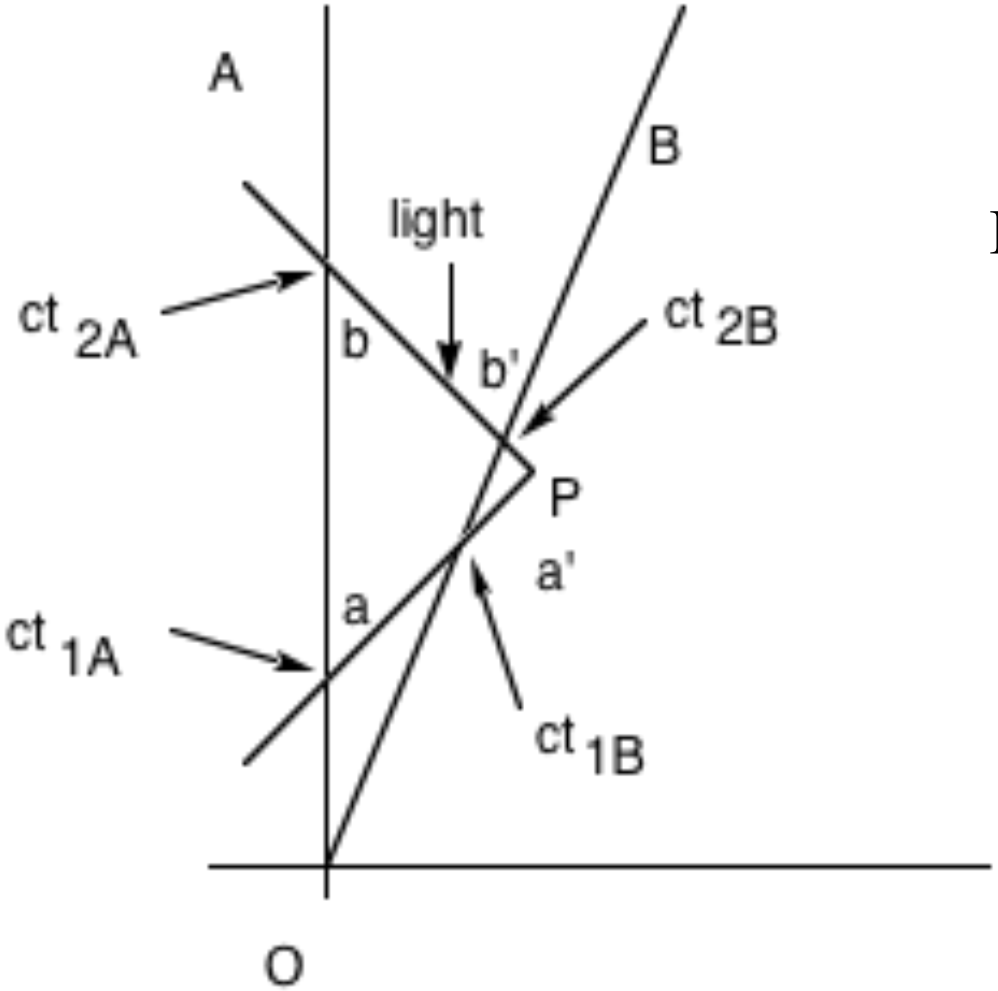


Now consider the experiment shown (2 observers measuring **same** event P). What happens?

A and B synchronize clocks to zero when worldlines cross at event O.

After time t_{1A} =T (according to A) A sends light signal to P -> event a (a is on A's worldline).

B receives light signal at event a$'$ (a$'$ is on B's worldline)



Signal reflected back to A from event P.

B receives reflected signal at event b′ (b′ is on B's worldline).

A receives reflected signal at event b (b is on A's worldline).

For event P observer, A(using radar method) ——->

$$x_P = \frac{c(t_{2A} - t_{1A})}{2} \quad , \quad ct_P = \frac{c(t_{2A} + t_{1A})}{2}$$

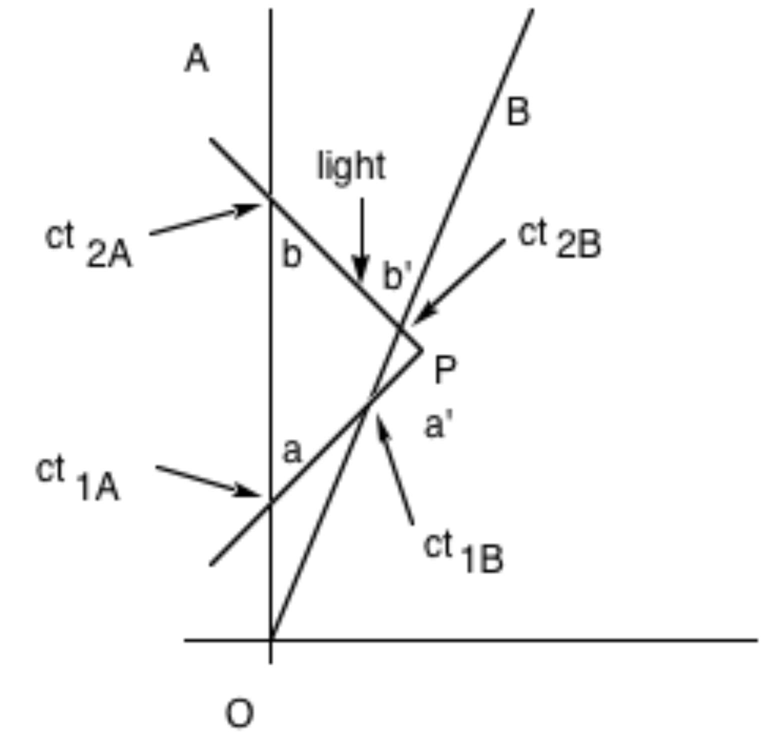
Observer B (using radar method) -> (same experiment and same equations for both A and B)

$$x'_P = \frac{c(t_{2B} - t_{1B})}{2} \quad , \quad ct'_P = \frac{c(t_{2B} + t_{1B})}{2}$$

Our assumption !!

Then, using $\Delta x = c\Delta t$ and $\Delta x' = c\Delta t'$ (assumption 1) we have that

$$c(t_{2A} - t_P) = x_P = c(t_P - t_{1A}) \text{ and } c(t_{2B} - t'_P) = x'_P = c(t'_P - t_{1B})$$



or

$$ct_{2A} = ct_P + x_P \text{ and } ct_{2B} = ct'_P + x'_P$$

$$ct_{1A} = ct_P - x_P \text{ and } ct_{1B} = ct'_P - x'_P$$

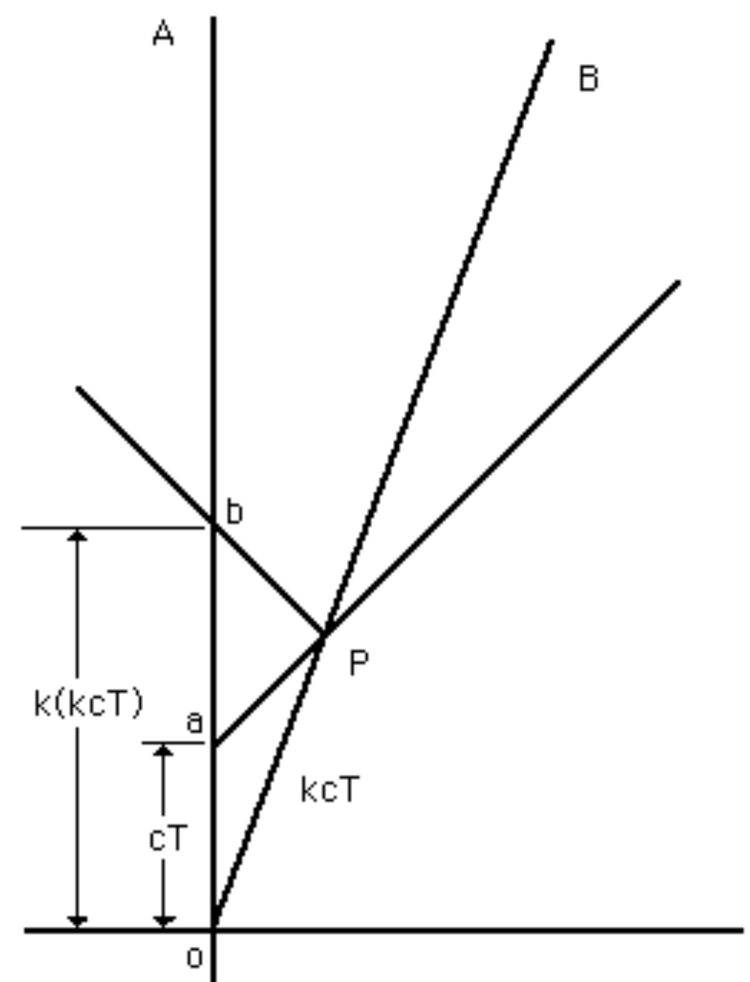
From earlier experimental results we have

$$ct_{1B} = kct_{1A} \text{ and } ct_{2A} = kct_{2B}$$

as shown

i.e., for observer B, interval OP = kcT (according to B's clock)

and for observer A, interval Ob = k(kcT) (according to A's clock).



Therefore, A has sent out signal to event P at $ct_{1A} = cT$

and received it back at $ct_{2A} = k^2 cT$.

Putting all stuff together + doing some algebra we get

$$ct'_P + x'_P = \frac{ct_P + x_P}{k}$$

$$ct'_P - x'_P = k(ct_P - x_P)$$

Further algebra then gives

(dropping subscript P - nothing special about that particular spacetime point)

and using value of k from our assumptions we get

$$ct' = \gamma(ct - \beta x) \text{ and } x' = \gamma(x - \beta ct) \qquad \text{where} \qquad \beta = \frac{v}{c} \quad , \quad \gamma = \sqrt{\frac{1}{1-\beta^2}}$$

——> **Lorentz transformations**.

These transformations allow 2 observers to relate their experimental results

—> they are translators between experiments done in different frames moving relative to each other with constant velocity in common x(x′) direction.

Note that for relative motion in x−direction (above) y and z coordinates are unchanged, i.e.,

$$y' = y \quad , \quad z' = z$$

so that we have the full set of Lorentz transformation relations

$$ct' = \gamma(ct - \beta x) \qquad x' = \gamma(x - \beta ct) \qquad y' = y \qquad z' = z$$

1st - note that as v $\rightarrow$ 0, $\gamma \rightarrow 1$ —> no difference between observers A and B

(correct because they will then be at rest relative to each other).

Also note the **mixing** of space and time so that they are **no longer independent of other**.

That is a very dramatic occurrence.

So

the principle of relativity together with two experimental results

allows us to derive these new relations

which constitute basic equations of theory of special relativity.

That is way theoretical physics works.

Take mixture of general principles (no one can argue with) **and** experimental results and create a set of assumptions about way world works.

Then derive consequences of assumptions -> Lorentz Transformations in this case.

Then have theory that agrees with our assumptions (will show that later).

If theory represents new paradigm in physics then it should be able to make new predictions, not related to assumptions, that agree with **all** future experiments.

Can make immediate prediction that nothing can travel faster than light.

Look at form of the γ–factor.

$$\beta = \frac{v}{c} \quad , \quad \gamma = \sqrt{\frac{1}{1-\beta^2}}$$

If it is possible for v > c, then 1 observer could measure two events separated by **real** time and space intervals while the 2nd observer would have to measure **imaginary(????)** intervals.

Since this has **never been observed** to happen

can confidently **predict** that all objects **must have v < c**

so that γ is **always real**.

This is corroborated by all known experiments.

This correct prediction then encourages the theorist to proceed further and see what other interesting features are lurking about.

Features of the Theory

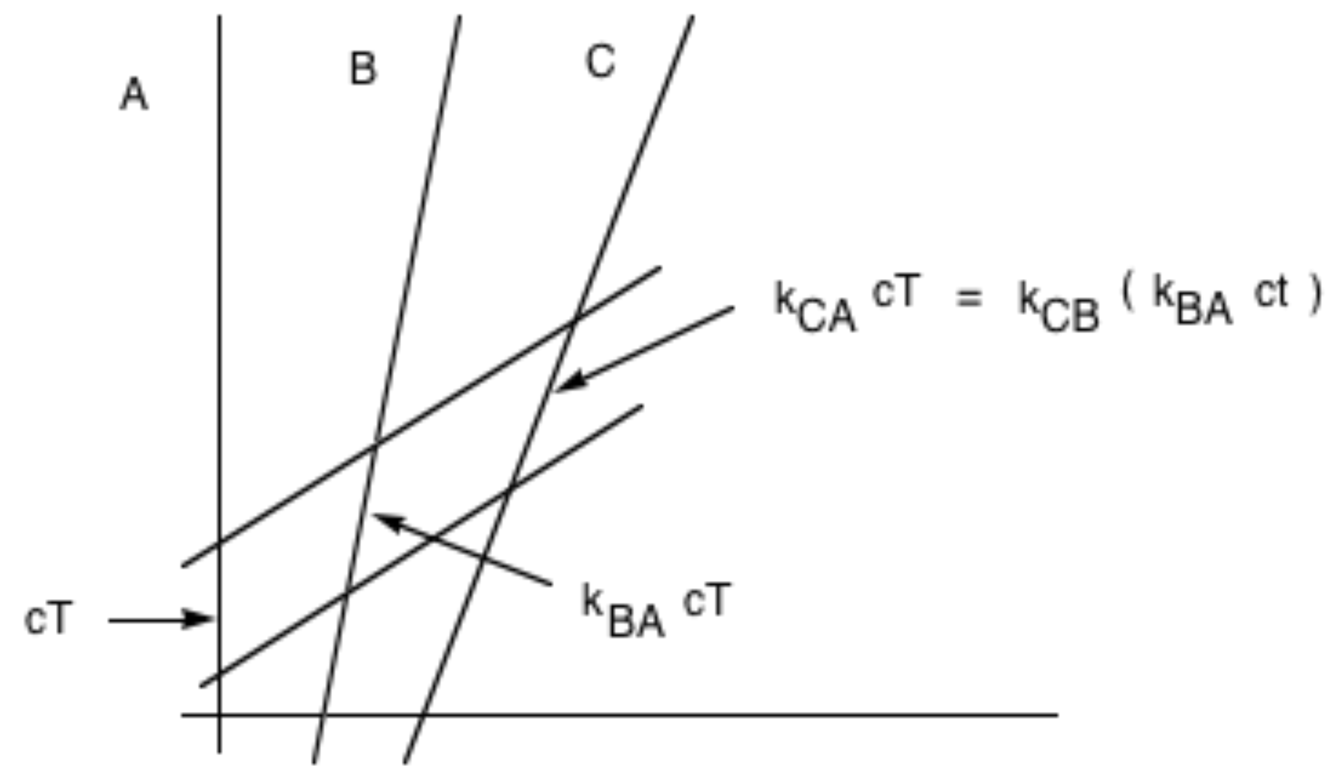
Suppose we have 3 observers A, B and C where

velocity of B relative to A is $v_{BA} > 0$

velocity of C relative to A is $v_{CA} > 0$

velocity of C relative to B is $v_{CB} > 0$.

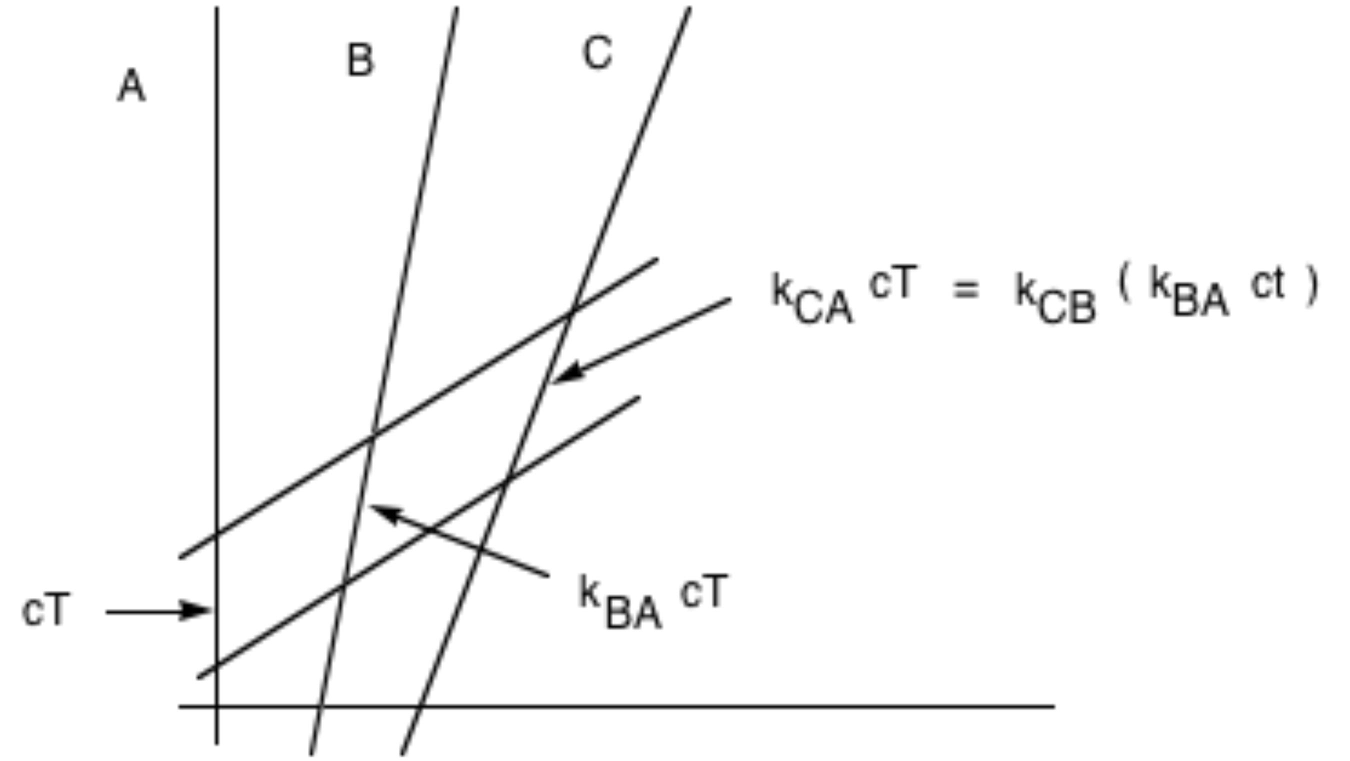
see diagram



A sends out 2 light signals

separated by interval cT (according to A)

that are received by both B and C (as shown).



Previous discussions —> B thinks interval between signals is $k_{BA}cT$ and C is $k_{CA}cT$, where

$$k_{BA} = \sqrt{\frac{c + v_{BA}}{c - v_{BA}}} \quad , \quad k_{CA} = \sqrt{\frac{c + v_{CA}}{c - v_{CA}}}$$

each uses their appropriate relative velocity in their calculation!

Similarly, C could assume that signals came from B and not A —> interval is $k_{CB}(k_{BA}cT)$, where

$$k_{CB} = \sqrt{\frac{c + v_{CB}}{c - v_{CB}}}$$

But 2 results must be identical(same interval according to C) —> must have

$$k_{CA}cT = k_{CB}(k_{BA}cT) \rightarrow k_{CA} = k_{CB}k_{BA}$$

—> relativistic velocity addition formula.

Converting back to velocities we have

$$v_{CA} = \frac{v_{CB} + v_{BA}}{1 + \frac{v_{CB}v_{BA}}{c^2}}$$

If all velocities are **small** compared to c, then we have

$$v_{CA} = \frac{v_{CB} + v_{BA}}{1 + \frac{v_{CB} v_{BA}}{c^2}} \to v_{CB} + v_{BA}$$

If we let

$$v_{CA} = v \quad , \quad v_{CB} = v' \quad , \quad v_{BA} = u$$

—-> It reduces back to Newton-Galileo result for v « c, as it must, i.e.,

$$v = v' + u \to v' = v - u$$

Finally, and most amazingly, if v_{CB} = c (i.e., B is now looking at light signal (C is light)) and v_{BA} = u (B is moving relative to A), then find

$$v_{CA} = \frac{v_{CB} + v_{BA}}{1 + \frac{v_{CB} v_{BA}}{c^2}} = \frac{u + c}{1 + \frac{uc}{c^2}} = c$$

-> prediction (also a verification of our assumption):

If **one** observer measures something moving with speed of light c

then **all** observers will also measure its speed to be c.

In new picture, space and time merge into a new 4-dimensional continuum.

Most important variables in any theory are those that are unchanged for different observers

-> **invariants**.

Speed of light = **invariant**.

Another very important invariant = spacetime interval constructed as follows.

Observers A and B independently measure spacetime coordinates for 2 events

$$\text{Observer A: } (ct_{A1}, x_{A1}, y_{A1}, z_{A1}) \text{ and } (ct_{A2}, x_{A2}, y_{A2}, z_{A2})$$

$$\text{Observer B: } (ct_{B1}, x_{B1}, y_{B1}, z_{B1}) \text{ and } (ct_{B2}, x_{B2}, y_{B2}, z_{B2})$$

Lorentz transformations relate coordinates by

$$ct_{B1} = \gamma(ct_{A1} - \beta x_{A1}) \; , \; x_{B1} = \gamma(x_{A1} - \beta ct_{A1}) \; , \; y_{B1} = y_{A1} \; , \; z_{B1} = z_{A1}$$

$$ct_{B2} = \gamma(ct_{A2} - \beta x_{A2}) \; , \; x_{B2} = \gamma(x_{A2} - \beta ct_{A2}) \; , \; y_{B2} = y_{A2} \; , \; z_{B2} = z_{A2}$$

Spacetime interval for observer **defined** for 2 events by

$$(\Delta s)^2 = c^2(\Delta t)^2 - (\Delta x)^2 - (\Delta y)^2 - (\Delta z)^2$$

Can now show (we do in detail later) using Lorentz transformations that corresponding spacetime intervals for 2 observer for 2 events above

$$(\Delta s_A)^2 = c^2(t_{A2} - t_{A1})^2 - (x_{A2} - x_{A1})^2 - (y_{A2} - y_{A1})^2 - (z_{A2} - z_{A1})^2$$

$$(\Delta s_B)^2 = c^2(t_{B2} - t_{B1})^2 - (x_{B2} - x_{B1})^2 - (y_{B2} - y_{B1})^2 - (z_{B2} - z_{B1})^2$$

are **invariant** or $(\Delta s_A)^2 = (\Delta s_B)^2$ **—> powerful consequences**

Before attempting to look at the actual construction of detailed spacetime diagrams, it will be useful for me go back and rederive the spacetime interval idea from a very different direction to give us a better understanding.

THREE KINDS OF TIME

A firecracker explodes.

Some time later and somewhere else, another firecracker explodes.

How much time passes between these two events?

How can we measure that time interval?

In a newtonian universe, such questions would be easy to answer.

We might measure the time between the events with a pair of synchronized clocks, one present at each event.

We might instead measure the time between the events with a single clock that travels in such a manner that it arrives at each event's location just as it happens.

Since all clocks in a Newtonian universe register the same universal, absolute time, these methods (and any of a number of other valid methods) will yield the same result.

It is unimportant what method is actually used.

But earlier we argued that in the real universe, universal and absolute time does not exist, and therefore the problem of measuring the time interval between two events is somewhat more problematic.

In this discussion we will discover that there are three fundamentally different ways to measure the time interval between two events in the theory of special relativity and that these different methods yield different results, even if applied to the same two events!

Because it is important in the real universe to distinguish between the various methods used to measure the time interval between two events, we refer to the time interval determined by each method using a special technical name: we speak of the coordinate time, the proper time, and the spacetime interval between the events.

The purpose of this discussion is to describe these distinct ways of measuring time and begin to uncover the relationships between them.

THE COORDINATE TIME BETWEEN TWO EVENTS

Once the clocks in an inertial reference frame have been satisfactorily synchronized, we can use them to measure the time coordinates of various events that occur in that frame.

In particular, we can measure the time between two events A and B in our reference frame by subtracting the time read by the clock nearest event A when it happened from the time read by the clock nearest event B when it happened: $\Delta t_{BA} \equiv t_B - t_A$

Note that this method of measuring the time difference between two events requires the use of two synchronized clocks in an established inertial reference frame.

Such a measurement therefore cannot be performed in the absence of an inertial frame.

In summary, the coordinate time between two events in a given frame is defined as follows:

Definition of Coordinate Time

The time measured between two events either by a *pair* of synchronized clocks at rest in a given inertial reference frame (one clock present at each event) or by a *single* clock at rest in that inertial frame (if both events happen to occur at that clock in that frame) is called the **coordinate time** between the events in that frame. The symbol Δt is used to represent the coordinate time between two events.

COORDINATE TIME IS FRAME-DEPENDENT

Imagine that the observer in some inertial reference frame (let us call this one the Other Frame: we will talk about a Home Frame in a bit) sets out to synchronize its clocks.

In particular, let us focus on two clocks ni that frame that lie on the x' axis and are equally separated from the master clock at x' = 0.

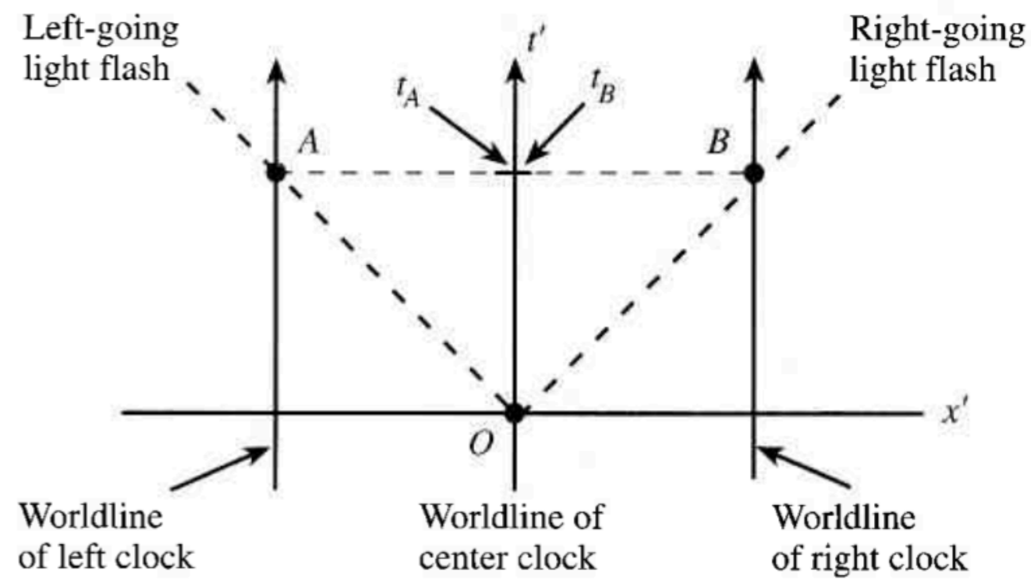
At t' = 0, the observer causes the center clock to emit two flashes of light, one traveling in the +x' direction and the other in the -x' direction.

Let the emission of these flashes from x' = 0 at t' = 0 be called the origin event O.

Now, since both the other clocks are the same distance from the center clock and since the speed of light is 1(light-) second/second in every inertial reference frame, the left-hand clock will receive the left-going light flash (call the event of reception event A) at the same time that the right-hand clock receives the right-going flash (event B).

By the definition of synchronization, both clocks should thus be set to read the same time at events A and B(a time in seconds equal to their common distance from the center clock).

This process is illustrated by the spacetime diagram in the Figure.



The synchronization of two clocks equally spaced from a center clock, as observed in the Other Frame. If the right and left clocks are set to agree at events A and B, they will be synchronized with each other.

Note that since all three clocks are at rest in this frame, their worldlines on the spacetime diagram are vertical.

Moreover, since the speed of light is 1ls/s in this (and every other inertial) frame, the worldlines of the light flashes will have slopes of ±1 on the spacetime diagram (i.e., they will make a 45° angle with each axis) as long as the axes have the same scale.

On this diagram it is clear that events A and B really do occur at the same time in this "other" reference frame.

Now consider a second inertial reference frame (the Home Frame) within which the Other Frame is observed to move in the +x direction at a speed β.

Let us look at the same events from the vantage point of the Home Frame.

For convenience, let us take the event of the emission of the flashes to be the origin event in this frame as well (i.e., event O is taken to occur at $t = x = 0$ in the Home Frame).

The observer in the Home Frame will agree that the right and left clocks in the Other Frame are always equidistant from the center clock in the Other Frame.

Moreover, at $t = 0$, when the center clock passes the point $x = 0$ in the Home Frame as it emits its flashes, the right and left clocks are equidistant from the emission event.

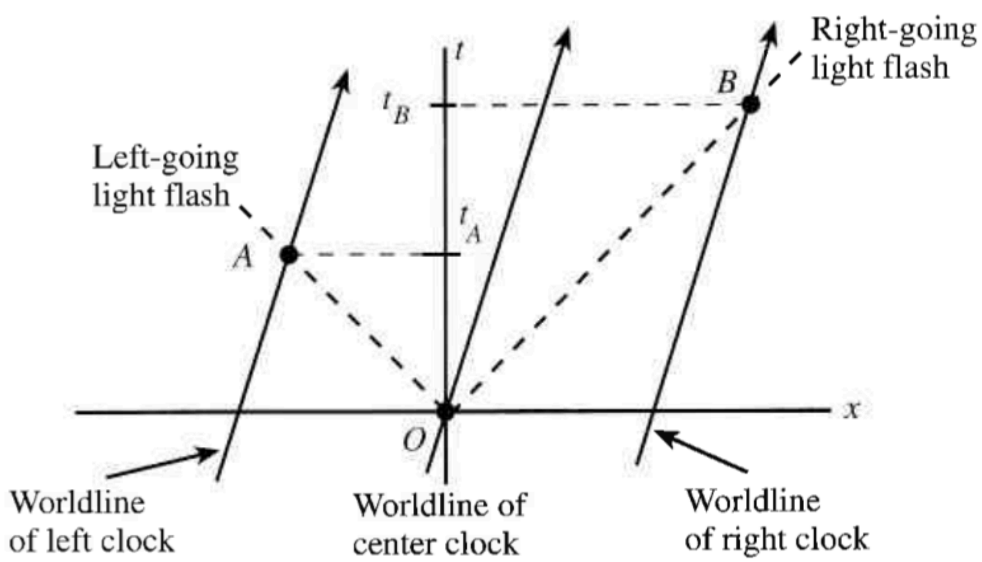
But during the time that the light flashes are moving to the outer clocks, the observer in the Home Frame observes the left clock to move up the x axis toward the flash coming toward it and the right clock to move up the x axis away from the flash coming toward it.

As a result, the left-going light flash has less distance to travel to meet the left clock than the right-going flash does to meet the right clock.

Since the speed of light is 1 in the Home Frame as well as in the Other Frame, this means that the left clock receives its flash first.

Therefore, event A is observed to occur before event B in the Home Frame.

A spacetime diagram of the process as observed in the Home Frame is shown in the Figure.



The same events as observed in the Home Frame. In this frame, event Bis measured to occur after event A.

In drawing the worldlines of the clocks in question, it is important to note that the clocks are not at rest in the Home Frame.

Their worldlines on a Home Frame spacetime diagram will be equally spaced lines with slopes of $1/\beta$ (we will see why shortly), indicating that the clocks are moving ot the right at a speed β.

The light flashes have a speed of 1 ls/s in the Home Frame (and every other frame), so their worldlines are drawn with a slope of ±1 on the spacetime diagram.

In summary, the coordinate time between events A and B as measured in the Other Frame is $\Delta t' = 0$ (by construction in this case), but the coordinate time between these events as measured in the Home Frame is $\Delta t \neq 0$.

We see that the coordinate times between the same two events measured in different reference frames are not generally equal.

Thus coordinate time differences are said to be relative (i.e., they depend on one's choice of inertial reference frame).

Why?

If each observer synchronizes the clocks in his or her own reference frame according to our definition, each will conclude that the clocks in the other's frame are not synchronized.

Notice that the Other Frame observer has set the right and left clocks to read the same time at events A and B.

Yet these events do not occur at the same time in the Home Frame.

Therefore the observer in the Home Frame will claim that the clocks in the Other Frame are not synchronized.

Of course, the observer in the Other Frame will make the same claim about the clocks in the Home Frame.

The definition of synchronization that we are using makes perfect sense within any inertial reference frame, but it does not allow us to synchronize clocks in different inertial frames.

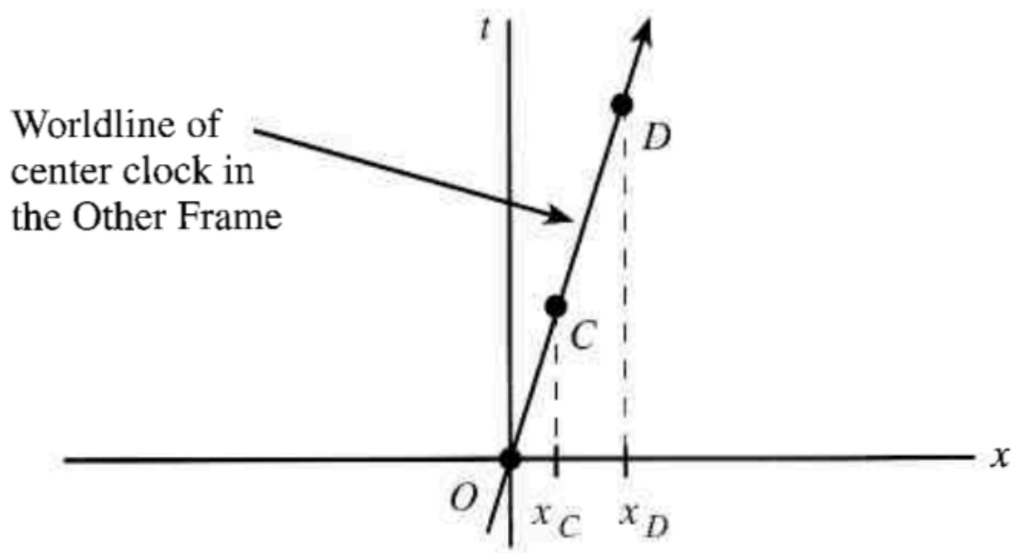
In fact, the definition requires that observers in different inertial frames measure different time intervals between the same two events, as we have just seen.

In general, two observers in different frames will also disagree about the spatial coordinate separation between the events.

Consider two events C and D that both occur at the center clock in the Other Frame, but at different times.

Since the center clock defines the location x' = 0 in the Other Frame, the events will be measured to have the same x' coordinate in that frame, implying that $\Delta x' = 0$.

But in the Home Frame, the center clock is measured to move in the time between the events, and so the two events do not occur at the same place: $\Delta x \neq 0$ (see Figure).



Events C and D both occur at the center clock in the Other Frame; therefore $\Delta x' = 0$ in that frame. But the center clock is measured to move with respect to the Home Frame between the events; therefore $\Delta x \neq 0$ in the Home Frame. The spacetime diagram shows events as they would be plotted by the Home Frame observer.

The point here is that the *coordinate differences between two events are frame-dependent quantities*, whether they are time coordinate differences or space coordinate differences: they will in general be measured to have different values in different inertial reference frames.

THE RADAR METHOD YIELDS THE SAME RESULTS

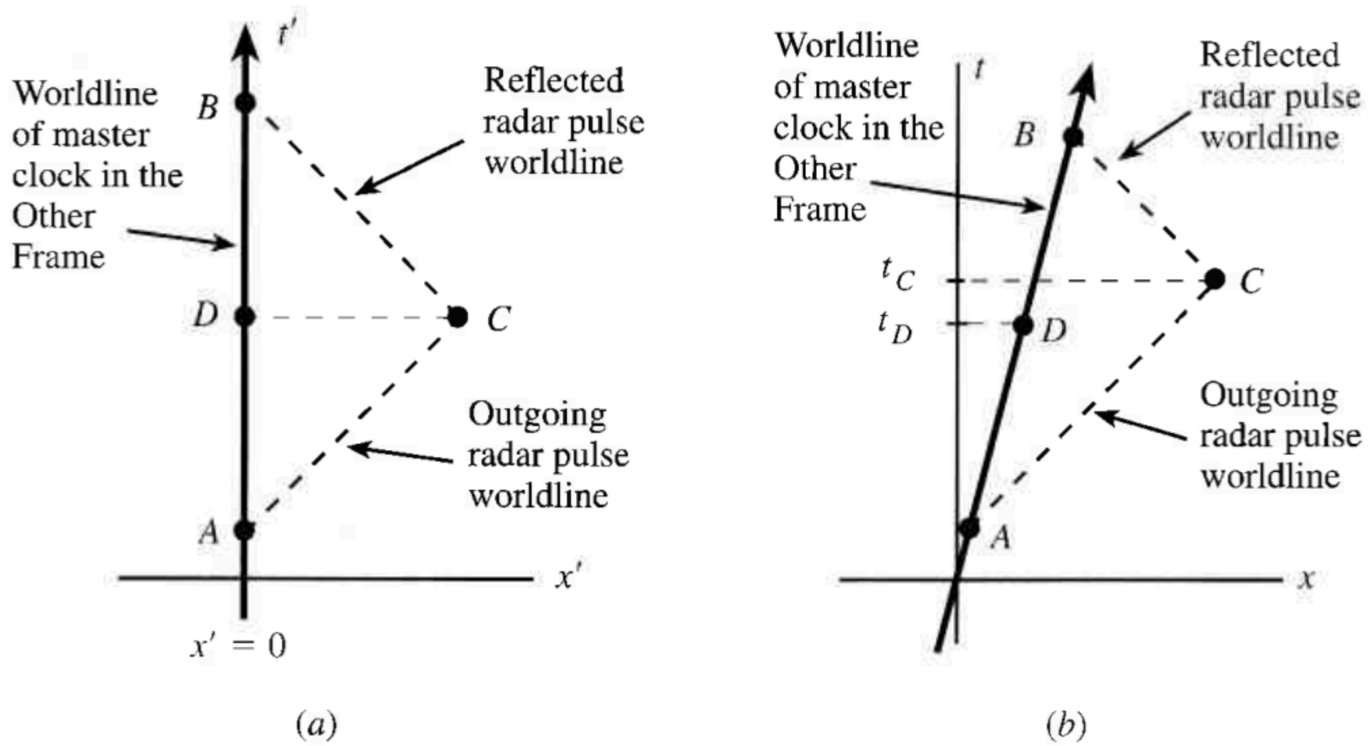
The basic reason why observers in different inertial frames disagree about whether the clocks in a given frame are synchronized or not is that synchronization is defined so that light flashes are measured to have a speed of 1 in every inertial frame: the frame dependence of coordinate time differences is a logical consequence of this assertion.

This can be illustrated by considering the radar method of assigning spacetime coordinates.

Though the radar method does not involve the use of synchronized clocks, it does depend on the assumption that the speed of light is the same in every inertial frame.

Does the radar method also imply that the coordinate time difference between two events is frame-dependent?

The Figure shows that it does.



(a) In the Other Frame, events C and D are simultaneous, where D occurs at the master clock at a time halfway between the emission event A and the reception event B. The coordinate time difference between events C and D in this frame is thus $\Delta t = 0$, according to the radar method. (b) In the Home Frame, the Other Frame's master clock moves to the right as time passes, so its worldline is slanted in a spacetime diagram based on measurements in that frame. On the other hand, radar pulse worldlines still have slope ± 1 as shown. This means that an observer in the Home Frame will conclude that event C happens after event D; the coordinate-time difference between the events is $\Delta t \neq 0$.

Figure Part (a) shows the observer in the Other Frame using the radar method to determine the space and time coordinates of event C.

The observer in that frame will conclude that events C and D occur at the same time, where D is the event of the master clock at x' = 0 reading $t'_D = ½(t'_B - t'_A)$, that is, the instant of time halfway between the emission of the radar pulse at t'_A and its reception at t'_B.

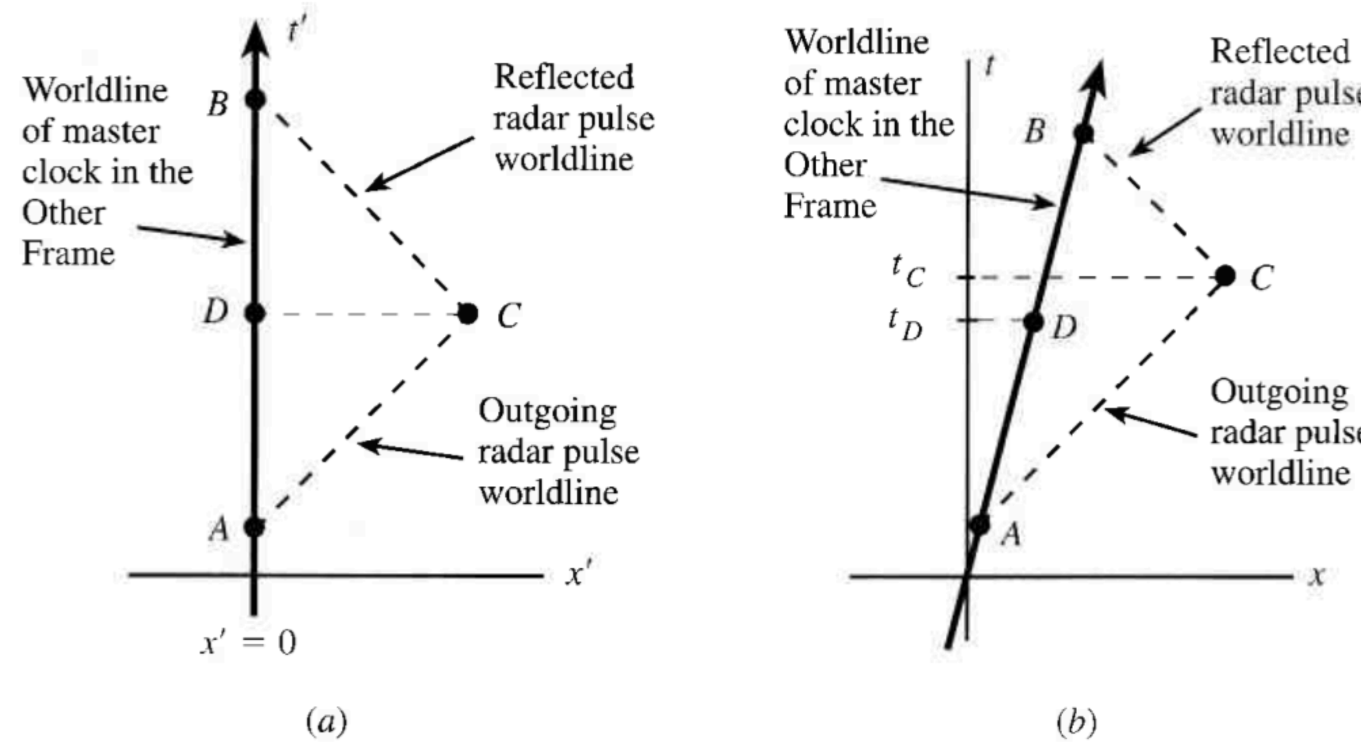
According to the radar method, then, the coordinate time interval between events C and D is $\Delta t' = 0$.

Note: Radar and visible light are both examples of electromagnetic waves: they just have different frequencies.

Both types of waves will thus move at a speed of 1(light-)second / second = 1.

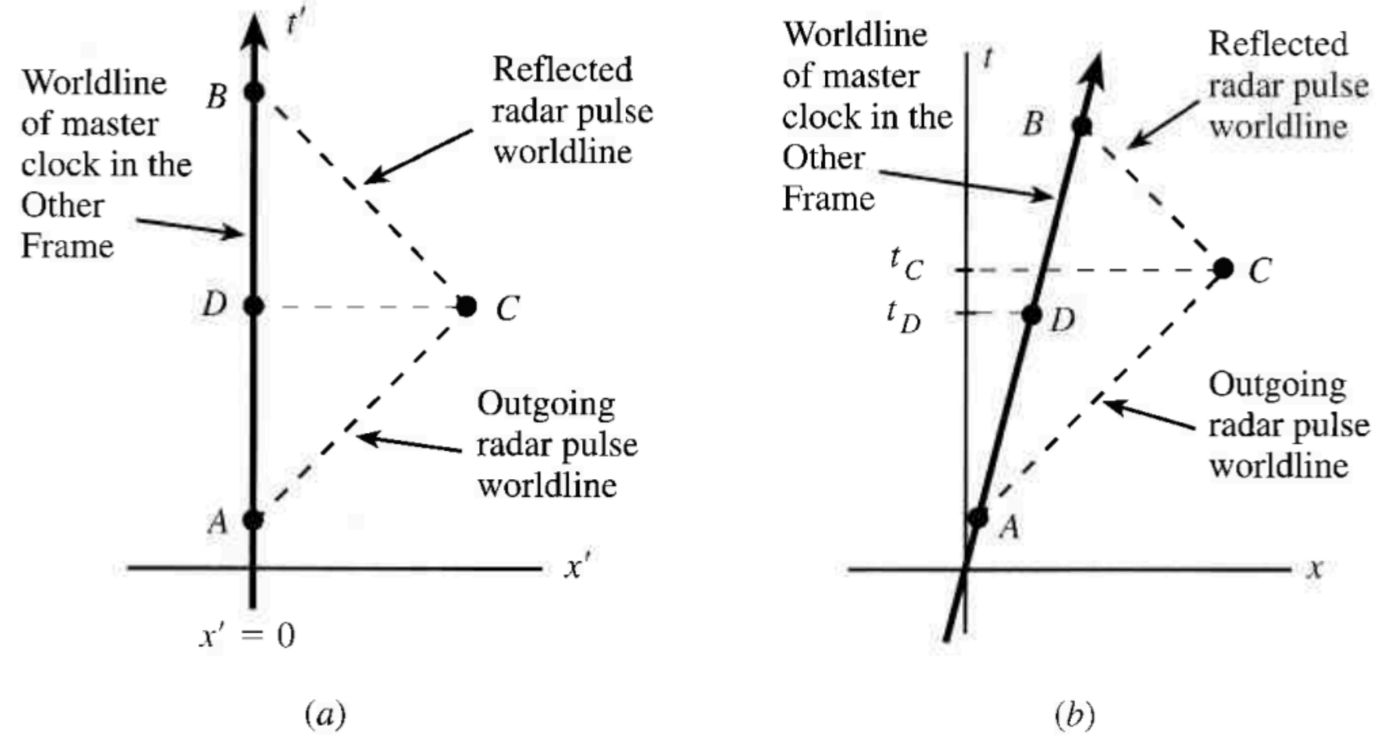
When the same sequence of events is viewed from the Home Frame, though, a different conclusion emerges (see Figure Part(b)).

According to observers in the Home Frame, the Other Frame's master clock is moving along the x axis with some speed β, so in a spacetime diagram based on measurements taken in the Home Frame, the worldline of that clock will appear as a slanted line (slope = $1/\beta$) instead of being vertical.



Radar pulse worldlines, on the other hand, still have slopes of ±1, just as they did in the Other Frame spacetime diagram.

The inevitable result (as you can se from the diagram) is that observers in the Home Frame are forced to conclude that event C occurs after event D does and, therefore, that the time difference between events C and D in the Home Frame is $\Delta t \neq 0$.



The point is that the relativity of coordinate time differences is a direct consequence of the fact that we are **defining** coordinate time by assuming that the speed of light is 1 in every inertial reference frame.

Remember, though, that we must make this assumption if the laws of electromagnetism are to be consistent with the principle of relativity!

A GEOMETRIC ANALOGY

It may be troubling that coordinate differences between events are not absolute but are instead frame-dependent.

This is particularly true of the time coordinate separation: it is not easy to let go of the Newtonian notion of absolute time!

But the fact is that we have no trouble at all with these ideas when they appear in a related but more familiar guise.

Consider the hypothetical town of Askew, North Dakota.

Most towns in rural America have streets that run north-south or east-west, but Askew has a problem.

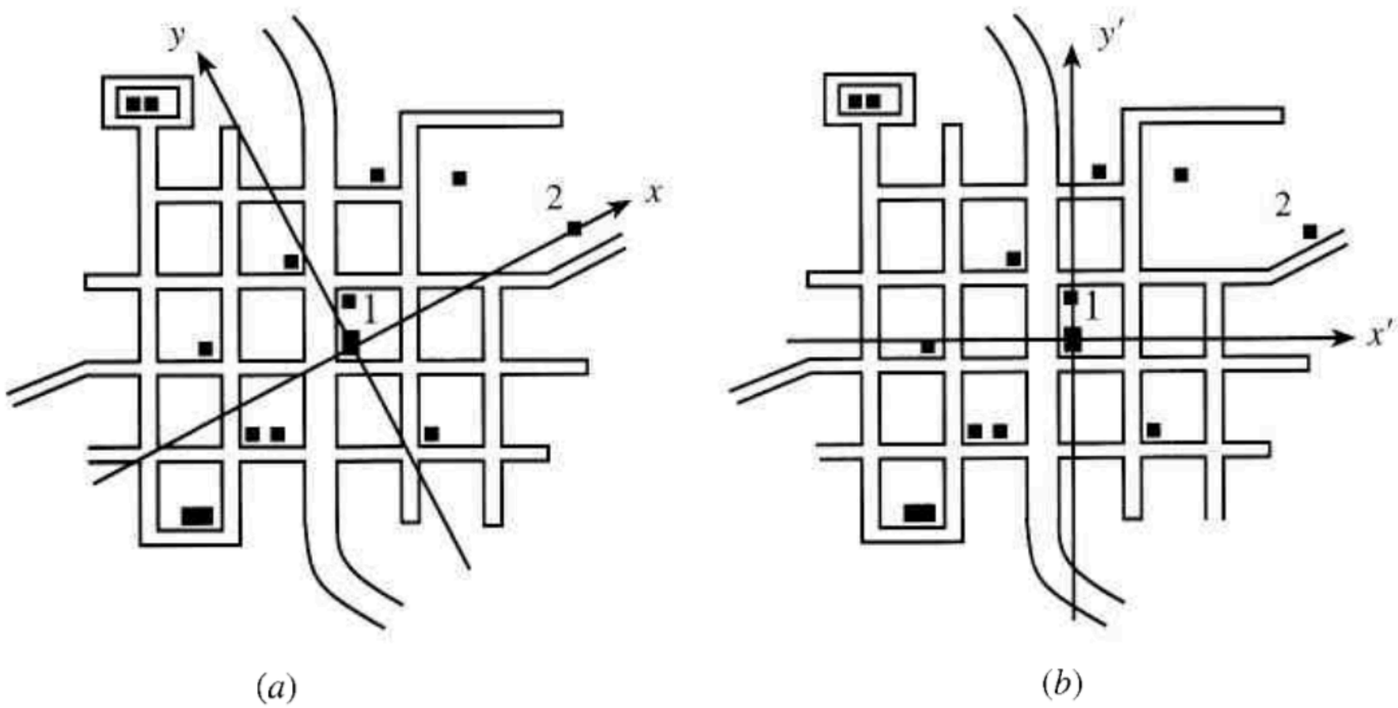
The surveyor who laid out the streets of this town in 1882 tried to calibrate his compass against the North Star the night before but in fact had forgotten exactly where the North Star was (it was a long time since he had this stuff in high school after all), and ended up choosing a star that turned out to be 28° east of the true North Star.

So all the streets of Askew are twisted 28° from the standard directions.

Now, if we would like to assign x and y coordinates to points of interest in this town (or any town), we have to set up a cartesian coordinate system.

It is conventional to orient coordinate axes on a plot of land so that the y axis points north and the x axis points east (see Figure Part (a)). This is usually convenient as well, since the streets will be parallel to the coordinate axes. But there is no reason why this has to be done, and in Askew's case, it is actually more convenient to use a coordinate system oriented 28° to the east (Figure Part (b)).

Note that the origin of both coordinate systems is chosen to be City Hall for the sake of simplicity.

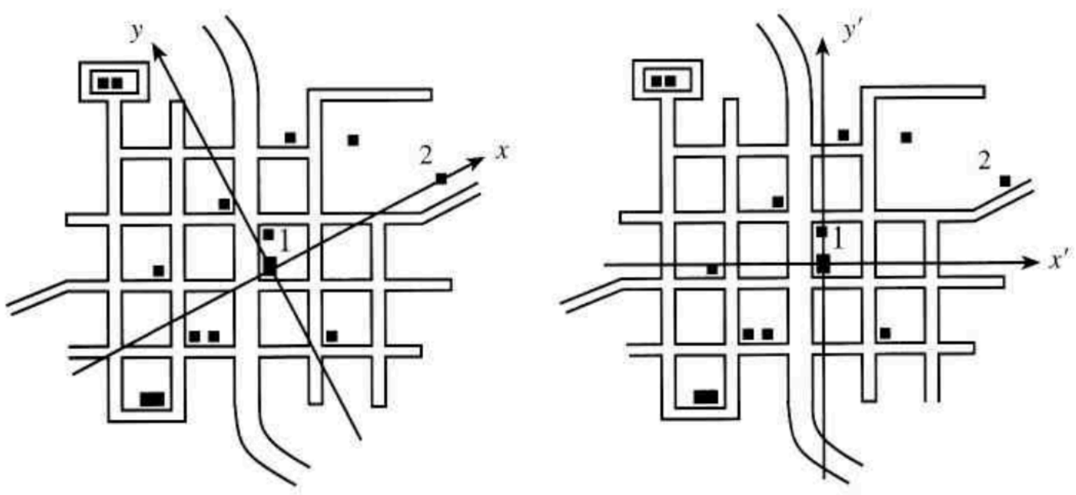


(a) Standard cartesian coordinate system superimposed on the town of Askew. (b) A more convenient coordinate system oriented 28° clockwise. On both maps, 1 is City Hall and 2 is the Statue of the Unknown Physicist.

We can use any coordinate system that we choose to quantify the positions of points of interest in the town: coordinate systems are arbitrary human artifacts that we impose for our convenience on the physical reality of the town.

But the coordinates we actually get for various points certainly do depend on the coordinate system used.

For example, the coordinate differences between City Hall and the Statue of the Unknown Physicist in Memorial Park might be $\Delta y = 0$, $\Delta x = 852.0$ m in the standard coordinate system, whereas in the "convenient" coordinate system, they might be $\Delta y' = 399.9$ m, $\Delta x' = 752.3$ m.



(a) (b)

Is it surprising that the results are different?

Do the differences in the results cause us ot suspect that one or the other coordinate system has been set up incorrectly?

Hardly!

We can accept the fact that both coordinate systems are perfectly correct and legal.

We already know and expect that differently oriented coordinate systems on a plane will yield different coordinate measurements.

This causes us no discomfort: we understand that this is the way that things are.

In an entirely analogous way, we have carefully and unambiguously defined a procedure for setting up an inertial reference frame and synchronizing its clocks.

This definition happens to imply that spacetime coordinate measurements in different frames yield different results.

This should be no more troubling to us than the fact that Askew residents who use different sets of coordinate axes will assign different coordinates to various points in town.

Coordinates have meaning only "relative" to the coordinate system or inertial frame being used.

The only reason that the relativity of time coordinate differences is a difficult idea is that we do not have common experience with inertial reference frames moving with high enough relative speeds to display the difference.

The kinds of inertial frames that we experience in daily life have relative speeds below 300 m/s, or about one-millionth the speed of light.

If for some reason we could only construct cartesian coordinate systems on the surface of the earth that differed in orientation by no more than a millionth of a radian, then we might think of cartesian coordinate differences as being "universal and absolute" as well!

So, to summarize, the coordinate differences between points on a plane (or events in spacetime) are "relative" because coordinate systems (or inertial reference frames) are human artifacts that we impose on the land (or spacetime) to help us quantify that physical reality by assigning coordinate numbers to points on the plane (or events in spacetime).

Because we are free to set up coordinate systems (or reference frames) in different ways, the coordinate differences between two points (or events) reflect not only something about their real physical separation but also something about the artificial choice of coordinate system (or reference frame) that we have made.

So, is it true then that everything is relative?

Is there nothing that we can measure about the physical separation of the points (or events) that is "absolute," i.e., independent of reference frame?

There is in fact a coordinate independent quantity that describes the separation of two points on a plot of land: the distance between those two points.

For example the distance between City Hal and the Statue of the Unknown Physicist in Askew, North Dakota, is $\Delta d = \sqrt{\Delta x^2 + \Delta y^2} = \sqrt{(852.0m)^2 + 0}$ = 852.0m in the standard coordinate system and $\Delta d' = \sqrt{(\Delta x')^2 + (\Delta y')^2} = \sqrt{(399.9m)^2 + (752.3m)^2}$ = 852.0m in the convenient coordinate system.

It does not matter what coordinate system you use to calculate the distance: you always will get the same answer.

The distance between two points on a plot of ground thus reflects something that is deeply real about the nature of the plot of ground itself, without reference to the human coordinate systems imposed on it.

This independence from coordinate systems arises because there is in fact a method of determining the distance between two points without using a coordinate system: lay a tape measure between the points!

No coordinate system is required to do this.

And since this method yields a certain definite result for the distance, calculations of this distance in any coordinate system should yield the same result if they are valid.

Of course, there are many ways that one could lay a tape measure between City Hall and the Statue of the Unknown Physicist.

One could lay the tape measure along a straight path between the two points: this would measure the distance "as the crow flies," which is what is usually meant by the phrase "the distance between two points."

But one might also lay the tape measure along other paths between the two points.

One might, for example, lay the tape measure two blocks down Elm Street from City Hall, then one block over along Grove Avenue, then up Maple Street, and so on.

This would measure a different kind of distance between the two points that we might call a *path length*.

Both the straight-line distance and the more general pathlength between two points can be measured directly with a tape measure and thus are quantities independent of any coordinate system.

But the distance and the pathlength between two points may not be the same. In general, the pathlength between two points will depend on the path chosen and will always be greater than (or at best equal to) the straight-line distance.

To summarize, we have at our disposal three totally different ways to quantify the separation of two points on a plane.

We can measure the coordinate separations between the points using a coordinate system.

The results will depend on one's choice of coordinate system.

We can measure the pathlength between the points with a tape measure laid along a specified path.

The result here will depend on the path chosen but is independent of coordinate system.

Or we can measure the distance between the points with a tape measure laid along a special path: a straight line.

Because in this last case the path of the tape is uniquely specified, the distance between two points is a unique number that quantifies in a very basic way the separation of the points in space.

THE PROPER TIME AND THE SPACETIME INTERVAL

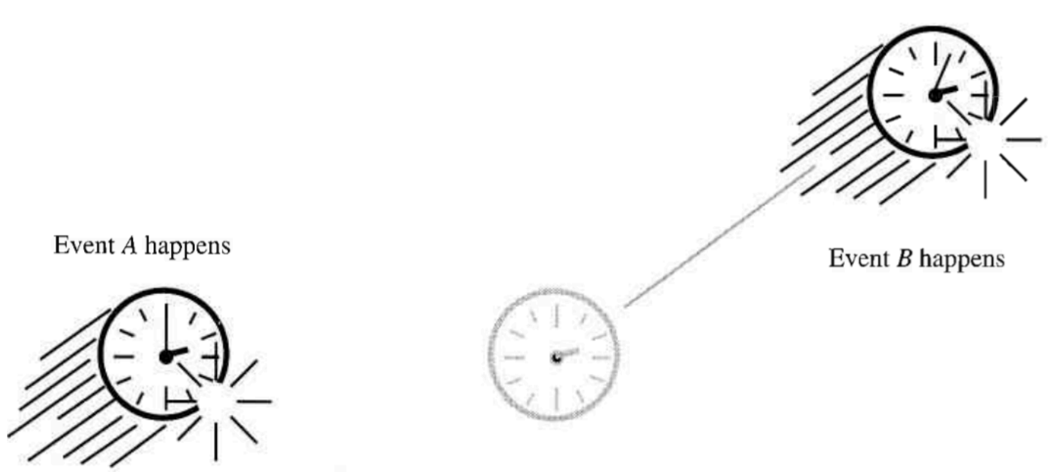
Consider two events.

Label them A and B.

Is there any way that we can measure the time between events A and B without using a reference frame, analogous to the way that we can measure the pathlength between two points on the plane without using a coordinate system?

We can avoid the use of a reference frame lattice if we measure the time interval between these events with a clock that is present at both events.

In a manner analogous to laying a tape measure between two points so that it passes close by each point, we send a clock between the events along just the right path so that it is very close to each event as it occurs (see Figure).



Measuring the time between two events with a clock that is present at both events (events A and B here are represented as firecracker explosions).

A tape measure stretched between two points marks off the distance between those points and presents a scale that can be laid right next to the two points for easy and unambiguous reading.

In an entirely similar manner, a clock that travels between two events marks off the time between those events and presents its face at each event for easy and unambiguous reading.

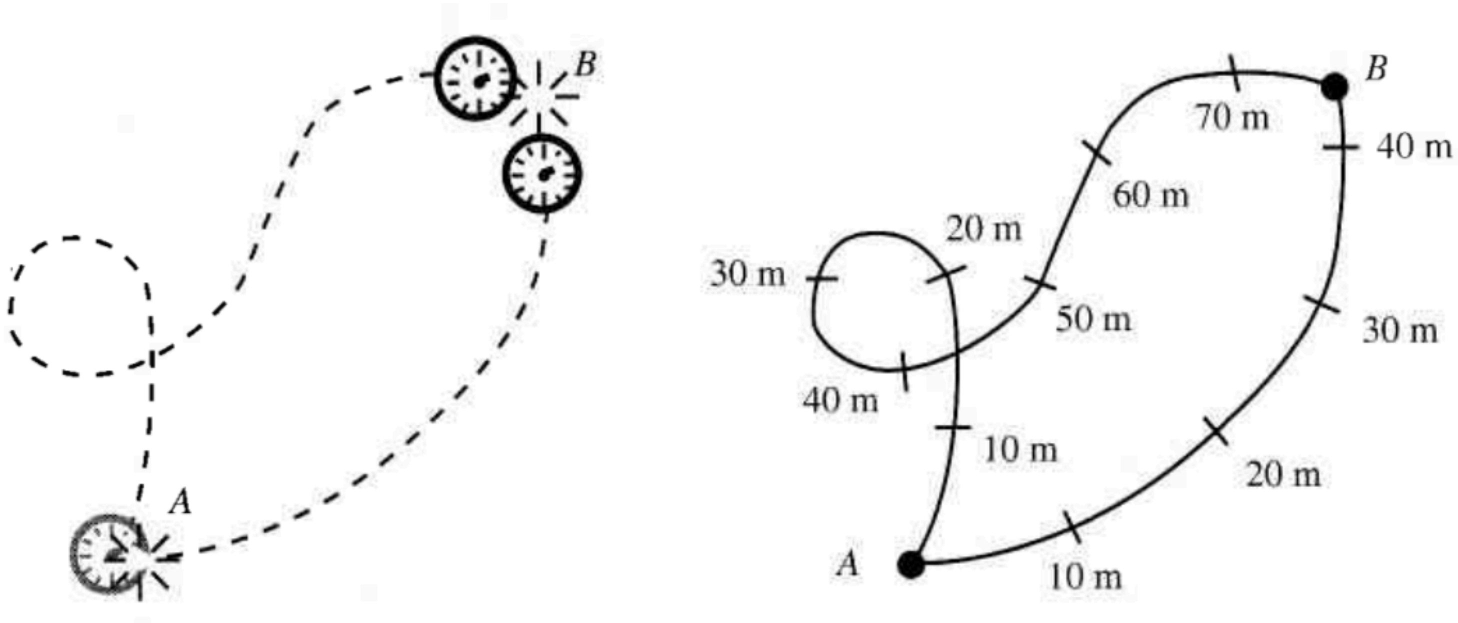
Since the clock's face is right there at each event, everyone looking at the clock will agree as to its reading as each event happens.

The quantity measured by this clock is therefore frame-independent: it is measuring some thing basic about the absolute physical relationship between the events.

Definition of Proper Time

The time between two events measured by any clock present at both events is called a **proper time**[1] between those events. We will use the symbol $\Delta\tau$ to represent a proper time between two events. A proper time measured by a given clock is an absolute quantity independent of reference frame.

But there is one thing that the proper time between two events might conceivably depend on other than the events themselves: it might depend on the worldline that the clock takes in traveling from one event to the other, just as a pathlength measured by a tape measure depends on the path along which ti is laid (see Figure).



The proper time measured between two events (diagram at left may depend on exactly how the clock travels between those events, just as the pathlength between two points (right) depends on the path along which it is measured

As we have seen earlier in our discussion from direct calculations the path dependence of proper time is a straightforward consequence of the principle of relativity.

Many experiments have been done and the point is that it is not merely possible that the time that a clock measures depends on the nature of its worldline, it is an established experimental fact.

When measuring distances on the plane, we distinguish between the pathlength between two points measured along a certain path and the distance between the points: the distance is measured along the special path that is the straight line between the two points.

Because the straight-line path is unique, the distance between two points along a straight line is a unique number that reflects something definite and unambiguous about the separation of those points in space.

Similarly, an inertial clock (a clock whose attached first-law detector registers no violation of Newton's first law) follows a unique and well-defined worldline through spacetime between two events.

Observed in an inertial reference frame, a clock would travel between the events in a straight line at a constant velocity.

Such a worldline defines a unique path ni space, and since there is only one value of a constant velocity that will be just right to get the clock from one event to the other, the clock's velocity along that worldline is also uniquely specified.

Definition of Spacetime Interval

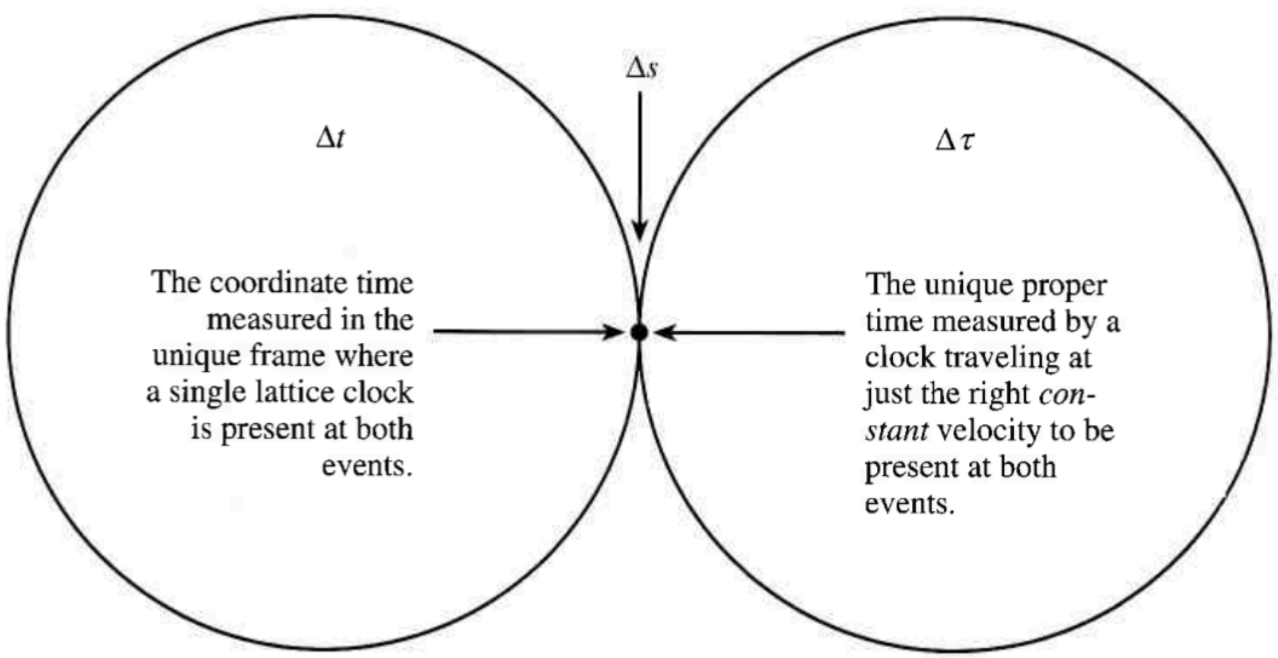
The **spacetime interval** between two events is defined to be the proper time measured by an *inertial* clock that is present at both events. This quantity is a unique, frame-independent number that depends on the separation of the events in space and time and *nothing else*. The symbol Δs is conventionally used to represent the spacetime interval between two events.

Now, it is important to note that the definitions of coordinate time, proper time, and the spacetime interval between two events overlap in certain special cases.

The definition above makes it clear that the spacetime interval between two events is a special case of a proper time between two events, just as the distance between two points is a special case of the more general pathlength between two points.

An inertial clock present at both events also measures the coordinate time between those events in the clock's own reference frame, since the time interval measured between two events by a clock or clocks at rest in any inertial reference frame is a coordinate time by definition.

So the spacetime interval between two events is a special case of a proper time and a special case of a coordinate time (see Figure).



Let various points in the left circle represent the various coordinate times that observers in inertial frames moving at various different velocities might measure between two events. Let various points in the right circle represent the various proper times measured between the events by clocks present at both events but moving between the events along various different worldlines. The single point in common between these sets of possible time intervals between the events is the spacetime interval As between the events.

Tables organize and summarize these ideas.

THE THREE KINDS OF TIME

	Coordinate time	Proper time	Spacetime interval
Definition	The time between two events measured in an inertial frame by a *pair of synchronized clocks,* one present at one event, the other present at the other event. (If both events happens to occur at the same place, a single clock suffices.)*	The time between two events as measured by a *single clock that is present at both events.* (Its value depends on the worldline that the clock follows in getting from one event to the other.)	The time between two events as measured by an *inertial* clock that is present at both events. (Because an inertial clock follows a unique worldline between the events, the spacetime interval is unique.)
Conventional symbol	Δt	$\Delta \tau$	Δs
Is value frame-independent?	No	Yes	Yes
Geometric analogy	Spatial coordinate differences	Pathlength	Distance

THE GEOMETRIC ANALOGY

Plane geometry	Spacetime geometry
Map	Spacetime diagram
Points	Events
Paths or curves	Worldlines
Coordinate systems	Inertial reference frames
Relative rotation of coordinate systems	Relative velocity of inertial reference frames
Differences between coordinate values	Differences between coordinate values
Pathlength along a path	Proper time along a worldline
Distance between two points	Spacetime interval between two events

Now let us spend more time discussing the **invariant** "spacetime interval".

$$(\Delta s)^2 = c^2(\Delta t)^2 - (\Delta x)^2 - (\Delta y)^2 - (\Delta z)^2$$

We discussed three different ways to measure the time interval between two events.

An observer in an inertial reference frame can measure the coordinate time Δt between the events by comparing the reading of the clock present at one event with the synchronized clock present at the other event.

Since observers in different reference frames will disagree about whether a given pair of clocks is synchronized, the coordinate time Δt measured between the events depends on the reference frame used.

One can avoid this problem by measuring the time between the events by a single clock that moves between the events so that it is present at both.

Such a clock measures a proper (proprietary) time $\Delta\tau$ between the events.

The value of such a proper time, while frame-independent (in the sense that all observers will agree on what a given clock present at both events actually registers between those events), is known experimentally to depend on the worldline traveled by the clock as it moves from one event to the other.

There is, however, one and only one worldline that takes a clock from one event to the other at a constant velocity.

The proper time between the events measured by an inertial clock is thus a unique, frame-independent number that depends on the spacetime separation of the two events and nothing else.

This unique proper time is called the *spacetime interval* Δs.

We also discussed an analogy that compared these different ways of measuring the time between events in spacetime with different ways of measuring the displacement of two points on a plane.

The coordinate time corresponds to the north-south (or east-west) coordinate displacement between the points.

Since surveyors using different coordinate systems will disagree about the exact direction of "north," the value of the north-south displacement between two points will depend on one's choice of coordinate system.

The proper time corresponds to the pathlength between the two points measured along a certain path using a tape measure.

Since measuring the pathlength does not require determining where north is, its value is independent of the coordinate system but does depend on the path chosen.

The spacetime interval corresponds to the straight-line distance between the points.

There is one and only one straight line between a given pair of points, so the pathlength measured along this line is a unique, coordinate-independent number that depends on the spatial separation of the points.

As discussed earlier, we can actually calculate the distance Δd between two points on the plane using the coordinate displacements Δx and Δy between the points (as measured in any given coordinate system) and the pythagorean theorem:

$$\Delta d^2 = \Delta x^2 + \Delta y^2$$

The amazing thing about this formula is that while the values Δx and Δy between two points depend on one's choice of coordinate system, the distance Δd calculated from these does not.

As I showed earlier by direct calculation, there is an analogous formula that links the coordinate time Δt and spatial displacements Δx, Δy, and Δz between two events measured in any given inertial reference frame with the frame-independent spacetime interval Δs between the events.

This equation, called the **metric equation**, provides the crucial key needed to escape the "relativity" of inertial reference frames and quantify the separation of the events in absolute (frame-independent) terms.

Now we will *derive* the metric equation and describe some of its immediate consequences.

Then, we will show how the metric equation can be used to compute proper times along more general worldlines and many other examples.

THE METRIC EQUATION or THE SPACETIME INTERVAL

The derivation that follows is the very core of the special theory of relativity.

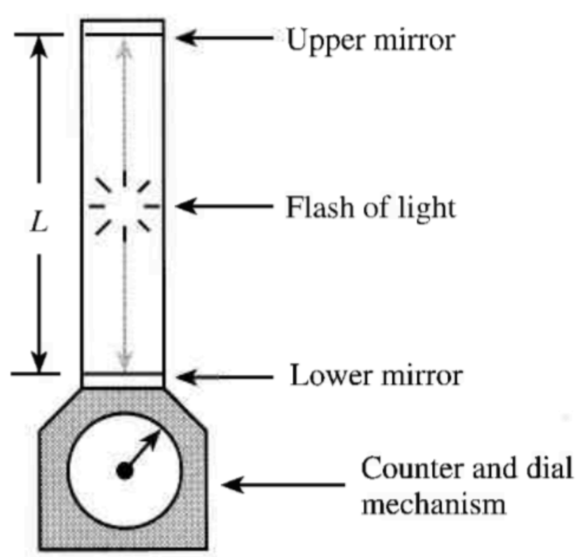
The metric equation is the key to understanding all the unusual and interesting consequences of the theory of relativity.

You should make a special effort to understand this argument thoroughly.

What we want to do is compare the time interval Δt between two events measured in some inertial frame (call it the Home Frame) with the time Δs measured by a clock moving at a constant velocity that is present at both events.

To make this argument easier, I want to consider a special kind of clock called a light clock.

An idealized light clock is shown in the Figure.



Schematic diagram of a light clock. Each "tick" of the light clock represents the passage of a time interval equal to 2L (if L is measured in seconds).

It consists of two mirrors a fixed distance L apart and a flash of light that bounces back and forth between the mirrors.

Each time the flash of light bounces off the bottom mirror, a detector in that mirror sends a signal to an electronic counter.

The clock dial thus essentially registers the number of round-trips that the light flash has completed.

Since the speed of light is defined to be 1 second of distance per 1 second of time in any inertial frame, this clock should be calibrated to register a time interval of 2L (where L is expressed in seconds) per "tick" (i.e., each time the light flash bounces of the bottom mirror): the clock will then read correct time as long as it is inertial.

Now consider an arbitrarily chosen pair of events A and B.

Let the coordinate time interval and spatial separation between these events (as measured in the Home Frame) be Δt and $\Delta d = \sqrt{(\Delta x)^2 + (\Delta y)^2 + (\Delta z)^2}$, respectively.

Also imagine that we have a light clock moving between these events (with its beam path oriented perpendicular to its direction of motion) at just the right constant velocity to be present at both events.

To simplify our argument, let us also imagine that the length L between the light-clock mirrors has just the right value so that events A and B happen to coincide with successive ticks of the light clock (in principle, we could always adjust L to make this true for the two events we want to examine).

In the inertial frame of the light clock, both events occur at the clock face, and the clock's light flash completes exactly one round-trip.

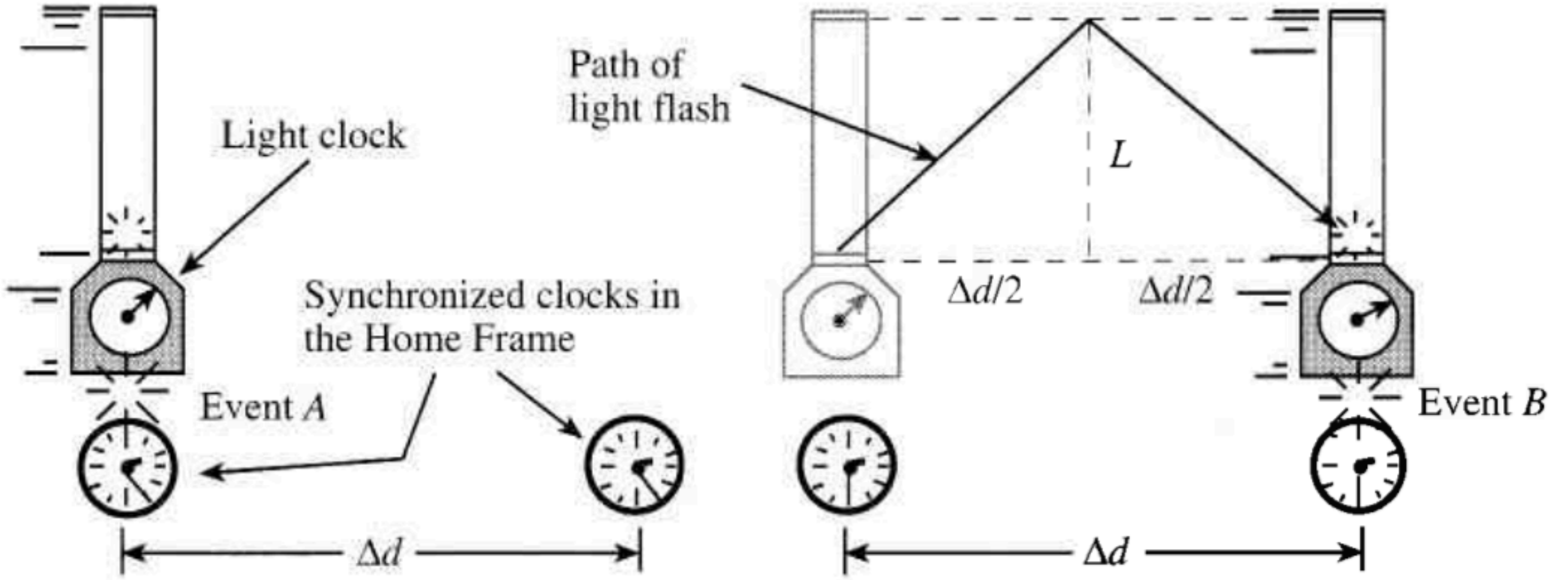
The time interval recorded by this clock between events A and B is thus exactly 2L.

Since this clock is present at both events, it registers the spacetime interval between these events, so $\Delta s = 2L$.

In the Home Frame, the time of each event is registered by the clock nearest the event; since the events occur at different places, the coordinate time interval between the events will be determined by taking readings from a pair of clocks.

In this frame, the light clock is observed to move the distance Δd in the time interval Δt.

This means that the light flash will be observed in the Home Frame to follow the zigzag shown in the Figure.



As the light clock moves from event A to event B in the Home Frame, its internal light flash will be observed to follow the zigzag path shown. The total distance covered by the flash in this frame is $2\sqrt{L^2 + (\Delta d/2)^2}$

The distance that the light flash travels in the Home Frame is (by the pythagorean theorem)

$$2\sqrt{L^2 + (\Delta d/2)^2} = \sqrt{4L^2 + \Delta d^2} = \sqrt{(2L)^2 + \Delta d^2}$$

Since the synchronized clocks in the Home Frame must (by definition of synchronization) measure the speed of light to be 1, the coordinate time interval Δt registered on the pair of synchronized clocks in the Home Frame must be equal to the distance that the light flash traveled between the events:

$$\Delta t = \sqrt{(2L)^2 + \Delta d^2}$$

But we saw above that the light clock registers the spacetime interval between the two events to be $\Delta s = 2L$.

Plugging this into the last equation and squaring both sides, we get

$$\Delta t^2 = \Delta s^2 + \Delta d^2 \quad \text{or} \quad \Delta s^2 = \Delta t^2 - \Delta d^2$$

Since $\Delta d^2 = \Delta x^2 + \Delta y^2 + \Delta z^2$ (where Δx, Δy, and Δz are the coordinate differences measured between the events in the Home Frame), we have, finally,

$$\boxed{\Delta s^2 = \Delta t^2 - \Delta x^2 - \Delta y^2 - \Delta z^2}$$

This extremely important equation (same as derived earlier) links the frame-independent spacetime interval Δs between any two events to the frame-dependent coordinate separations $\Delta t, \Delta x, \Delta y, \Delta z$ measured between those events in any arbitrary inertial reference frame!

Note that we have not sacrificed anything by using a light clock instead of some other kind of clock in this argument: the speed of light is 1 in any inertial frame, and thus any decent clock that we construct must agree with what the light clock says.

The only real limitation to our argument is that Δt must be greater than Δd for the two events in question so that it is possible for a light flash to travel between the events.

Note that if $\Delta t < \Delta d$, this equation yields an imaginary number for Δs, an absurd result indicating that the conditions of the proof have been violated.

Just as the spacetime interval Δs between two events is analogous to the distance Ad between two points on a plane, the formula $\Delta s^2 = \Delta t^2 - \Delta x^2 - \Delta y^2 - \Delta z^2$ is directly analogous to the pythagorean relation $\Delta d^2 = \Delta x^2 + \Delta y^2 + \Delta z^2$.

Note that the pythagorean relation also relates a coordinate-independent quantity (the distance Δd between two points) with quantities whose values depend on the choice of coordinate system (the coordinate differences Δx, Δy, and Δz).

Indeed, the formula for the space-time interval would be just like a four-dimensional version of the pythagorean theorem if it were not for the minus signs that appear.

We will see that these minus signs have a variety of interesting and unusual consequences.

This result is called the **metric equation**.

It is the link between our human-constructed reference frames and the absolute physical reality of the separation between two events in space and time.

It is difficult to overemphasize the importance of this equation: virtually al the rest of our study of the theory of relativity will revolve around the implications of this equation.

The previous argument assumes that the vertical length L between the light-clock mirrors is the same in both the light-clock frame (where it was used to compute the spacetime interval) and the Home Frame (where it was used to compute the coordinate time).

But how do we know that this is true?

Since coordinate differences between events are generally frame-dependent, what gives us the right to assume that the displacement between the mirrors has the same value in both frames?

We will not show the derivation in this class, but it can be shown that if we have two reference frames in relative motion along a given line, any displacement measured *perpendicular* to that direction of motion will have the same value in both reference frames.

This follows directly from the principle of relativity.

In addition, many experiments have confirmed the metric equation validity.

THE GEOMETRY OF SPACETIME IS NOT EUCLIDEAN

We have found the analogy between ordinary euclidean plane geometry and spacetime geometry to be very illuminating, and this basic analogy will remain fruitful as we continue to develop the consequences of the theory of relativity.

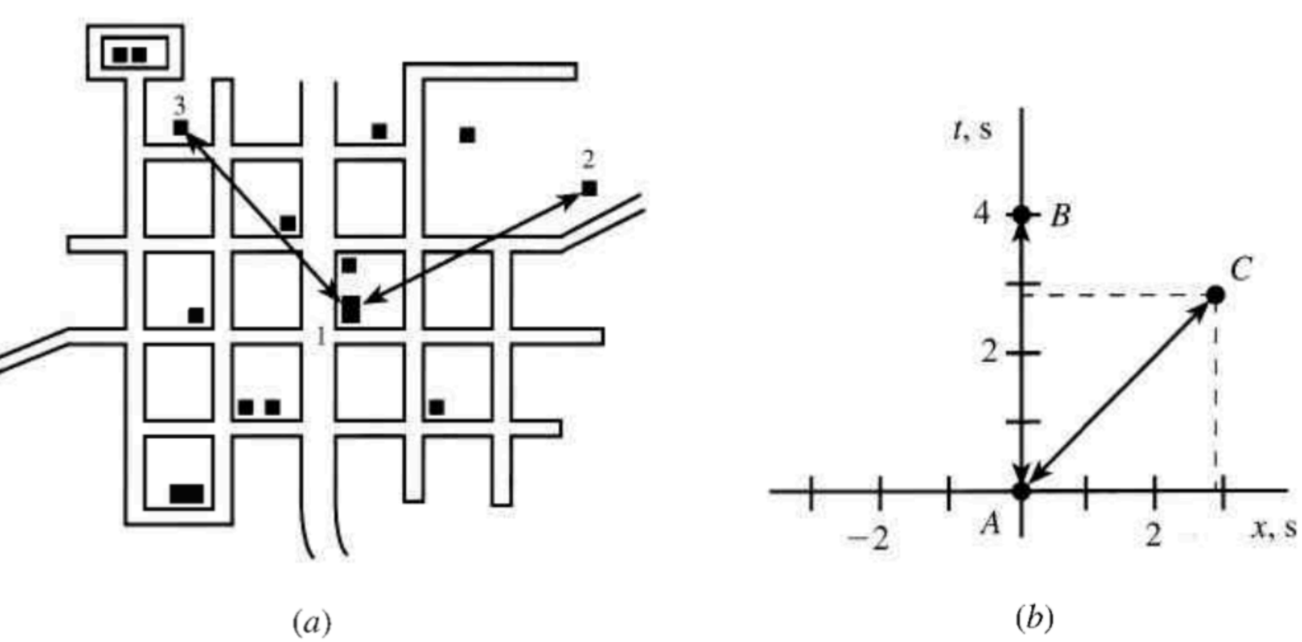
Nevertheless, it is important at this point to describe some of the important differences between euclidean geometry and spacetime geometry that are a result of the negative signs in the metric equation $\Delta s^2 = \Delta t^2 - \Delta x^2 - \Delta y^2 - \Delta z^2$ that do not appear in the corresponding pythagorean theorem $\Delta d^2 = \Delta x^2 + \Delta y^2$.

One important difference concerns the representation of distances on a map and spacetime intervals on a spacetime diagram.

If one prepares a scale drawing (e.g., a map) of various points on a plane, the distance between any two points on the map is proportional to the actual distance between those points in space.

That is, distances on the drawing directly correspond to distances in the physical reality being represented.

In the Figure Part (a) below, for example, to determine the distance between City Hall and the Statue of the Unknown Physicist, one need merely measure the distance (in inches) between the two sites on the map shown above and multiply by the conversion factor 1000m = 1in.

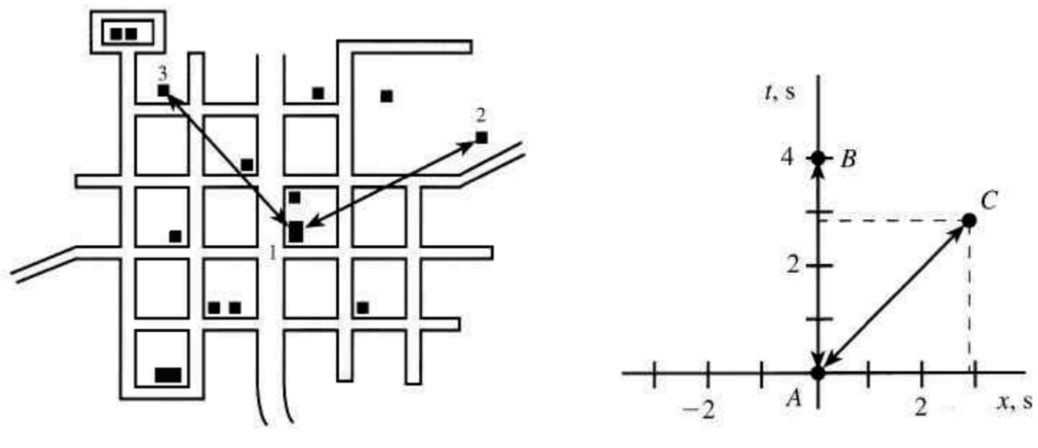


(a) As cale drawing of Askew, North Dakota (1 in = 1 km). The actual distance between (1) City Hal and (2) the Statue of the Unknown Physicist is 852 m. The distance between City Hal and the bandshell at (3) Higgenbottom Park is also 852 m. The length of both arrows on the drawing is 0.852 inch. (b) Both event Band event Care the same distance from event A on the spacetime diagram, but the spacetime interval between A and B is 4s, while the spacetime interval between A and Cis zero (since $\Delta t = \Delta x$)!

It does not matter how the line between the two sites is oriented or where the sites are located on the drawing: the distance in the physical space being represented by the map is always proportional to the distance measured on that map.

However, it is not true that the displacement between two points on a spacetime diagram is proportional to the spacetime interval between the corresponding events.

This is illustrated in Figure Part (b).

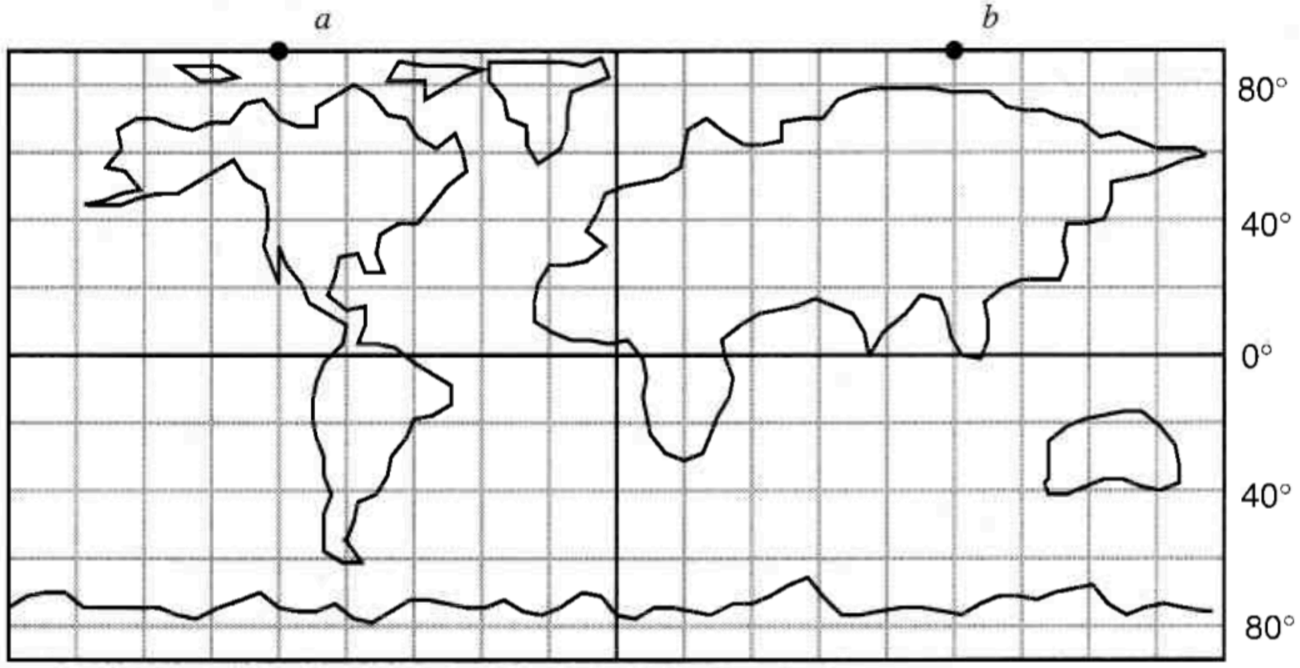


(a) (b)

We see that a spacetime diagram accurately displays the spacetime coordinates of various events, but the distances between the points representing those events on the diagram are not proportional to the actual spacetime intervals between those events in spacetime.

This is very strange, and it may seem particularly strange that two events (such as A and C in Figure Part (b)) can occur at different places and times and yet have zero spacetime interval between them.

Nonetheless, there exists a useful analogy with something you may have seen before. Imagine a map of the world where the lines of longitude and latitude are drawn as equally spaced straight lines (as shown in the Figure).



A pseudo-Mercator projection map of the world, where the lines of longitude and latitude are mapped to equally spaced straight lines.

Have you ever noticed how the shapes and sizes of the continents appear very warped near the north and south poles on such maps?

For example, look at the continent of Antarctica on the map shown.

It looks huge and seems to be shaped like a strip.

But in fact it is not so large, and it has a nearly circular shape: its size and shape are quite distorted by the nature of the map.

The shapes of Greenland and northern Canada are distorted as well.

As a matter of fact, the two points marked a and b on the map are both at 90° north latitude, i.e., at the north pole.

Thus though these points are separated by a significant distance on the map, the physical distance between these points on the surface of the earth is zero!

Why does this map not accurately represent the distances between points on the earth's surface?

The problem is that the surface of the earth as a whole is the surface of a sphere, which has a very different geometry than the euclidean geometry of a flat sheet of paper.

For example, on a sheet of paper, the interior angles of a triangle always add up to 180°, parallel lines never intersect, and so on.

But on the surface of the earth, the interior angles of a triangle can add up to more than 180° (consider a triangle with one vertex at the north pole and two vertices at the equator), initially parallel lines do not remain parallel, and so on.

Because of these fundamental geometric differences between the surface of the earth and the sheet of paper, any flat map of the earth will necessarily be a distorted representation: one cannot make a map of the surface of the earth on a flat sheet of paper and have distances on the sheet correspond to actual distances on the earth.

Similarly, one cannot draw a spacetime diagram in such a way that distances between points on the drawing are proportional ot the spacetime intervals between the corresponding events.

Like the surface of the earth, spacetime has a different geometry than the flat sheet of paper on which a spacetime diagram is drawn.

The fact that the metric equation of spacetime has negative signs where the corresponding pythagorean relation has positive signs is symptomatic of this difference.

The moral of the story is this: do not expect a spacetime diagram to give you direct information about the spacetime interval between various events, any more than you would expect a flat map of the earth to give you accurate information about distances on the earth's surface.

A spacetime diagram is meant to visually represent the coordinates of events and the worldlines of particles, nothing more.

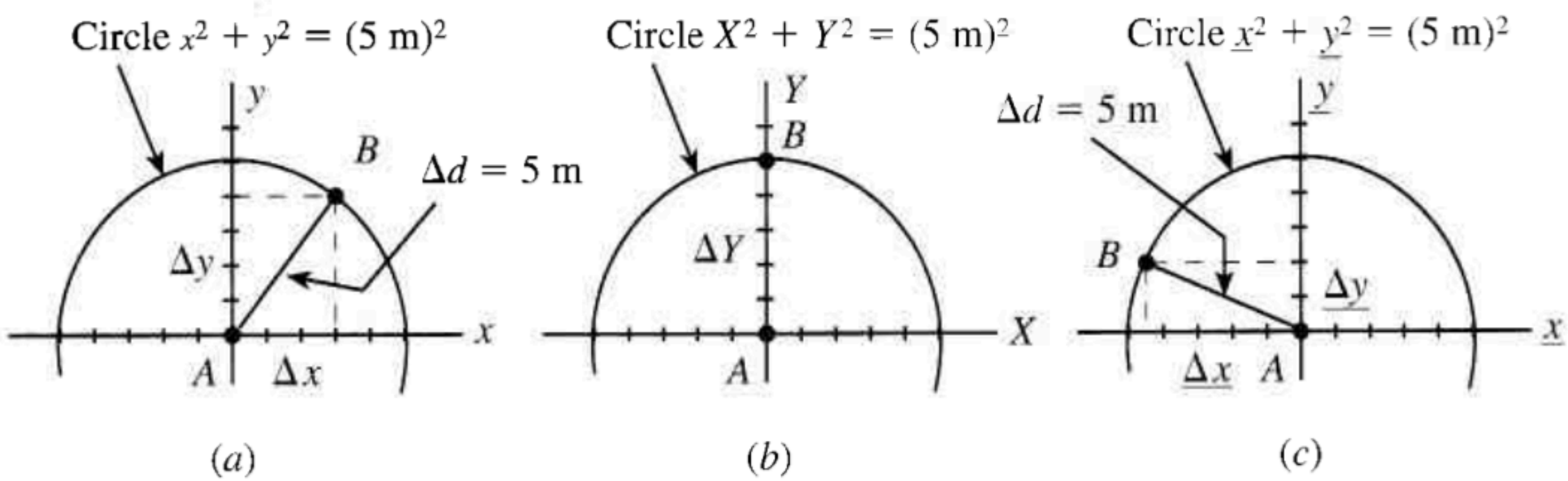
You can always compute the spacetime interval between two events from the coordinates of those events if necessary.

THE METRIC EQUATION AND THE PYTHAGOREAN THEOREM

In spite of the issue just raised, we can extend the geometric analogy to spacetime by exploring the basic similarities (as well as differences) in how the metric equation describes the geometry of spacetime and how the pythagorean theorem describes the geometry of a plane.

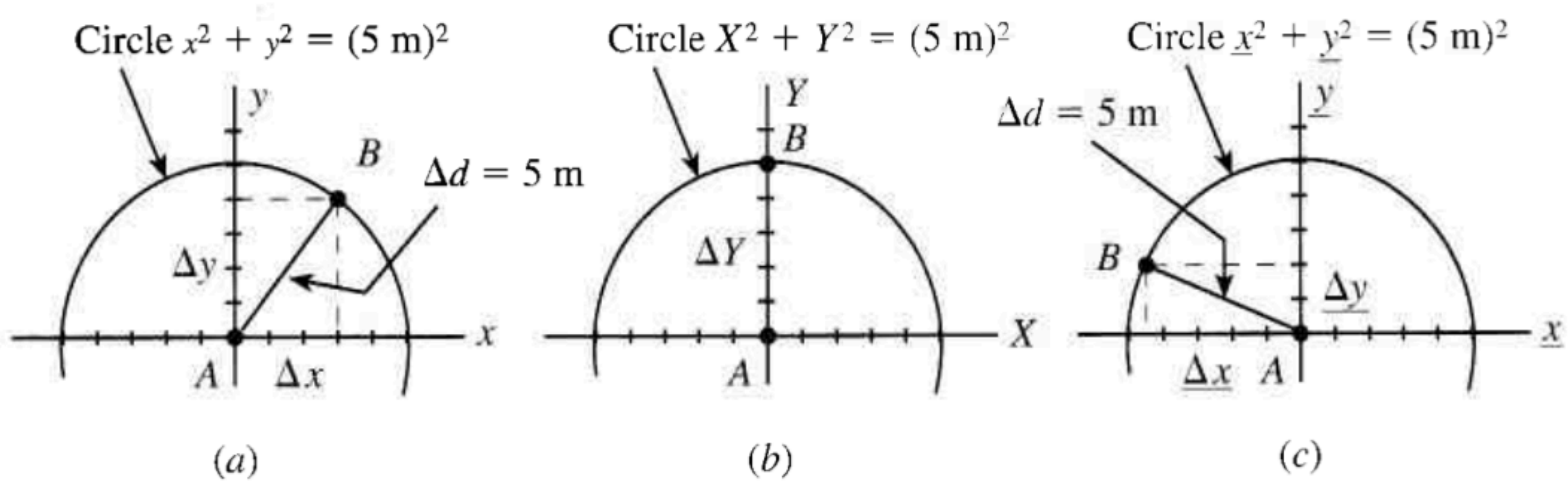
The most important thing about both these equations is that they enable us to calculate an absolute quantity (Δs or Δd) in terms of frame-dependent coordinate differences measured in an arbitrary inertial frame or coordinate system.

This similarity is illustrated in the next two Figures.



(a) A pair of points on the plane that are separated a distance of 5m. In this coordinate system, these points happen to have coordinate separations of $\Delta y = 4$ m and $\Delta x = +3$ m. (b) We can find a coordinate system in which A and B lie along the vertical axis ($\Delta X = 0$). In this unique system, the coordinate separation ΔY is equal to the distance between the points. (c) If we twist the axes farther clockwise, we can find a coordinate system where the coordinate separations are $\Delta \underline{y} = 2$m and $\Delta \underline{x} = -4.6$.m. In all circumstances, though, B will lie somewhere on the circle shown.

The first Figure shows the same pair of points on the plane A(and B) as viewed in coordinate systems having various different orientations with respect to "north."



Note that if we set up the coordinate systems so that point A is at the origin, then point B in each coordinate system lies somewhere on the *circle* defined by the equation $x^2 + y^2 = \text{constant} = \Delta d^2$, where Δd^2 is the squared distance between the points (since Δd is the distance between the points in all coordinate systems).

Similarly, the next Figure shows the same pair of events (A and B) plotted on spacetime diagrams drawn by observers in different inertial frames.

If we choose these frames so that event A occurs at $t = x = 0$, then event B lies somewhere on the curve defined by $t^2 - x^2 = \text{constant} = \Delta s^2$, where Δs^2 is the squared spacetime interval between the events (since Δs has a frame-independent value): such a curve is a *hyperbola*, as shown.

Note that we are assuming $\Delta y = \Delta z = 0$ for these two events.

The point is that the set of all points a given distance from the origin on the plane form a circle and the set of all events a given spacetime interval from the origin event in spacetime form a hyperbola.

The reason both curves are not circles is because of the negative sign in the metric equation that does not appear in the pythagorean relation.

But there is a nice one-to-one correspondence between circles in plane geometry and hyperbolas in spacetime geometry.

Note that one consequence of difference between the metric equation and the pythagorean relation is that in the first Figure we see that the y-coordinate separation between a pair of points is always less than or equal to the distance between the points (the hypotenuse on the diagram): $\Delta y \leq \Delta d$.

In the second Figure, though, we see that the coordinate time Δt between a pair of events is always greater than the spacetime interval Δs between them: $\Delta t \geq \Delta s$ even though the hypotenuse that represents Δs on the diagram looks larger.

Now let us spend more time discussing the details(the nitty-gritty) of the construction of these diagrams and their properties because they are so important.

Earlier, we explored consequences of the metric equation that links the spacetime interval Δs between two events to coordinate differences between those events measured in any inertial frame.

Our focus in the earlier work was to link coordinate measurements (which are frame-dependent) to the frame-independent quantities Δs and $\Delta \tau$.

To delve further into the implications of the principle of relativity, we need to go a step further: we need to find a way of linking an event's t, x, y, z coordinates measured in one inertial frame with the same event's , t',x',y',z' coordinates measured in another inertial frame.

These equations that link coordinates in one frame with those in another are called the Lorentz transformation equations and are the relativistic generalization of the Galilean transformation equations as we saw.

We needed these equations to find the relativistic generalization of the Galilean velocity transformation equations, to explore the phenomenon of "length contraction," to explain why nothing can go faster than the speed of light, and so on.

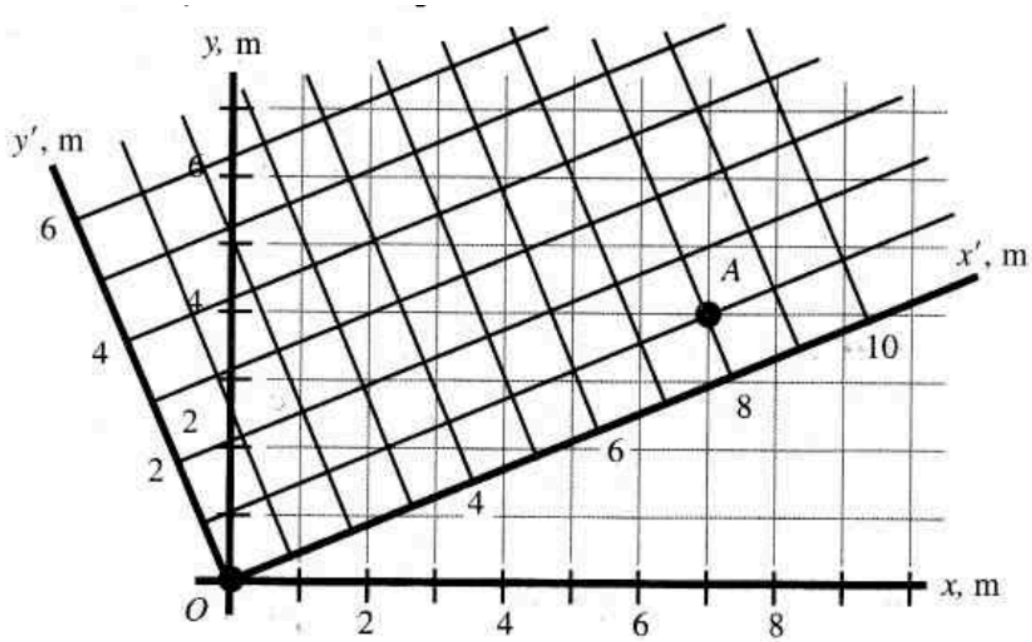
The derivation of the Lorentz transformation equations was direct but somewhat abstract and mathematical, and that abstraction can blunt one's intuition about what is really going on.

Therefore, now we addressed the same problem using a more visual, intuitive tool called a two-observer spacetime diagram or Minkowski diagram.

We now want to spend more time with this important tool.

In a two-observer spacetime diagram, we superimpose the coordinate axes for two different observers on the same spacetime diagram.

What we will end up with will be analogous to the drawing shown in the Figure, which shows two ordinary cartesian coordinate systems (one rotated with respect to the other) superimposed upon the plane.

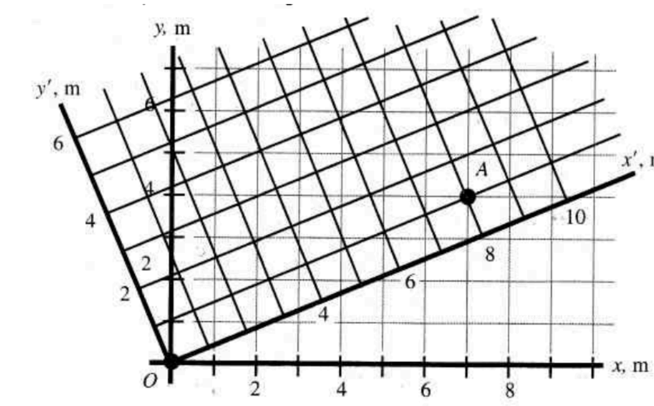


A drawing showing two sets of cartesian coordinate axes superimposed on the same plane. The coordinates of point A in both coordinate systems can be read easily from such a diagram.

Once we have set up the Cartesian two-observer diagram shown in the Figure, if we know the coordinates of a point A in the xy coordinate system (shown as having coordinates $x_A = 7$ m, $y_A = 4$ m); we can plot point A relative to point O.

Then we can just read the coordinates of A in the x'y' coordinate system from the diagram (the coordinates are x'_A, = 8m, y'_A= 1m in this case).

We do not have to use any equations or do any calculations at all!



Setting up such a diagram is easy enough for plane cartesian coordinate systems: it merely involves drawing two sets of perpendicular axes (one rotated with respect to the other), scaling the axes, and drawing coordinate grid lines for each set of axes.

Setting up the two sets of coordinate axes representing different inertial frames on a spacetime diagram is a similar process, but the peculiarities of spacetime geometry relative to plane geometry lead to some surprising dissimilarities as well.

Therefore we need to develop the procedure in a careful step-by-step manner so that we are sure to catch all these dissimilarities.

To make the task of constructing two-observer spacetime diagrams easier, it is convenient to make several assumptions.

First, we assume that the two inertial reference frames are in standard orientation with respect to each other.

That is, we assume that their corresponding spatial axes point in the same directions in space and the relative motion of the frames is directed along the common x direction.

Since one can choose whatever orientation desired for the axes of a spatial coordinate system, we do not really lose anything by choosing the frames to have this orientation, and we gain much in simplicity.

Second, we will work with only those events that occur along the common x and x' axes of these frames (i.e., those having coordinates $y = y' = z = z' = 0$).

This is a substantial concession to convenience: we would really like to be able to handle any event.

But plotting an event with arbitrary coordinates on a spacetime diagram would require that the diagram have four dimensions, which is impossible to represent on a sheet of paper.

We choose therefore to limit our attention to events that can be easily plotted on a two-dimensional spacetime diagram.

We will see that most interesting problems can still be treated within this restriction.

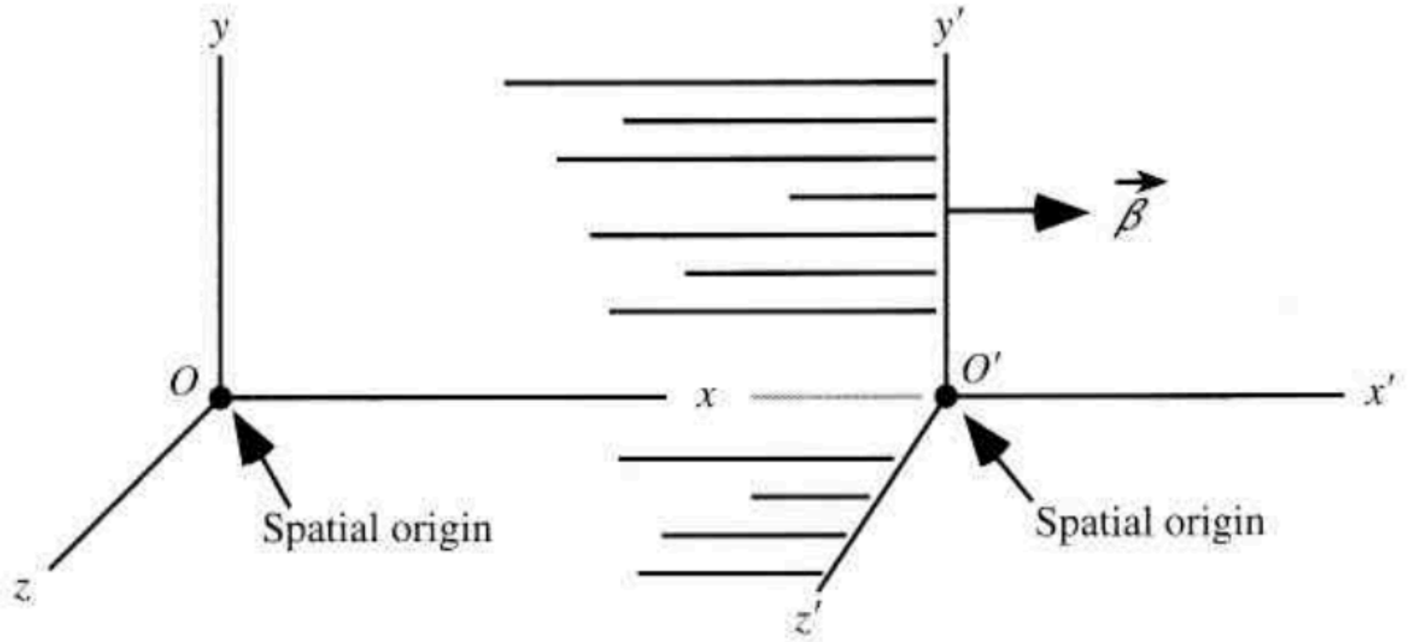
So until you are told otherwise, you should assume that $y = y' = z = z' = 0$ for all events under discussion.

The first step in actually drawing a two-observer spacetime diagram is to pick one of the two frames to be the **Home Frame**.

The other frame can be called the **Other Frame** (easy to remember, right?).

Remember that the terms Home Frame and Other Frame are capitalized in our discussions to emphasize that these phrases are actually names of inertial frames.

It is conventional (but not absolutely necessary) to pick the Other Frame to be the frame of the two that moves in the +x direction with respect to the Home Frame, as shown in the Figure (alternatively, one can think of the Home Frame as the one that moves in the -x direction with respect to the Other Frame): some signs in equations that follow assume we are following this convention.



Two inertial reference frames in standard orientation. The frames are represented schematically here by bare orthogonal axes. Note that when frames are in standard orientation, the Other Frame is always taken to be the frame moving in the positive direction along the common x direction.

It is also conventional to distinguish each frame's axes and coordinate measurements by using t, x, y, z for Home Frame axes and coordinates and t', x', y,' and z' for the Other Frame axes and coordinates.

Now, our choice of a Home Frame does not necessarily mean we are considering that frame to be at rest and the Other Frame to be moving: we still want to reserve the freedom to consider either frame as being at rest.

What this choice does imply is simply that we will represent the Home Frame t and x axes in the usual manner in a space-time diagram (i.e., its t axis will be vertical and its x axis will be horizontal).

DRAWING THE TIME AXIS FOR THE OTHER FRAME

The time axis for any frame on a spacetime diagram is by definition the line connecting all events that have x coordinate = 0 in that frame.

This means that the time axis is the worldline of the clock at the spatial origin of the reference frame (all events happening at the spatial origin of a reference frame have spatial coordinate x = 0 by definition).

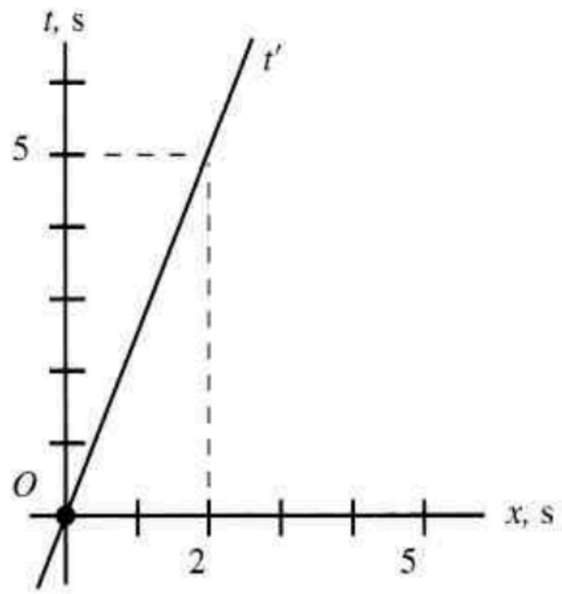
The t axis of the Home Frame is drawn as a vertical line by convention.

How should the t' axis of the Other Frame be drawn?

The Other Frame moves with speed β along the +x direction with respect to the Home Frame by hypothesis.

This means that as measured in the Home Frame, the Other Frame's spatial origin moves β units in the +x direction every unit of time.

The worldline of the Other Frame's origin as plotted on the spacetime diagram is thus a straight line of slope $1/\beta$ as shown in the Figure.



The time axes for both the Home and Other frames. The t axis connects all events occurring at x =0 in the Home Frame, while the t' axis connects all events occurring at x' = 0 in the Other Frame. The diagram is drawn assuming that the Other Frame is traveling with a speed β= 2/5 with respect to the Home Frame: therefore the slope of the t' axis is $1/\beta$ = 5/2.

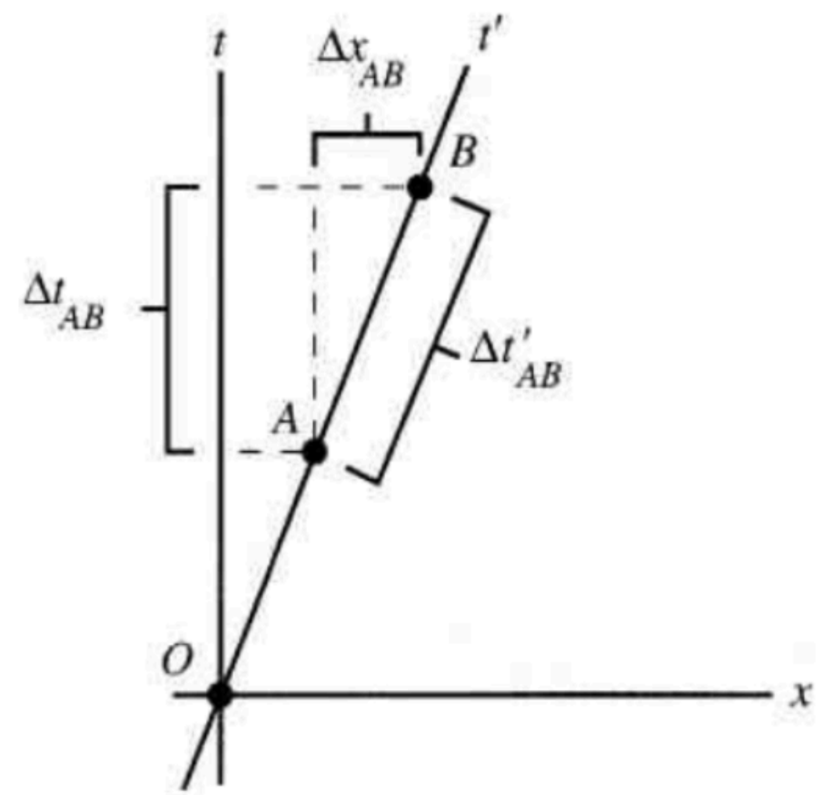
Note that this line goes through the origin event O since the spatial origins of both frames coincide at t = t' = 0 if the frames are in standard orientation.

CALIBRATING THE t' AXIS

The next step is to put an appropriate scale on the Other Frame's t' axis.

It is (unfortunately) not correct to simply mark this axis using the same scale as used for the t and x axes. The purpose of this section is to describe how to correctly mark a scale on the t' axis.

Consider the marks on the t' axis labeled A and B on the Figure.



This figure shows two arbitrary scale marks on the t' axis, separated by a time $\Delta t'_{AB}$. Since the axis represents the worldline of the spatial origin of the Other Frame as it moves to the right with speed β, in the time Δt_{AB} between events A and B, the origin will move a distance $\Delta x_{AB} = \beta \Delta t_{AB}$.

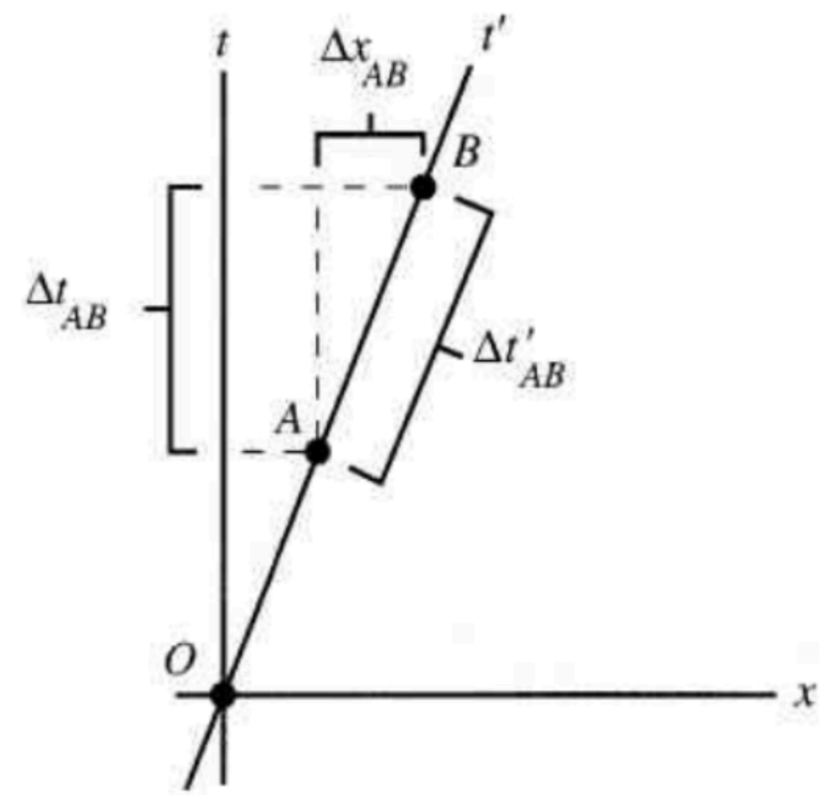
Each of these marks corresponds to an event in spacetime: in the Other Frame, these events are separated by the coordinate differences $\Delta t'_{AB} = 1$ s and $\Delta x'_{AB} = 0$ (since all events along the t' axis occur at the spatial origin of the Other Frame by definition).

Let Δt_{AB} and Δx_{AB} be the coordinate differences between the same two events as measured in the Home Frame.

Since the spacetime interval between these two events is frame-independent, we have

$$(\Delta t_{AB})^2 - (\Delta x_{AB})^2 = (\Delta s_{AB})^2 = (\Delta t'_{AB})^2 - (\Delta x'_{AB})^2 = (\Delta t'_{AB})^2$$

because $\Delta x'_{AB} = 0$ in this case.



earlier, we decided that the slope of the t' axis was $1/\beta$ (where β is the velocity of the Other Frame with respect to the Home Frame).

This means that Δt_{AB} and Δx_{AB} must be related as follows:

$$\frac{\Delta x_{AB}}{\Delta t_{AB}} = \beta \qquad \text{or} \qquad \Delta x_{AB} = \beta\, \Delta t_{AB}$$

NOTE: Another way to think of this is to remember that the t' axis is the worldline of the master clock at the spatial origin of the Other Frame, which in the time Δt_{AB} between the events travels a distance $\Delta x_{AB} = \beta\, \Delta t_{AB}$, since it is moving at speed B by definition.

Putting the last two equation together, we get

$$(\Delta t'_{AB})^2 = (\Delta t_{AB})^2 - (\beta\, \Delta t_{AB})^2 = (\Delta t_{AB})^2(1 - \beta^2)$$

and solving for Δt_{AB}, we get

$$\Delta t_{AB} = \frac{\Delta t'_{AB}}{\sqrt{1 - \beta^2}}$$

where we have already defined

$$\gamma \equiv \frac{1}{\sqrt{1 - \beta^2}}$$

Note that γ is a number that is always larger than 1.

Thus, we can write

$$\Delta t_{AB} = \gamma\, \Delta t'_{AB}$$

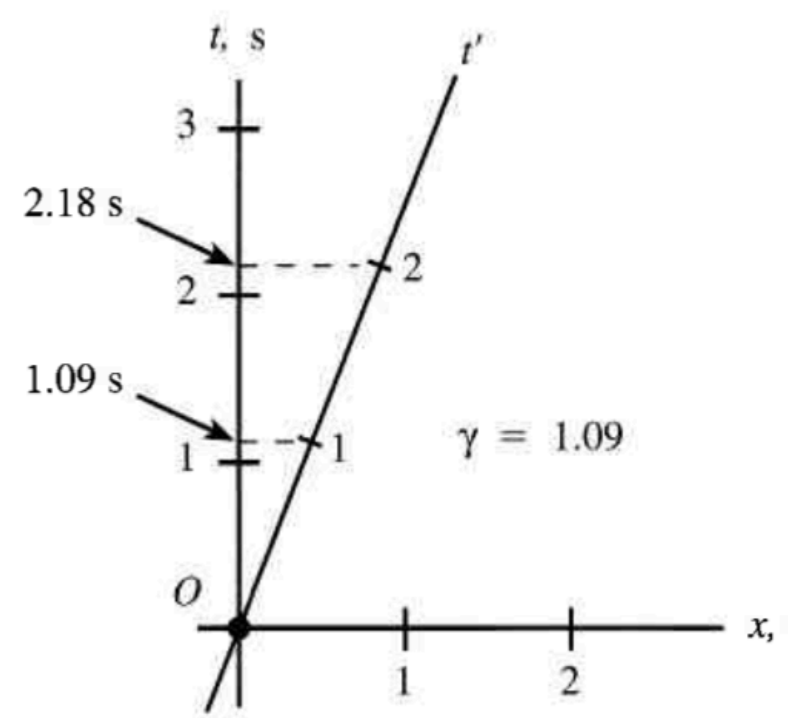
The point of all this is that if you want to draw marks on the t' axis of the graph that are separated by some time interval $\Delta t'$, then you must draw these marks so that they have a vertical separation of $\Delta t = \gamma\, \Delta t'$.

To give a concrete example, imagine that your Other Frame moves at a speed of $\beta = 2/5$ and you want to draw marks on the t' axis for that frame that are separated by $\Delta t' = 1$s.

According to our results, you need to draw these marks so that they have a vertical separation of

$$\Delta t = \gamma\, \Delta t' = \frac{1\text{ s}}{\sqrt{1-(2/5)^2}} \approx 1.09\text{ s}$$

as shown in in the Figure.



Calibrating the t' axis of an Other Frame moving at a speed of 2/5 relative to the Home Frame. Marks separated by a time $\Delta t' = 1$s in the Other Frame are separated by $\Delta t = 1.09$ s in the Home Frame. Note that the distance between the marks on the t' axis is "stretched out" compared to the distance between the marks on the t axis.

DRAWING THE DIAGRAM x′ AXIS

The t' axis on the spacetime diagram is defined to be the line connecting all events that occur at x' = 0 (that is, the spatial origin) of the Other Frame.

Analogously, the diagram x' axis is defined to be the line connecting al events that occur at t' = 0 (that is, at the same time as the origin event).

The diagram x axis for the Home Frame is conventionally drawn as a horizontal line on a spacetime diagram. How should we draw the diagram x' axis of the Other Frame?

The natural thing to do would be to draw the diagram x' axis perpendicular to the t' axis.

Unfortunately, this approach is not consistent with the definition of the diagram x' axis given above.

To figure out the right way to draw this axis, we have to carefully consider the implications of the fact that *the diagram x' axis connects events that are simultaneous in the Other Frame*.

Note: The phrase "diagram x' axis" or "diagram x axis" is going to refer to a line drawn on a spacetime diagram that connects all events that occur at zero time.

This is to be sharply distinguished from the line in physical space that goes through the spatial origin and connects all points having $y = z = 0$: when I want to talk about the latter, I will always speak of the”x direction" or the "spatial x axis".)

We begin by considering a set of events that illustrates the use of the radar method to determine coordinates in the Other Frame.

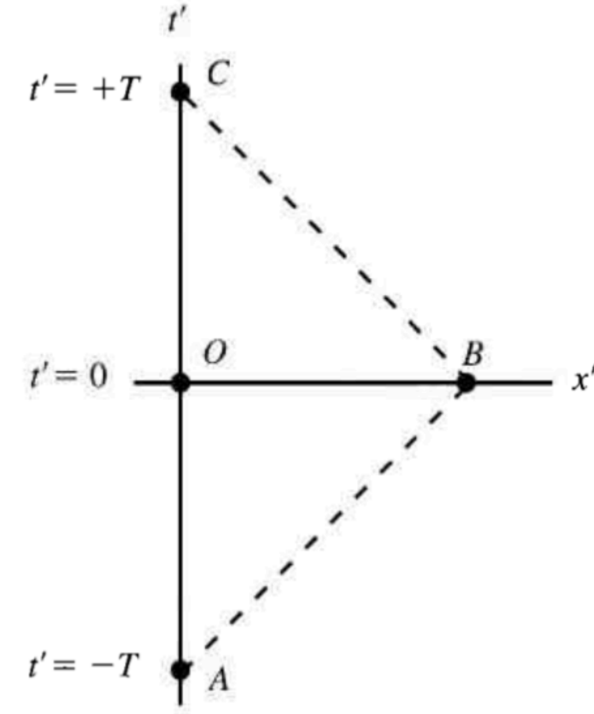
Imagine that at time $t' = -T$ (where T is some arbitrary number) a light flash is sent from the master clock (located at the spatial origin of the Other Frame) in the +x direction (call the emission of the flash event A).

This flash reflects from some event B and returns to the Other Frame's master clock at time $t' = +T$ (call the reception of the flash by the master clock event C).

Since the light flash takes an equal amount of time to return from event B as it took to get there, we conclude that event B happened at a time halfway between $t'_A = -T$ and $t'_C = +T$, that is, $t'_B = 0$.

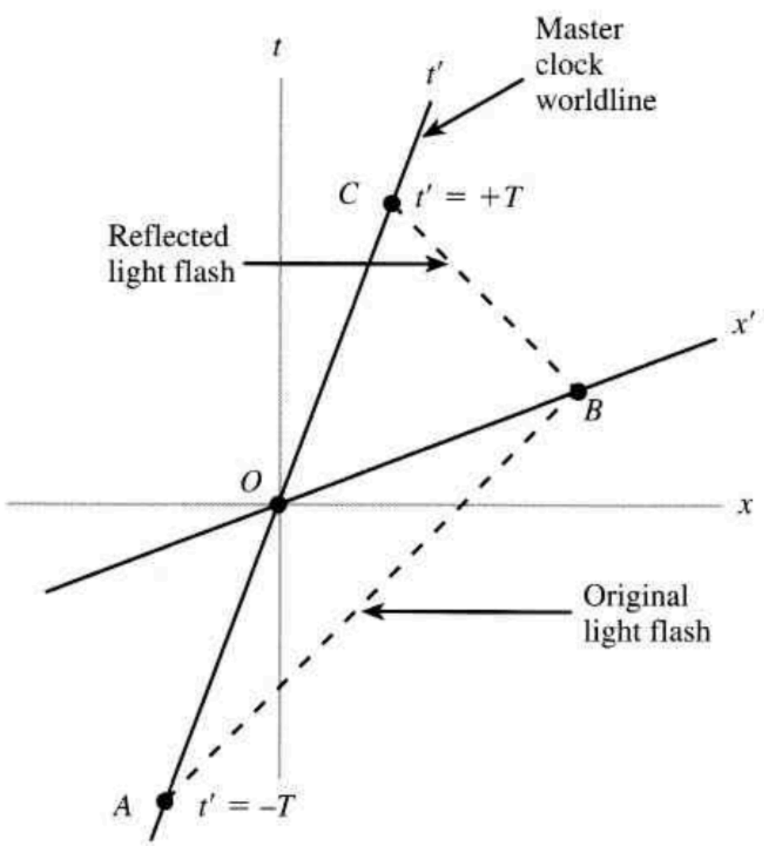
This means that event B will be simultaneous with the origin event O.

This is illustrated in the Figure.



Spacetime diagram illustrating the use of the radar method to determine the time coordinate of event B. A flash of light is emitted from the spatial origin of the Other Frame at time t' = - T(event A.) This flash is reflected at event B and returns to the spatial origin at time t' = +T(event C). Because the light flash travels the same distance down and back in the spatial x direction, event B must occur halfway between B times -T and + T that is, it must occur at time t' = 0. Thus event B lies on the diagram x' axis of this spacetime diagram by definition.

Now let us draw this set of events on a spacetime diagram based on the Home Frame (see Figure).

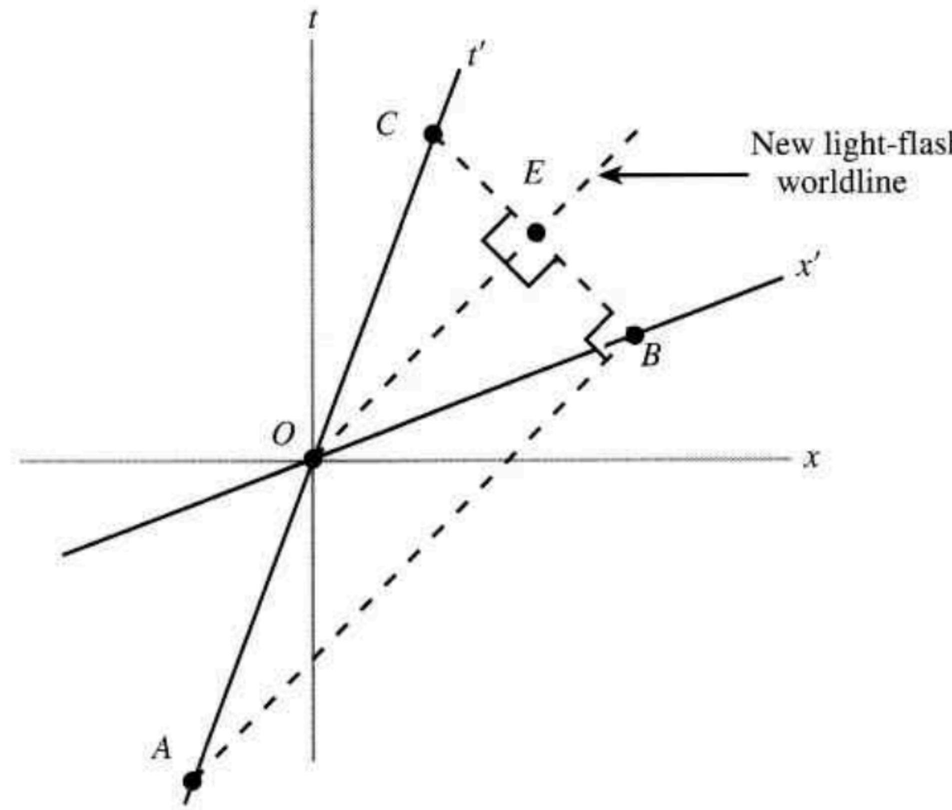


A spacetime diagram of the set of events illustrated in the last Figure as plotted by an observer in the Home Frame. Note that A and C are symmetrically placed on the t' axis on opposite sides of the origin event O. Note that both event O and event B occur at time t' = 0 by definition of clock synchronization in the Other Frame. Therefore, these events must lie along the diagram x' axis, implying that the diagram x' axis must have an upward slope as shown.

Events A, O, and C all occur at x' = 0, so they al lie on the t axis of the spacetime diagram.

Moreover, events A and C occur at t' = - T and t' = +T, so they must be symmetrically spaced on opposite sides of the origin event, as shown.

The worldline of the clock present at event B must be parallel to and just far enough away from the t' axis so that the light-flash worldlines for the right- and left- going flashes meet at the worldline of the second clock.



Note that since the speed of light is 1 in every reference frame, these light-flash worldlines must have aslope of ±1 on this diagram.

Now, by the definition of clock synchronization in the Other Frame, events O and B both occur at time t' = 0.

The diagram x' axis is defined to be the line connecting all events that occur at time t' = 0.

Therefore, the diagram x' axis must go through events O and B.

This means that the diagram x' axis must angle upward, as shown in the above Figure.

In fact, by considering the geometric relationships implicit in the above Figure, I can show you that the diagram x' axis makes the same angle with the diagram x axis that the t' axis makes with the t axis.

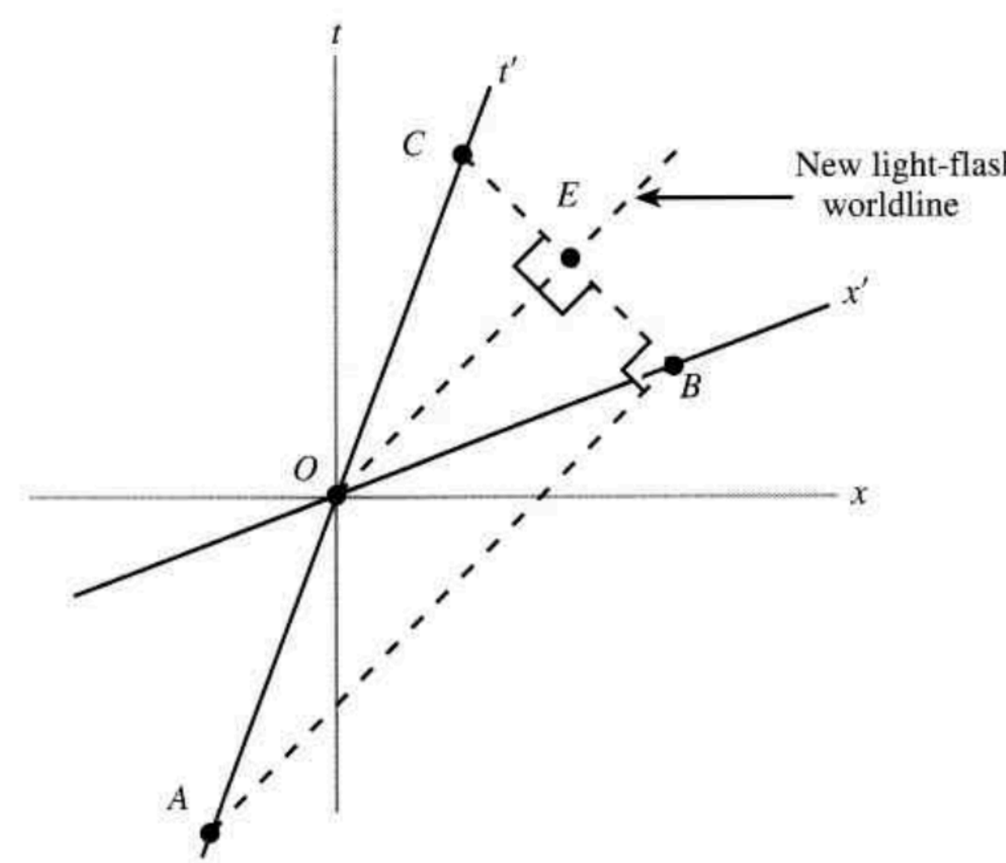
The argument goes like this.

Imagine that the master clock emits a right-going light flash at the origin event 0, and let event E be this flash meeting the incoming reflected flash.

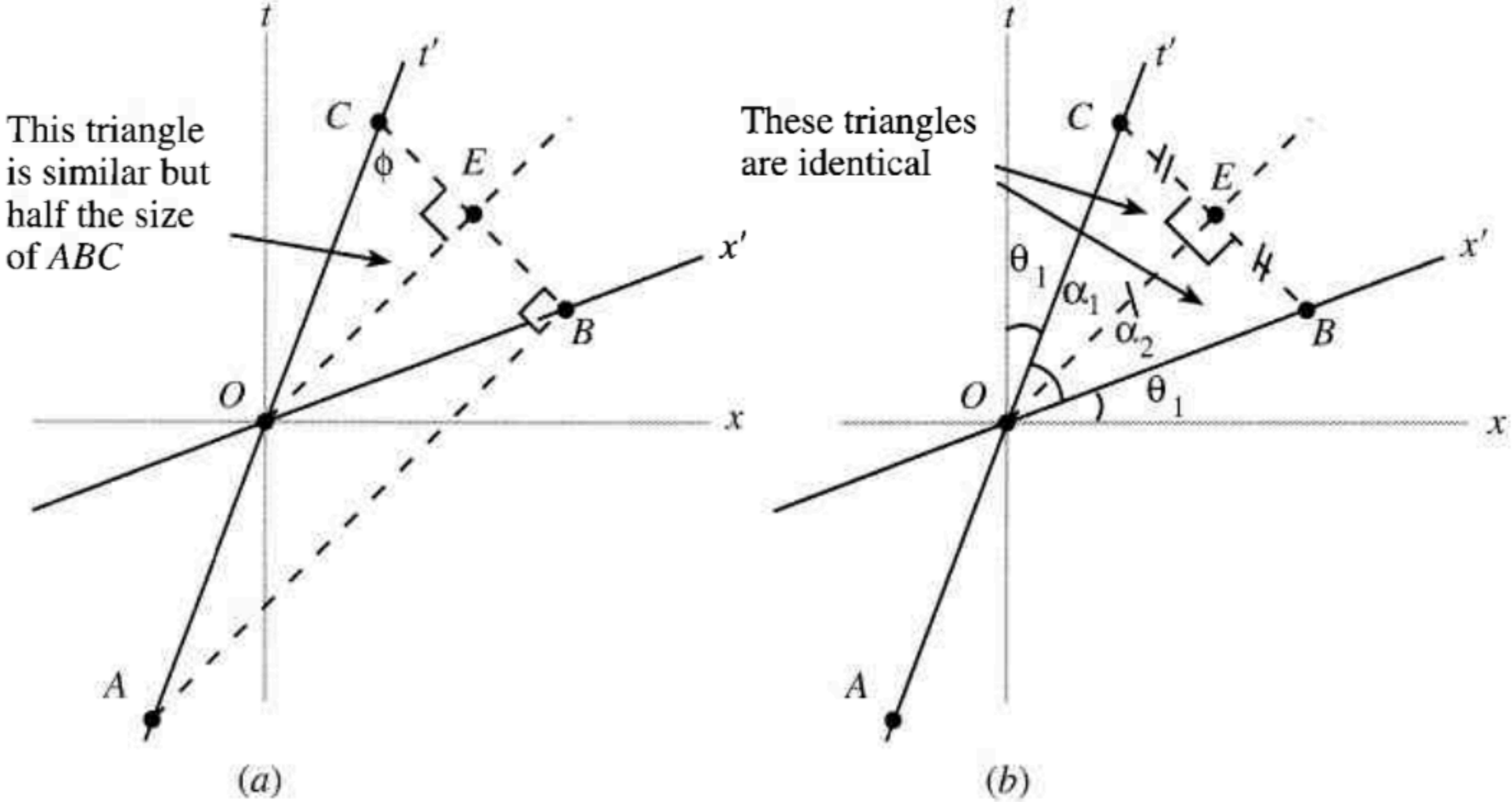
This new light flash has no physical importance: it just makes the following argument simpler.

Since light-flash worldlines always have a slope of ±1, they all make a 45° angle with respect to the vertical or horizontal directions.

This means that if light-flash worldlines cross at all, they always cross at right angles (see Figure).



The spacetime diagram shown with the addition of a right-going light flash emitted by the master clock at the origin event O. Because the light-flash worldlines all have slopes of ±1, they al make an angle of 45° with respect to the vertical. This means that these worldlines always cross at right angles



(a) Triangles ABC and OEC are similar (they are right triangles and share the common angle ϕ). OEC is half the size of ABC, since line AO has the same length as line OC. But this means that line BC is twice as long as line EC, implying that line BE has the same length as line EC. (b) But this implies that triangles OEC and OEB are identical since they are right triangles whose corresponding legs are equal in length. Therefore the two a angles are equal, implying that the two e angles are equal. Note also that the length of line OC is equal to the length of line OB.

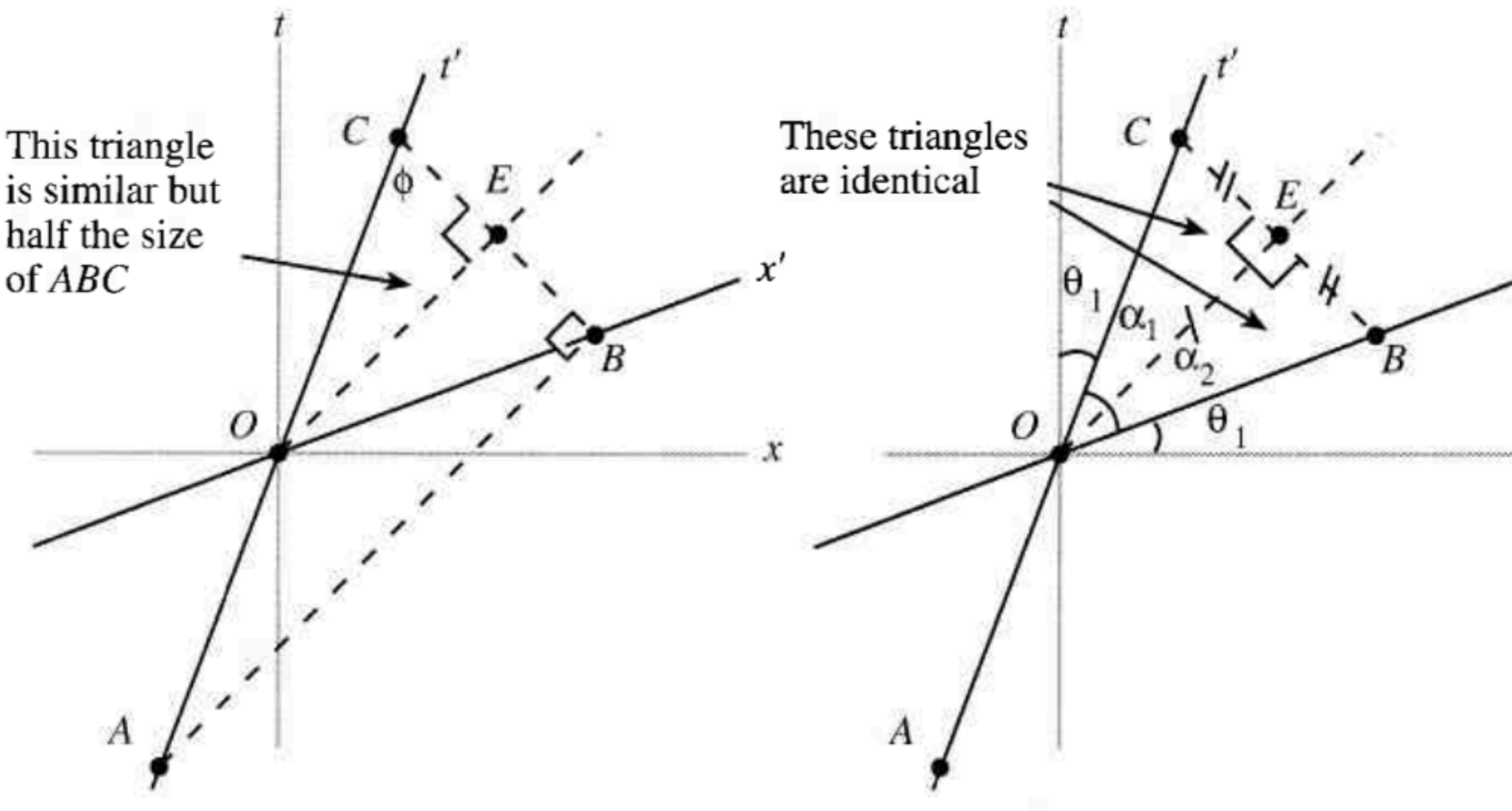
Now, I claim that triangles ABC and OEC are similar triangles: they are right triangles that share the common angle ϕ(see Figure Part (a) above).

Moreover, the hypotenuse of ABC is twice as long as that of OEC, since A and C are symmetrically placed about the event O.

This means that the triangle ABC must be exactly twice as large as the triangle OEC, implying that the line BC must also be twice as large as the line EC (remember that if two triangles are similar, the lengths of their corresponding sides are proportional).

But if line BC is twice as large as EC, then the length of line BE must be equal to the length of line EC (see above Figure Part (b))

But this means that triangles OEC and OEB must be identical, since they are both right triangles and their corresponding legs are equal in length.



(a) (b)

This means that $\alpha_1 = \alpha_2$, which in turn means that $\theta_1 = 45° - \alpha_1$ is equal to $\theta_2 = 45° - \alpha_2$.

Thus the diagram x' axis makes the same angle with the diagram x axis that the t' axis makes with the t axis, as previously asserted.

Another important consequence is that the length of the line OC (which represents the coordinate time interval $\Delta t' = +T$) is the same as the length of the line OB (which represents the coordinate displacement $\Delta x' = T$, the distance that the light signal had to travel to get to event B).

This means that the scale of both axes must be the same; that is, the spacing of marks on the diagram x' axis will be exactly the same as the spacing of marks on the t' axis!

Note that $\tan \theta_1$ = run/rise for the t' axis = 1/slope of t' axis = β. Note also that $\tan \theta_2$ = rise/run for the diagram x' axis = slope of diagram x' axis.

Since we have just seen that $\tan \theta_1 = \tan \theta_2$, we have

$$\text{Slope of } x' \text{ axis} = \beta$$

So, to be consistent with the principle of relativity, we must draw the Other Frame t' and diagram x' axes with slopes $1/\beta$ and β, respectively.

READING THE TWO-OBSERVER DIAGRAM

In summary, what have we discovered?

To construct a two-observer spacetime diagram, we first construct the Home Frame's t axis and diagram x axis perpendicular with each other (with the t axis vertical).

Then we draw the t axis of the Other Frame with slope $1/\beta$ and the diagram x' axis of the Other Frame with slope β.

We then calibrate the Other Frame time axis with marks that are separated vertically by $\Delta t = \gamma \Delta t'$ (where $\Delta t'$ is the time interval between the marks in the Other Frame and $\gamma = 1/\sqrt{1-\beta^2}$.

Finally, we calibrate the diagram x' axis with marks separated by the same distance as marks on the t' axis (i.e., the marks should be separated horizontally by $\Delta x = \gamma\, \Delta x'$, where $\Delta x'$ is the spatial separation between the marks in the Other Frame).

We can now find the t' and x' coordinates of any event on the diagram as follows.

The t' axis is by definition the line on the spacetime diagram connecting al events that occur at x' = 0.

The line connecting al events that have coordinate x' = 1s (or any given value not equal to zero) will be a line parallel to the t' axis, because the Other Frame's lattice clock at x' = 1s moves at the same velocity as the master clock at x' = 0, and the latter's worldline defines the t' axis.

Similarly, the line on the diagram connecting al events that have the same t' coordinate must be a line parallel to the diagram x' axis (the line connecting al events having t' = 0).

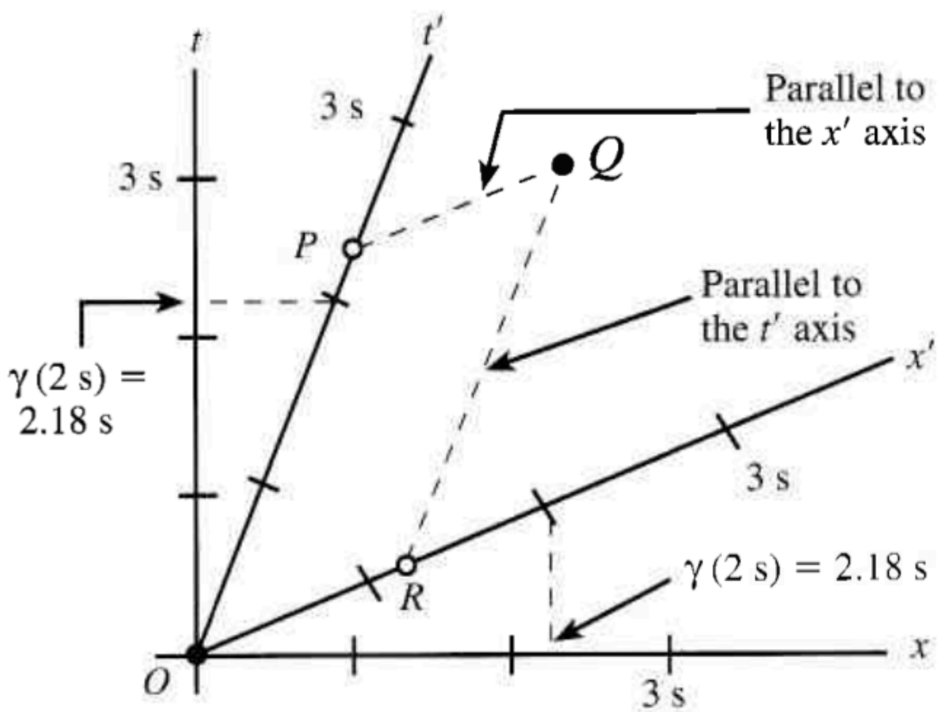
Here is an argument for this statement: if the line connecting al events that occur at t' = 1 s for example, were not parallel to the line connecting all events that occur at t' = 0, then these lines would intersect at some point on the diagram.

The event located at the point of intersection would therefore occur at both t' = 1 s and t' = 0, which is absurd.

Therefore a line connecting events having the same t' coordinate must be parallel to another such line.

So if the line connecting all events occurring at the same time in the Other Frame is parallel to the diagram x' axis and the line connecting all events occurring at the same place in that frame is parallel to the t' axis, we find the coordinates of an event in the Other Frame by drawing lines through the event that are parallel to the t' and diagram x' axes (and not perpendicular to them).

The places where these lines of constant x' and t' cross the coordinate axes indicate the coordinates of the event in the Other Frame (see Figure)

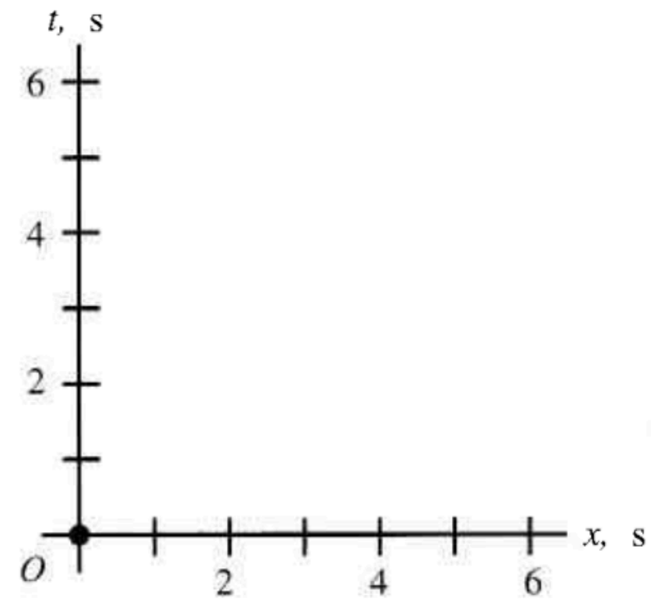


Events Q and R occur at the same place in the Other Frame, since all events that occur at the same place in that frame lie along a line parallel to the t' axis. Similarly, events P and Q occur at the same time in the Other Frame. Thus the time of Q is the same as the time of P in that frame (that is, $t'_Q = 2.3$ s in this case), and the position of Q is the same as the position of R (that is, $x'_Q = 1.2$ s in this case). The relative speed of the frames is $\beta = 2/5$ in the case shown.

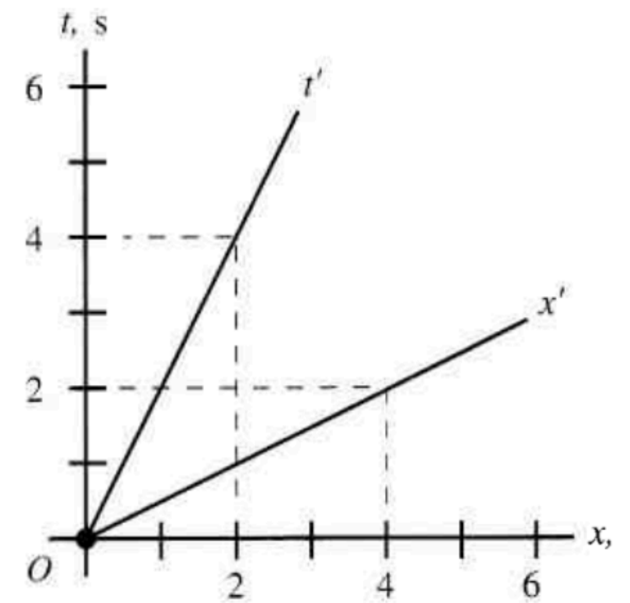
Finding the coordinate values by dropping parallels instead of perpendiculars may seem strange to you, and will probably take some getting used to.

Nonetheless, I hope you see from the argument above that dropping "parallels" is the only way to read the coordinates that makes any sense in this case.

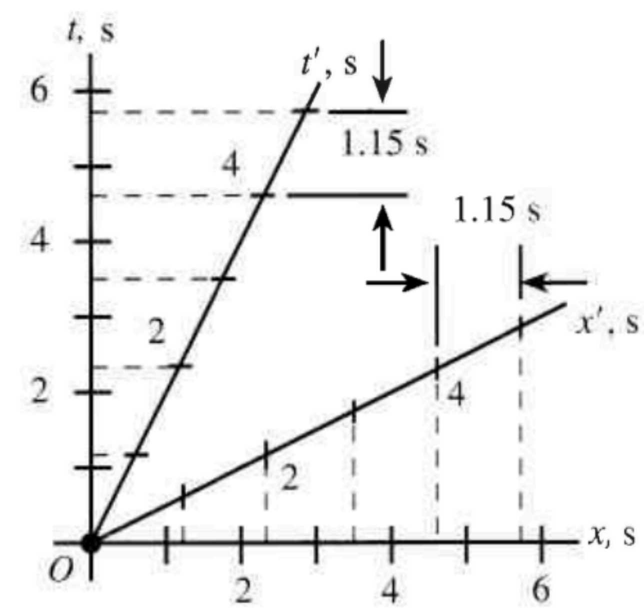
SUMMARY: DRAWING A TWO-OBSERVER SPACETIME DIAGRAM



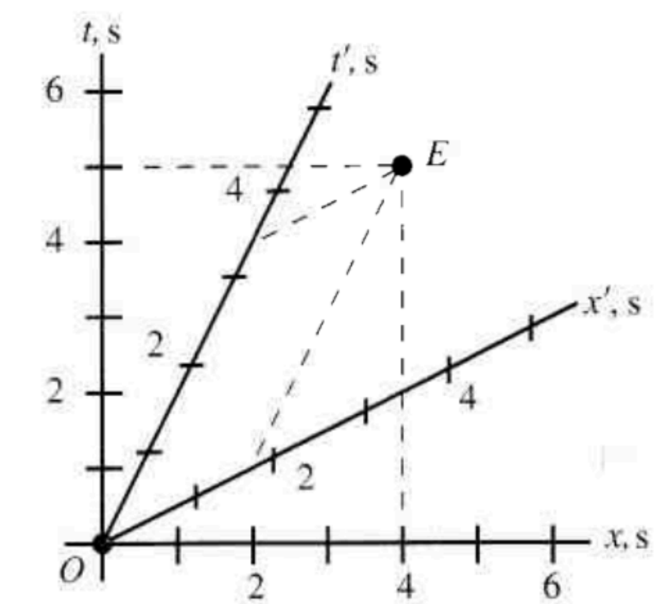
(a) Choose one frame to be the Home Frame. Draw its axes in the usual manner, and indicate the position of the origin event O. Calibrate the axes with some convenient scale.



(b) Draw the t' axis of the Other Frame from the origin event O with a slope $1/\beta$ (where β is the x velocity of the Other Frame with respect to the Home Frame). Draw the diagram x' axis of the Other Frame with slope β. This part and subsequent parts are drawn assuming that $\beta = 1/2$.



(c) Calibrate the t' axis with marks that are vertically separated by $\Delta t = \gamma \Delta t'$, where $\Delta t' = 1$ s in this case. Calibrate the diagram x' axis with marks having the same spacing. When $\beta = 1/2$, $\gamma \approx 1.15$.



(d) Read the coordinates (in either frame) of any event E by dropping parallels from the event to the appropriate axes. (In this case, E has the coordinates $t \approx 5$ s, $x \approx 4$ s in the Home Frame and $t' \approx 3.4$ s, $x' = 1.7$ s in the Other Frame.)

THE LORENTZ TRANSFORMATION EQUATIONS

The two-observer spacetime diagram discussed in the preceding discussions provides a very visual and intuitive tool for linking the coordinates of an event measured in one inertial reference frame with the coordinates of the same event measured in another inertial frame.

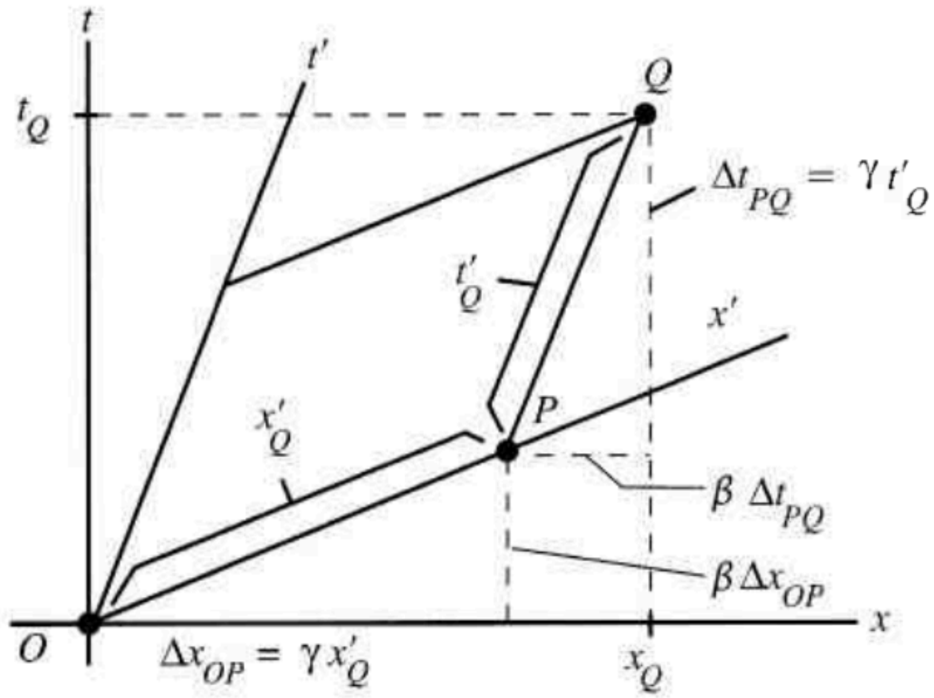
Because it is visual in nature, it is much more immediate and less abstract than working with equations. But this tool does lack one thing: the quantitative precision that only equations can provide.

We now develop (rederive) a set of equations that link the coordinates of an event measured in the Home Frame with the coordinates of the same event measured in the Other Frame.

These equations will do mathematically exactly what the two-observer diagram does visually.

These two tools together will enable us to discuss problems in relativity theory with both clarity and precision.

Now, consider an arbitrary event Q, as illustrated in the Figure.



Pick an arbitrary event Q. Then choose event P to occur at t' =0 (that is, on the x' axis) and at the same place as Q in the Other Frame. Note that since the line connecting events P and Q is parallel to the t' axis, its slope must be equal to $1/\beta$. Note also that the x' axis has slope β.

Imagine that we know the coordinates t'_Q and x'_Q of this event in the Other Frame.

This means that we can locate an event P which occurs at t' = 0(that is, on the diagram x' axis) and at the same place as Q in the Other Frame (that is, $x'_Q = x'_P$).

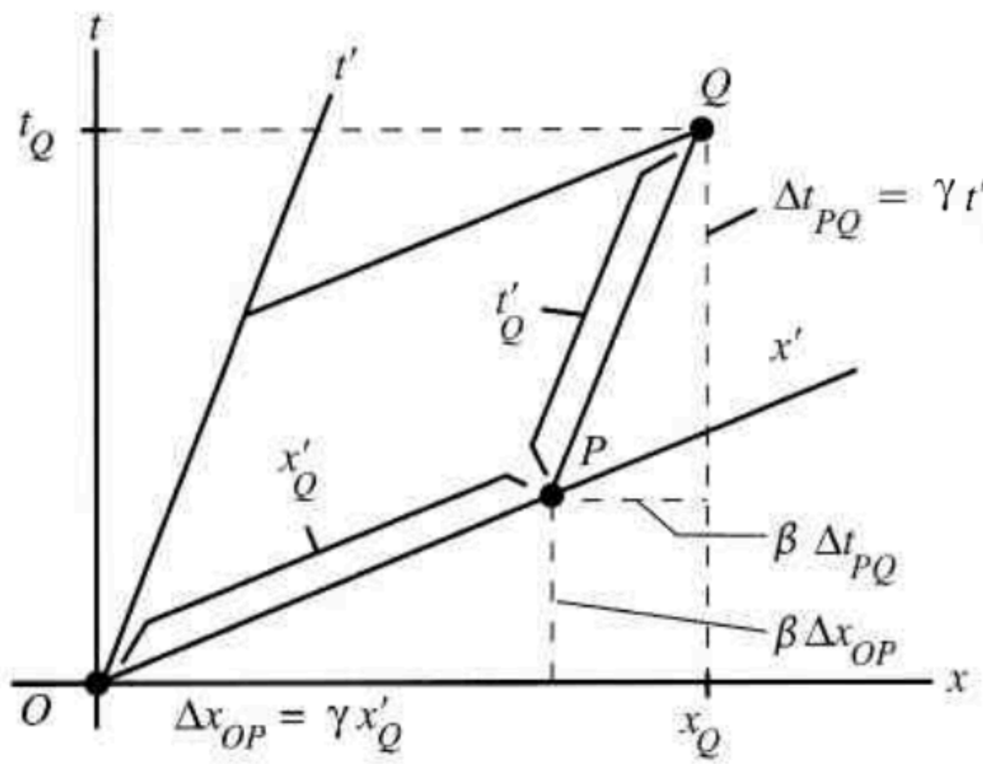
Let the time coordinate separation between events P and Q be Δt_{PQ} and the spatial coordinate separation between O and P be Δx_{OP} in the Home Frame.

Proper calibration of the Other Frame axes requires that $\Delta t_{PQ} = \gamma t'_Q$ and $\Delta x_{OP} = \gamma x'_Q$.

Also, since the line connecting events P and Q is parallel to the t' axis, its slope must be $1/\beta$, implying that the bottom leg of the triangle involving points P and Q has to have length $\beta\Delta t_{PQ}$.

Similarly, the slope of the diagram x' axis is β, so the vertical leg of the triangle involving points 0 and P must have length $\beta\Delta x_{OP}$.

All these things are shown in the Figure below.



Pick an arbitrary event Q. Then choose event P to occur at t' =0 (that is, on the x' axis) and at the same place as Q in the Other Frame. Note that since the line connecting events P and Q is parallel to the t' axis, its slope must be equal to $1/\beta$. Note also that the x' axis has slope β.

Now, you can see from the diagram that

$$t_Q = \Delta t_{PQ} + \beta\,\Delta x_{OP} = \gamma t'_Q + \gamma\beta x'_Q$$
$$x_Q = \Delta x_{OP} + \beta\,\Delta t_{PQ} = \gamma x'_Q + \gamma\beta t'_Q$$

Since the event Q is purely arbitrary, we can drop the subscript and simply say that the Home Frame coordinates t, x of any event can be expressed in terms of the Other Frame coordinates t' and x' of the same event as follows:

$$t = \gamma(t' + \beta x')$$
$$x = \gamma(\beta t' + x')$$

These equations are called the inverse **Lorentz transformation equations**.

The "just plain" Lorentz transformation equations, which express the Other Frame coordinates t', x' of an event in terms of the Home Frame coordinates t, x can be easily found by solving these equations for t' and x'.

The results are

$$t' = \gamma(t - \beta x)$$
$$x' = \gamma(-\beta t + x)$$

These equations can be easily generalized to handle events having nonzero coordinates y and z.

We learned earlier that if two inertial reference frames are in relative motion along a given line, any displacement measured perpendicular to that line has the same value in both frames.

Since frames in standard orientation move relative to each other along their common x axis, this means that

$$y' = y$$
$$z' = z$$

The Lorentz transformation equations are the relativistic generalization of the Galilean transformation equations.

THE TRANSFORMATION OF COORDINATE DIFFERENCES

Often we are not so much interested in the raw coordinates of an event as we are in the coordinate differences between two events.

Consider a pair of events A and B that are separated by coordinate differences $\Delta t = t_B - t_A$ and $\Delta x = x_B - x_A$ as measured in the Home Frame.

What are the corresponding differences $\Delta t' = t'_B - t'_A$ and $\Delta x' = x'_B - x'_A$ as measured in the Other Frame?

Applying Lorentz transformation equations to t'_A and t'_B separately, we get

$$\Delta t' \equiv t'_B - t'_A = \gamma(t_B - \beta x_B) - \gamma(t_A - \beta x_A)$$
$$= \gamma(t_B - \beta x_B - t_A + \beta x_A) = \gamma[(t_B - t_A) - \beta(x_B - x_A)]$$

Thus

$$\Delta t' = \gamma(\Delta t - \beta\, \Delta x)$$

Similarly,

$$\Delta x' \equiv x'_B - x'_A = \gamma(-\beta t_B + x_B) - \gamma(-\beta t_A + x_A)$$
$$= \gamma(-\beta t_B + x_B + \beta t_A - x_A) = \gamma[-\beta(t_B - t_A) + (x_B - x_A)]$$

Thus

$$\Delta x' = \gamma(-\beta\, \Delta t + \Delta x)$$
$$\Delta y' = y'_B - y'_A = y_B - y_A = \Delta y$$
$$\Delta z' = z'_B - z'_A = z_B - z_A = \Delta z$$

These are the Lorentz transformation equations for coordinate differences.

Note that they have the same form as the ordinary Lorentz transformation equations: one simply replaces the coordinate quantities with the corresponding coordinate differences.

The inverse Lorentz transformation equations for coordinate differences are analogous:

$$\Delta t = \gamma (\Delta t' + \beta \, \Delta x')$$
$$\Delta x = \gamma (\beta \, \Delta t' + \Delta x')$$
$$\Delta y = \Delta y'$$
$$\Delta z = \Delta z'$$

Minkowski Spacetime Diagrams or Visualizing the world of Special Relativity

You have already learned how to do this, so this will be a review of the ideas using a different approach to make sure we understand all aspects, add a few ideas and hopefully a reinforcement of the significance of the diagrams.

We can visualize the Lorentz transformation

by superposing the (x,ct) and (x',ct') planes into a common diagram

= a Minkowski or spacetime diagram :

Choose (x,ct) axes perpendicular(**always free** to do this for **one** set of axes).

Calibrate these axes(**arbitrary choice**).

Locate x' and ct' axes within framework of (x,ct) axes.

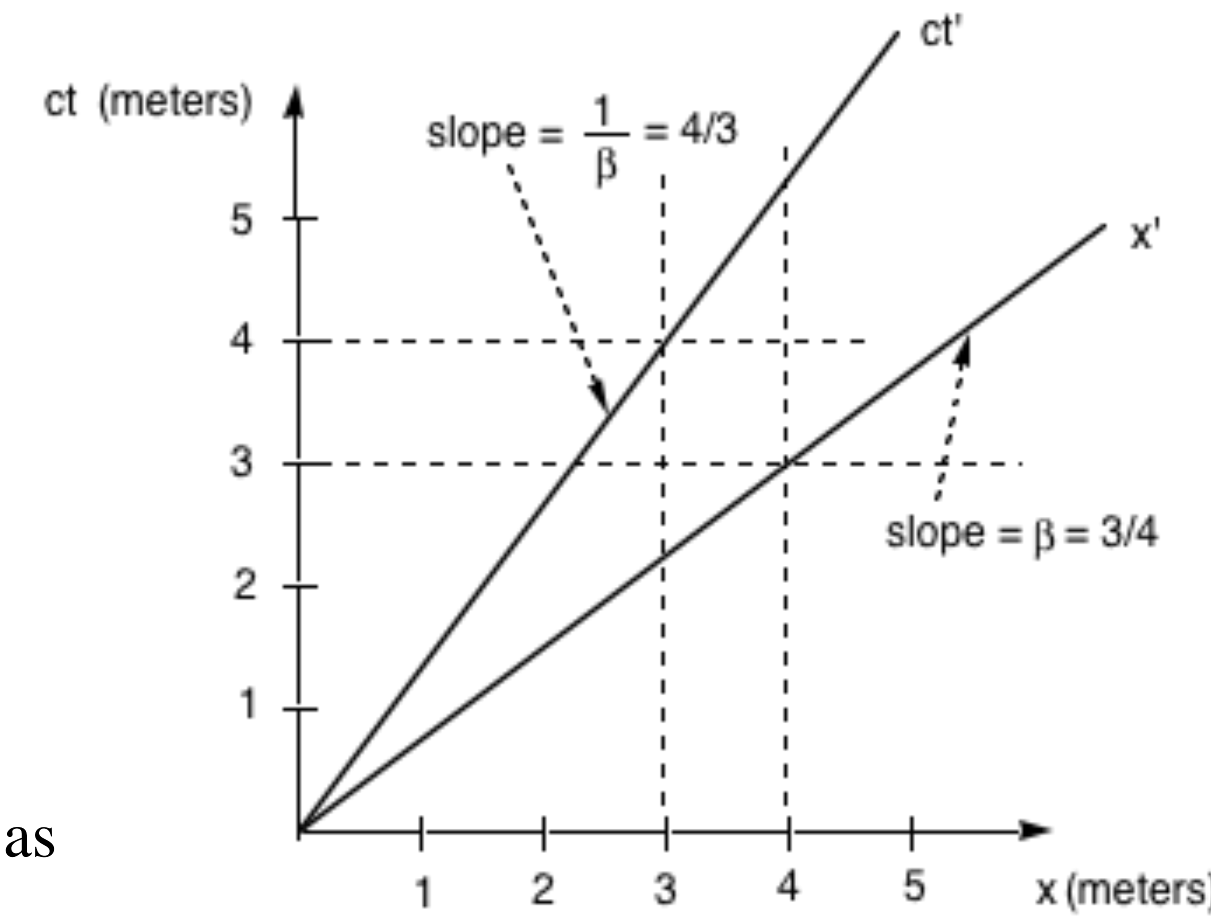
x' axis is line $ct' = 0$ and ct' axis is line $x' = 0$.

Diagram => Case $\beta = 3/4$

From Lorentz transformations these lines(axes) correspond to

$$x' = \gamma(x - \beta ct) = 0 \rightarrow ct = \frac{1}{\beta}x \rightarrow ct' - axis$$

$$ct' = \gamma(ct - \beta x) = 0 \rightarrow ct = \beta x \rightarrow x' - axis$$



x'–axis = straight line with slope β in (x,ct) plane

ct'–axis is straight line with slope $1/\beta$ in (x,ct) plane

Important result: We can can only choose 1 set of axes as perpendicular!!

We have no choice for 2nd set of axes if they are to coexist on same diagram!!

Now can see why we needed to make correct choice about parallel versus perpendicular for determining coordinates of event.

Wrong choice gives very different results for non-perpendicular axes.

We calibrate primed axes using invariance of interval:

Consider 2 events, (0,0), (x,ct) s.t.

$$(\Delta S)^2 = c^2(\Delta t)^2 - (\Delta x)^2 = c^2 t^2 - x^2 = -1$$

For 2nd observer, events = (0,0), (x',ct') s.t .

$$(\Delta S)^2 = c^2(\Delta t')^2 - (\Delta x')^2 = c^2 t'^2 - x'^2 = -1$$

where

$$ct' = \gamma(ct - \beta x)$$

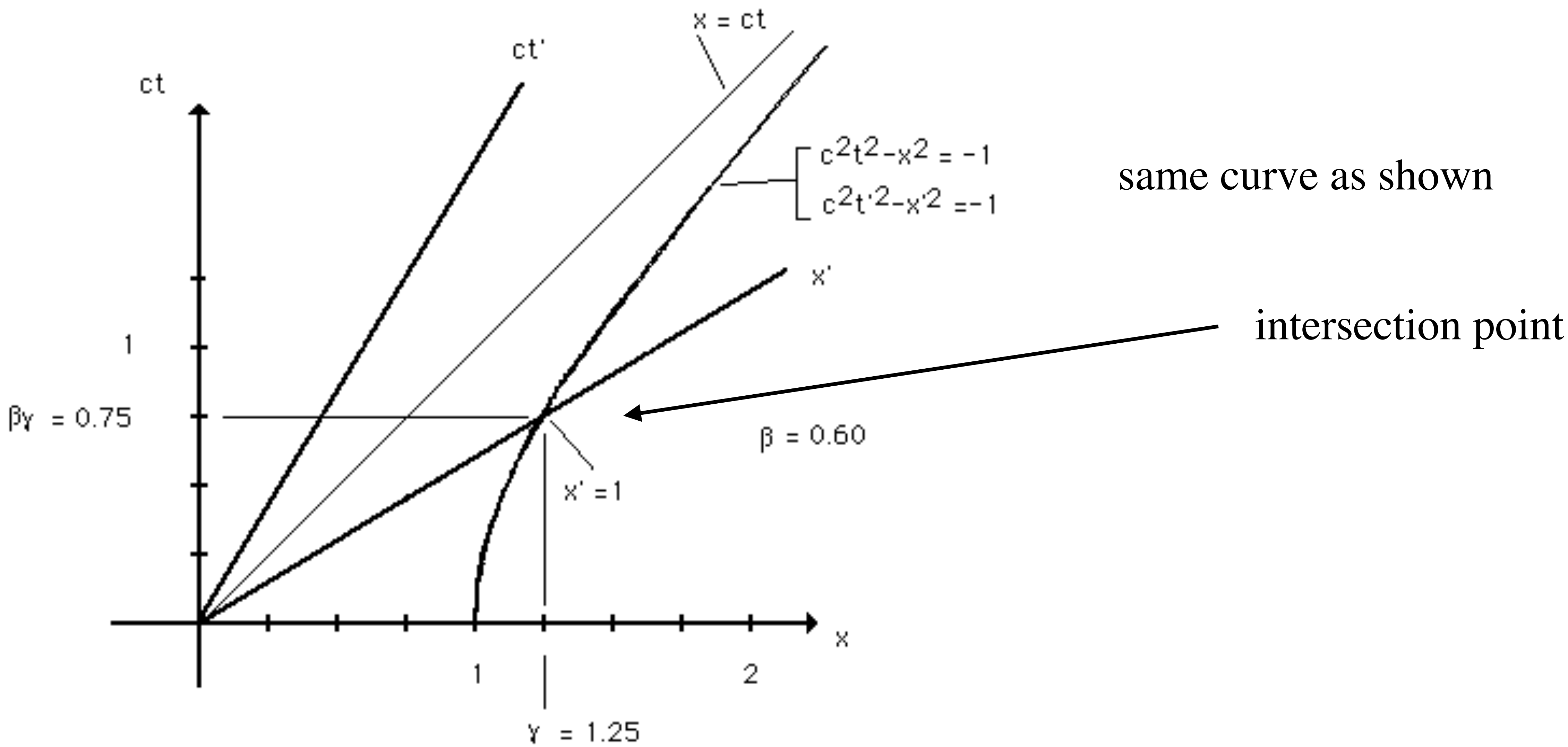
$$x' = \gamma(x - \beta ct)$$

We have used the invariance of spacetime interval in this calculation.

The set of all events that satisfy these equations is a curve on spacetime diagram.

The curve is hyperbola (as shown).

We are not surprised base on our earlier discussions.



It intersects x–axis at x = 1 (when ct = 0) and x'–axis at $x' = 1$ (when $ct' = 0$) and thus, it allows us to calibrate x'–axis once have calibrated the x–axis (or vice versa).

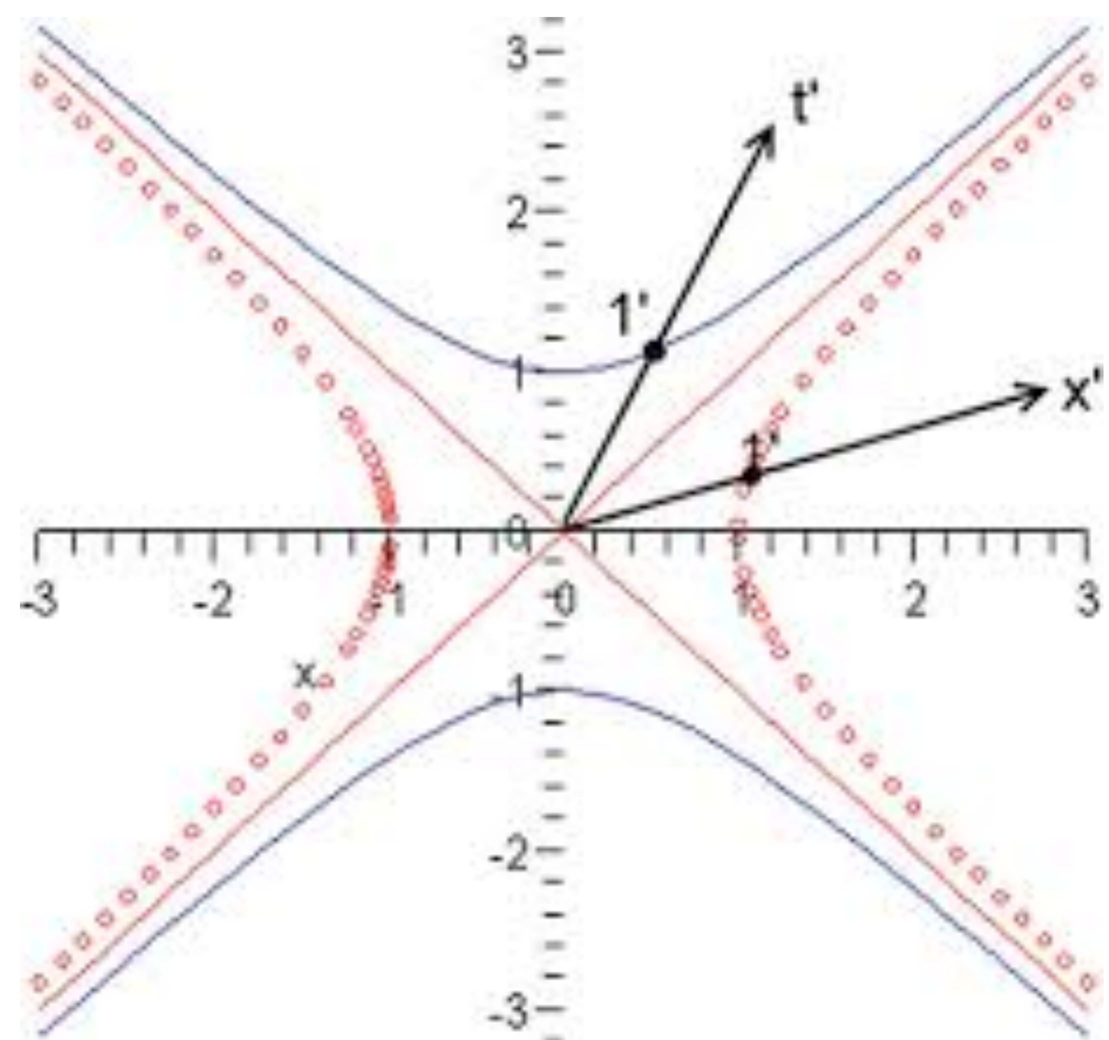
For diagram construction convenience, note that point (ct = βγ,x = γ) corresponds to intersection determining point $x' = 1$ as shown.

No need to draw curve!

Similarly, the ct'–axis is calibrated in terms of ct–axis using curves

$$(\Delta S)^2 = c^2(\Delta t)^2 - (\Delta x)^2 = c^2t^2 - x^2 = +1$$

$$(\Delta S)^2 = c^2(\Delta t')^2 - (\Delta x')^2 = c^2t'^2 - x'^2 = +1$$



It intersects ct–axis at ct = 1 (when x = 0) and ct'–axis at $ct' = 1$ (when $x' = 0$) and allows us to calibrate ct'–axis once have calibrated ct–axis (or vice versa).

For diagram construction convenience note that point (ct = γ, x = βγ) corresponds to intersection determining point $ct' = 1$.

Note that light rays are 45° lines on Minkowski diagram(because of scale choice).

One Can also Use an Experiment to Calibrate the Axes ……..

Alternatively, we can use experimental results to calibrate time axis

and

then assume by symmetry that space axis calibrates in same manner.

This experiment involves decay of elementary particle called mu-meson.

A Mu-meson is short-lived elementary particle produced

in large numbers at top of atmosphere

when the atmosphere is struck by a high-energy cosmic ray particle.

Mu-mesons can also produced in large numbers in an accelerator laboratory.

Experimentally, if mu-mesons produced in laboratory at rest ($v = 0$)

then live for very short time $\tau_0 = 2 \times 10^{-6}$ sec = 2 microseconds = 2μs = **lifetime at rest**.

Since no object can have speed greater than $c = 3 \times 10^8$ m/sec,

the maximum distance mu-mesons can travel during their lifetime

before they decay into electron and neutrino is about $c\tau_0 = 600$m.

In that calculation, we have explicitly assumed absolute time,

which says that lifetime of moving mu-meson

is same as that of mu-meson at rest (which we know is false).

1st experimental indication that absolute time was false concept

came from mu-mesons produced by cosmic rays.

Since they are produced at the top of the atmosphere

and

can only live to travel maximum of 600 m

and

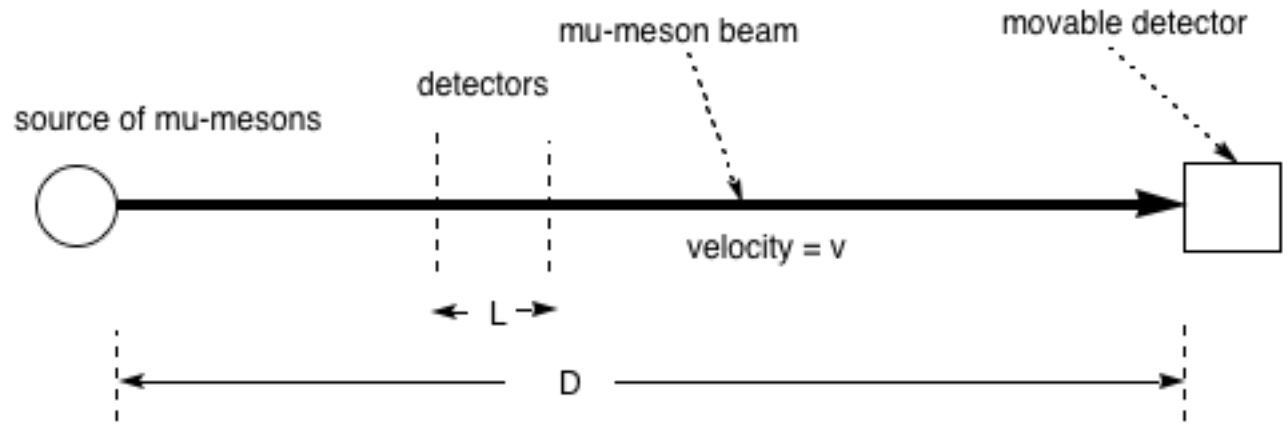
since atmosphere about 10000 m thick, **no mu-mesons should be observed on ground**

Experimentally, however, number at top is same as number at bottom.

Something is extending lifetime of mu-mesons.

In laboratory we can do this experiment with precision.

Setup shown:



Beam of mu-mesons sent from source to movable detector distance D away.

Along the way 2 detectors a distance L apart measure time Δt takes mu-mesons to travel the distance L

—-> determines muon velocity

$$v = \frac{L}{\Delta t}$$

If absolute time correct,

then after distance $d = v\tau_0$ all mu-mesons should decay

and

none should be seen in movable detector if D > d.

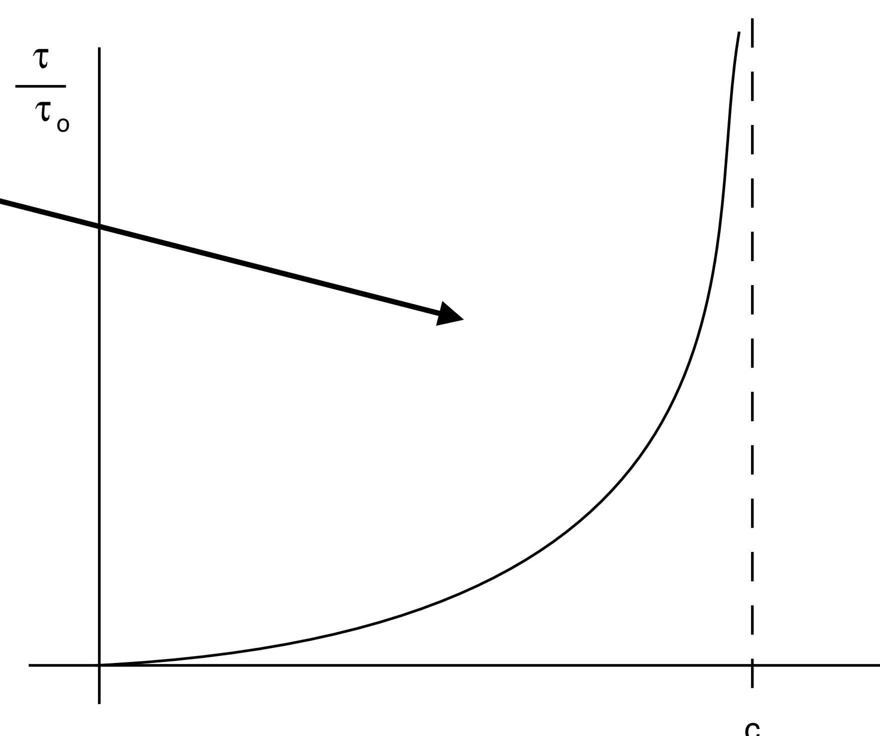
Experimental result is that mu-mesons actually travel maximum distance = $v\tau$

where τ is **lifetime of moving mu-meson**.

Experiments found that

Plot of result looks like:

$$\tau = \gamma\tau_0 = \frac{1}{\sqrt{1-\frac{v^2}{c^2}}}\tau_0$$



Lifetime observed to get larger and larger as velocity approaches speed of light!!!!

Muon speed never equals or exceeds c !!

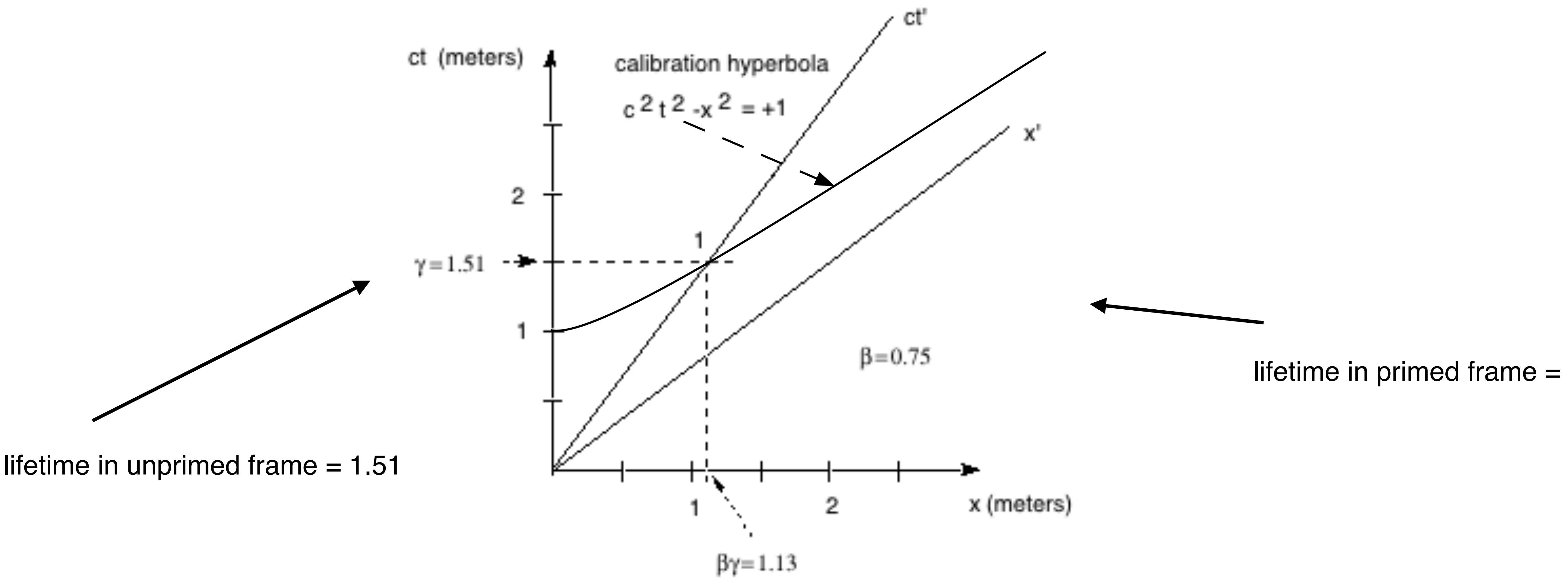
If let τ_0 = 1 tick of clock

(i.e., clock vanishes after single tick!)

and

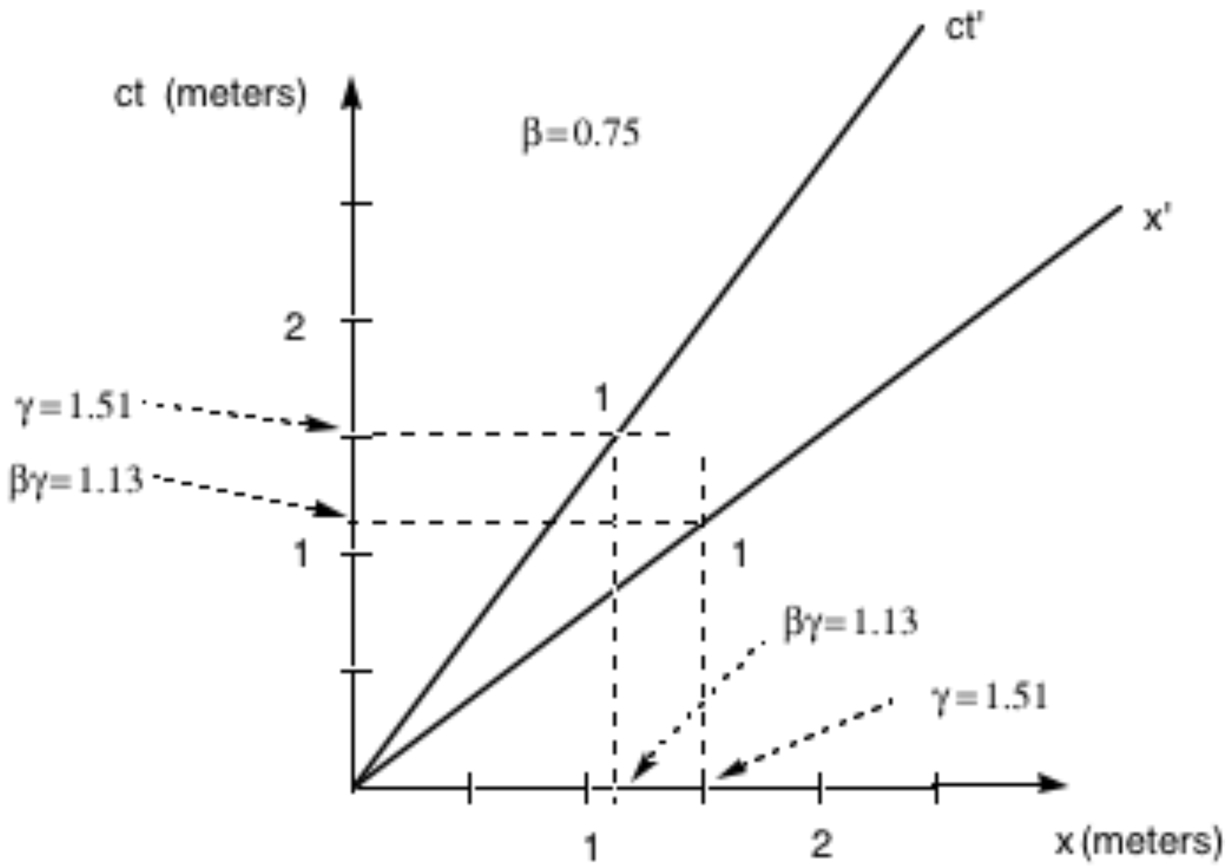
let mu-meson travel with the primed observer (at rest in primed frame),

then experimental results are represented as shown:



Clearly, calibration procedure using invariance of spacetime interval agrees with experimental result 1(primed) ——> 1.51 (unprimed).

Short cut calibration diagram shown:



Calibrations are identical!!!

We now have the theory = Special Relativity.

Can represent it either by

Lorentz transformations

and invariance of interval

or

by a Minkowski spacetime diagram.

All representations of theory equivalent. Einstein 1905.

What is a theory?

A theory is set of assumptions that agree with set of known experiments

and lead to predictions (correct) for all new experiments

Newton-Galileo Relativity lasted over 250 years before any experiment was sophisticated enough to show invalid.

Special Relativity has now lasted 119 years.

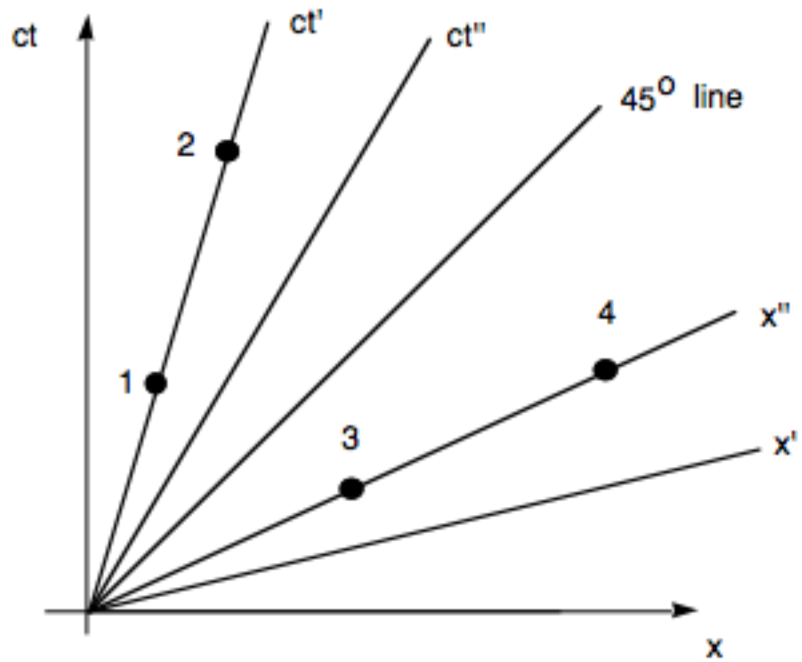
It has been subjected to significantly more experiments than was case for Newton-Galileo theory.

These experiments are also significantly more sophisticated and more precise.

What are the feature of the new theory?

The Strange World of Special Relativity

Relationships between Events



From diagram at left it is clear that:

For any pair of events like 1 and 2, i.e., $(\Delta s)^2 > 0$

it is always possible to find some observer

(—> new ct$'$–axis as shown)

such that 2 events takes place at same location

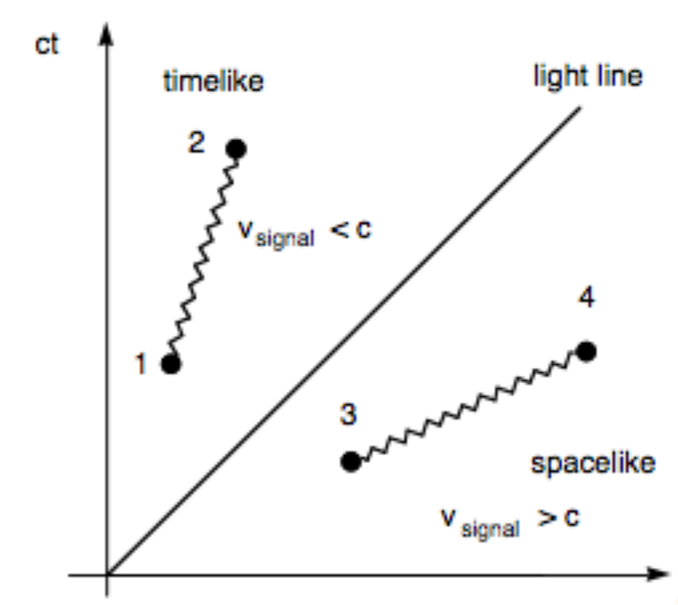
-> a pure time interval -> timelike events (reason for name).

For pair of events like 3 and 4 it is always possible to find some observer, i.e., $(\Delta s)^2 < 0$

(-> new x'-axis) such that 2 events take place simultaneously

-> a pure space interval -> spacelike events (reason for name).

Time-like and spacelike events are radically different as diagram right clearly shows:

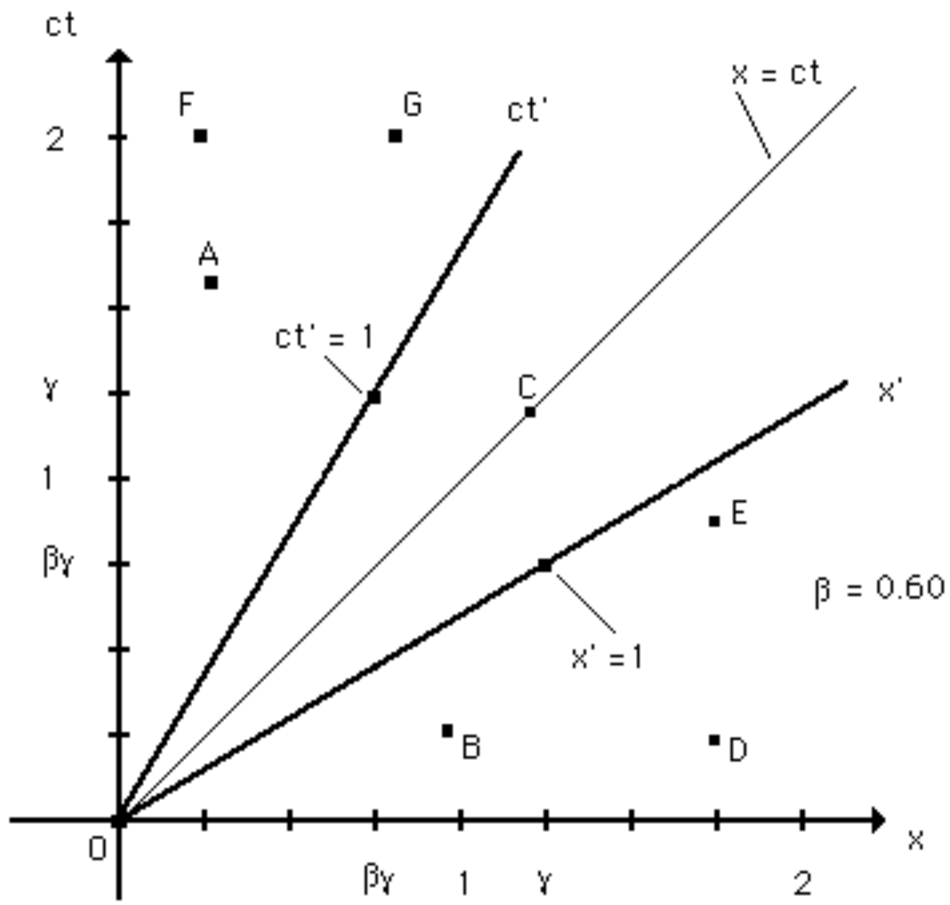


Event #2 -> timelike relative to event #1 -> is in future(later time) of event #1.

Event #4 -> spacelike relative to event #3 -> in future(later time) of event #3.

Events 1 and 2 can be connected with signal traveling with speed less than light. But, events 3 and 4 would require signal speed greater than light.

Now consider events labeled O, A, B, C, D, E, F, and G on spacetime diagram below:



Corresponding intervals have following properties:

$$(\Delta S)^2_{AO} = c^2(t_A - t_O)^2 - (x_A - x_O)^2 > 0 \rightarrow \text{ a } \textbf{timelike} \text{ interval}$$

$$(\Delta S)^2_{DO} = c^2(t_D - t_O)^2 - (x_D - x_O)^2 < 0 \rightarrow \text{ a } \textbf{spacelike} \text{ interval}$$

$$(\Delta S)^2_{CO} = c^2(t_C - t_O)^2 - (x_C - x_O)^2 = 0 \rightarrow \text{ a } \textbf{lightlike} \text{ or } \textbf{null} \text{ interval}$$

$$(\Delta S)^2_{FG} = c^2(t_F - t_G)^2 - (x_F - x_G)^2 < 0$$

Note that F and G are **simultaneous** in the (x,ct) frame

$$(\Delta S)^2_{ED} = c^2(t_E - t_D)^2 - (x_E - x_D)^2 > 0$$

Note that E and D are at the **same** place in the (x,ct) frame

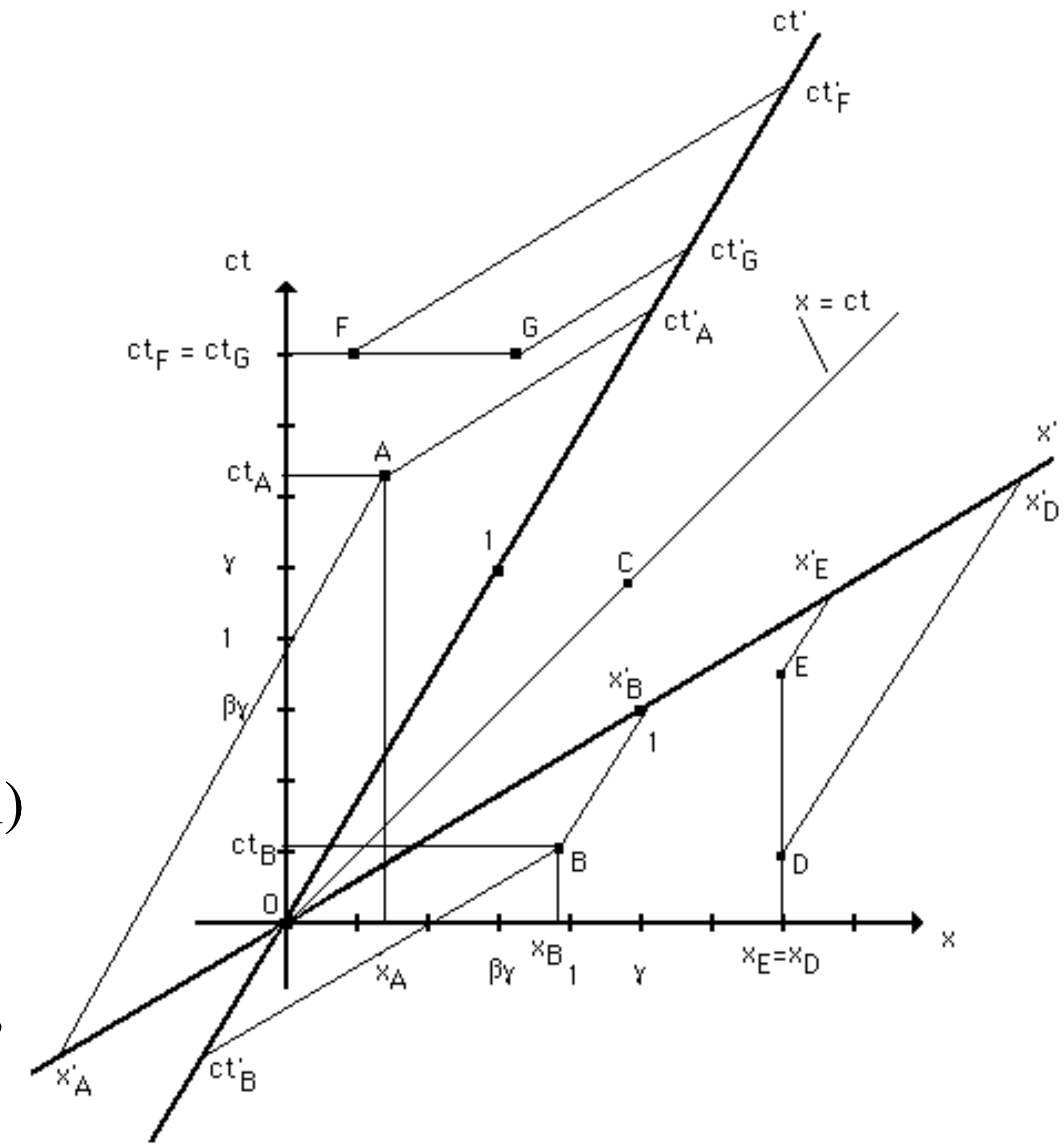
Consider the same set of events from viewpoint of an (x', ct') frame.

Look at diagram right where we have marked all coordinates.

Looking carefully at diagram we can draw following conclusions:

Events simultaneous in one frame are not necessarily simultaneous in other frames (events F and G)

simultaneity is relative concept!



Events occurring at same place in one frame **do not** occur at same place in other frames (events E and D)

Time order of timelike events (events with timelike interval) **does not change** between frames (events 0 and A)

Time order of spacelike events (events with spacelike interval)

can be reversed (events O and B);

In (x,ct) frame B occurs after O, but in (x',ct') frame O occurs after B

Numerical values of spatial separations and time separations are **different** in different frames

Note that line $x = ct$, which represents light ray starting at origin in unprimed frame is line $x' = ct'$, which represents light ray starting at origin in primed frame - Light is **only** physical object that both observers see in **identical** fashion.

Consider in more detail the **reversal** in time order of 2 events.

Seems to be very serious problems since could possibly lead to a violation of idea of **causality**.

Concept of **causality** is connected with idea of **cause and effect**,

i.**e., that an event should not occur before own cause,**

for example, a firecracker should not explode before we light its fuse!

Suppose we have 2 events in (x,ct) frame with coordinates (x_1, ct_1) and (x_2, ct_2)

and suppose, in addition, that

$$\Delta x = (x_2 - x_1) > 0 \quad , \quad \Delta t = (t_2 - t_1) > 0$$

so that event 2 comes **after** event 1 in the unprimed frame.

Then Lorentz transformations give result (in (x', ct') frame) that

$$\Delta t' = \frac{1}{c}(ct'_2 - ct'_1) = \frac{1}{c}\left(\gamma(ct_2 - \beta x_2) - \gamma(ct_1 - \beta x_1)\right)$$

$$= \gamma\left((t_2 - t_1) - \frac{\beta}{c}(x_2 - x_1)\right) = \gamma\left(\Delta t - \frac{\beta}{c}\Delta x\right)$$

It is easy to see that $\Delta t'$ can be **negative**,

—> time order of 2 events would be **reversed** if 2 events are related such that

$$\Delta t - \frac{\beta}{c}\Delta x < 0 \text{ or } \frac{\Delta x}{\Delta t} > \frac{c}{\beta} > c$$

or events must be **connected** by signal with v > c

—> must be **spacelike** separated events!

For all timelike related pairs of events we have

$$\frac{\Delta x}{\Delta t} < c$$

—-> **cannot** reverse time order.

Important to note that **only** for timelike related events can event #1 cause event #2

(since all real world signals **must** have v < c).

Thus, all cause/effect related events cannot have time order reversed

preserving idea of causality.

Special relativity is **consistent** with causality **without** us having to impose consistency!

This usually the **sign** of a valid theory!

All spacelike related pairs of events have

$$\frac{\Delta x}{\Delta t} > c$$

—> time order might be reversed in different frames.

Since cannot be cause/effect related, this **does not** affect idea of causality.

It does, however, lead to number of strange **paradoxes**(as we will see later).

Earlier, we derived the metric equation, which links the spacetime coordinate differences between two events measured in any inertial frame with the frame-independent spacetime interval between those events.

Now we will use the metric equation to calculate the proper time that a clock traveling along various worldlines might measure between events along its worldline.

In a given inertial reference frame, the motion of a clock (and thus its worldline through spacetime) can be completely specified by stating its initial position and then describing its velocity $\vec{v}(t)$ as a function of time subsequently.

What we want to do now is express the proper time that would be measured by that clock between events along its worldline in terms of the clock's velocity $\vec{v}(t)$ in a given inertial frame and the coordinate time Δt between those events in the same frame.

We begin with the simplest case where $\vec{v}$ is constant and then move on to deriving a more general formula valid for any worldline.

PROPER TIME ALONG A STRAIGHT WORLDLINE

The metric equation links the spacetime interval Δs between two events (call them A and B) with the coordinate differences Δt, $\Delta x, \Delta y, \Delta z$ measured between these events in any given inertial reference frame (call this the Home Frame):

$$\Delta s^2 = \Delta t^2 - \Delta x^2 - \Delta y^2 - \Delta z^2$$

We can express this equation in a different form if we divide both sides by Δt:

$$\frac{\Delta s^2}{\Delta t^2} = 1 - \frac{\Delta x^2}{\Delta t^2} - \frac{\Delta y^2}{\Delta t^2} - \frac{\Delta z^2}{\Delta t^2} = 1 - \frac{\Delta d^2}{\Delta t^2}$$

Now, Δd is the spatial distance between the two events as measured in the Home Frame, and Δt is their separation in time.

The clock that measures Δs must be present at both these events by definition, so in the Home Frame this clock will be observed to cover the distance Δd between the events in the time Δt.

This means that speed (i.e., the magnitude of this clock's constant velocity vector), as determined by observers in the Home Frame, must be $v = \Delta d/\Delta t$.

Plugging this into the last equation and taking the square root of both sides and multiplying by Δt, we get

$$\Delta s = \sqrt{1 - v^2}\,\Delta t$$

This very useful equation links the spacetime interval Δs measured by an inertial clock present at two events with the coordinate time separation Δt between those events in some inertial frame and the speed v of the clock *as measured in the same inertial frame*.

One can see several things immediately from this equation.

(1) The spacetime interval between two events is always less than (or at best equal to) the coordinate time measured between those events in any given reference frame.

(2) As the speed of the clock measuring the spacetime interval approaches 1(the speed of light) in the inertial frame, the discrepancy between the spacetime interval and the coordinate time between the events can become quite large.

(3) Conversely, if the speed of the clock present at both events is small compared to the speed of light ($v \ll 1$), that clock will register almost the same time between them as measured in the inertial frame: $\Delta s = \Delta t$.

When applying this equation, it is important to remember two things.

First of all, coordinate time Δt and spacetime interval Δs represent the time interval between two events measured in two fundamentally different ways (just as the northward displacement Δy and the distance Δd represent two fundamentally different ways of measuring the spatial separation of two points on the earth's surface).

The coordinate time between events is measured with a pair of synchronized clocks in an inertial frame, while the spacetime interval is measured by an inertial clock present at both events.

One cannot use the equation to link readings on just any old set of clocks.

Second, the quantities Δs and Δt appearing on both sides of $\Delta s = \sqrt{1 - v^2}\Delta t$ in fact refer to the time interval between a specific pair of events measured in two different ways.

Perhaps the most common error made by beginners in applying the result is implicitly using different pairs of events to delimit the time intervals Δs and Δt.

To avoid making this error, you must always think carefully about exactly what events delimit Δs and Δt.

Note that in this case the spacetime interval Δs between the events is the same as the proper time $\Delta\tau$ that the inertial clock present at both events registers between the events (by definition).

Our result thus allows us to compute the proper time between the events in the special case where the clock moves between the events at a constant velocity.

A CURVED FOOTPATH

The equation we derived

$$\Delta s = \sqrt{1 - v^2}\,\Delta t$$

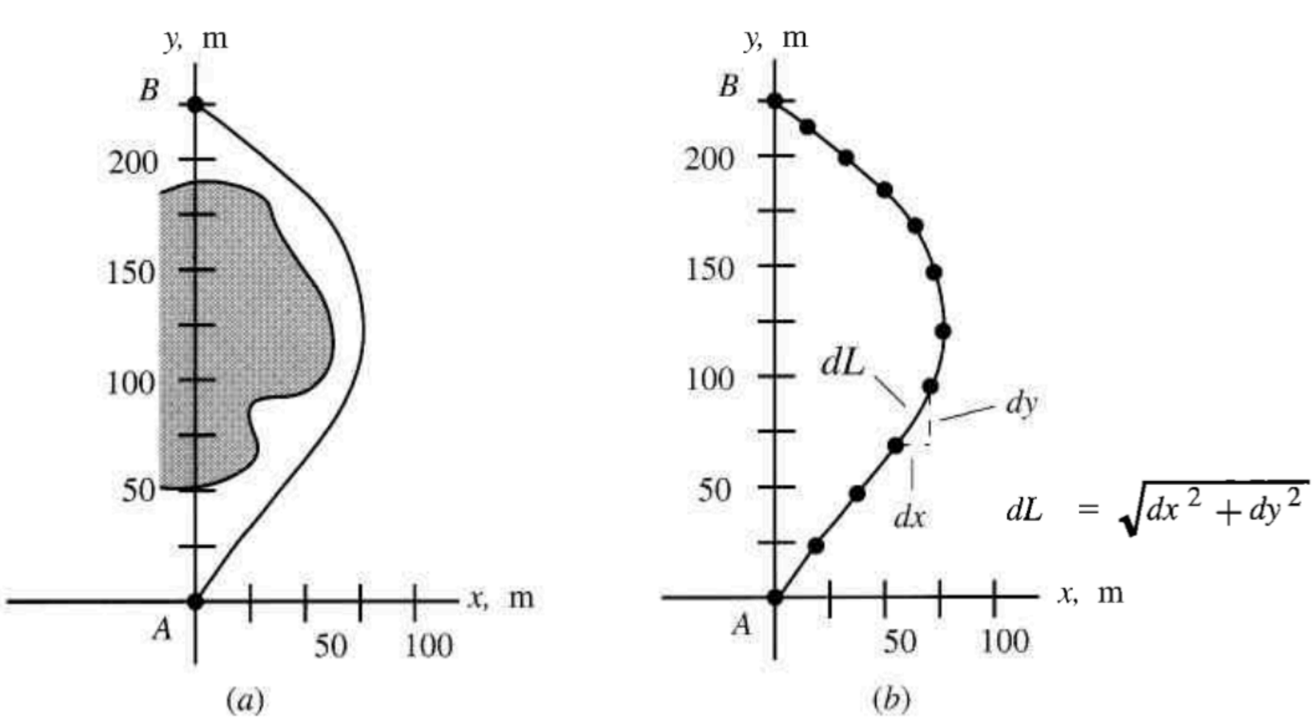
compares the coordinate time Δt between two events as measured in any given inertial reference frame with the proper time $\Delta\tau$ measured by an inertial clock present at both events ($\Delta\tau = \Delta s$ in this case).

Now, we will generalize this result to connect coordinate time to the proper time Δt measured by any clock present at both events, inertial or not.

It has sometimes been said that one must use the theory of general relativity to properly analyze the behavior of accelerating clocks.

In fact, we can quite adequately analyze the behavior of such clocks using only the metric equation if we simply remember the analogy between the proper time between events in spacetime and the pathlength between two points on a plane.

Consider a footpath around a small pond, which is illustrated in the Figure Part (a) by a scale drawing with a superimposed coordinate system.

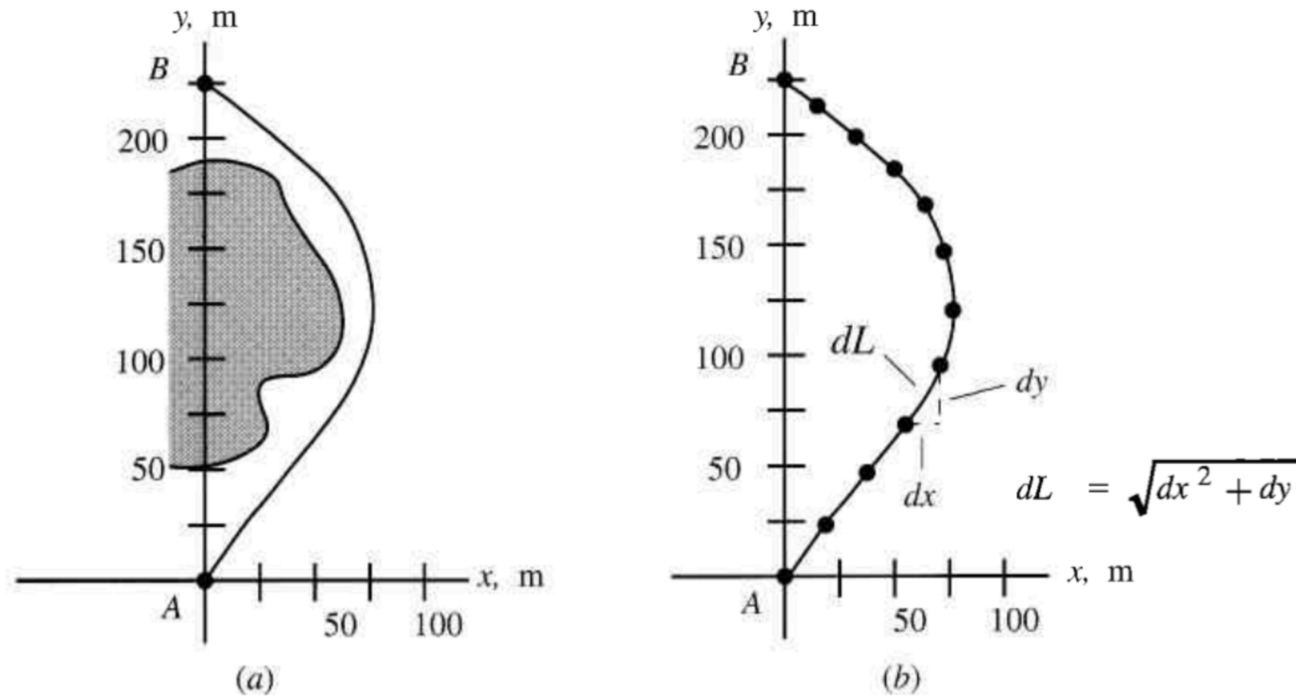


(a) Map of a path around a small pond, with a superimposed coordinate system shown. (b) We can compute the length of the path by subdividing the path into many infinitesimal, almost straight sections, finding the length dL of each section and summing to find the total length.

We could measure the length of this path from point A to point B with a long, flexible tape measure.

But once we have set up a coordinate system, we can also compute the length of the path in the following manner.

Imagine dividing up the path into a large number of infinitesimally small sections, as shown in Figure Part (b).



If we make these sections small enough, each will be approximately straight.

In this limit, the pathlength dL of a given segment as measured by a flexible tape measure will be almost equal to the straight length computed (using the pythagorean theorem) from the coordinate differences of the segment's endpoints:

$$dL^2 \approx dx^2 + dy^2 \qquad \text{or} \qquad dL = \sqrt{dx^2 + dy^2}$$

The total length ΔL_{AB} of the path from A to B is the sum of all the segment lengths, which in the infinitesimal-segment limit becomes the integral

$$\Delta L_{AB} = \int_{\text{path}} dL = \int_{\text{path}} \sqrt{dx^2 + dy^2}$$

Note that since the length dL of each segment is greater than its northward extension dy, the total pathlength between points A and B will be greater than the straight-line northward distance of 250m between A and B; quite generally, therefore, $\Delta L_{AB} \geq \Delta d_{AB}$.

If we think of the path as being specified by giving the x coordinate of each point on the path as a function of y [that is, the path is specified by the function x(y)], then we can write the integral above as a single-variable integral over x by pulling a factor of dy out of the square root:

$$\Delta L_{AB} = \int_{y_A}^{y_B} \sqrt{1 + \left(\frac{dx}{dy}\right)^2}\, dy$$

As we have discussed before, though this equation uses the coordinates x and y measured in a given coordinate system, the pathlength itself is an invariant quantity: we will get the same answer (the answer that a flexible tape measure would give) no matter what coordinate system we use.

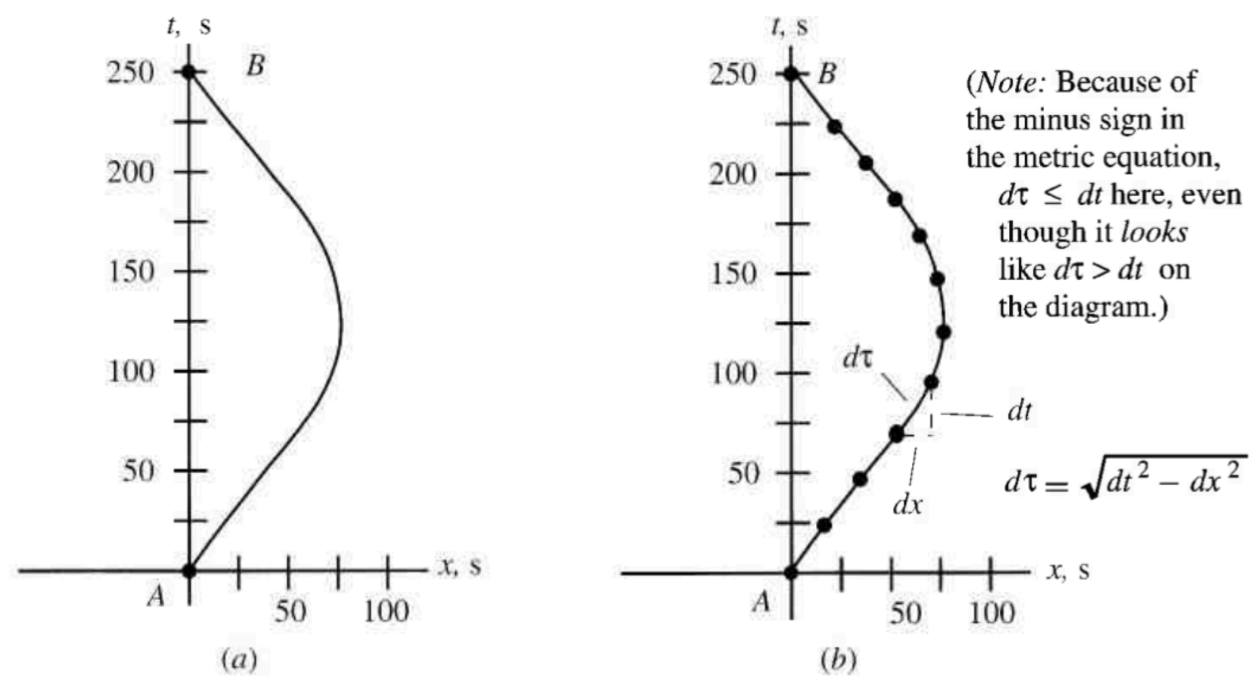
CURVED WORLDLINES IN SPACETIME

The analogy to worldlines in spacetime is direct.

Consider the worldline of a particle that travels out from the origin of some inertial reference frame a certain distance along the x axis and then returns.

Such a worldline is shown on the spacetime diagram in Figure Part (a) with the coordinate axes of that frame superimposed.

(a) Spacetime diagram of the motion of a particle's worldline based on measurements obtained in some inertial frame. (b) Computing the proper time along the worldline by subdividing it into many infinitesimal and almost straight segments, finding the proper time $d\tau$ elapsed during each segment, and then summing to find the total proper time.



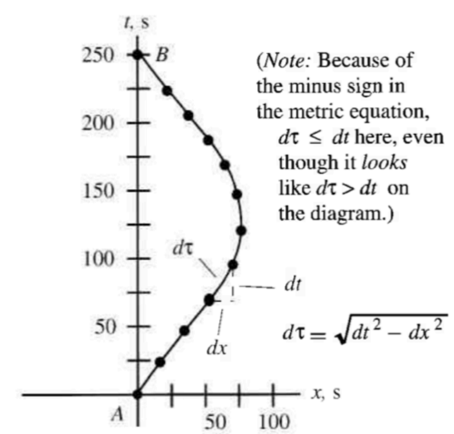
Such a worldline describes an accelerating particle; we can see from the graph that the particle's x velocity $v_x = dx/dt$ (which is the inverse slope of its worldline on the diagram) changes as time progresses.

A clock traveling with the particle measures the proper time $\Delta\tau_{AB}$ as between events A and B along this worldline (by definition of proper time).

But once we have measured the worldline of the particle in an inertial reference frame (any inertial frame), we can calculate what this clock will read between events A and B by using the metric equation in a manner analogous to our determination of the pathlength between points A and B earlier.

Imagine that we divide the particle's worldline up into many infinitesimal segments, each of which si nearly a straight line on the spacetime diagram (Figure Part (b)).

That is, each segment is chosen to be short enough so that the particle's velocity is approximately constant as it traverses that segment.



(b)

If a given segment is short enough so that the particle's velocity is almost constant along it, the proper time $d\tau$ that a clock would measure along that segment will be almost equal to the spacetime interval ds between the events that mark the ends of the segment, since the clock is present at both these events and travels between them with an almost constant velocity.

Therefore, by the metric equation

$$d\tau^2 \approx ds^2 = dt^2 - dx^2 - dy^2 - dz^2$$

Taking the square root and pulling out a factor of the coordinate time dt, we get

$$d\tau = \sqrt{dt^2 - dx^2 - dy^2 - dz^2} = \sqrt{1 - \left(\frac{dx}{dt}\right)^2 - \left(\frac{dy}{dt}\right)^2 - \left(\frac{dz}{dt}\right)^2}\, dt = \sqrt{1 - v^2}\, dt$$

In some sense, this result is just a differential version of earlier result, but its meaning is a bit more general.

It expresses the infinitesimal proper time $d\tau$ measured by a clock traveling between *two infinitesimally separated events* in terms of the coordinate time dt between those events measured in some inertial frame and the clock's instantaneous speed v measured in that frame.

The clock may be moving along any smooth worldline (with v not necessarily constant).

To find the total proper time measured between events A and B by a clock traveling along the worldline, we sum up the proper times measured for each segment of the worldline, which in the infinitesimal segment limit amounts to integrating the equation:

$$\Delta\tau_{AB} = \int_{\text{worldline}} \sqrt{1 - v^2}\, dt$$

If the speed v of the clock is expressed as a function of time v(t), then the integral is simply an ordinary one-variable integral with respect to t, which can be evaluated in principle.

This equation links the total proper time $\Delta\tau_{AB}$ as between two events measured by a clock traveling between those events with t and v, the coordinate time between the events and the speed of the clock, both as measured in some given (but arbitrary) inertial reference frame.

Though we are using an inertial frame to measure t and v, remember that the result above is frame-independent, since the proper time is measured by the clock in question directly without reference to any frame.

Note that since $\sqrt{1-v^2}$ is always ≤ 1, the proper time measured between two events along any path will be smaller than or equal to the coordinate time between those events measured in any inertial frame: $\Delta\tau_{AB} \leq \Delta t_{AB}$ under all circumstances.

If the speed v of the clock is constant, the integral can be done very easily:

$$\Delta\tau_{AB} = \sqrt{1-v^2}\int_{t_A}^{t_B} dt = \sqrt{1-v^2}\,\Delta t_{AB} \qquad \text{if } v = \text{constant}$$

Please note that "constant speed" here does not necessarily imply "constant velocity,"as the direction of a particle's velocity may change without changing its speed.

The equation can be applied to clocks traveling along straight or curved worldlines, as long as the speed of the clock remains fixed.

It may be used when the speed changes.

THE SPACETIME INTERVAL IS THE LONGEST PROPER TIME

Note that the result

$$\Delta\tau_{AB} = \int_{\text{worldline}} \sqrt{1 - v^2}\, dt$$

implies that generally the proper time measured by a clock between two events will indeed depend on the worldline that the clock follows between the events: specifically, the proper time depends on the particular way that the speed v of the clock varies with time.

This is analogous to the way that the pathlength between two points on a plane depends on the curvature of the path along which it is measured.

In euclidean geometry, the straight-line distance between two points is the shortest possible pathlength between the two points.

Now, we will prove that an inertial clock that travels between two events (which thus measures the spacetime interval between them) measures the *longest* proper time between those events, i.e., a longer time than any noninertial clock.

Theorem

The worldline of an inertial particle represents the path of *greatest* proper time between two events in spacetime: $\Delta s \geq \Delta\tau$ for all possible worldlines between the events.

Proof: Consider an arbitrary pair of events, A and B.

The time between these events is measured by two clocks that are present at both events: Clock I follows an inertial worldline between the events, while clock NI follows a noninertial worldline.

Since clock I is inertial, it can be used to define a Home Frame in which it is at rest.

Since it is inertial, its proper time $\Delta\tau_I$, will be equal (by definition) to the spacetime interval Δs_{AB} between the events.

We will take advantage of the fact that we can calculate the proper time for a given worldline using any inertial reference frame we please: the result will be frame-independent.

Let us choose to evaluate the proper times for clocks I and NI in the particular frame where clock I is at rest.

To compute the proper time along any path from event A to event B in this frame, we integrate the factor $\sqrt{1-v^2}$ (with v measured in the Home Frame) from t_A to t_B (also as measured in the Home Frame).

For the inertial clock I the integrand since that clock is at rest in the reference frame we are using (so v = 0).

Since clock NI travels along a different worldline by hypothesis, it must at least sometimes have a nonzero velocity in our chosen frame, implying that $\sqrt{1-v^2} < 1$ for at least part of the range of the integration.

Therefore, $\Delta s_{AB} = \Delta\tau_I = \int [1]dt > \int \sqrt{1-v^2}dt = \Delta\tau_{NI}$, (since both integrals have the same endpoints).

Since the values of the proper times are frame-independent, the same inequality must apply no matter what inertial frames we use to make the calculation.

Note that in a spacetime diagram based on measurements made ni an inertial reference frame, an inertial clock will have a straight worldline (since it moves with constant velocity with respect to any inertial frame) whereas a noninertial clock will have a curved worldline.

This theorem thus says that a straight worldline between any two events on a spacetime diagram is the worldline of greatest proper time between the events.

That the spacetime interval represents the longest proper time between events while the distance represents the shortest pathlength between points on a plane is a direct consequence of the negative signs that appear in the metric equation while only positive signs appear in the corresponding pythagorean relation.

This is another of the basic differences between spacetime geometry and the euclidean geometry of points on a plane.

Even so, there remains a similarity in that a straight worldline (or path) leads to an extreme value for the proper time (or pathlength).

The equation

$$\Delta s = \sqrt{1 - v^2}\, \Delta t$$

also implies that the coordinate time Δt between two events mea- sured ni any inertial reference frame is greater than (or equal to) the spacetime interval Δs between the points.

The three kinds of time interval that you can measure between two events thus stand in the strict relation

$$\Delta t \geq \Delta s \geq \Delta \tau$$

where the first inequality becomes an equality if the events occur at the same place in the inertial reference frame where Δt is measured (so that a clock in that frame can be present at both events) and the second inequality becomes an equality if the clock measuring the proper time follows an inertial path between the events (so that its proper time is the spacetime interval as well).

Light Cones

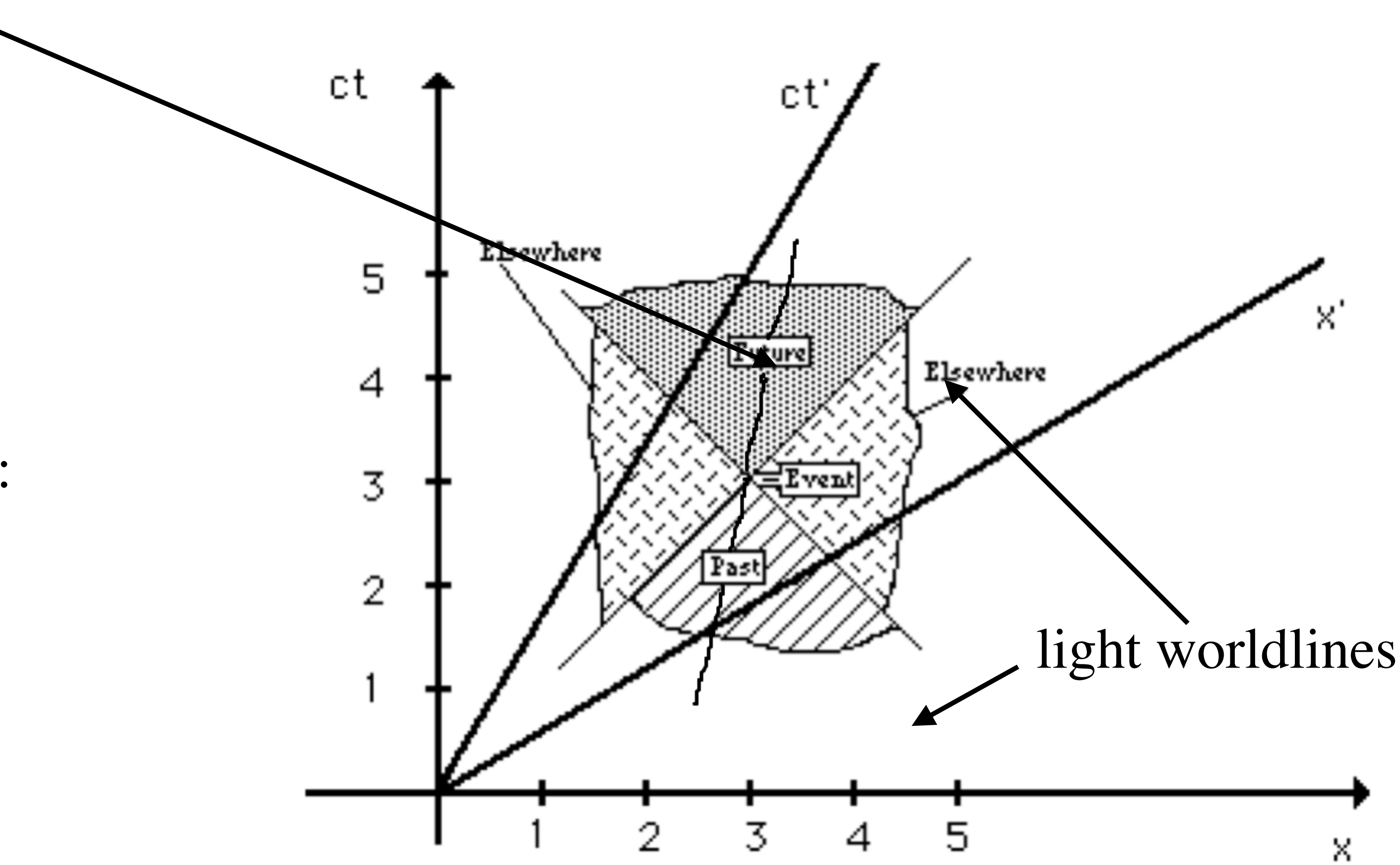
Another way to look at these ideas is via the concept of a **light cone**.

We will again make a quick pass through the idea so you are familiar with all the things involved and then go over everything again in much greater detail.

Since the maximum allowed speed for any physical object = speed of light c

we can use world lines of light emanating from event to delineate distinct regions of spacetime for any object **having that event on its worldline**.

Consider the diagram:



where you experience an event (on your worldline) as indicated.

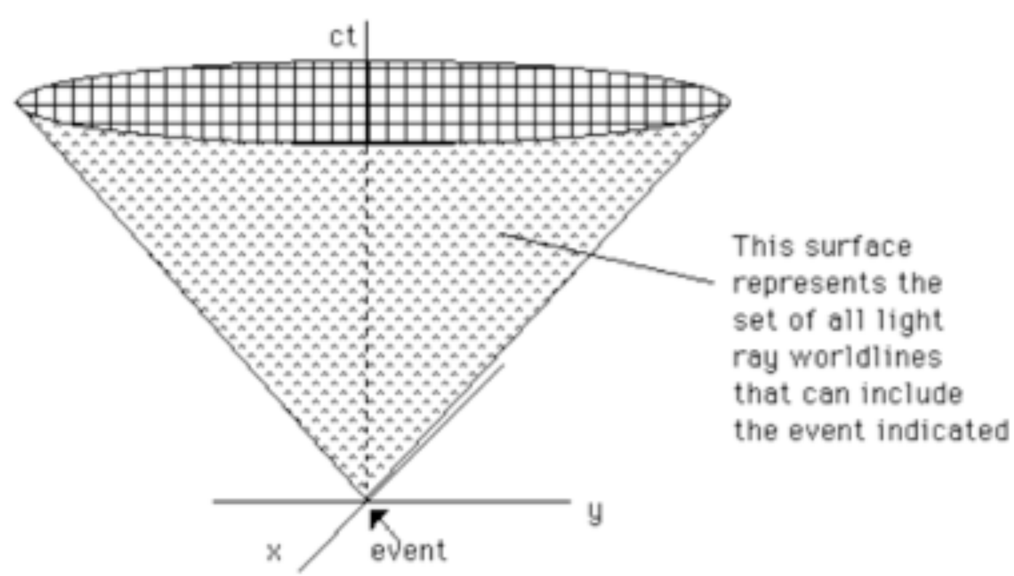
Since neither you nor any signals you send or receive can travel faster than light and since light ray worldlines containing this event are 45° lines as shown, region labeled future -> all events that you can either experience or influence with signal at later time(all events in this region are timelike separated from event you experienced), region labeled past -> all events you could have experienced or that could have influenced you(all events in this region are timelike separated from event you experienced).

Regions labeled elsewhere

are such that you can neither experience them nor influence them with any signal

(all events in these regions are spacelike separated from event you experienced).

If draw picture in 3-dimensional world (x, y, ct) then corresponding regions would look like:



then inside

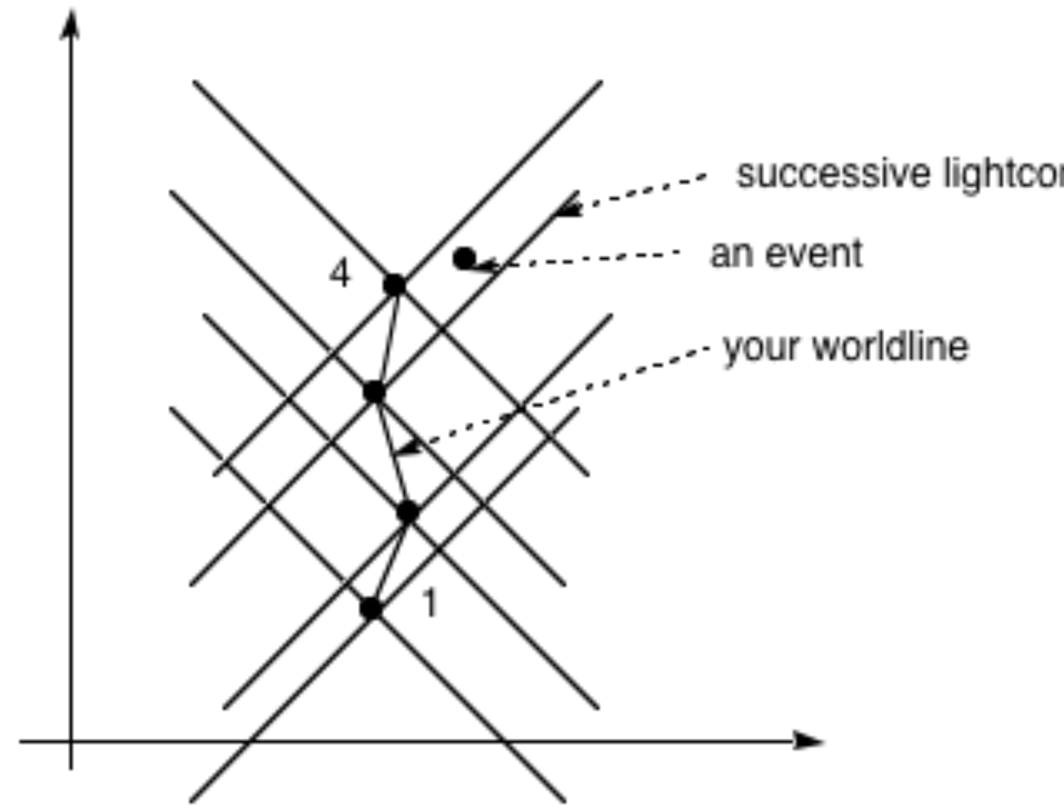
= future light cone!

-> name light cone.

What has happened to your possible future while we have been discussing these light cones?

Event labelled on diagram was in your possible future when you were experiencing event #1, but is no longer in your possible future when you are experiencing event #4.

So be careful about wasting time doing nothing!!



The Cosmic Speed Limit

Now we approach the light cone concept more formally.

"Nothing can go faster than the speed of light."

This fact is a commonly known consequence of the theory of special relativity.

But what is it about special relativity that requires this to be true?

Are there any loopholes in the argument that might make faster-than-light travel possible?

Earlier, it clear that the metric equation and the proper time equation break down if one attempts to apply them to a clock traveling faster than the speed of light: both equations would predict that the time registered between two events by such a clock would be an imaginary number, which is absurd.

This absurdity results from the violation of the $\Delta t^2 > \Delta d^2$ restriction necessary for the derivation of the metric equation.

Thus the metric equation really says nothing about what a clock traveling at faster than the speed of light will measure.

In fact, the question of what a clock traveling faster than light would measure is a moot issue in special relativity.

Now we will see (with the help of some two-observer diagrams) that it is a consequence of the principle of relativity that not only can no clock travel faster than the speed of light, but it is not possible for any object or even a message to travel faster than light!

The speed of light thus represents a true "cosmic speed limit."

In this context, we will also have to rethink the Galilean velocity transformation.

For example, if a particle is traveling at a speed of 0.9c in the +x direction in a frame that is itself traveling at a speed of 0.9c in the +x direction with respect to our frame, the Galilean velocity transformation equations imply that the particle would be measured to have a speed of 1.8c in our frame, a clear violation of the cosmic speed limit.

The Galilean velocity transformation equations are thus clearly wrong (as we already knew from the fact that the speed of light has the same value in every inertial frame).

But how can we arrive at the equations that take their place?

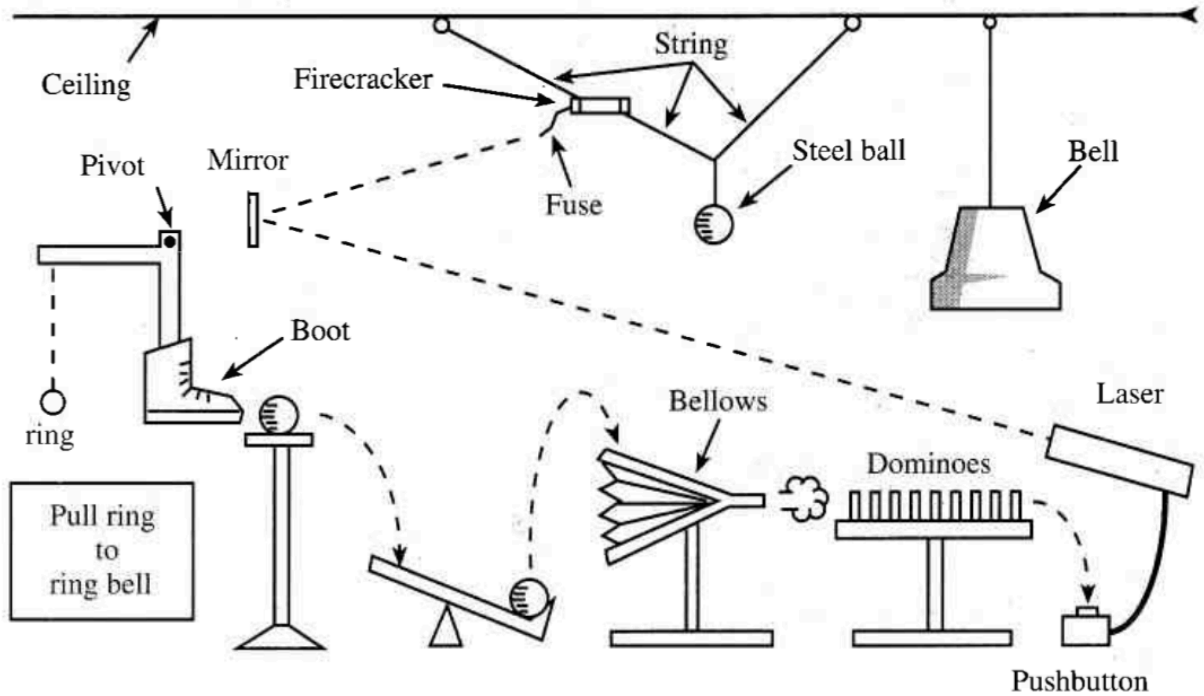
In our discussion of these issues, we will see that two-observer spacetime diagrams play a crucial role in vividly illustrating the basic physical issues involved.

CAUSAL CONNECTIONS AND THE ULTIMATE SPEED

The problem with faster-than-light travel is that it violates causality.

What do I mean by causality?

In physics (and more broadly, in daily life), we know that certain events cause other events to happen (Figure).



Causal connections.

For example, even couch potatoes know that if you press the appropriate button on the remote control, the TV channel will change.

These kinds of causal connections imply that certain causally connected events have an invariant temporal order: for example, we would be deeply disturbed if the TV channel changed just before we pressed the remote control button.

Consider two distinct events (call them P and Q, respectively) such that event P causes event Q.

That is, Q is caused to occur upon reception of some kind of information that P has occurred.

This information can be transmitted from P to Q in any number of ways: via some mechanical effect (such as the movement of an object or the propagation of a sound wave), via a light flash, via an electrical signal, via a radio message, etc.

Basically, the information can be carried by any object or effect that can move from place to place and is detectable.

Let us consider the TV remote control again as a specific example.

Imagine that you press a button on your TV remote control handset (event P).

The information that the button has been pressed is sent to the TV set in some manner, and in response, the TV set changes channels (event Q).

Keep this basic example in mind as we go through the argument that follows.

Now let us pretend that the "causal influence" that connects event P to event Q can flow between them at a constant speed v_{ci} faster than the speed of light as measured in your inertial reference frame, which we will call the Home Frame - perhaps the TV manufacturer has found some way to convey a signal from the remote to the TV using "Z waves" that travel faster than light.

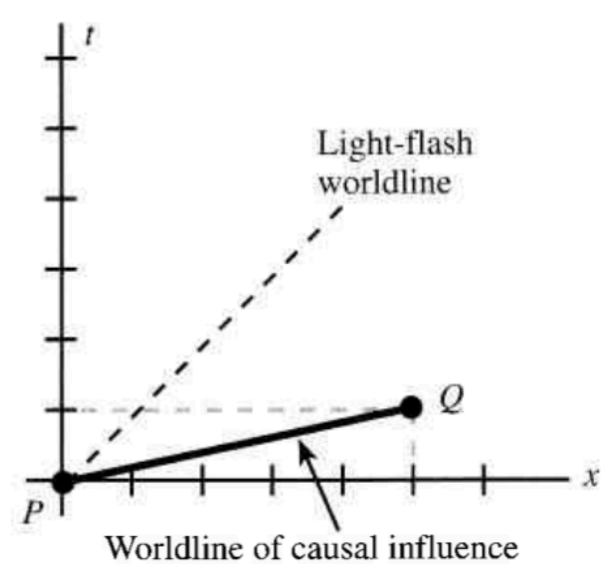
We will show that this leads to a logical absurdity.

Choose event P to be the origin event in that frame, and choose the x axis of the frame so that both events P and Q lie along it.

We can always do this: it is just a matter of choosing the origin and orientation of our reference frame.

Choosing the frame to be oriented this way is just a matter of convenience.

Figure shows a spacetime diagram (drawn by an observer in the Home Frame) of a pair of events P and Q fitting the description above.

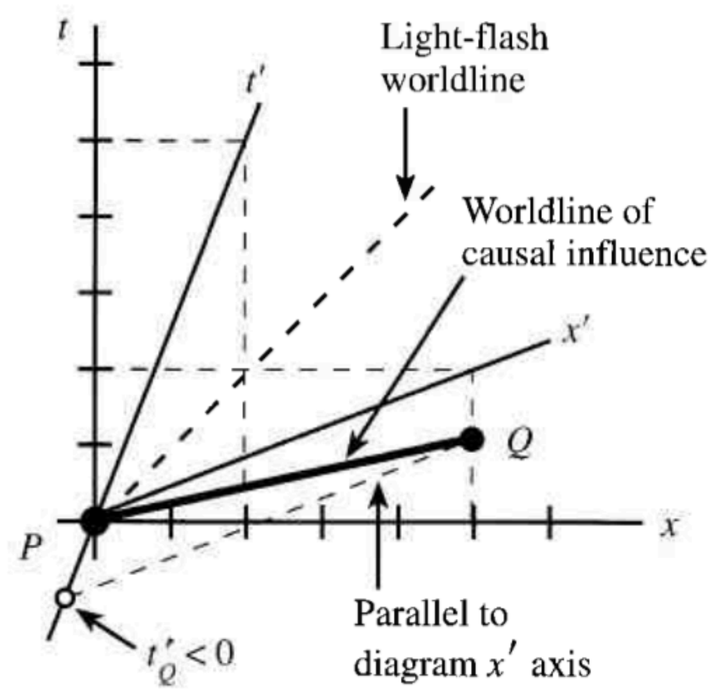


Imagine that events P and Q are connected by a hypothetical causal influence traveling with a speed v_{ci} faster than the speed of light. In the case shown, P has been chosen (for the sake of convenience) to be the origin event, and the influence travels 5s of distance for every second of time, so its speed is $v_{ci} = 5$ (that is, five times the speed of light).

Note that if the causal influence flows from P to Q faster than the speed of light, its worldline on the diagram will have aslope $1/v_{ci} < 1$, that is, less than the slope of the worldline of a light flash leaving event P at the same time (which is also shown on the diagram for reference).

Now consider the next Figure.

In this two-observer spacetime diagram, I have drawn the t' and x' axes for an Other Frame that travels with a speed β with respect to the Home Frame.



This two-observer spacetime diagram is the same as in the last Figure with the addition of the axes for an Other Frame that is moving with a speed β with respect to the Home Frame. In the drawing, β has been chosen to be 2/5, so the slope of the diagram x axis is 2/5 > 1/5, which is the slope of the causal influence worldline. Note that in the Other Frame, event Q will be measured to occur before event P.

Note that according to our earlier work, the slope of the diagram x' axis in such a diagram is β.

Note also that since the slope of the causal influence worldline is $1/v_{ci} < 1$, it is always possible to find a value of β such that $1/v_{ci} < \beta < 1$, meaning that the slope of the x' axis lies between the light-flash worldline and the causal connection worldline, as shown.

In such a frame, event Q will be measured to occur *before* event P, as one can see by reading the time coordinates of these events from the diagram.

Here is the absurdity: in the Other Frame, event P is observed to occur after event Q does.

This is absurd, because event P is supposed to cause event Q.

How can an event be measured to occur before its cause?

This is not merely a semantic issue, nor is it mere appearance.

According to any and every physical measurement that one might make in the Other Frame, event Q will really be observed to occur before its "cause" P.

To vividly illustrate the absurdity, consider our TV remote example.

If the signal could go from your remote control to the TV faster than light, in certain inertial reference frames, you would observe the TV set to change channels before the button was pushed.

If this were to happen in your reference frame, you would consider this a violation of the laws of physics (presuming your TV set was not broken).

But the laws of physics are supposed to hold in every inertial reference frame.

Therefore this observed inversion of cause and effect violates the principle of relativity!

We have only three options at this point.

We can reject the principle of relativity and start over at square one.

We can radically modify our conception of causality in a way that is yet unknown.

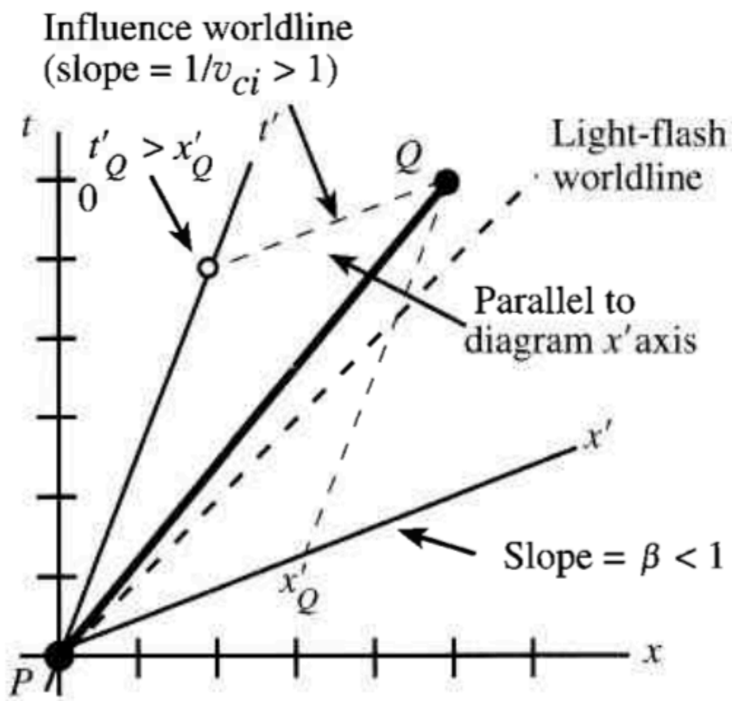
Or we can reject the assumption that got us into this trouble in the first place, namely, that a causal influence can flow from P to Q faster than light, i.e., with $v_{ci} > 1$.

The latter option is clearly the least drastic.

If information can only flow from P to Q with a speed $v_{ci} \leq 1$in the Home Frame, then the worldline of the causal influence connecting event P to Q will have a slope $1/v_{ci} > 1$,

Any Other Frame must travel with $\beta < 1$, by this hypothesis (since the parts of the reference frame, like any material object, could be the agent of a causal influence).

Therefore, the slope β of the Other Frame's diagram x' axis on a spacetime diagram must always be less than the slope $1/v_{ci}$ of the causal connection worldline connecting P and Q, and thus Qwil occur after P in every Other inertial reference frame (Figure).



Assume that the speed v_{ci} of the causal influence between event P and Q satisfies the restriction $v_{ci} < 1$ in the Home Frame. Consider any Other inertial frame. The slope of the Other Frame's diagram x axis is equal to the speed of the Other Frame with respect to the Home Frame, which must be less than 1. Thus the slope of line PQ, which is $1/v_{ci}$, will always be greater than the slope of the axis, so P will be measured to occur before Q in any Other inertial frame.

In short, if the speed of reference frames and causal influences is limited to $v_{ci} < 1$, then effects will occur after their causes in every inertial reference frame, as required by the principle of relativity when applied to the concept of causality.

At the most basic physical level, the problem with causes following effects is that this would violate the second law of thermodynamics, which requires that the entropy of the universe always increase (or at least remain the same) during a physical process.

This law implies that events ni certain physical processes will occur in one temporal order but not in the reverse order.

Therefore, if the second law of thermodynamics is to be true in al inertial frames, as required by the principle of relativity, then the temporal order of all events that might be linked by that law must be preserved in all inertial frames.

“Cause" and "effect" is thus really a colloquial way to talk about the invariant temporal order imposed on events by the second law of thermodynamics.

In summary:

Theorem: The Cosmic Speed Limit

In order for causality (i.e., the idea that one event can *cause* another event to happen) to be self-consistent, information (i.e., *any* effect representing a causal connection between two events) cannot travel between two events with a speed v_{ci} greater than that of light.

In order for causality (i.e., the idea that one event can cause another event to happen) to be self-consistent, information (i.e., any effect representing a causal connection between two events) cannot travel between two events with a speed v_{ci} greater than that of light.

Since anything movable and detectable can carry information (i.e., cause things to happen), this consequence of the principle of relativity applies not only to all physical objects (waves, particles, and macroscopic objects) but indeed to any trick or means of conveying a message that might be imagined (e.g., instantaneous changes in a gravitational field, telepathy, magic, whatever.)

So, with a straightforward argument using two-observer spacetime diagrams, we have proved the existence of a cosmic speed limit, an idea having profound physical and and philosophical implications.

As usual, this prediction is amply supported by experiment.

No particle, object, or signal of any kind has ever been definitely observed to travel at faster than the speed of light in a vacuum.

Science fiction fans and space travel buffs who hope for the discovery of faster-than-light travel may hope forever: both the argument (based as it is on the firmly accepted and fundamental ideas of the principle of relativity and the physical reality of causality) and the experimental evidence present a pretty ironclad case for this cosmic speed limit.

TIMELIKE, LIGHTLIKE, AND SPACELIKE INTERVALS

We are now in a position to understand more fully the true physical nature of the spacetime interval between any two events in spacetime.

Earlier, we saw that for two events whose coordinate differences in a given inertial reference frame are Δt, Δx, Δy and Δz the quantity $\Delta s^2 = \Delta t^2 - \Delta x^2 - \Delta y^2 - \Delta z^2$ had a frame-independent value, and that value was equal to the time registered by an inertial clock present at the two events.

But to make the proof of the metric equation work, it was necessary to assume that $\Delta d = \sqrt{\Delta x^2 + \Delta y^2 + \Delta z^2}$ was smaller than Δt so that there was more than sufficient time for a light flash to travel from one event to the other along the length of the light clock.

The purpose of this discussion is to study the meaning of the spacetime interval when $\Delta d > \Delta t$ that is, when this condition is violated.

We have exploited the analogy between spacetime geometry and euclidean plane geometry extensively in the earlier discussions.

We have noted, though, that the negative signs in the metric equation (which do not appear in the corresponding pythagorean relation) lead to some subtle differences between spacetime geometry and euclidean geometry.

Yet another one of these differences is the following.

In Euclidean geometry the square distance Δd^2 between two points on a plane is necessarily positive:

$$\Delta d^2 = \Delta x^2 + \Delta y^2 \geq 0$$

But taken at face value, the metric equation allows the squared spacetime interval between two events to be positive, zero, or negative, depending on the relative sizes of the coordinate separations Δd and Δt between those events:

$$\Delta s^2 = \Delta t^2 - \Delta d^2$$

Therefore

$$\text{If} \quad \Delta d > \Delta t \qquad \text{then} \qquad \Delta s^2 < 0!$$

We see that while there is only one kind of distance between pairs of points on a plane, the possible spacetime intervals between pairs of events in spacetime fall into three distinct categories depending on the sign of Δs^2.

These categories are(as we indicated earlier):

If $\Delta s^2 > 0$, the interval between the events is said to be **timelike.**
If $\Delta s^2 = 0$, the interval between the events is said to be **lightlike.**
If $\Delta s^2 < 0$, the interval between the events is said to be **spacelike.**

The reasons for these names will become clear shortly.

The peculiar category here is the spacelike category-there is nothing corresponding to it in ordinary plane geometry (where the squared distance between two events is always positive).

What does it mean for two events to have a spacelike spacetime interval between them?

First of all, note that events separated by spacelike spacetime intervals certainly do exist.

Consider, for example, the case of two events that are measured in a certain inertial frame to occur at the same *time* but at different *locations*.

Since the time separation between these events is zero in that reference frame, we have $\Delta s^2 = 0 - \Delta d^2 = - \Delta d^2 < 0$, so the interval between these events is necessarily spacelike.

Therefore, the spacelike interval classification is needed if we are to be able to categorize the spacetime interval between arbitrarily chosen events.

As we have already discussed, the squared spacetime interval between two events Δs^2 that appears in the metric equation $\Delta s^2 = \Delta t^2 - \Delta d^2$ has been linked with the frame-independent time measured by an inertial clock present at both events only in the case that $\Delta t^2 > \Delta d^2$.

For two events for which $\Delta t^2 < \Delta d^2$, it is not clear how one can directly measure the squared spacetime interval between the events at all.

For example, for an inertial clock to be present at both events if $\Delta d > \Delta t$, it would have to travel at a speed $v > 1$ in that frame.

We have just seen that this is impossible; thus a spacelike spacetime interval cannot be measured by a clock or anything else that travels between the events.

Since the proof of the metric equation given earlier does not handle the case of spacelike intervals, it is not even completely clear that the squared interval $\Delta s^2 = \Delta t^2 - \Delta d^2$ is frame-independent when it is less than zero.

In fact, the squared spacetime interval Δs^2 does have a frame-independent value, no matter what its sign is.

This can easily be demonstrated by using the Lorentz transformation equations for coordinate differences given by the relations:

$$\Delta t' = \gamma(\Delta t - \beta\,\Delta x)$$

$$\Delta x' = \gamma(-\beta\,\Delta t + \Delta x)$$

$$\Delta y' = y'_B - y'_A = y_B - y_A = \Delta y$$

$$\Delta z' = z'_B - z'_A = z_B - z_A = \Delta z$$

The argument goes like this.

Let $\Delta t, \Delta x, \Delta y, \Delta z$ be the coordinate separations of two events measured in the Home Frame, and let $\Delta t'$, $\Delta x', \Delta y', \Delta z'$ be the coordinate separations of the same two events measured in an Other Frame moving with speed β in the +x direction with respect to the Home Frame.

Then, the equations imply that

$$
\begin{aligned}
(\Delta t')^2 - (\Delta x')^2 - (\Delta y')^2 - (\Delta z')^2 \\
&= [\gamma(\Delta t - \beta\,\Delta x)]^2 - [\gamma(-\beta\,\Delta t + \Delta x)]^2 - \Delta y^2 - \Delta z^2 \\
&= \gamma^2(\Delta t^2 - \beta\,\Delta x\,\Delta t + \beta^2\,\Delta x^2) - \gamma^2(\beta^2\,\Delta t^2 - \beta\,\Delta x\,\Delta t + \Delta x^2) - \Delta y^2 - \Delta z^2 \\
&= \gamma^2(\Delta t^2 + \beta^2\,\Delta x^2 - \beta^2\,\Delta t^2 - \Delta x^2) - \Delta y^2 - \Delta z^2 \\
&= \gamma^2(1 - \beta^2)(\Delta t^2 - \Delta x^2) - \Delta y^2 - \Delta z^2 \\
&= \Delta t^2 - \Delta x^2 - \Delta y^2 - \Delta z^2
\end{aligned}
$$

where the last step uses the fact that

$$\gamma = 1/\sqrt{1-\beta^2}$$

The sign of $\Delta s^2 = \Delta t^2 - \Delta d^2$ is irrelevant to this derivation: thus Δs^2 always has the same numerical value in every inertial reference frame, no matter whether it is spacelike, timelike, or lightlike.

How can we measure the value of the spacetime interval between two events separated by a spacelike interval?

We cannot use a clock, as we have noted above.

In fact, a spacelike spacetime interval is measured with a *ruler*, as we will shortly see.

Let us define the spacetime separation $\Delta\sigma$ between two events as follows:

$$\Delta\sigma^2 = \Delta d^2 - \Delta t^2$$

The spacetime separation, so defined, is conveniently real whenever the interval between the events is spacelike.

Now note that if we can find an inertial reference frame where the events are simultaneous ($\Delta t = 0$), we have

$$\Delta\sigma^2 = \Delta d^2 \Rightarrow \Delta\sigma = \Delta d \qquad \text{(in a frame where } \Delta t = 0\text{)}$$

Now, I claim that we can always find a frame in which $\Delta t = 0$ if the events are separated by a spacelike interval.

This can be seen as follows.

Imagine that two events occur with coordinate differences Δt and $\Delta d > (\Delta t)$ as measured in the Home Frame.

Reorient and reposition the axes of the Home Frame os that the events in question both occur along the spatial x axis, with the later event located in the +x-direction relative to the earlier event.

This can be done without loss of generality: we are always free to choose the orientation of our coordinate system to be whatever we find convenient.

Once this is done, $\Delta d = \Delta x$ in the Home Frame.

Now, consider an Other Frame in standard orientation with respect to the Home Frame and traveling in the +x direction with speed β with respect to the Home Frame.

According to the transformation equations written down earlier, the time coordinate difference between these events in the Other Frame is

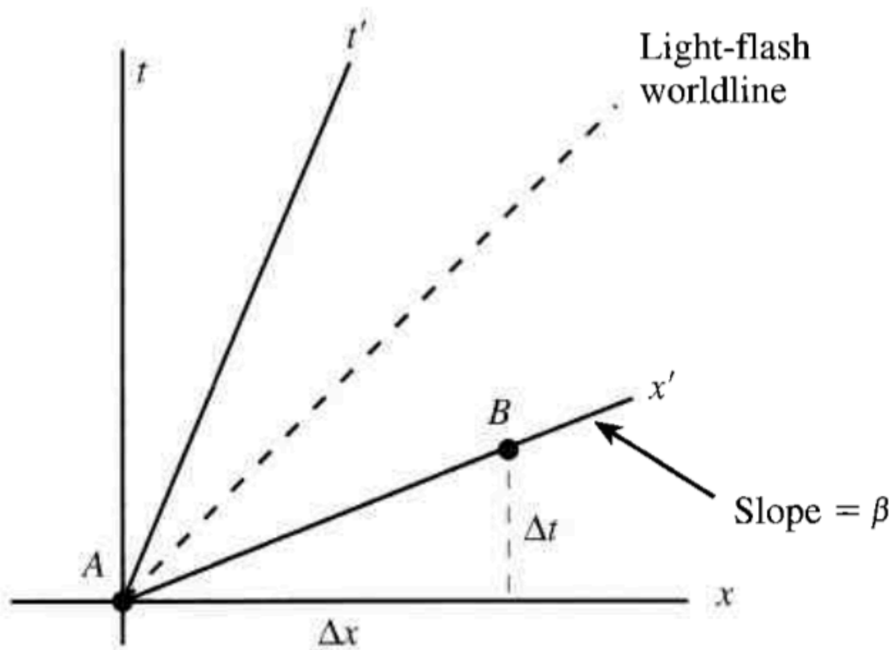
$$\Delta t' = \gamma(\Delta t - \beta\,\Delta x)$$

These events will be simultaneous in the Other Frame (that is, $\Delta t'$ will equal zero) if and only if the relative speed of the frames is chosen to be $\beta = \Delta t/\Delta x = \Delta t/\Delta d$.

This relative speed β will be less than 1 since $\Delta d > \Delta t$ for our events by hypothesis.

In short, given any pair of events that are separated by a spacelike interval in some inertial frame (which we are calling the Home Frame), it is possible to find an Other inertial frame moving with speed $\beta < 1$with respect to the Home Frame in which the two events will be simultaneous (Figure).

QED



Given any pair of events A and B separated by a spacelike interval ($\Delta x > \Delta t$) in the Home Frame, it is possible to find an Other Frame in which the events are simultaneous. The speed of this special Other Frame with respect to the Home Frame simply has to have the right value to give the slope of the diagram x' axis the value $\Delta t/\Delta x$, that is, $\beta = \Delta t/\Delta x$. Since $\Delta x > \Delta t$ this $\beta < 1$, so it is always possible to find such a frame.

To summarize, then, if the spacetime interval between two events is spacelike, then:

1. It is possible to find an inertial frame in which these events occur at the *same time.*
2. $\Delta\sigma$ is the *distance* between the events in that special frame.
3. Observers in all other inertial reference frames can use Eq. 8.4 to *calculate* $\Delta\sigma$: they will all get the same value for $\Delta\sigma$ that is measured directly in the special frame.

These statements are directly analogous to statements that can be made about events separated by a timelike spacetime interval.

If the spacetime interval between two events is timelike, then:

1. It is possible to find an inertial frame in which these events occur at the *same place* (this is the frame of the inertial clock that is present at both events).
2. Δs is the *time* between the events in this special frame (i.e., the time measured by that inertial clock).
3. Observers in all other inertial reference frames can use the ordinary metric equation to calculate Δs: they will all arrive at the same value measured directly in the special frame.

Thus there is a fundamental symmetry between spacelike and timelike spacetime intervals, a symmetry that arises because both reflect the same underlying physical truth.

It is possible to describe the separation of any two events in space and time with a frame-independent quantity Δs^2 (which we call the **squared spacetime interval**) that is analogous to the *squared distance* between two points in plane geometry.

It is simply a peculiarity of the geometry of spacetime that the quantity in spacetime that corresponds to ordinary (unsquared) distance on the plane comes in three distinct flavors (the spacetime *interval* Δs if $\Delta s^2 > 0$ for the events, the *spacetime* separation $\Delta\sigma$ if $\Delta s^2 < 0$, and the *lightlike* interval $\Delta s = \Delta\sigma = 0$ when $\Delta s = 0$) which are measured in different ways using different tools.

But it is important to realize that these three quantities are only different aspects of the same basic frame-independent quantity Δs^2

We see that timelike intervals are directly measured with atime-measuring device (i.e., an inertial clock present at both events), while spacelike intervals are directly measured with a space-measuring device (i.e., a ruler in the special inertial frame where Δt between the events is zero).

This is why these interval classifications have the names timelike and spacelike: the names are intended to tell us whether we have to measure the interval with a clock (because it is timelike) or with a ruler (because it is spacelike).

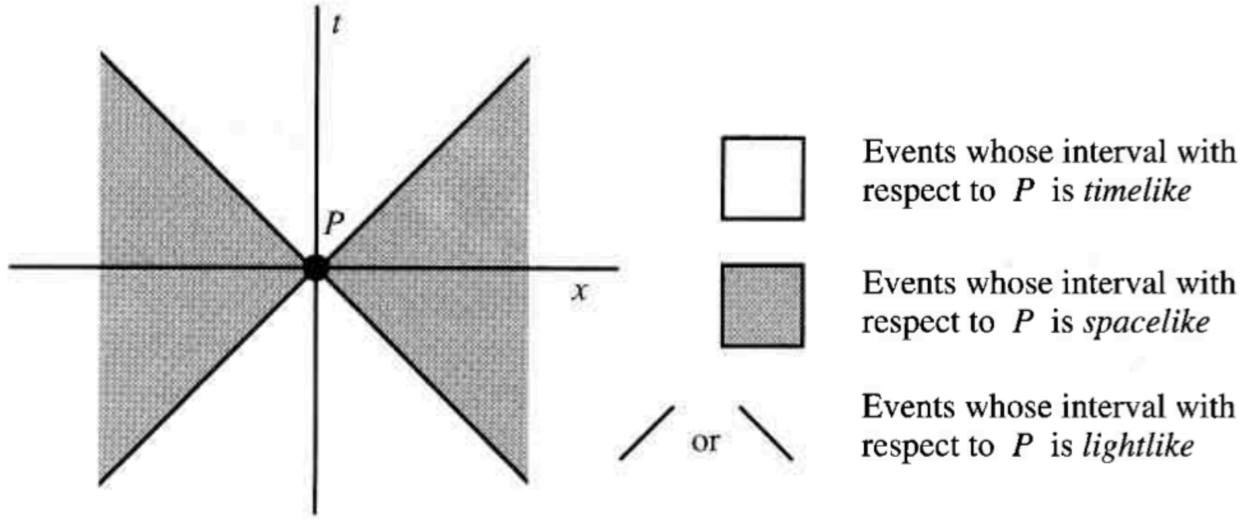
The lightlike interval classification stands between the other classifications.

When the interval between two events is lightlike, we have $\Delta d = \Delta t$, which implies that these events could be connected by a flash of light.

THE CAUSAL STRUCTURE OF SPACETIME

Now, because it is true that the value of the squared spacetime interval Δs^2 is frame-independent no matter what its sign, all inertial observers will agree as to whether the interval between a given pair of events is timelike, lightlike, or spacelike (since if they all agree on the value of Δs^2, they will surely all agree on its sign).

This means that the spacetime around any event P can be divided up into the distinct regions shown in Figure, and every observer will agree about which events in spacetime belong to which region.



The regions of spacetime relative to an event P.

Because these regions can be defined in a frame-independent manner, it is plausible that they reflect something absolute and physical about the geometry of spacetime.

In fact, these regions distinguish those events that can be causally connected to P from those that cannot.

Remember earlier we found that two events can be causally linked only if $\Delta d \leq \Delta t$ between them: otherwise the causal influence would have to travel between the events faster than the speed of light.

This means that every event that can be causally linked with P must have a timelike (or perhaps lightlike) interval with respect to P: such events will lie in the white regions shown in the Figure above.

We can be more specific yet.

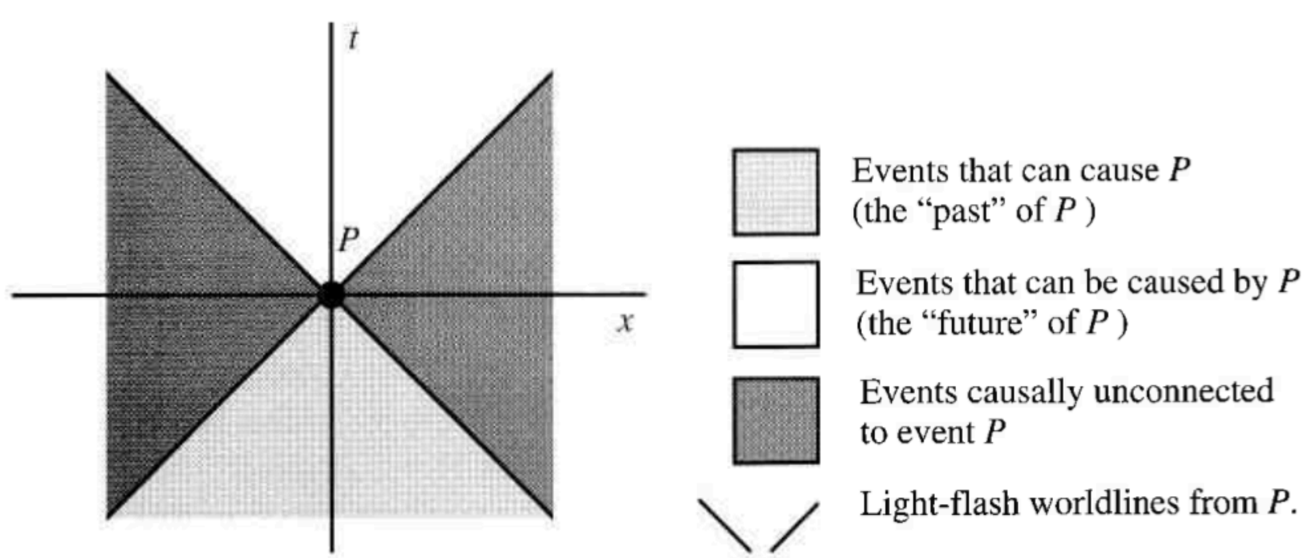
Since the temporal order of events is preserved in all inertial frames if $\Delta d > \Delta t$ (as we showed earlier), all events in the upper white region in the above Figure will occur after P in every frame (and thus could be caused by P), and all events in the lower white region of the above Figure will occur before P in every frame (and thus could cause P).

These regions are referred to as the future and past of P, respectively.

Events whose spacetime interval with respect to P is spacelike ($\Delta d > \Delta t$) cannot influence P or be influenced by it.

We say that these events (which inhabit the shaded region of the above Figure) are causally unconnected to event P.

With this in mind, we can relabel the regions in the above Figure as shown in the Figure below.



The causal structure of spacetime with regard to an event P.

Remember that every observer agrees on the value of the spacetime interval between event P and any other event, so every observer agrees as to which event belongs in which classification.

The structure illustrated is thus an intrinsic, frame-independent characteristic of the geometry of spacetime.

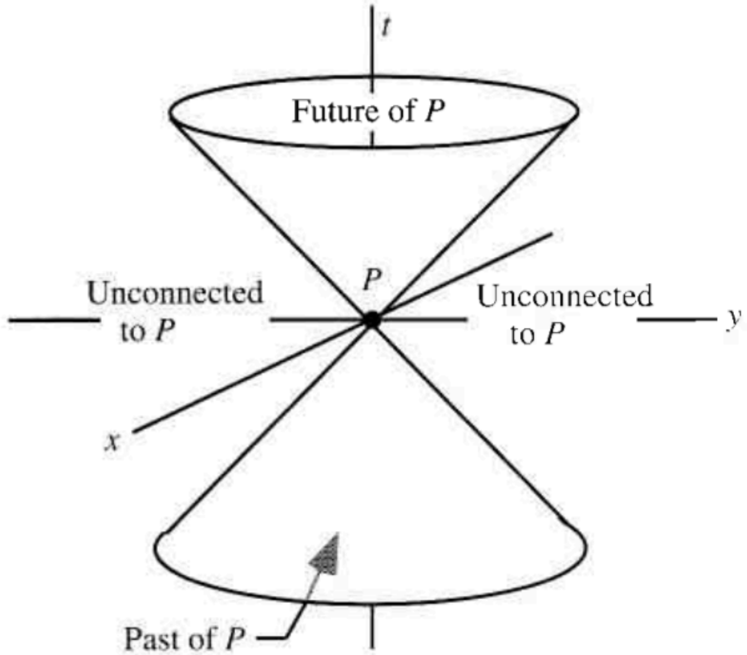
Now, the boundaries of the regions illustrated in the last Figure are light-flash worldlines.

If we consider two spatial dimensions instead of one, an omnidirectional light flash is seen as an ever-expanding ring, like the ring of waves formed by the splash of a stone into a still pool of water.

If we plot the growth of such a ring on a spacetime diagram, we get a cone.

The boundaries between the three regions described are then two tip-to- tip cones, as shown in the next Figure.

This boundary surface is often called the light cone for the given event P.



The light cone and regions of spacetime for the event P illustrated in am spacetime diagram with two spatial axes.

The Figure still leaves one spatial dimension unexpressed.

The difficulty of representing (or even imagining) how the z dimension fits into such a diagram is obvious.

This diagram represents the best that we can do in visualizing the four-dimensional reality of the light cone.

To summarize, the point of this discussion is that the spacetime interval classifications, which are basic, frame-independent features of the geometry of spacetime, have in fact a deeply physical significance: The sign of the squared spacetime interval between two events unambiguously describes whether these events can be causally connected or not.

The light cone shown in the Figure effectively illustrates this geometric feature of spacetime.

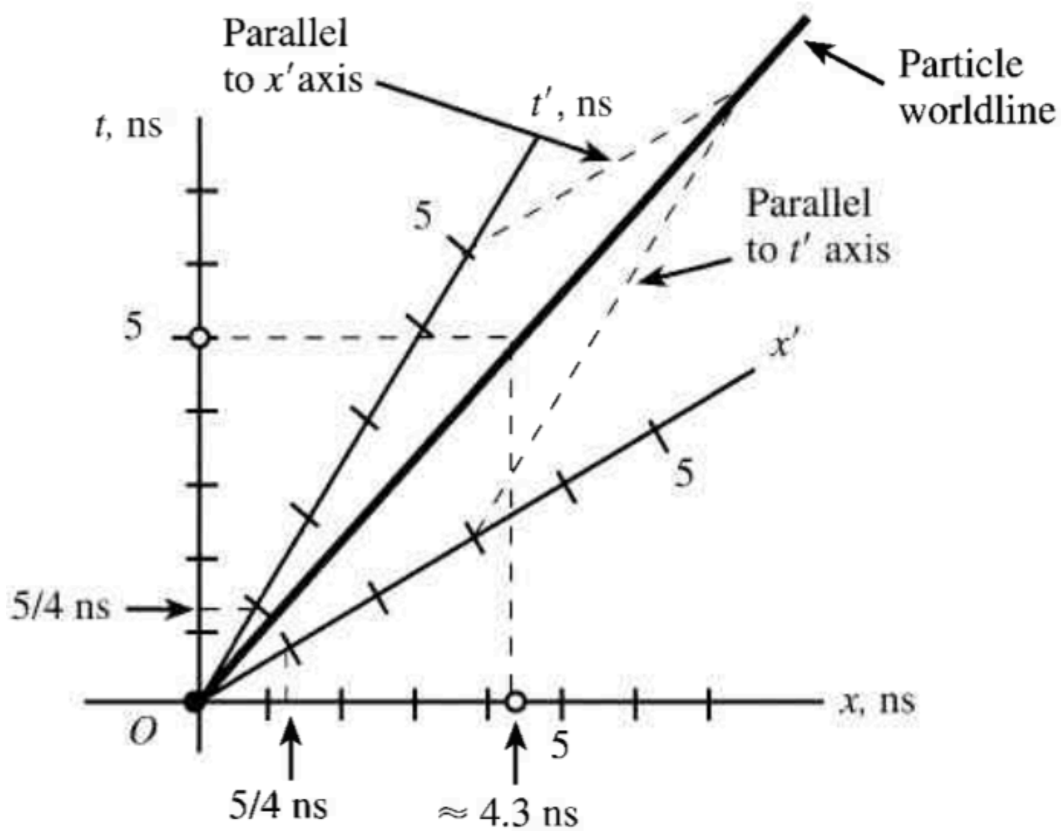
THE EINSTEIN VELOCITY TRANSFORMATION

Now we rederive the relativistic generalization of the Galilean velocity transformation equations.

Imagine a particle that is observed to move along the spatial x axis with a constant x velocity v, in the Other Frame, which itself is moving with a speed β in the positive x direction with respect to the Home Frame.

What is the x velocity v_x, of the particle as observed in the Home Frame?

Figure shows how to construct a two-observer spacetime diagram that we can use to answer this question.



The particle shown here has an x velocity of v'_x = 3/5 as measured in the Other Frame, which itself has a velocity of β= 3/5 as measured in the Home Frame. The x velocity v_x, of the particle in the Home Frame is seen from the diagram to be about 4.3 ns / 5 ns = 0.86. Note that when β= 3/5, γ = 5/4.

After drawing both sets of coordinate axes, one simply draws the worldline of the particle in such a manner that its slope in the Other Frame is $1/\, v'_x$.

The slope of the line in the Home Frame can then be easily determined by picking an arbitrary rise (5 ns in the diagram) and determining the run.

One can then compute v_x, by taking the inverse of the slope.

In the case where β= 3/5 and v'_x = 3/5, the Figure implies that the value of v_x = 0.86, not the value 3/5 + 3/5 = 6/5 that the Galilean velocity transformation equations predict.

Now let us see if we can derive an exact formula using the Lorentz transformation equations that does the same thing that this diagram does.

We are trying to compute the x velocity v_x, of a particle at a given instant of time as measured in the Home Frame knowing its x velocity v'_x as measured in the Other Frame, assuming that the frames are in standard orientation.

Pick two infinitesimally separated events along the worldline of the particle in question.

Let the coordinate differences between these events as measured in the primary frame be dt and dx.

Let the coordinate differences between the same two events as measured in the Other Frame be dt' and dx'.

The x velocity of the particle as it travels between these events is

$$v_x \equiv \frac{dx}{dt} = \frac{\gamma(\beta\, dt' + dx')}{\gamma(dt' + \beta\, dx')}$$

where I have used the inverse Lorentz transformation equations (β -> - β or x' -> x instead of x -> x').

Dividing the right side top and bottom by dt' and using dx'/dt' = v_x' , we get

$$v_x = \frac{\beta + v_x'}{1 + \beta v_x'} \qquad \text{(relativistically valid)}$$

In a similar fashion, the y and z components of the particle's Home Frame velocity are found to be

$$v_y = \frac{v'_y\sqrt{1-\beta^2}}{1+\beta v'_x}$$

$$v_z = \frac{v'_z\sqrt{1-\beta^2}}{1+\beta v'_x}$$

These equations are called the inverse Einstein velocity transformation equations and express algebraically what the last Figure expresses graphically.

Note that the result given here is different from what the Galilean velocity transformation predicts:

$$v_x = \beta + v'_x \qquad \text{(from the galilean transformation)}$$

Note that when the velocities β and v'_x, are very small the Einstein equations reduce to the Galilean equations.

You can easily show that these equations never predicts that the x velocity of an object in the Home Frame will exceed the speed of light: even if both β and $v'_x = 1$(which are their maximum possible values),

Moreover, you can easily use this equation to show that the speed of light = 1 in al frames: if the x velocity of any object in the Other Frame is $v'_x = 1$, its speed in the Home Frame will be $v_x = 1$!

The direct Einstein velocity transformation equations (x -> x') express the Other Frame velocity components in terms of the Home Frame velocity components:

$$v'_x = \frac{v_x - \beta}{1 - \beta v_x} \qquad v'_y = \frac{v_y\sqrt{1-\beta^2}}{1 - \beta v_x} \qquad v'_z = \frac{v_z\sqrt{1-\beta^2}}{1 - \beta v_x}$$

Measurements in Special Relativity

We now consider measurement of length and time in special relativity.

1st, we need to restate definitions from earlier:

Length of object

= spatial separation of 2 events representing endpoints of object measured simultaneously

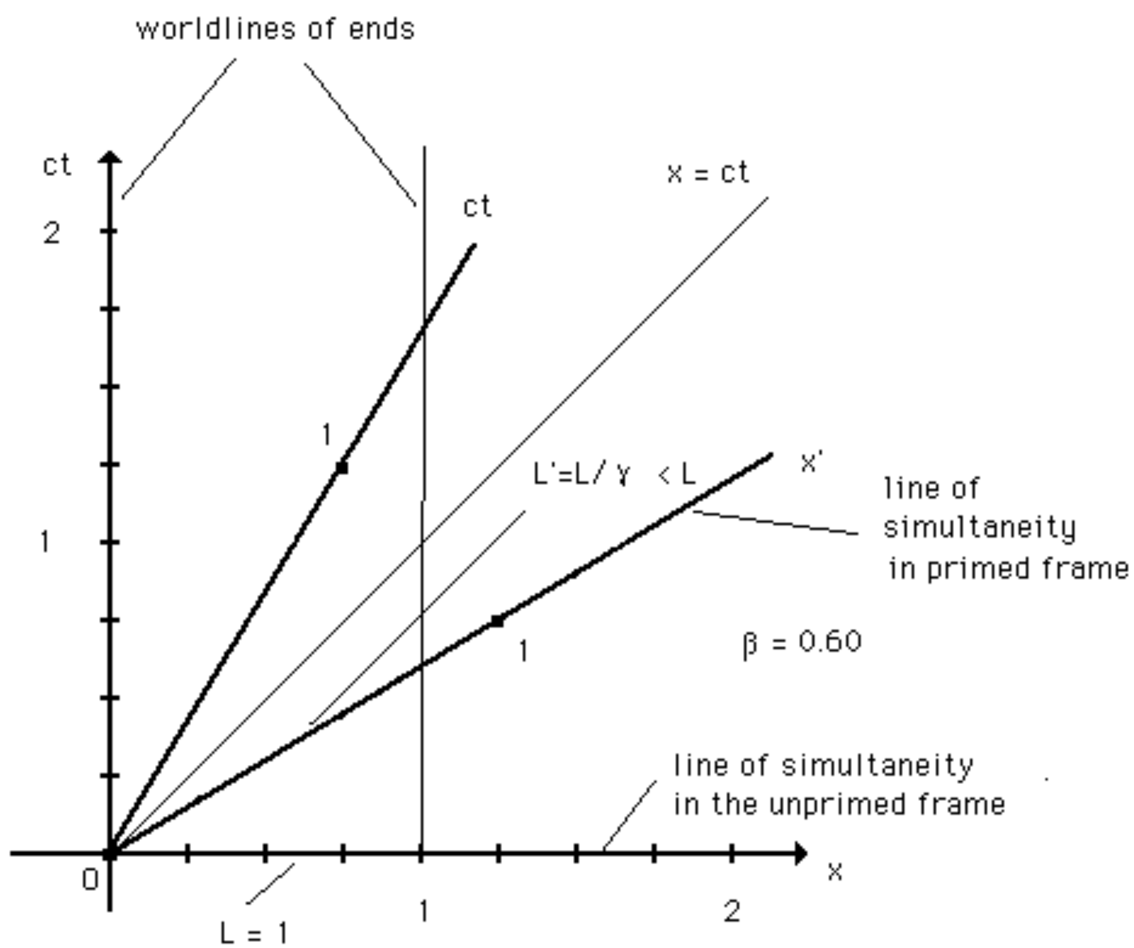
(2 events are on line of simultaneity in a given frame).

Time interval between 2 events

= time separation of 2 events measured by clock at rest with respect to 2 events

(2 events are on the worldline of the clock).

Suppose have 2 events (ct_1,x_1) and (ct_2,x_2) that correspond to events on worldlines of endpoints of object being measured, crossing line of simultaneity (same time) (see diagram).



-> length of object = $L = x_2 - x_1$.

As shown in diagram that is not length as measured in the other reference frame.

In fact, $L' = x'_2 - x'_1 = L/\gamma < L$, which is the famous length contraction(will discuss shortly).

Do not be deceived by it looking longer, remember the scales on different axes are not same.

The **proper length** of object is length measured in object's rest frame.

Unprimed frame -> where endpoint worldlines are parallel to time axis

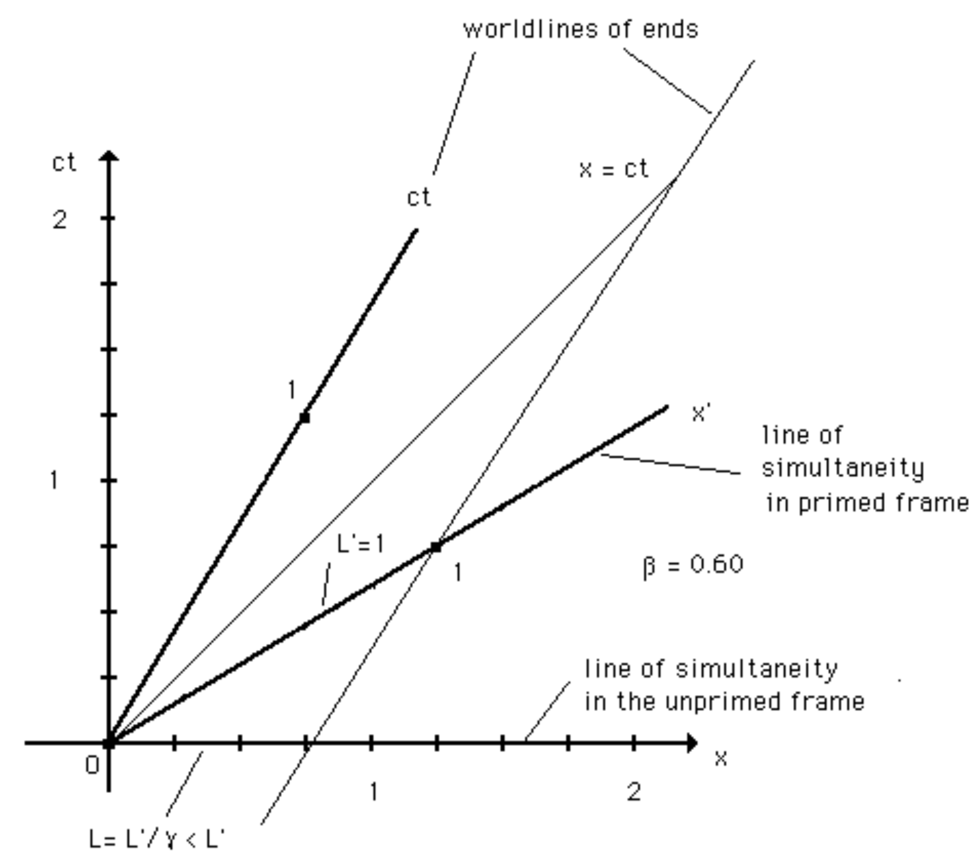
-> definition of being at rest.

Proper length is **maximum** measured length.

Measured length is less in other frame because two observers do not agree about simultaneity,

i.e., have different lines of simultaneity.

Even though we use the word contraction, you must understand that effect is due to a disagreement about simultaneity and no physical contraction has actually occurred

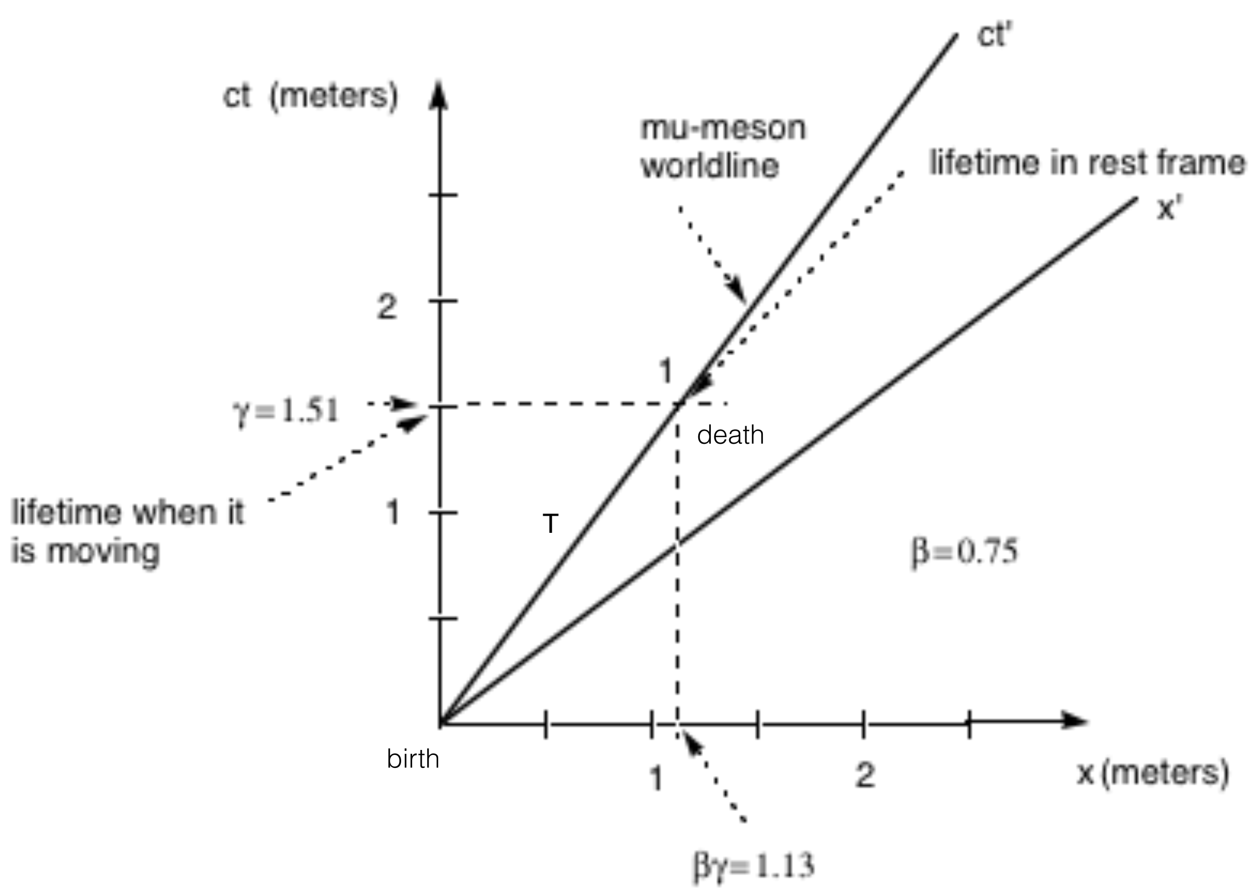


If object is at rest in primed frame, get identical result just exchanging roles of two frames.

Diagram, in this case -> $L = x_2 - x_1 = L'/\gamma < L'$.

Time measurements and time dilation are handled in same way.

Consider diagram(next slide) representing system at rest in primed frame that only lives for finite amount of time(like mu-mesons).



Proper time interval for this system is time separation T between events (birth(event #1) and death(event #2)) as measured by clock at rest with respect to system or, in this case, at rest in primed frame.

As can be seen from diagram time separation for an observer in unprimed frame is

$$T' = \gamma T > T.$$

Proper time is **shortest** time interval.

Result is identical to mu-meson experiment discussed earlier

-> example of **time-dilation**

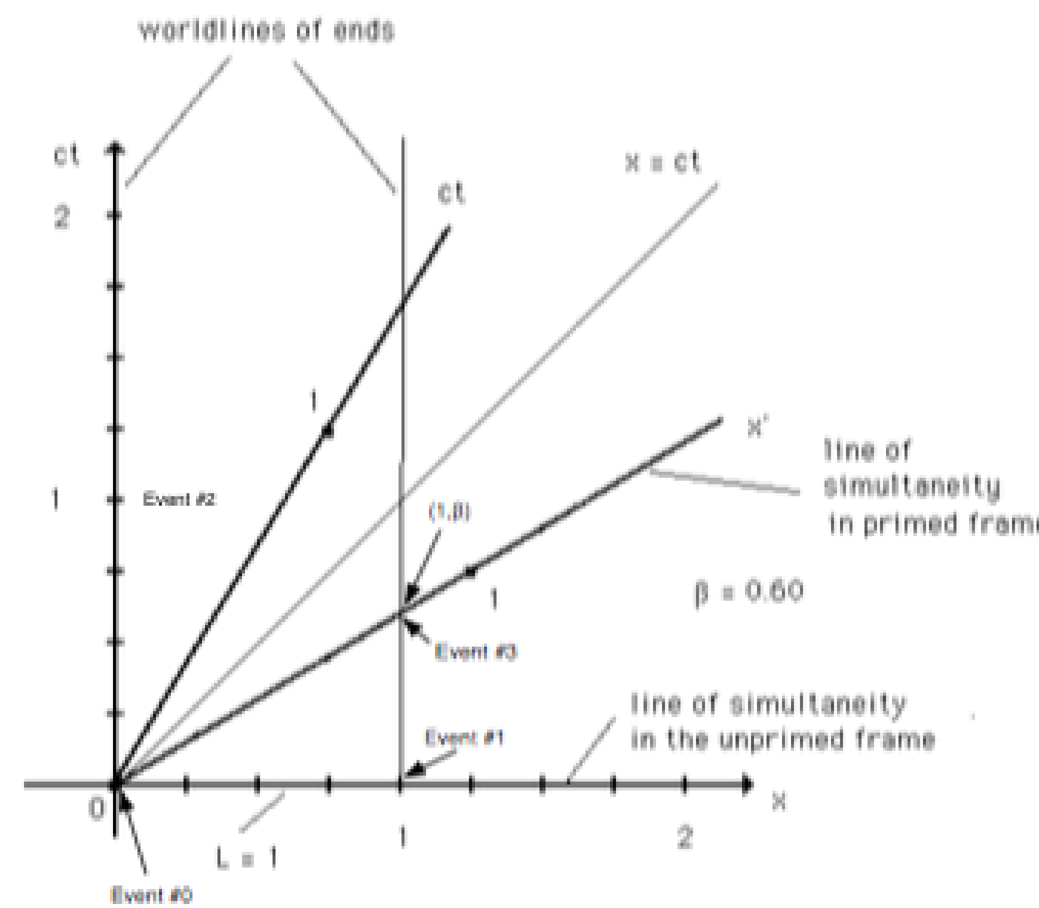
(see diagram above, where $T' = 1$ and $T = \gamma T' = \gamma$).

Can also see both of these results directly using Lorentz transformations or invariance of interval.

Using Lorentz Transformations - Deal with equation just once!!

Length Contraction

Relevant events are on worldlines representing ends of an object as shown:



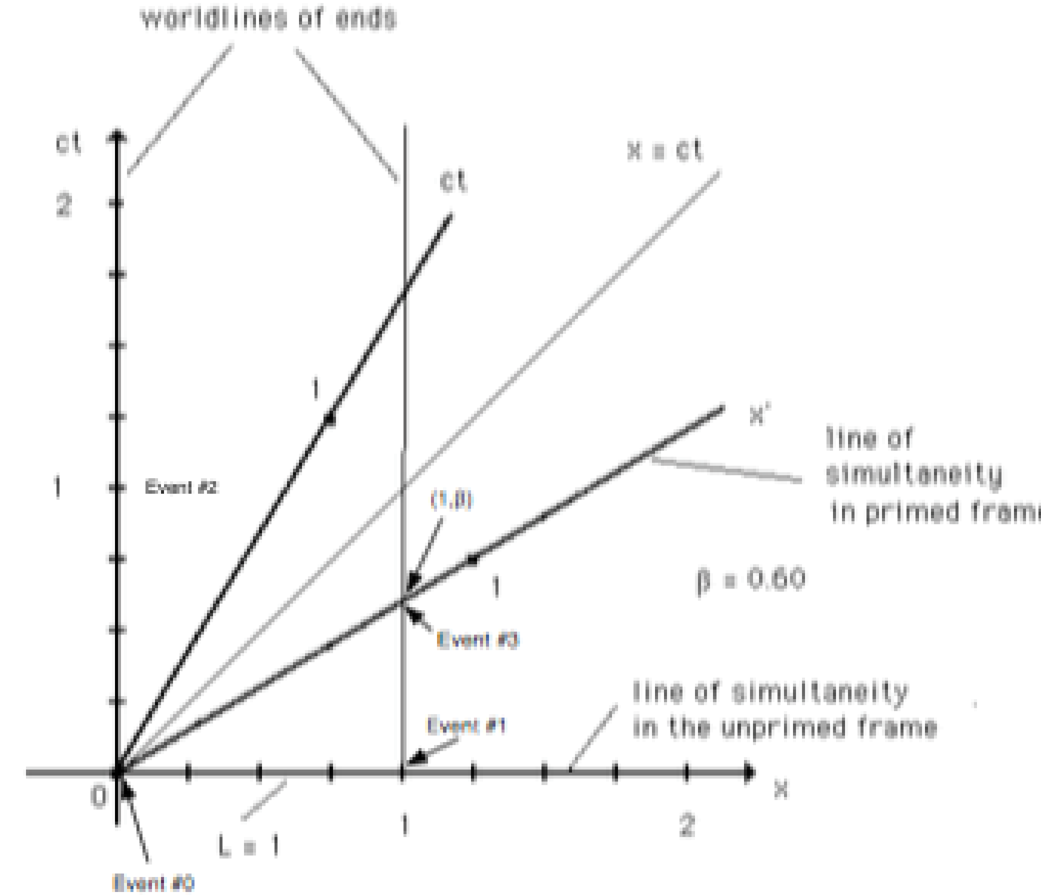
$(x_0, ct_0) = (0.0, 0.0)$ and $(x_1, ct_1) = (1.0, 0.0)$ for the unprimed observer

$(x'_0, ct'_0) = (0.0, 0.0)$ and $(x'_3, ct'_3) = (1.0, \beta) = (1.0, 0.6)$ for the primed observer

Length this object, by definition,

= spatial separation along line of simultaneity for unprimed observer

$$L = x_1 - x_0 = \Delta x = 1.$$



For other observer, length =

spatial separation along line of simultaneity for primed observer

$$L' = x'_3 - x'_0 = \gamma(x_3 - x_0) - \beta c(t_3 - t_0) = 1.25(1.0 - 0.6(0.6)) = 0.8 = \frac{1}{\gamma} = \frac{L}{\gamma}$$

Time Dilation

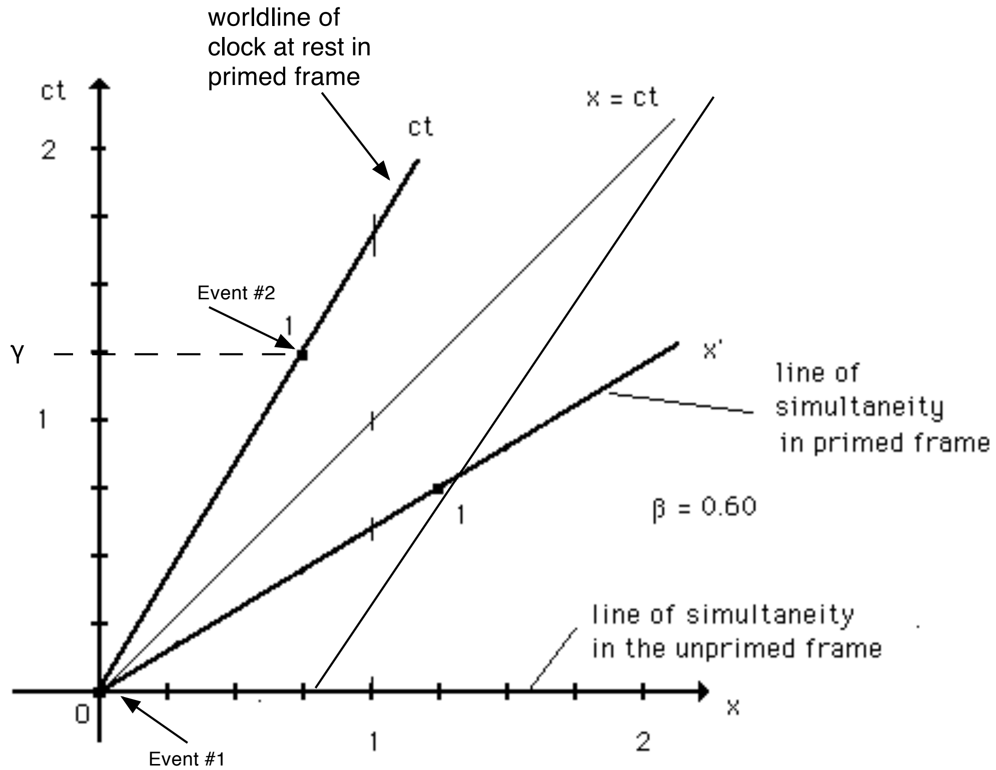
Suppose clock at rest in primed frame.

Then relevant events representing worldlines of ends of object are

$$(x'_1, t'_1) = (0.0, 0.0) \quad , \quad (x'_2, t'_2) = (0.0, 1.0)$$

Then for unprimed observer have

$$\Delta x = \gamma(\Delta x' + \beta c \Delta t') = \gamma\beta \quad , \quad c\Delta t = \gamma(c\Delta t' + \beta \Delta x') = \gamma$$



Note change in signs in Lorentz transformations when we go from primed to unprimed coordinates. Why?

Now return to k–factor.

Our original k–factor assumption

-> if unprimed observer sending out signals every T seconds and primed observer receiving them every T $'$ seconds where T $'$ = kT , then have the relationship

$$f' = \frac{1}{T'} = \frac{1}{kT} = \frac{f}{k} = \sqrt{\frac{c-v}{c+v}}\, f$$

between frequency f measured in unprimed frame and frequency f' measured in primed frame.

Above result corresponds to 2 observers moving away from each other

—-> $f' < f$

and hence primed observer sees wavelength increase (wavelength = c/f)

-> famous red shift.

If they move towards each other,

then $v \rightarrow -v$ or $k \rightarrow 1/k$

and

frequency increases (wavelength decreases)

-> blue shift.

-> relativistic Doppler effect for light.

The Doppler Effect

Sound and the Acoustic Doppler Effect

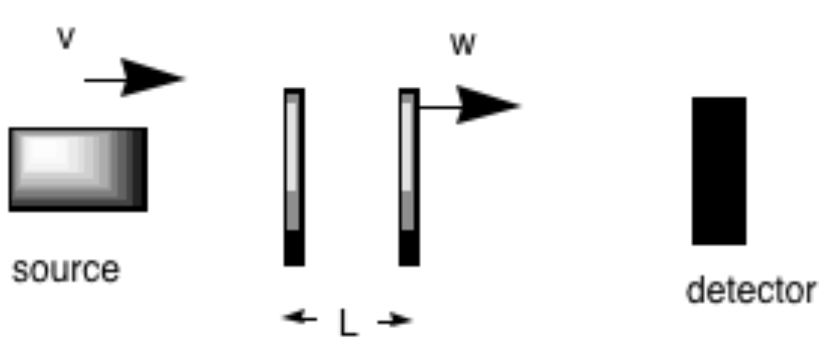
Sound travels through medium such as air with speed w (with respect to the medium).

Speed is determined by properties of medium

-> independent of motion of source.

Consider source of sound moving with velocity v through medium towards observer at rest.

Observer (the detector) lies along line of motion of source.



As shown, represent sound wave as regular series of pulses.

Pulses are separated in space by distance L and in time by an amount $\tau_0 = 1/f_0$, where f_0 is frequency of sound from source.

In time T sound travels distance wT and

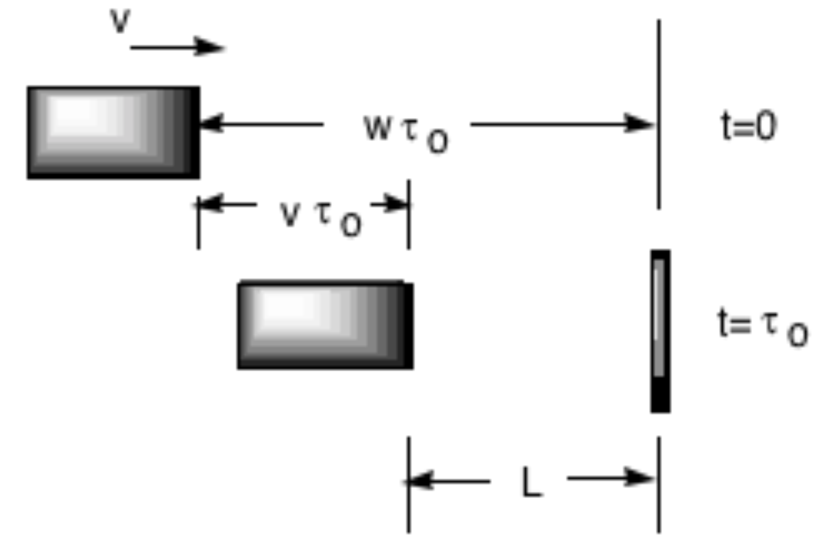
if pulses separated by distance L the number reaching detector is wT /L.

Rate at which pulses arrive is

$$\frac{w}{L} = \text{frequency} \left(\frac{\text{number}}{T}\right) \text{ of sound at the detector } = f_0$$

To determine L, consider a pulse emitted at t = 0 and

2nd pulse emitted at t = τ_0.



During interval τ_0 1st pulse travels distance $w\tau_0$ in medium and source travels distance $v\tau_0$ as shown —> distance between pulses is given by

$$L = w\tau_0 - v\tau_0 = (w-v)\tau_0 = \frac{(w-v)}{f_0}$$

and

$$f_D = \frac{w}{L} = \text{ frequency at detector } = f_0 \frac{1}{1-\frac{v}{w}} \text{ for a moving source}$$

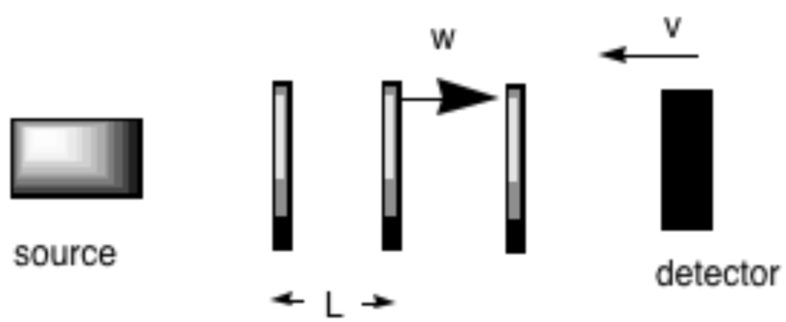
For an approaching source v > 0 and thus

$$f_D > f_0.$$

For a receding source, v < 0 and thus

$$f_D < f_0$$

If source at rest and detector moving (as shown) —> situation different.



Speed of pulses relative to detector is w + v.

Rate at which pulses arrive is

$$f_D = \frac{w+v}{L}$$

Since source at rest, $L = w\tau_0 = w/f_0$ and thus

$$f_D = f_0\left(1 + \frac{v}{w}\right) \text{ for a moving detector}$$

Two results **not symmetric**.

Approximately same for **small** v/w.

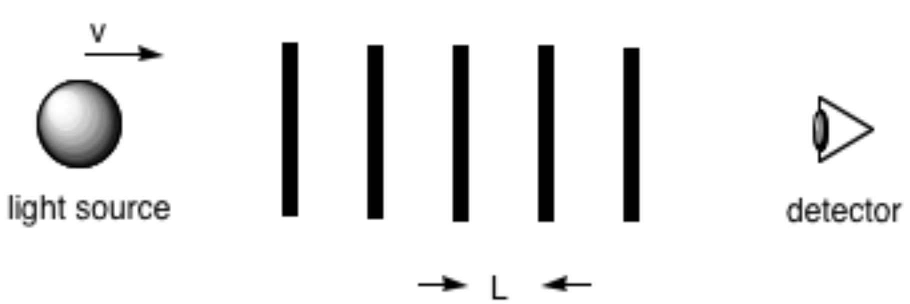
If know f_D, then can tell whether it is source or detector that is moving!!

This is so because speed of sound is not a universal constant but only has a definite value **relative** to medium where it is propagating.

Light and the Relativistic Doppler Effect

Suppose light source flashes with period $\tau_0 = 1/f_0$ in its rest frame and

that source is moving towards observer(detector) with velocity v as shown below left

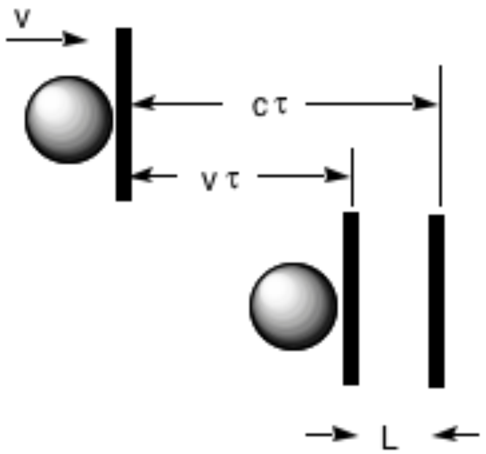


Due to time dilation, period in detector rest frame is $\tau = \gamma\tau_0$.

Since speed of light is universal constant, pulses arrive at detector with speed c.

As shown below frequency of pulses is $f_D = c/L$ where L is pulse separation in detector frame.

Since source is moving towards detector we have (as shown right)



$$L = c\tau - v\tau = (c-v)\tau = (c-v)\gamma\tau_0 = \gamma\frac{(c-v)}{f_0}$$

and

$$f_D = f_0\frac{\sqrt{1-\frac{v^2}{c^2}}}{1-\frac{v}{c}} = f_0\sqrt{\frac{c+v}{c-v}}$$

Here f_D is frequency in detector frame and v is relative velocity of source and detector.

Does not matter which one is actually moving!!

Result -> red shift formula started with earlier, as expected!

Now consider a spacetime diagram -> Doppler effect in spacetime.

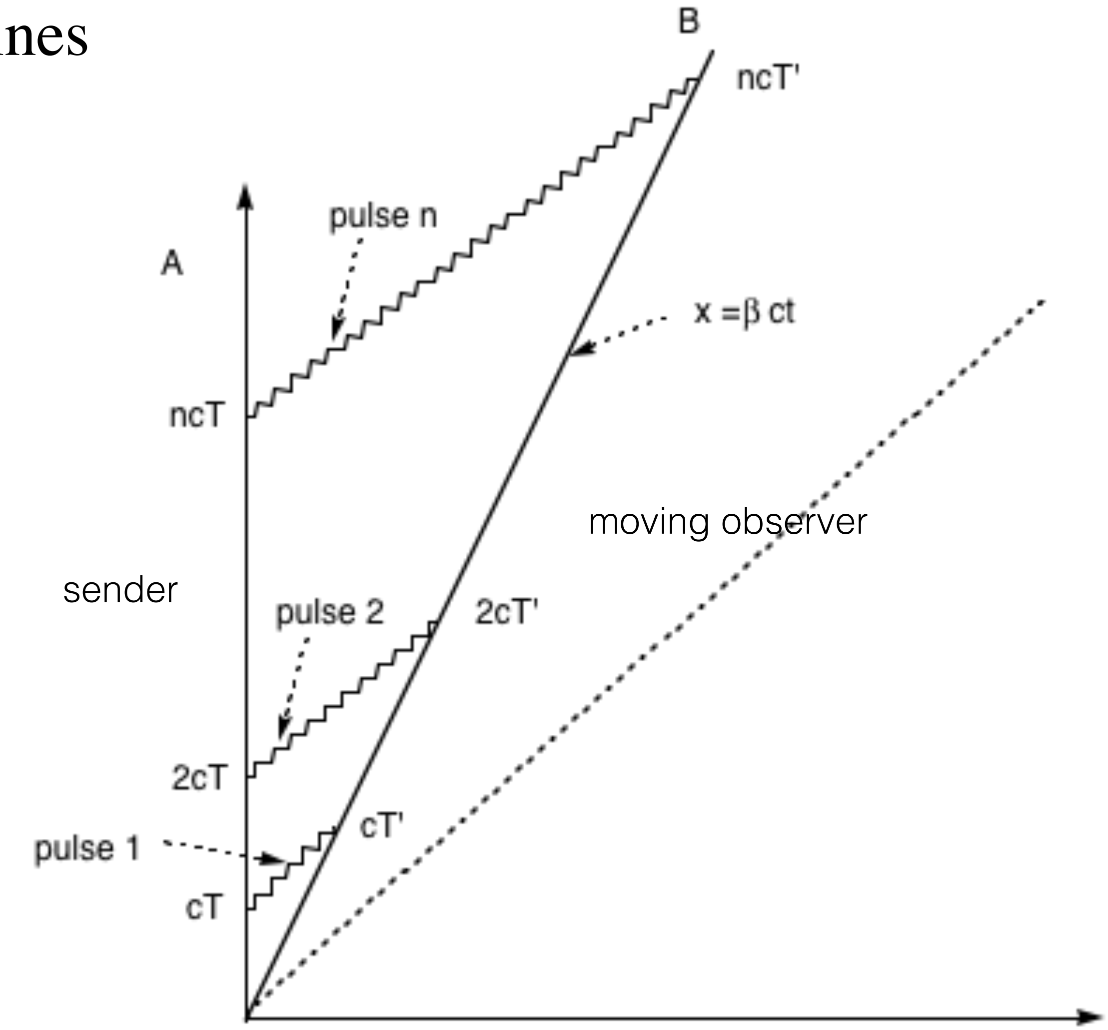
Reception of **last** pulse occurs at point of intersection of lines

$$x = c(t - nT) \text{ and } x = \beta ct$$

(as shown) or at event (all in umpired frame)

$$ct = \frac{cnT}{1-\beta} \quad , \quad x = \frac{\beta cnT}{1-\beta}$$

Solve above equations



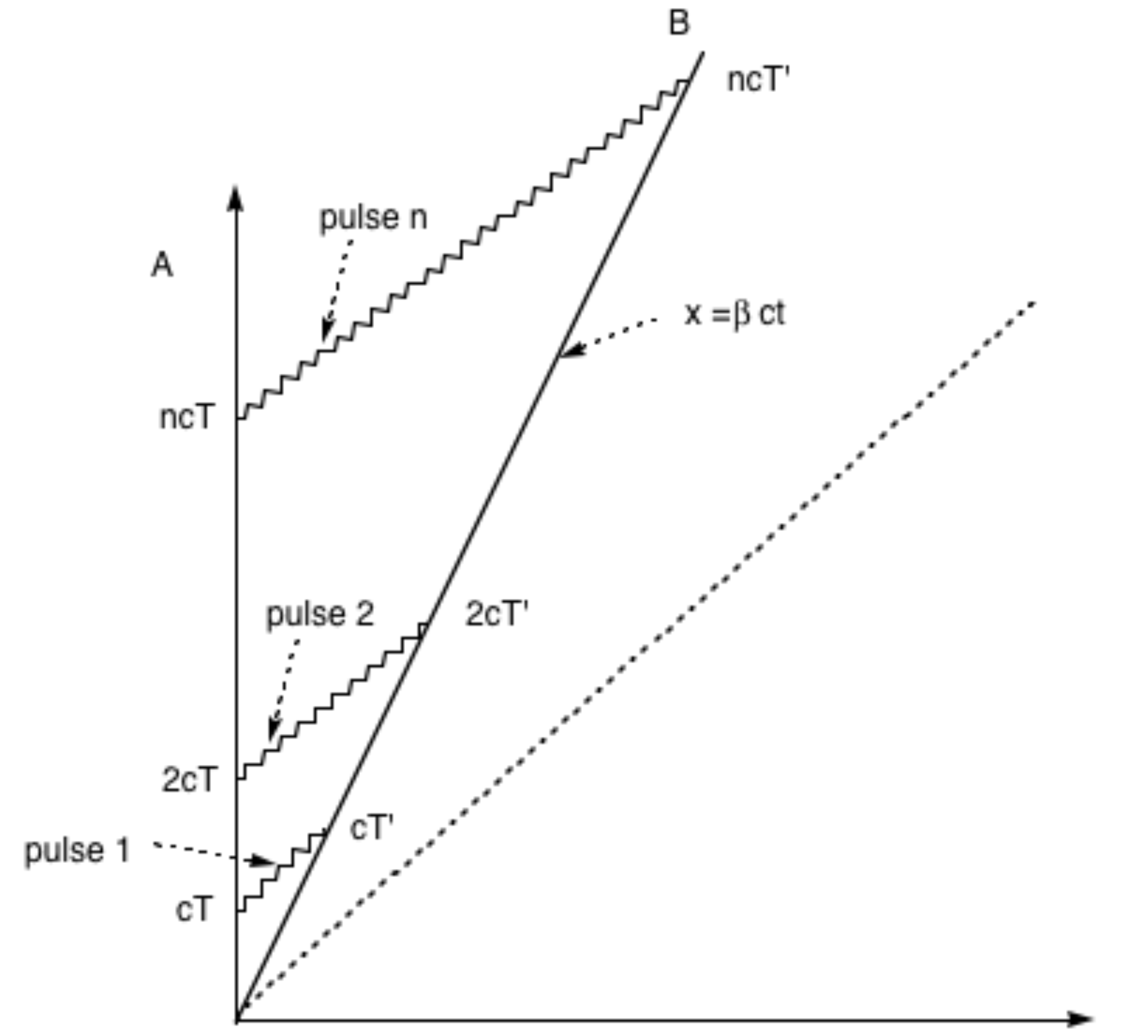
n pulses sent out by unprimed observer in nT seconds

-> period = T seconds and frequency is 1/T.

n pulses(same number) received by primed observer in nT $'$ seconds

-> period = T $'$ seconds and frequency is 1/T $'$.

Now, reception point also corresponds to



$$ct' = \gamma(ct - \beta x) = \gamma\left(\frac{cnT}{1-\beta} - \frac{\beta^2 cnT}{1-\beta}\right)$$

$$= \gamma cnT\frac{1-\beta^2}{1-\beta} = ncT' = \gamma cnT(1+\beta)$$

Using

$$\gamma = \frac{1}{\sqrt{1-\beta^2}}$$

get

$$T' = \sqrt{\frac{1+\beta}{1-\beta}}T$$

-> standard Doppler effect formula for light as we assumed at the beginning.

How Do We Talk to Each Other in this New Relativistic World?

In this new world what happens if we try to tell a story?

In particular, these are some of the words that are no longer usable?

where, when, speed, distance, time interval, simultaneous, same place, length, etc ...

If we want to use such words, then each reader (other observers) must first use the Lorentz transformations to translate the story before trying to read it!

The only words (concepts) that we are allowed to use if we do not want to do any translations are

interval, c, number of events

Not having grown up in this new world, we would find it very difficult to tell such a story.

The Famous Paradoxes

The Twin Paradox

State this problem in bad way, i.e., the way that leads to the so-called paradox.

Then state it correctly and paradox will disappear -> able to draw correct conclusions.

Might be lesson for life also!

Statement #1

Two twins are traveling relative to each other with speed v.

Time dilation says that the clock of the moving twin should tick slower (the time between ticks is larger).

Since each twin considers herself to be at rest, the other twin should have a clock that runs slower and hence the other twin should be younger.

Which twin is younger?

There is no definite answer to the question as posed since we do not know which twin is moving (changed reference frames - has accelerated) and hence we have a supposed paradox.

Statement #2

Two twins have been together since birth (they have been on the same worldline - in the same frame of reference).

At one point in time, one of the twins, gets into a rocket ship and changes her frame of reference changes her velocity experiences a period of acceleration

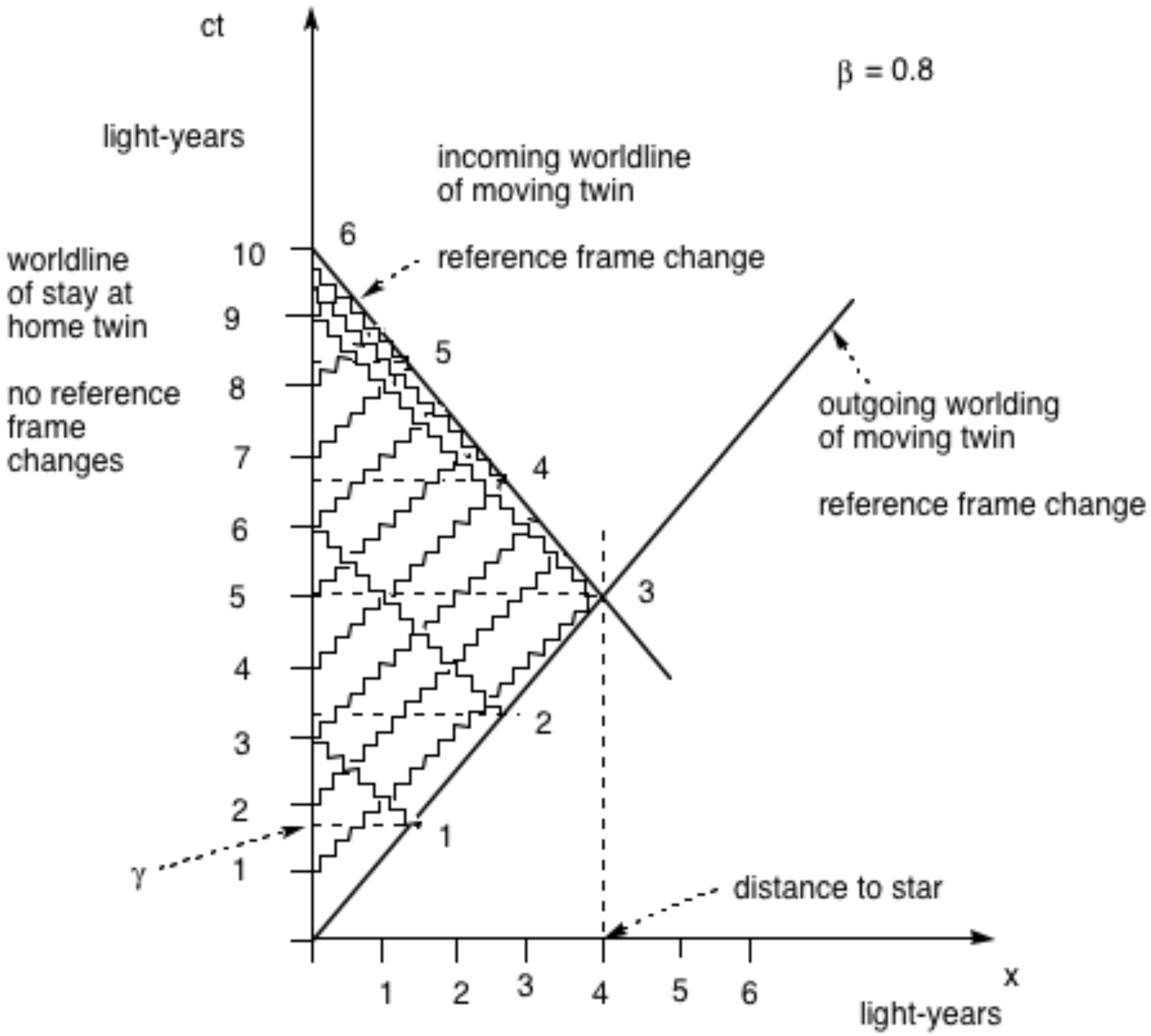
The twin in the rocket ship travels to a distant star and then changes her frame of reference again (reverses her velocity - accelerates for a period of time)

The twin in the rocket ship travels back to the earth and then changes her frame of reference again (come to rest on the earth - accelerates for a period of time).

Finally, the two twins remain together again (in the same frame of reference - on the same worldline).

Which twin, if any, is younger?

Description represented by spacetime diagram:



On diagram β = 0.8 and γ = 1.67.

Respective time axes have been calibrated.

Each twin sends one signal per year (by their own clock) to other twin.

While separating, k–factor says that

$$f_{observed} = f_{reduced} = \sqrt{\frac{1-\beta}{1+\beta}} 1(\text{per year}) = \frac{1}{3} \text{ per year}$$

While coming back together, k–factor says that

$$f_{observed} = f_{increased} = \sqrt{\frac{1+\beta}{1-\beta}} 1(\text{per year}) = 3 \text{ per year}$$

Clear from diagram that both twins see these different rates during designated periods.

For both twins reduced rate starts immediately.

However, switch over to increased rate takes place at different times according to each observer.

They are not identical observers and thus should not expect identical results from their measurements

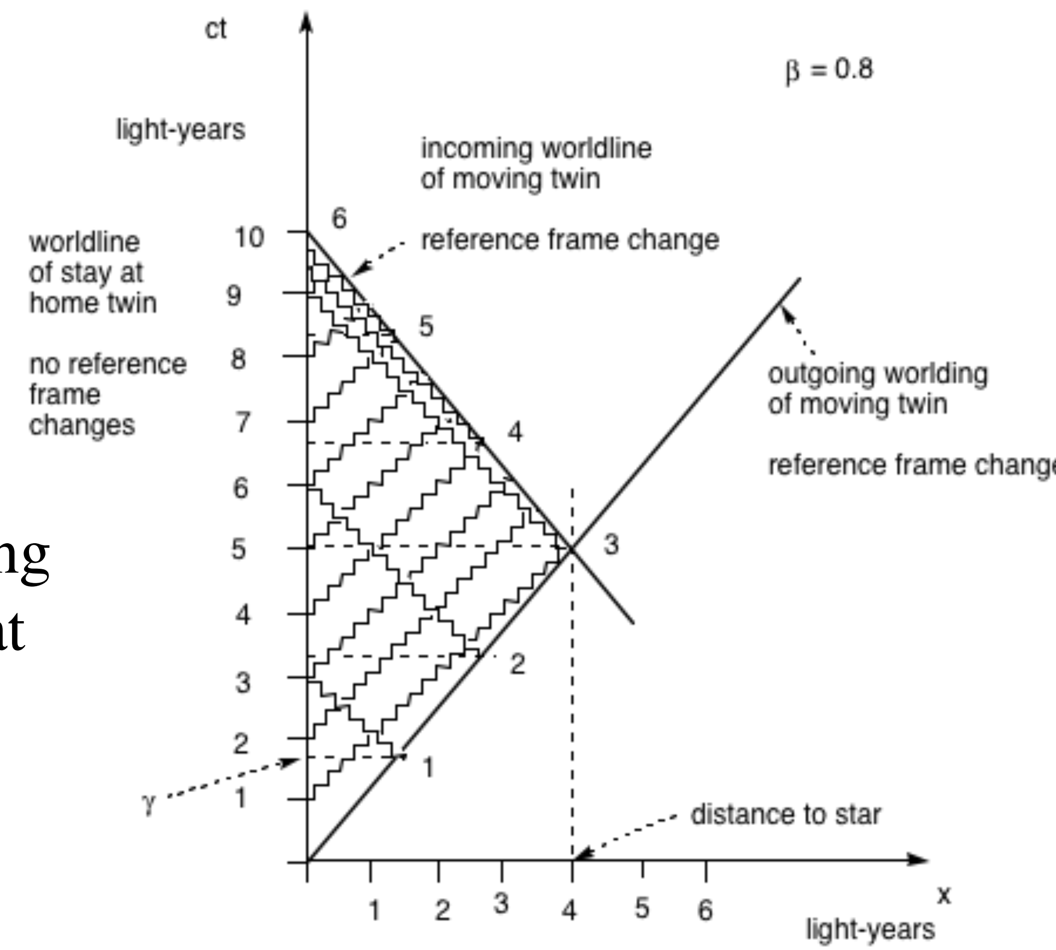
For moving twin switchover takes place exactly at midpoint of trip or at year 3 (see diagram).

For stay-at-home twin, however, switchover take place at year 9.

Thus moving twin sees 3 x 1/3 + 3 x 3 = 10 signals from stay-at-home twin

and thus knows that stay at home twin is 10 years older

and she is only 6 years older



Stay-at-home twin sees 9 x 1/3 + 1 x 3 = 6 signals from moving twin and thus knows that moving twin is 6 years older and that she is 10 years older

Both agree and this there is no paradox.

Traveler ages less because moving clocks

(clocks that have changed frames of reference) run slower .

The Pole in the Barn Paradox

Following situation:

2 farmers have barn which is 10 meters long in their rest frame (unprimed).

Farmers are standing at left and right doors of barn (doors are open).

Pole carrier has pole of length 12 meters in her rest frame and is carrying it horizontally while she runs towards barn with speed given by $\beta = 0.8$ and $\gamma = 1.67$.

If believe relativity and length contraction stuff then farmers think pole is

$$L_{pole} = \frac{L'_{pole}}{\gamma} = 9.8\, meters$$

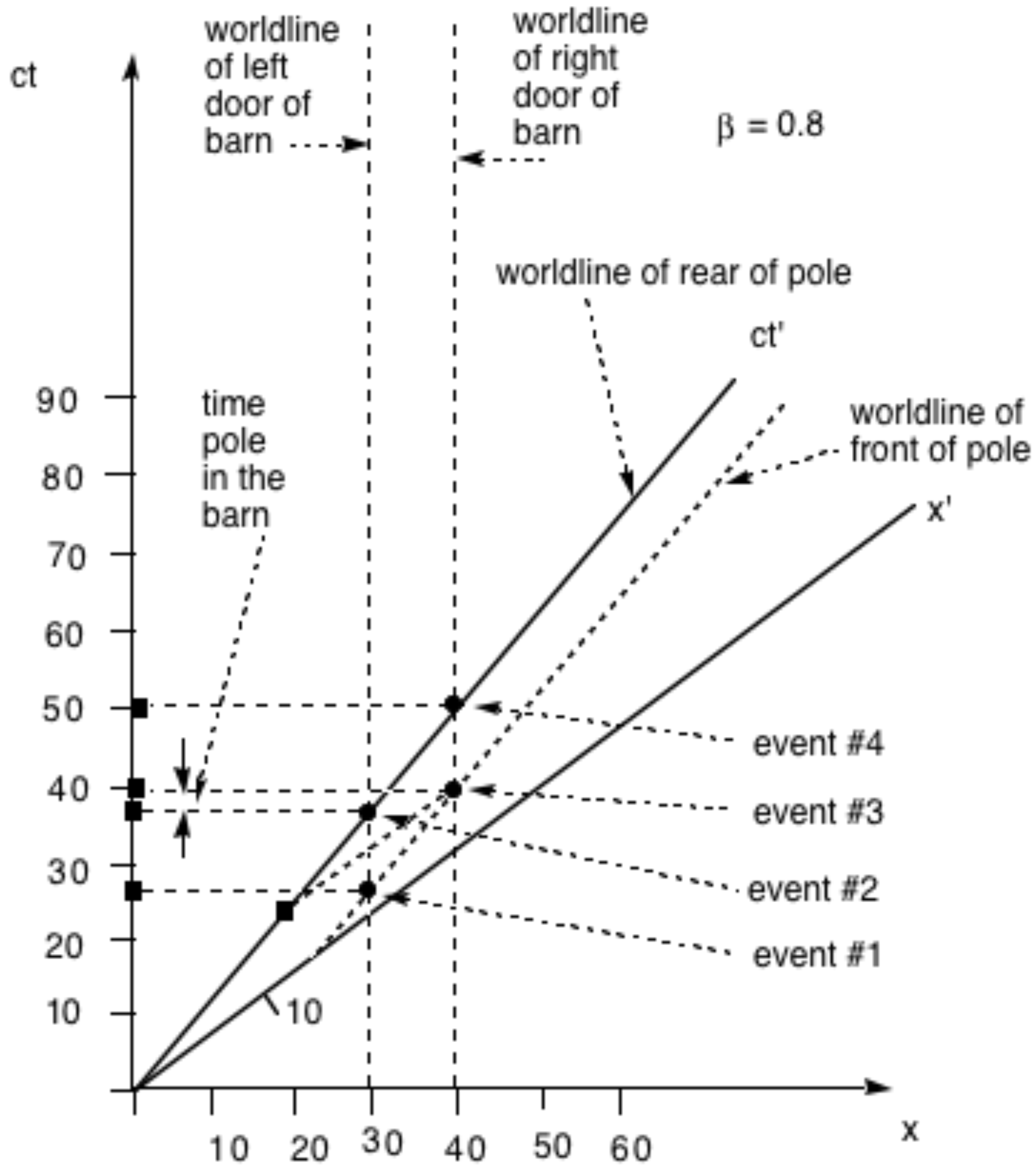
However, pole carrier thinks barn is only

$$L'_{barn} = \frac{L_{barn}}{\gamma} = 8.0\, meters$$

This means that, according to farmers, pole should be able to fit into barn.

Pole carrier, however, say no way, barn is much too small.

Possible spacetime diagram for experiment is shown below.



Is there any correct answer to dilemma?

To answer question, label **4 crucial events**:

Event #1: front of the pole enters the barn

Event #2: ear of the pole enters the barn

Event #3: front of the pole leaves the barn

Event #4: rear of the pole leaves the barn

Now if $t_3 > t_2$, then pole is completely within barn for period of time $t_3 - t_2$.

Clear from diagram, that according to farmers

pole is within barn for short period of time!

Pole carrier disagrees, however.

For pole carrier, $t'_2 > t'_3$ and therefore pole is never completely within barn.

There is disagreement between two sets of observers because the time order of two crucial events (namely 2 and 3) has **reversed (they were space like separated)**.

Thus, both are correct.

Pole is within barn and not within barn depending on your frame of reference.

Relativity is subjective, that is, dependent on observer information in certain cases.

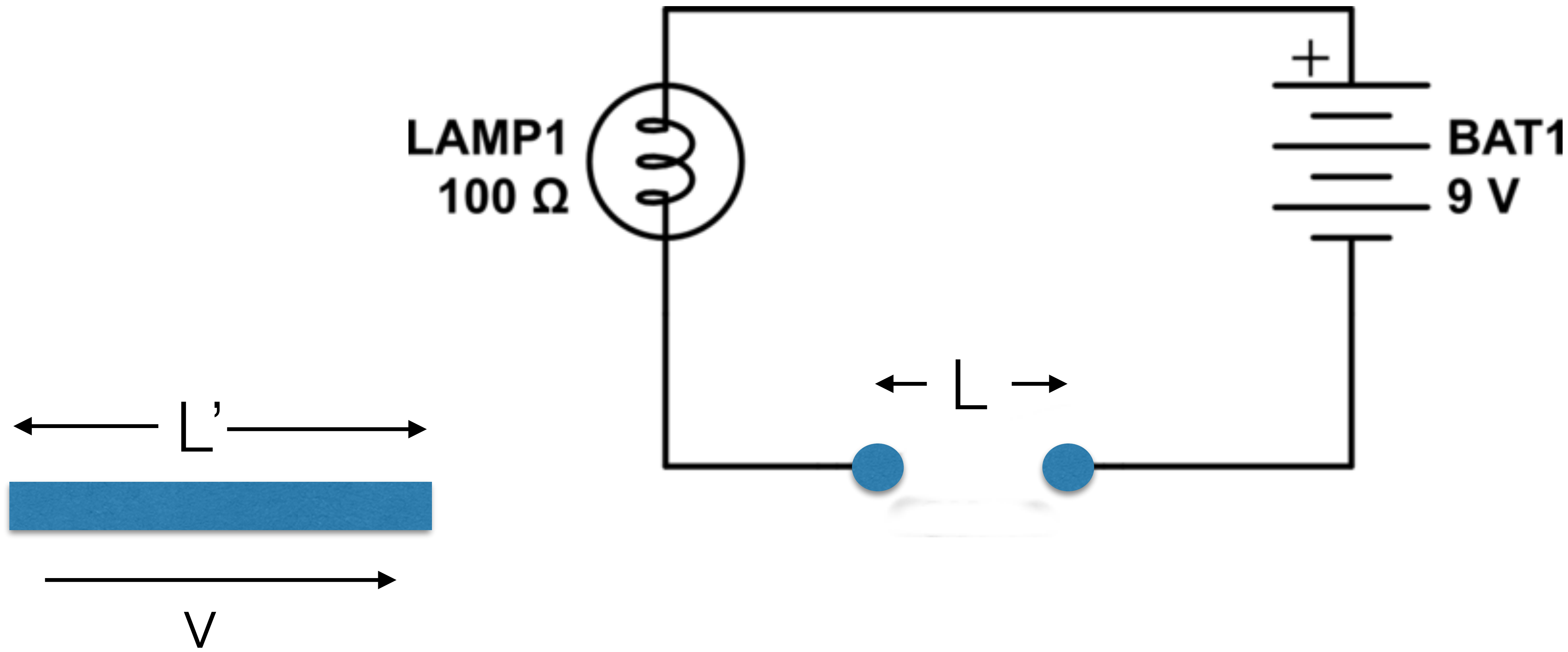
Relativity allows different observers to tell differing stories like this when time order reverses.

Time order reversal is OK in this case because events 2 and 3 are spacelike separated and thus reversing their time order cannot upset causality.

There is no paradox!

Now let us consider the following - for fun!

Variation of the Pole in the Barn



Gap is L = 10 meters long in their circuit frame (unprimed).

Rod has length L' = 12 meters in its rest frame and it is moving towards the Gap with speed given by $\beta = 0.8$ and $\gamma = 1.67$.

If believe relativity and length contraction stuff then circuit thinks rod is

$$L_{Pole} = L'_{Pole}/\gamma = 9.8 \text{ meters}$$

However, rod thinks gap is only

$$L'_{Gap} = L_{Gap}/\gamma = 8.0 \text{ meters}$$

This means that, according to the circuit, the rod should not make contact simultaneously with the two blue electrodes in the circuit and thus there will never be a closed circuit and the light bulb will never light.

The rod, however, says no way, the Gap is small enough for it to make contact simultaneously with the two blue electrodes in the circuit and thus there will be a closed circuit and the light bulb will light.

But, in this case, they cannot disagree about the end result

i.e., the light bulb either light of it does not light!

Signals faster than Light Paradox

What happens if allow some signal to go faster than speed of light?

Consider following story.

Sam is walking down a path towards a clock tower.

As he passes near near the tower a stone block falls off the tower and lands on his head, killing him.

So Sam is now lying in heap at the base of the tower.

Soon after that incident, Sally comes along.

Sam is Sally’s good friend and she is distraught when she sees Sam lying in a heap.

Sally is walking past Sam with some speed u (she is in a different frame of reference).

Now, Sally understands Special Relativity.

Sally has in her possession a special device that can send a signal to someone on the other side of the universe at a speed > c if they are in the same frame of reference.

So Sally sends out a signal indicating what happened to Sam.

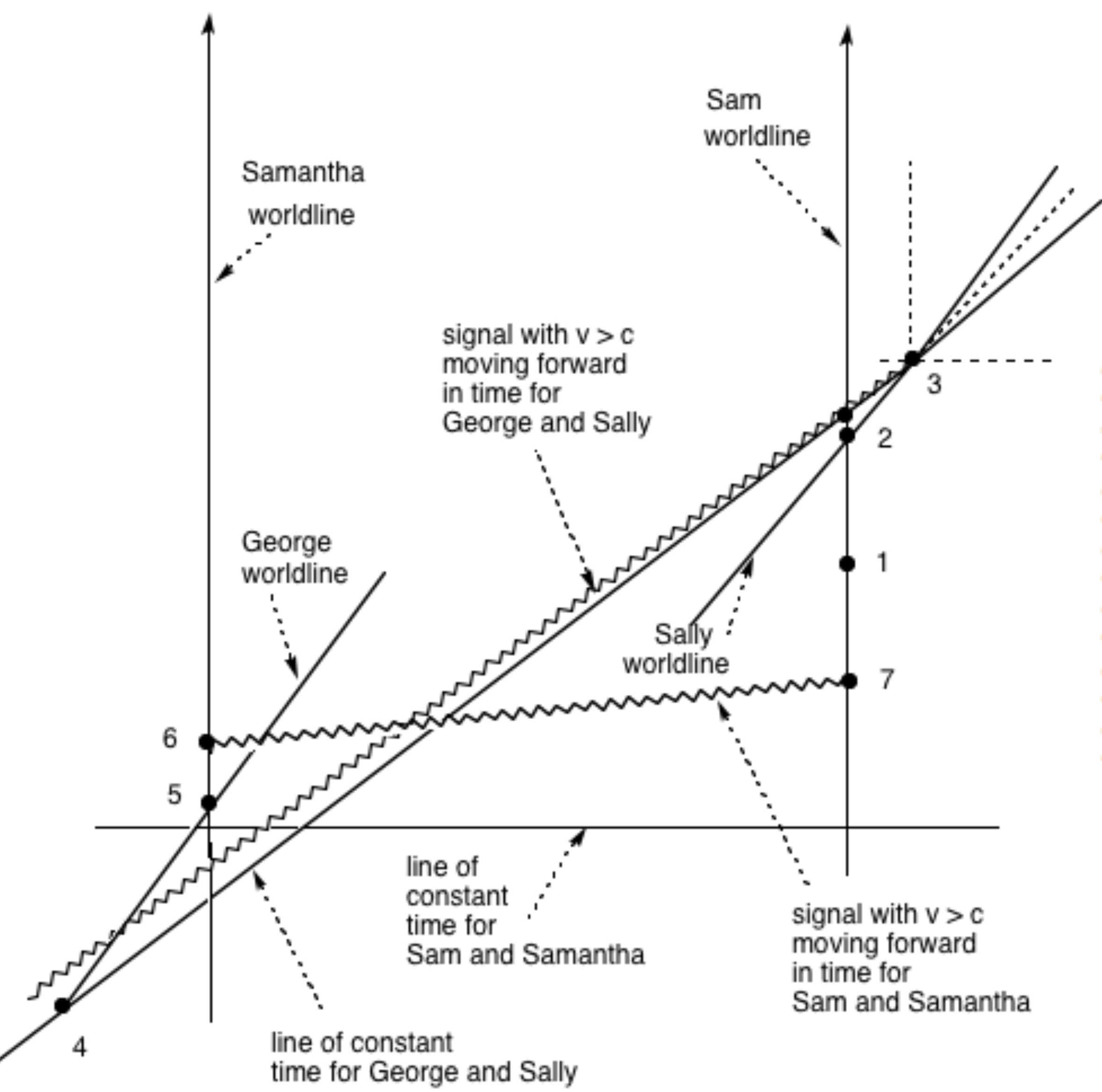
The signal is received on the other side of the universe by George (in the same frame of reference as Sally).

He is now desperate to tell Sam so he can avoid the stone block, but Sam is in a different frame and cannot receive his signal.

So he tells the story to someone in Sam's frame, namely, Samantha.

Samantha also happens to have one of those devices that sends the speedy signal and she sends a signal to Sam.

Entire sequence of worldlines with associated events shown in diagram:



Important events are:

Event #1: Sam gets killed
Event #2: Sally sees Sam
Event #3: After patiently waiting Sally sends a $v > c$ signal to George
Event #4: George receives the signal
Event #5: George tells Samantha
Event #6: Samantha patiently waits and then sends a $v > c$ signal to Sam
Event #7: Sam receives the signal from Samantha, realizes he is about to die and stops walking, thus avoiding the block and subsequent death

Questions

If Sam is not dead, why would Sally send any signal?

If Sally does not send signal making all other stuff happens
then why would Sam stop?

If Sam has no reason to stop
then he gets killed and Sally has reason to send signal.

Which is it?

We have what is called a **closed causal loop** here.

No logical way out of this loop.

Does that mean it cannot occur, i.e., that no signal can travel faster than light?
or is there some other explanation?

What about free will?

Now we shift gears.

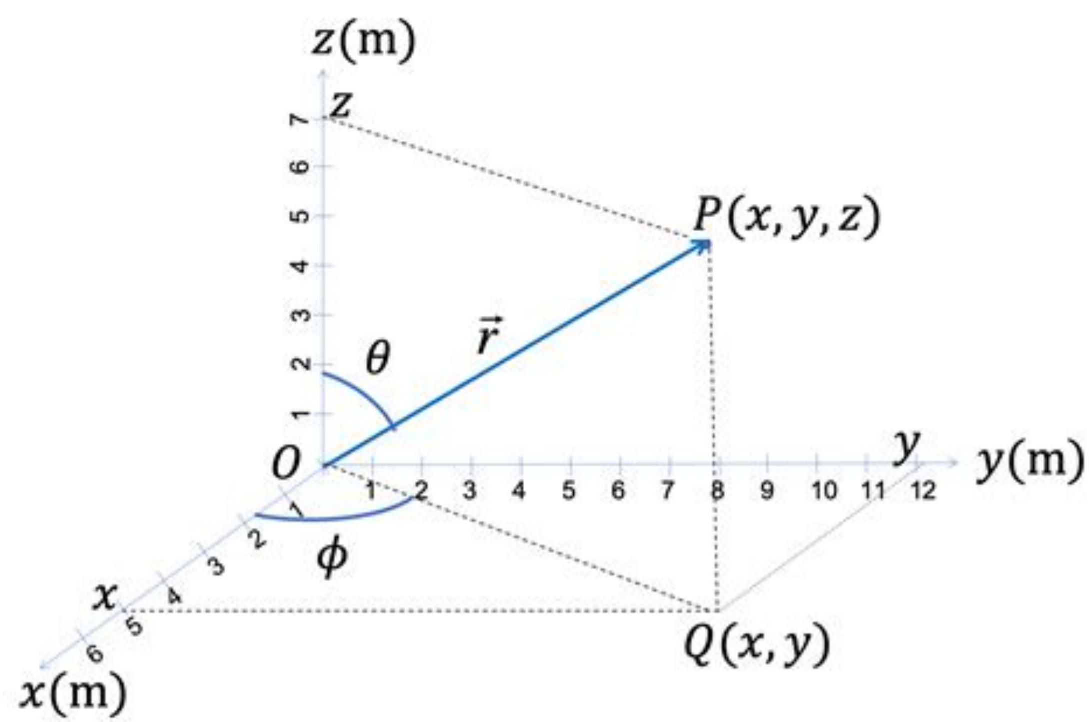
Dynamics in Special Relativity

Basic Ideas of Kinematics/Dynamics (study of motion in time) - Quick Tour

We will again make a quick pass through the idea so you are familiar with all the things involved and then go over everything again in much greater detail(and more mathematics).

Position ($\vec{r}(t)$) defined as vector from coordinate origin to 3-dimensional point where object located.

1-dimension -> x(t).



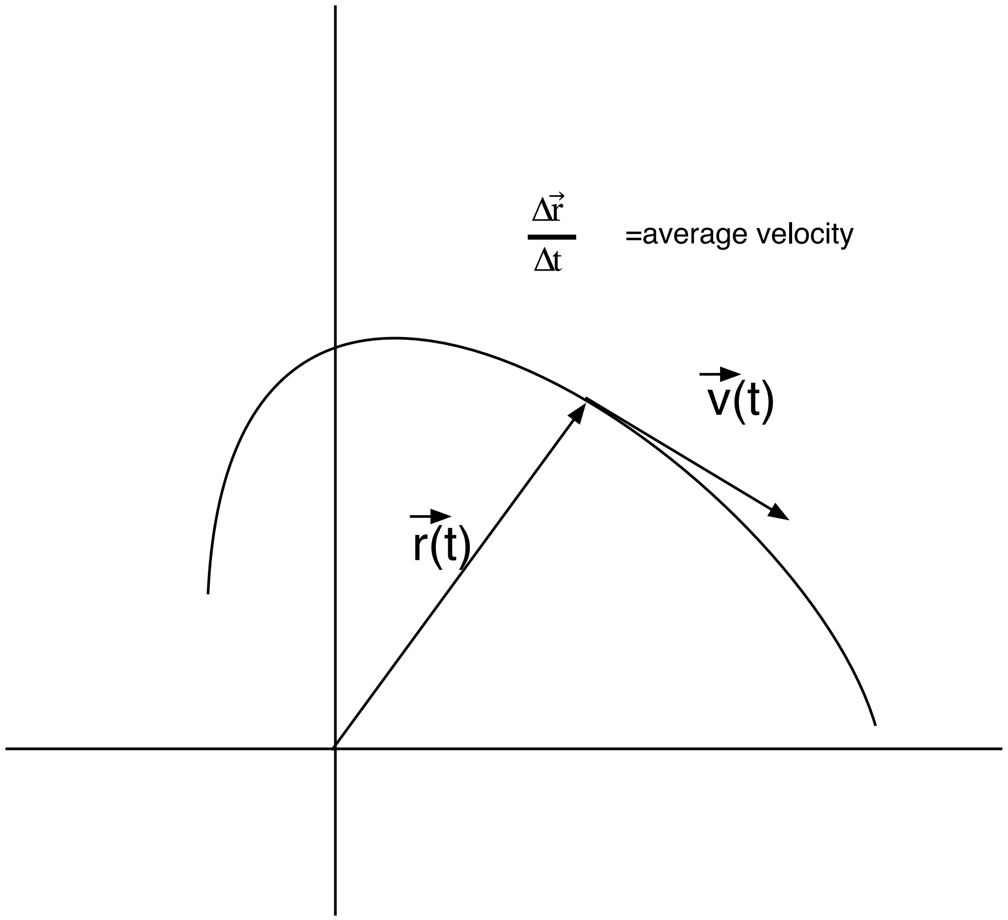
Goal of classical physics -> determine position of an object as function of time.

Position answers

Where question for events.

Velocity ($\vec{v}(t)$) defined as vector in direction of change of position vector and having magnitude given by

$$\vec{v}(t) = \lim_{\Delta t \to 0} \frac{\Delta \vec{r}}{\Delta t} \text{ or } v(t) = \lim_{\Delta t \to 0} \frac{\Delta x}{\Delta t} \text{ in 1-dimension}$$



Direction of velocity always tangent to path of motion (diagram).

Velocity -> how fast object is moving and in what direction.

If velocity constant -> both magnitude (speed) **and** direction are constant because it is vector.

Simple example in one dimension(± signs replace vectors for direction information).

Suppose particle has constant velocity $v = +10$m/s (+ means towards $+\infty$) and $x = 2$ m.

Where will particle be 1 sec later?

Clearly answer is $x = 12$m since $12 = 2 + 10(1) = x(t=0) + v\Delta t$.

If velocity is not constant situation is more **complicated**.

If, however, if I can tell you average velocity over next second = 10 m/s, then rule $12 = 2 + 10(1) = x(t=0) + v\Delta t$ still works ->

$$x(t + \Delta t) = x(t) + v(t)\Delta t \text{ in 1-dimension}$$

where $v(t)$ = average velocity during interval t to $t + \Delta t$

Acceleration is defined as vector in direction of change of velocity vector and having magnitude given by

$$\vec{a}(t) = \lim_{\Delta t \to 0} \frac{\Delta \vec{v}}{\Delta t} \text{ or } a(t) = \lim_{\Delta t \to 0} \frac{\Delta v}{\Delta t} \text{ in 1-dimension}$$

-> $v(t + \Delta t) = v(t) + a(t)\Delta t$ in 1-dimension

Now, suppose I interact with 2 different bodies in same manner, i.e., hang same object over pulley and attach it to 2 different bodies with a string (2 experiments).

Define "amount of stuff" in body by seesaw balancing

-> "amount of stuff" in 2 bodies is **identical** if seesaw balances

and

ratio of stuff in 2 bodies given by inverse ratio of distance from pivot

when seesaw balances.

Stuff = mass = m!

Experiment says that when one interacts with body (mass) or one **exerts a force**

-> body will **accelerate** (change velocity -> change position)

and we find that when we interact with 2 different bodies in same manner (exert same force)

$$\frac{a_1}{a_2} = \text{constant} = \frac{\text{stuff in 2}}{\text{stuff in 1}} = \frac{m_2}{m_1}$$

Whenever a simple results come out of an experiment, physicists(Newton and Galileo in this case) say that something profound must be going on here........

Turned equations around we have -> $m_1 a_1 = m_2 a_2$

-> this must have something to do with fact that I exerted ***same*** interaction!!

So given the acceleration, can calculate velocity and then calculate position and get answer .... process uses calculus ...... —> why Newton invented it.

But how do we find acceleration from 1st principles .... —> subject of Dynamics ....

Newton's Laws (crowning achievement of classical physics)

Body at rest is not moving!

No difference between body at rest and body moving with constant velocity since can always change frame of reference -> body with constant velocity looks like it is at rest and body that was at rest now looks like it has a constant velocity.

Body is interacting with surroundings when we observe changing velocity or acceleration.

Now push(or pull) on object -> accelerates.

Clear body has smaller/larger acceleration depending on strength of interaction.

Clear interaction directional ... produces directional or vector quantity -> acceleration.

-> concept of force.

Force = vector quantity that represents and quantifies interaction with body.

Earlier experiment, when I said the interaction in 2 cases was the same

-> I was exerting the same force.

Led Newton to postulate relationship $\vec{F} = m\vec{a}$

-> in earlier experiment exerting same force in two cases!

Be careful here!

Is there any new physical content to introduction of concept of force or is all the physics contained in acceleration?

I can **measure** acceleration!

Can I measure force or do I just **infer** it from measured acceleration?

Newton's laws are:

(1) an isolated body has no acceleration (Any real content?)

(1') body at rest or moving with constant velocity remains at rest or moving with constant velocity unless it interacts with something (Any real content?)

(2) $\vec{F} = m\vec{a}$ (is this simply a definition of force?)

(3) If body A exerts force on body B , then body B exerts equal and opposite force on body A (finally, some real content!)

Energy

Most dynamics problems of everyday world can be solved using Newton's laws.

But is not suitable for generalization beyond the realm of everyday experience.

In order to find rules and laws appropriate in other regimes of interest like very high speeds (Special Relativity(SR)) we must find a different way of thinking about universe.

This new way is based on Newton's laws -> no new physical content, but it will be possible to extend meaning of new laws so that new physical content and thus new physical theories can be formulated.

Energy -> new concepts that allows generalization.

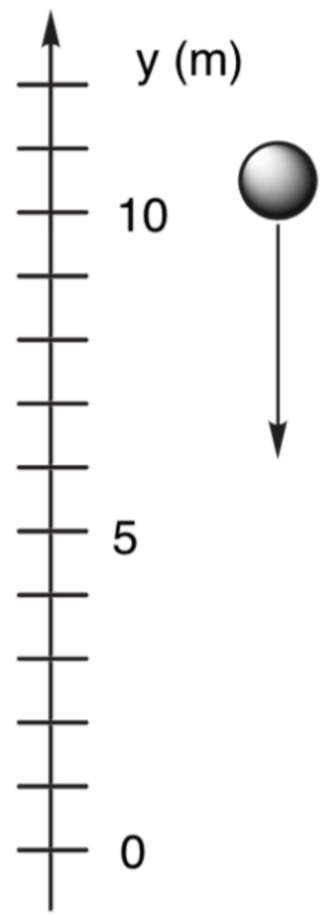
Define kinetic energy or energy due to motion as

$$K = \frac{1}{2}mv^2$$

An experiment.

Take any object raise it up to some height h above ground and then release it.

Find following relationships:



$$v_{ground}^2 = 2gh \quad g = 9.8\,m/s^2$$

$$v(t) = gt$$

$$y(t) = h - \frac{1}{2}gt^2$$

$$v^2(t) + 2gy(t) = \text{ constant}$$

Last result is the key.

As said earlier, much of theoretical physics is search for invariants.

Saw couple in SR.

When studying dynamics, invariants are quantities that are constant in time.

Last experimental relation can be written

$$\frac{1}{2}mv^2(t) + mgy(t) = \text{ constant}$$

$$K + V = E$$

where have defined two new energies

$$V = mgy(t) = \text{ potential energy}$$

$$E = K + V = \text{ total energy}$$

The experimental result then allows us to postulate

The total energy is a constant of the motion

Kinetic energy K and potential energy V are not constant during motion.

Constantly changing into one another, i.e., exchange between K and V during motion.

Law that energy is constant > example of conservation law for some invariant quantity.

Derive conservation laws from simple experiments and then generalize their validity to a much wider range of phenomena.

Momentum is another one of new concepts -> vector quantity.

Velocity is not important dynamic variable.

How can we see this?

Suppose have hill with two dump trucks at top.

One of dump trucks is filled with sand and other is empty.

Know from experiment that velocity is same for different trucks when they reach bottom of hill or they have same acceleration ...

Acceleration seems to be independent of amount of stuff in trucks (property of gravitational interaction).

Now ask this question?

Which of these 2 trucks would you attempt to stop at bottom of hill?

Clearly answer is truck with least stuff or smaller mass.

This tells us to define anew dynamical quantity

$$\vec{p} = m\vec{v} = \text{ linear momentum}$$

Newton's second law then becomes

$$\vec{F} = \lim_{\Delta t \to 0} \frac{\Delta \vec{p}}{\Delta t} = \lim_{\Delta t \to 0} \frac{\Delta (m\vec{v})}{\Delta t} \qquad \text{which for constant mass system becomes}$$

$$\vec{F} = \lim_{\Delta t \to 0} \frac{\Delta \vec{p}}{\Delta t} = \lim_{\Delta t \to 0} m\frac{\Delta \vec{v}}{\Delta t} = m\vec{a} \qquad \text{as before.}$$

Therefore now restate Newton's laws (using momentum) as:

(1) The linear momentum is conserved for an isolated body; $\Delta\vec{p} = 0$.

(2) $$\vec{F} = \lim_{\Delta t \to 0} \frac{\Delta \vec{p}}{\Delta t}$$

(3) The total momentum of an isolated system is a constant

$$\vec{p_1} + \vec{p_2} = \text{ constant}$$

$$\Delta\vec{p_1} + \Delta\vec{p_2} = 0$$

$$\frac{\Delta\vec{p_1}}{\Delta t} = -\frac{\Delta\vec{p_2}}{\Delta t}$$

$$\vec{F_{12}} = -\vec{F_{21}}$$

So for an isolated system have conservation law for linear momentum which one can generalize to a much wider range of phenomena.

So, here is way scientists of that day thought..

Classical universe followed well-defined laws.

Everything was, and is, predictable.

If find force, know masses, and velocities of all objects under consideration at one single time

then all is predictable from then on!!

Universe is gigantic Newtonian clockwork.

Cause and effect rule.

Nothing is by chance.

Everything is ultimately accountable.

Perfect determinism.

Laws of physics are to be obeyed, because it is impossible to disobey them.

The is no room for free will, salvation and damnation, or love and hate.

Even most trifling thought has been determined long ago.

Might have imagined that you are free-thinking person, but even that imagination is nothing but universal clockwork turning in some yet-to-be-discovered way.

So now you are probably thinking…glad they found out those ideas were wrong and got rid of them!

Just remember it is always dangerous to make quick judgements like that especially when you are not sure what will come along to replace it.

And then there was light.......... and Special Relativity.

Derivation of SR has shown that:

(1) Lose position and time as separate quantities
-> everyday experience may not carry over into these new realms.

Why didn't physicists notice before?

Generally matter of accuracy and precision available to experimentalists, i.e., prior to the 20th century, experimental measurements of speed of light could not say that it was not infinite.

If it were infinite, then special relativity would reduce to Galilean relativity and Newtonian physics would still be valid. Look at formulas.

(2) We must choose our observables with some care.....

(3) We must use conservation laws to give us the physical quantities that represent what we can really know about systems.

(4) We can fully extend classical physics validity to all speeds.

(5) We must rethink our world view (happens all the time in physics)

Everyday experience cannot be our guide

We must be prepared to give up preconceived ideas because they are based on our experiences

Everyday experience is fine for world of everyday objects

We must trust measurements to tell us what is going on but we must define them carefully

But classical physics still hangs on, albeit modified.......

Everything works so well ....

What does that statement really means to a physicist?

We must adopt an "only know what we measure" ....philosophy.

In this world motion is a continuous blend of changing positions.

The object moves in a flow from one point to another.

Science is a reasonable, orderly process of observing nature and describing the observed objectively.

There is a conviction that whatever one observed as being out there was really out there.

The idea of objectivity being absent from science is abhorrent to any rational physicist.

One firmly believes in the passive(non-disruptive) observer.

Humans are creatures of the eye.

They believe what they see.

So summarizing, classically

(1) Things moved in a continuous manner.

(2) Things move for reasons. The reasons are earlier causes and all motion was determined and predictable.

(3) All motion could be analyzed or broken down into its component parts. Each part played a role in the giant machine called the universe.

The complexity of this machine could be understood in terms of the simple movement of its various component parts.

(4) The observer observed and never disturbed. All experimental errors could be analyzed and understood.

All of these ideas turn out to be **false** in modern theoretical physics!!!!!

Now back to SR.

What happens to momentum and energy when we enter the realm of SR?

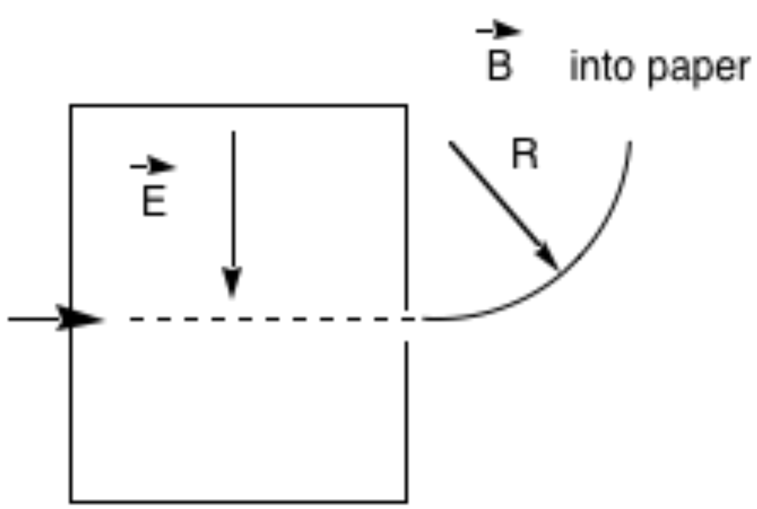
At this level, we must rely on experiments to point proper way to proceed.

Following result confirmed by experiment.

Force felt by charged particle in electric and magnetic fields is given mathematically by Lorentz force law

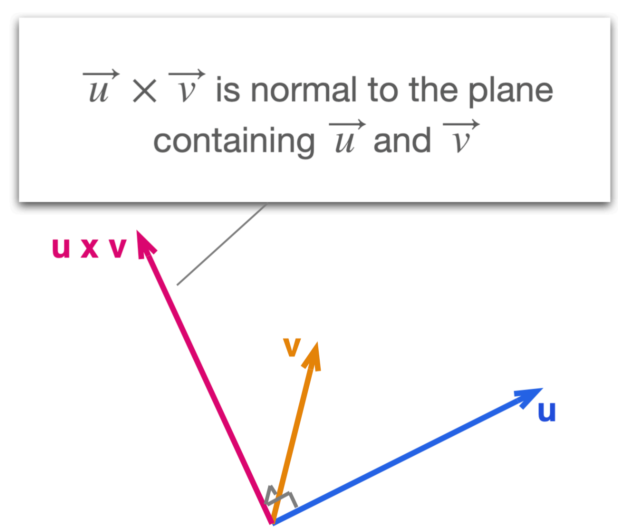
$$\vec{F} = q\left(\vec{E} + \frac{\vec{v} \times \vec{B}}{c}\right)$$

Consider experimental setup



where $\vec{\mathrm{v}}$ is particle velocity and $\vec{\mathrm{E}}$ and $\vec{\mathrm{B}}$ are electric and magnetic field.

$\vec{\mathrm{v}} \times \vec{\mathrm{B}}$ relationships as shown:



In box region electric and magnetic fields adjusted

-> $\vec{F} = 0$ for particle moving along dotted line with definite velocity.

Electric force points downward

and magnetic force perpendicular to velocity direction

(upward in box for particle moving along dotted line).

Means particles with particular velocity, namely,

$$q\left(-E + \frac{v}{c}B\right) = 0 \rightarrow \frac{v}{c} = \frac{E}{B}$$

pass undeflected though box. —> Box = velocity selector.

Outside box

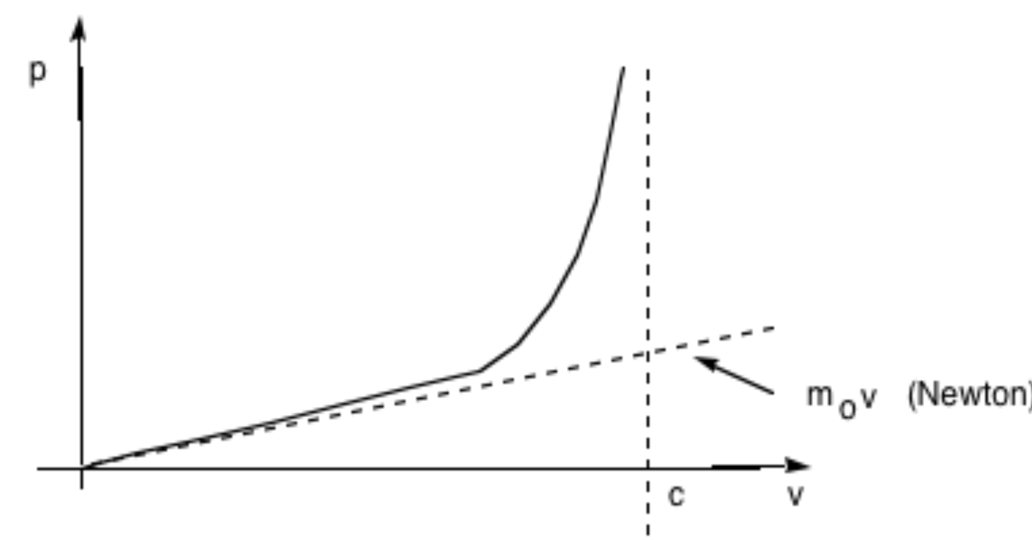
-> no electric field, so particle moves on circular path

(force always perpendicular to velocity) with radius

$$R = \frac{pc}{qB}$$

-> measuring radius corresponds to measuring relativistic momentum.

Thus, in **same** experiment can measure both velocity and momentum independently and thus **determine** the relationship between them.



Plot of the experimental results looks like

Corresponds to result

$$p = \gamma m_0 v \text{ where } \gamma = \frac{1}{\sqrt{1-\beta^2}}$$

instead of Newtonian assumption that p = m_0v, where m_0 = so-called **rest mass**.

Note that "rest mass" is the only valid mass for particle since measure mass when body at rest.

Any measurement of mass when particle moving = measurement of momentum and this -> it is incorrect to assume that any different mass value can be used for a moving object.

There is no such thing as relativistic mass.

What is relativistically correct form of energy of particle?

Generalize concept of energy

-> use Newtonian definition of kinetic energy in conjunction with relativistically correct definition of momentum.

Derivation uses calculus.

Do not worry if cannot follow mathematical steps - included for mathematically inclined.

For this class only the results are important.

Proceed (just once for those who know calculus) as follows.

Formal definition of kinetic energy is

$$\Delta K = K - K_0 = \text{ work done by force } = \int_{\vec{r}_0}^{\vec{r}} \vec{F} \cdot d\vec{r} = \int_{\vec{r}_0}^{\vec{r}} \frac{d\vec{p}}{dt} \cdot d\vec{r}$$

Experiment -> $\vec{p} = m_0\gamma(v)\vec{v}$ where $\gamma(v) = (1 - \beta^2)^{-1/2}$ and $\beta = v/c$.

Therefore we have

$$K - K_0 = \int_{\vec{r}_0}^{\vec{r}} \frac{d(m_0 \gamma(v) \vec{v})}{dt} \cdot \vec{v} dt = m_0 \int_0^v \vec{v} \cdot d(\gamma(v)\vec{v})$$

Since kinetic energy = zero when velocity = zero have

$$K = m_0 \int_0^v \vec{v} \cdot d(\gamma(v)\vec{v})$$

Now

$$d(\gamma v^2) = d(\gamma \vec{v} \cdot \vec{v}) = \vec{v} \cdot d(\gamma \vec{v}) + \gamma \vec{v} \cdot d\vec{v}$$

-> can write

$$K = m_0 \int_0^v (d(\gamma v^2) - \gamma \vec{v} \cdot d\vec{v}) = m_0 \int_0^v d(\gamma v^2) - m_0 \int_0^v \gamma \vec{v} \cdot d\vec{v})$$

$$= m_0 \int_0^v d(\gamma v^2) - \frac{1}{2} m_0 \int_0^v \gamma d(v^2) = m_0 \gamma v^2 - \frac{1}{2} m_0 c^2 \int_0^{v^2/c^2} \frac{du}{\sqrt{1-u}}$$

$$= m_0 \gamma v^2 + m_0 c^2 \left(\frac{1}{\gamma} - 1 \right) = m_0 c^2 \left(\frac{1}{\gamma} + \gamma \beta^2 \right)$$

$$= m_0 c^2 (\gamma - 1)$$

Does this result makes sense?

We know ordinary definition of kinetic energy.......

Let us compare.

What is low velocity (non-relativistic) limit of the new expression?

Using $\gamma(v) = (1-\beta^2)^{-1/2} \approx 1 + \frac{1}{2}\beta^2 = 1 + \frac{1}{2}\frac{v^2}{c^2}$ we have

$$K = m_0 c^2(\gamma - 1) \approx m_0 c^2 \frac{1}{2}\frac{v^2}{c^2} = \frac{1}{2} m_0 v^2$$ as expected.

So new expression is the relativistic form of kinetic energy

If rearrange our result we have

$$\gamma m_0 c^2 = K + m_0 c^2 = \text{ Energy(motion) } + \text{ Energy(rest) } = \text{ Total Energy } = E$$

Now, it is **only** total energy that is conserved!

As a bonus, we obtain Einstein's famous relation

$$E_{rest} = m_0 c^2$$

What is the connection to momentum?

Some algebra gives following important results for relativistic objects:

$$\frac{pc}{E} = \frac{m_0 \gamma v c}{\gamma m_0 c^2} = \frac{v}{c} = \beta \text{ and } \left(\frac{E}{c}\right)^2 - p^2 = (m_0 c)^2$$

Some questions arise:

How come we do not notice rest energy in everyday experiences?

Some numbers for 1 Kg mass moving at 1 m/s:

$$\text{typical kinetic energy } = 0.5(1)(1)^2 \approx 1\, Joule$$

$$\text{typical rest energy } = (1)(3 \times 10^8)^2 \approx 10^{17}\, Joule$$

In most beginning physics classes, we typically ignore the significantly larger quantity!!

Reason for this is that in everyday situations rest energy does not change; same amount of mass remains in system at all times.

Thus, rest energy is not a source of possible energy to do other things.

However, in microscopic systems like atoms and nuclei, etc, rest mass changes in many interactions and thus this energy becomes available for other purposes.

Two examples are nuclear fission and fusion.

Are there any new predictions we can make from these results?

The two relations

$$\frac{pc}{E} = \frac{m_0 \gamma v c}{\gamma m_0 c^2} = \frac{v}{c} = \beta \text{ and } \left(\frac{E}{c}\right)^2 - p^2 = (m_0 c)^2$$

make following interesting prediction:

$$v = c \rightarrow \beta = 1 \rightarrow E = pc$$

$$\left(\frac{E}{c}\right)^2 - p^2 = 0 = (m_0 c)^2$$

or objects that can travel at speed of light must have rest mass equal to **zero**!

However, even though light particles have zero rest mass, still possess energy and momentum thus defying classical ideas and equations!

Such a particle has been observed .... it is the **photon** or the particle of light.

Now for a second pass on these ideas with much greater detail and the appropriate mathematics.

Up to this point, we have been studying **relativistic kinematics** (*kinematics* is the study of how the motion of objects is to be measured and described mathematically.)

But physicists are not only interested in *describing* motion, they are interested in explaining how objects interact and how the forces that they exert on each other *determine* their motion.

The study of object interactions is called *dynamics*.

In the remainder of this class, we will explore the basic principles of **relativistic dynamics**.

The basic principles of Newtonian dynamics are expressed by Newton's famous three laws of motion.

The laws of Newtonian dynamics are consistent with the principle of relativity if the Galilean velocity transformation equations are true —> **belief in absolute time**.

But as we saw earlier that the Galilean velocity transformation equations are in fact not true; they only represent the low-velocity limit of the relativistically correct Einstein velocity transformations.

This means that the laws of Newtonian dynamics are not generally consistent with the principle of relativity; the laws of Newtonian dynamics likewise represent only low-velocity approximations to the laws of relativistic dynamics, laws that are the same in all inertial frames (as the principle of relativity requires).

Well, what are these laws of relativistic dynamics, and how can we find them?

We could address this question by searching for a relativistic generalization of Newton's second law, then Newton's third law, then the law of universal gravitation, and so on.

This approach(used earlier) seems on the surface to be logical and straightforward.

But it turns out that the most fruitful way to address this problem is to focus our attention on the law of conservation of momentum.

Not only is the correct relativistic generalization of this law fairly easy to find, but this law proves to be extremely rich in implications and applications.

Indeed, virtually everything that is useful to know about relativistic dynamics can be learned by a close examination of the law of conservation of momentum.

The basic argument in this discussion can be outlined as follows.

1 First, I will show you that conservation of ordinary newtonian momentum (defined as mass × velocity) is *not* consistent with the principle of relativity: if the total newtonian momentum of a system of interacting objects is conserved in one inertial reference frame, it will *not* be conserved in other inertial frames. This means that we need to look for an appropriate relativistic modification of the idea of momentum that *can* be made consistent with the principle of relativity.

2 I will then propose a natural relativistic generalization of the idea of momentum called the *four-momentum*. This four-momentum is a *four*-component vector having a *time component* as well as the usual x, y, and z components.

3 I will then show you that a law of conservation of four-momentum *is* consistent with the principle of relativity (and thus represents a reasonable candidate as a relativistic generalization of the newtonian law of conservation of momentum). But for this to be true, the t component of a system's total four-momentum vector *must be conserved* along with its x, y, and z components. So the law of conservation of four-momentum not only generalizes the idea of conservation of newtonian momentum, it tells us that *something else is conserved as well.*

4 This fourth conserved quantity (i.e., the t component of the four-momentum vector) turns out to be a relativistic generalization of the idea of *energy*. Thus the law of conservation of four-momentum not only represents a relativistic generalization of the newtonian law of conservation of momentum but the newtonian law of conservation of energy as well! Special relativity thus ends up giving us two conservation laws for the price of one!

Let us see how this all works out.

NEWTONIAN MOMENTUM IS NOT CONSERVED IN ALL FRAMES

The law of conservation of momentum must satisfy the principle of relativity if it is to be a valid physical law.

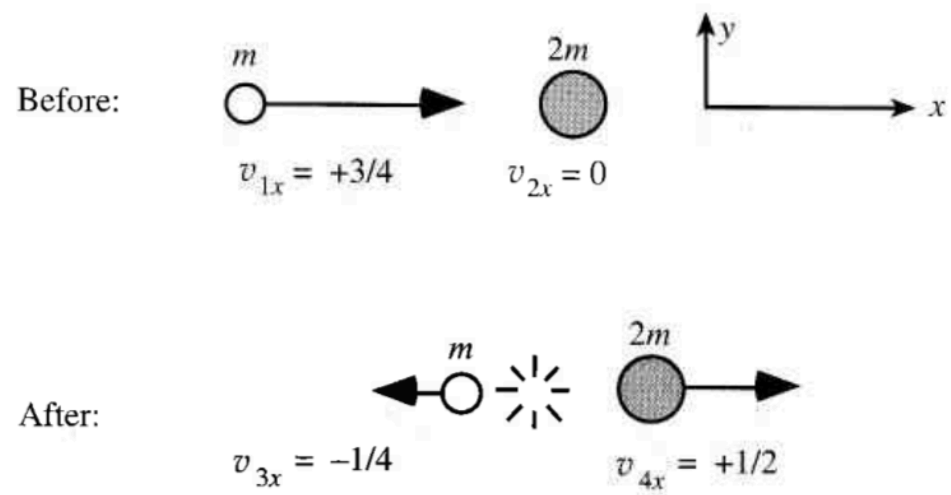
This implies that if the total momentum of a system of interacting particles is conserved in one inertial reference frame (and the law of conservation of momentum is a valid physical law), then it should be conserved in all inertial reference frames.

Now, I will show you that the law of conservation of newtonian momentum is not consistent with the principle of relativity and thus cannot be a valid law of physics.

To illustrate the problem, it is sufficient to demonstrate a single instance of the inconsistency.

For the sake of simplicity, I will illustrate the difficulty with the ordinary newtonian definition of momentum using a simple one-dimensional collision.

Figure shows such a collision as viewed in the Home Frame.



A hypothetical collision of two particles as viewed in the Home Frame. The total newtonian momentum of this system is conserved in this frame.

In this frame, an object of mass m is moving in the +x direction with an x velocity $v_{1x} = +3/4$.

It then strikes an object of mass 2m at rest ($v_{2x} = 0$).

The lighter mass rebounds from the collision with an x velocity of $v_{3x} = -1/4$, while the heavier object rebounds with a velocity of $v_{4x} = +1/2$.

The system can be considered to be isolated from external effects.

The system's total Newtonian x momentum is conserved in the Home Frame in this case:

Total x momentum *before*: $mv_{1x} + 2mv_{2x} = m(+3/4) + 2m(0) = +3m/4$

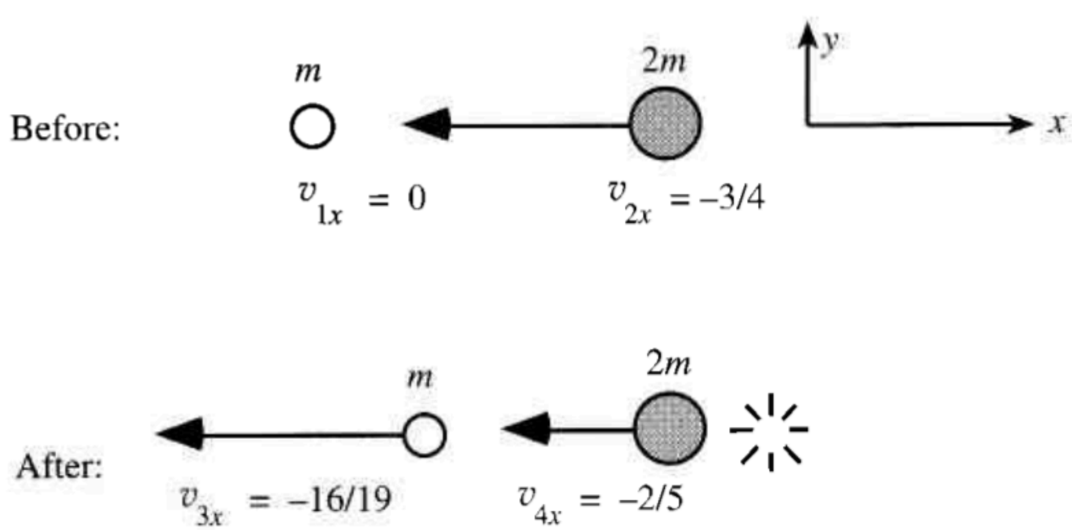
Total x momentum *after*: $mv_{3x} + 2mv_{4x} = m(+1/4) + 2m(+1/2) = +3m/4$

Now consider how this collision looks when observed in an Other Frame that is moving with a speed β = 3/4 in the +x direction.

Since this frame essentially moves along with the lightweight particle, it appears to be at rest in the Other Frame: $v'_{1x} = 0$.

Since the larger object is at rest in the Home Frame and the Home Frame appears to be moving backward with respect to the Other Frame at a speed of 3/4, the x velocity of the more massive object must be $v'_{2x} = -3/4$ as well.

In this frame, then, the collision will be measured to happen as shown in the Figure.



A hypothetical collision of two particles as viewed in the Home Frame. The total newtonian momentum of this system is conserved in this frame.

The objects' final x velocities are not so easy to intuit: we need to compute these velocities using the first of the Einstein velocity transformation equations:

$$v'_{3x} = \frac{v_{3x} - \beta}{1 - \beta v_{3x}} = \frac{-1/4 - 3/4}{1 - (3/4)(-1/4)} = -\frac{16}{19}$$

$$v'_{4x} = \frac{v_{4x} - \beta}{1 - \beta v_{4x}} = \frac{+1/2 - 3/4}{1 - (3/4)(1/2)} = -\frac{2}{5}$$

Note that these must be the final velocities of the two objects if the Einstein velocity transformation equation is true and the collision actually takes place as shown in the last Figure.

The system's total Newtonian x momentum is *not* conserved in this case!

$$\begin{aligned}
&\text{Total } x \text{ momentum } \textit{before:} \quad mv'_{1x} + 2mv'_{2x} = m(0) + 2m(-3/4) = -3m/2 \\
&\text{Total } x \text{ momentum } \textit{after:} \quad mv'_{3x} + 2mv'_{4x} = m(-2/5) + 2m(-16/19) \\
&\qquad\qquad\qquad\qquad\qquad\qquad\qquad\quad = -156m/95
\end{aligned}$$

We can see in this frame that the x component of the momentum is somewhat larger after the collision than it was before the collision, since 156/95 > 3/2.

The law of conservation of newtonian momentum therefore does not hold in the Other Frame, even though it did hold in the Home Frame.

Note that the law of conservation of momentum requires that each component of the system's total momentum be conserved separately, so a violation of the conservation of even one component, the x component in this case, is a violation of the entire law.

The principle of relativity requires that the laws of physics be the same in all inertial reference frames.

The conclusion is inescapable: If the Einstein velocity transformation equations are true, then the law of conservation of newtonian momentum is not consistent with the principle of relativity.

THE FOUR-MOMENTUM VECTOR

Ordinary Newtonian momentum $\vec{p}$ of an object is defined to be mass times velocity:

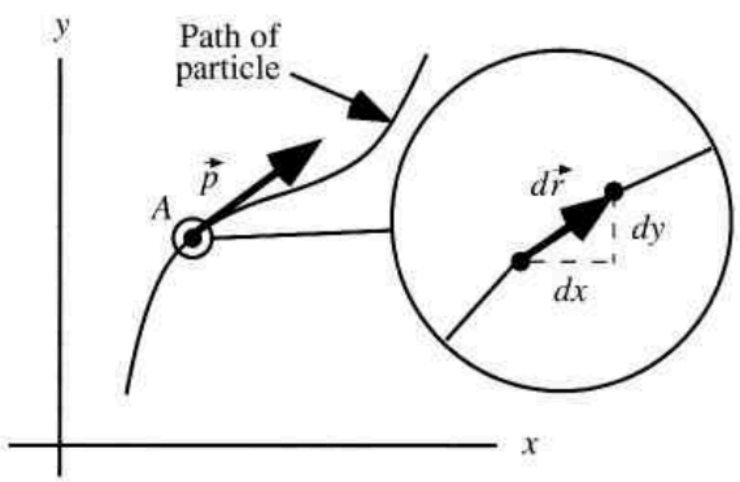
$$\vec{p} \equiv m\vec{v} = m\frac{d\vec{r}}{dt}$$

The vector $d\vec{r}$ represents an infinitesimal displacement in space, which is divided by an infinitesimal time dt to get the object's velocity vector $\vec{v}$.

The components of the infinitesimal displacement vector $d\vec{r}$ are [dx, dy, dz], so the components of the Newtonian momentum are

$$p_x = mv_x = m\,\frac{dx}{dt} \qquad p_y = m\,\frac{dy}{dt} \qquad p_z = m\,\frac{dz}{dt}$$

Notice that the momentum vector is parallel to the infinitesimal displacement $d\vec{r}$ and so will be tangent to the path of the object through space, as shown in the Figure.



The path of a particle through space is shown in the graph at left. The ordinary momentum $\vec{p}$ of the particle at point A is defined to be a vector parallel to the displacement vector $d\vec{r}$ that connects two infinitesimally separated points surrounding A.

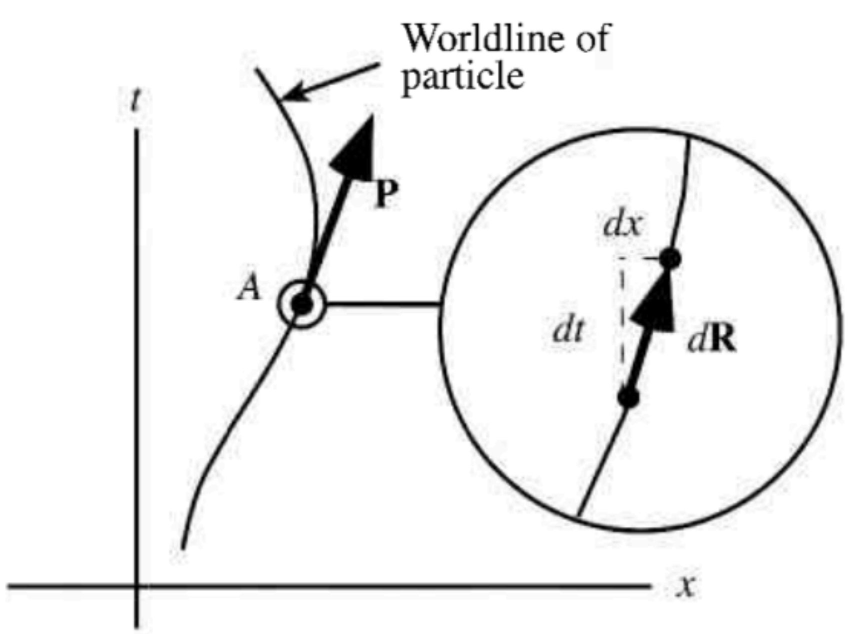
How can we arrive at a relativistic generalization of this process?

In special relativity, space and time are considered to be equal parts of a whole called spacetime; thus we describe the motion of an object not merely by describing its path through space but by describing its worldline through spacetime.

The appropriate relativistic generalization of an infinitesimal displacement in space $d\vec{r}$ between two infinitesimally separated points on an object's path in space is a displacement $d\mathbf{R}$ in spacetime between two infinitesimally separated events on the object's worldline (see Figure).

Note that the displacement $d\mathbf{R}$ in spacetime between two events is specified by four numbers:

$$d\mathbf{R} = [dt, dx, dy, dz]$$



The worldline of an object through space is shown in the spacetime diagram at left. The relativistic momentum $\mathbf{P}$ of the object at event A is defined to be an arrow parallel (on the diagram) to the displacement arrow $d\mathbf{R}$ that connects two infinitesimally separated events surrounding the event A. (The y and z dimensions are not shown on this diagram.)

Including the time displacement dt on an equal footing with the spatial displacements dx, dy, dz makes the displacement $d\mathbf{R}$ a four-component vector called a four-vector.

In these notes, I will always use boldface capital letters ($d\mathbf{R}$, $\mathbf{P}$, etc.) to represent such four-vectors.

Given the components [dt, dx, dy, dz] of the displacement four-vector $d\mathbf{R}$ that stretches between two infinitesimally separated events on the worldline of our object, how do we define its relativistic momentum?

By analogy with the newtonian momentum, we want to divide $d\mathbf{R}$ by a quantity that somehow represents the time between the two events and then multiply by the mass of the object.

But in the theory of relativity, the time measured between two events depends on who does the measuring!

Whose time shall we use?

There is one kind of time that can be measured between the events that has a frame- independent value and is uniquely related to the motion of the object in question: this time is the infinitesimal proper time $d\tau$ between the events that would be measured by a clock traveling with the object itself.

When you think about it, it makes a certain kind of sense to use the object's own time to characterize its momentum.

Therefore our proposed relativistic generalization of the newtonian momentum $\vec{p}$ of an object is the relativistic four-momentum $\mathbf{P}$, defined as follows:

$$\mathbf{P} = m \frac{d\mathbf{R}}{d\tau}$$

The four-momentum is a four-dimensional vector having components [P_t, P_x, P_y, P_z) such that

$$P_t = m \frac{dt}{d\tau}$$

$$P_x = m \frac{dx}{d\tau}$$

$$P_y = m \frac{dy}{d\tau}$$

$$P_z = m \frac{dz}{d\tau}$$

Just as the newtonian momentum is a vector tangent to the object's path through space, the four-momentum (displayed as an arrow on a spacetime diagram) is tangent to the object's worldline, as shown in the last Figure.

We can express the components of the four-momentum in a given inertial frame in terms of the ordinary velocity of the object measured in that frame.

Earlier results tell us that the infinitesimal proper time between two infinitesimally separated events measured by a clock traveling between two events at a speed v in a given reference frame is related to the coordinate time dt measured between those events by

$$d\tau = \sqrt{1-v^2}\, dt$$

This means that

$$P_t \equiv m\frac{dt}{d\tau} = \frac{m}{\sqrt{1-v^2}}\frac{dt}{dt} = \frac{m}{\sqrt{1-v^2}}$$

$$P_x \equiv m\frac{dx}{d\tau} = \frac{m}{\sqrt{1-v^2}}\frac{dx}{dt} = \frac{mv_x}{\sqrt{1-v^2}}$$

$$P_y \equiv m\frac{dy}{d\tau} = \frac{m}{\sqrt{1-v^2}}\frac{dy}{dt} = \frac{mv_y}{\sqrt{1-v^2}}$$

$$P_z \equiv m\frac{dz}{d\tau} = \frac{m}{\sqrt{1-v^2}}\frac{dz}{dt} = \frac{mv_z}{\sqrt{1-v^2}}$$

These equations allow one to find the components of the four-momentum of an object in a given frame knowing the object's velocity vector $\vec{v}$ in that frame.

When the speed v of an object becomes very small compared to the speed of light ($v << 1$), the square roots in the denominators in above equations become almost equal to 1, and we have

$$\begin{aligned} P_t &\approx m \\ P_x &\approx mv_x \\ P_y &\approx mv_y \\ P_z &\approx mv_z \end{aligned} \qquad (\text{when } v << 1)$$

Thus in the limit where the velocity of an object is very small (i.e., ordinary, everyday velocities), the spatial components of the four-momentum reduce to being the same as the corresponding components of the object's Newtonian momentum.

Note also that since velocity in the SR unit system is a unitless number, all four components of an object's four-momentum have units of mass in the SR system.

PROPERTIES OF THE FOUR-MOMENTUM VECTOR

Why define an object's relativistic four-momentum this way?

The definition has several attractive features.

One feature has already been mentioned: on a spacetime diagram, the four-momentum is represented by an arrow tangent to the worldline of the object through spacetime, just as the ordinary momentum vector is represented by an arrow tangent to the path of the object through space.

It is also nice that the definition of the four-momentum treats the time coordinate in the same manner as the spatial coordinates: all four coordinate displacements dt, dx, dy, dz appear on an equal footing in the definition of the four-momentum as given.

We have already seen how it is important in relativity theory to treat time and space as being equal participants in the larger geometric whole that we call spacetime.

The definition of the four-momentum given above maintains this symmetry between the different spacetime coordinates.

All this symmetry has a certain beauty about it which a physicist like Einstein might take as corroborating evidence that we are on the right track with this definition.

But we will see that the most important feature of the definition is that, given the components of the four-momentum in a particular reference frame, there is a very simple and straightforward method for calculating what its components will be measured to be in any other inertial reference frame.

Let us find out what this transformation rule is.

The components of the four-momentum of an object are frame-dependent quantities, because the values of the coordinate differences dt, dx, dy, dz that appear in the numerators of defining equations are frame-dependent.

The differential proper time dt appearing in the denominator, on the other hand, is a frame-independent quantity.

In this discussion, we will also consider the mass m of the object to be a frame-independent measure of the amount of "stuff" in the object.

You may have heard in another context that special relativity implies that the mass of an object depends on its velocity.

This is an old-fashioned way of looking at mass that obscures some of the simplicity and beauty of relativity theory.

Let us imagine that we know the components [P_t, P_x, P_y, P_z] of a given object's four-momentum in the Home Frame and that we want to find the corresponding components in an Other Frame moving with speed β in the +x direction with respect to the Home Frame.

The time component P'_t of the object's four-momentum in the Other Frame can be calculated as follows:

$$P'_t = m\,\frac{dt'}{d\tau} = m\,\frac{\gamma(dt - \beta\,dx)}{d\tau} = \gamma m\,\frac{dt}{d\tau} - \gamma\beta m\,\frac{dx}{d\tau}$$

$$\Rightarrow \gamma(P_t - \beta P_x)$$

where I have used the Lorentz transformation to express dt' as measured in the Other Frame in terms of dt and dx as measured in the Home Frame.

Similarly, the transformation equation for the four-momentum x component is (as you can show)

$$P'_x = \gamma(-\beta P_t + P_x)$$

We also have

$$P'_y = m\frac{dy'}{d\tau} = m\frac{dy}{d\tau} = P_y$$

$$P'_z = m\frac{dz'}{d\tau} = m\frac{dz}{d\tau} = P_z$$

If you compare these with the Lorentz transformation equations, you will see that these equations are exactly the same as the Lorentz transformation equations except that the four-momentum components P_t, P_x, P_y, P_z, P'_t, P'_x, P'_y, and P'_z have been substituted for the coordinate displacement components Δt, Δx, Δy, Δz, $\Delta t'$, $\Delta x'$, $\Delta y'$, and $\Delta z'$, respectively.

Thus the components of the four-momentum transform from frame to frame according to the Lorentz transformation equations, just as coordinate differences do!

The transformation equations for the four-momentum come out so nicely because (1) the time coordinate appears on an equal footing with the spatial components in the definition of the four-momentum, (2) we have divided the displacement by the frame-independent differential proper time $d\tau$ instead of the frame-dependent differential coordinate time dt, and (3) we are considering the mass m to be a frame-independent quantity.

In fact, the technical definition of a four-vector requires this kind of transformation law:

Definition of a Four-Vector

A **four-vector** is a physical quantity represented by a vector having four components that transform according to the Lorentz transformation equations (i.e., just as the coordinate differences Δt, Δx, Δy, Δz do) when we go from one inertial reference frame to another.

Now we know that although the coordinate differences $\Delta t, \Delta x, \Delta y, \Delta z$ between two events are frame-dependent quantities, the spacetime interval Δs between the events given by

$$\Delta s^2 = \Delta t^2 - \Delta x^2 - \Delta y^2 - \Delta z^2$$

is a frame-independent quantity.

Similarly, we can define a frame-independent number called the **four-magnitude** of a four-vector as follows: If a four-vector **A** has components $[A_t, A_x, A_y, A_z]$, then its squared four-magnitude is defined to be

$$|\mathbf{A}|^2 = A_t^2 - A_x^2 - A_y^2 - A_z^2$$

This is analogous to using the pythagorean theorem to find the magnitude of an ordinary vector.

You can easily show that the four-magnitude of an object's four-momentum vector is equal to its frame-independent mass:

$$m = |\mathbf{P}| = \sqrt{P_t^2 - P_x^2 - P_y^2 - P_z^2}$$

The easiest way to do this involves using the fact that for infinitesimally separated events, there is no distinction between proper time and spacetime interval: $d\tau = ds = \sqrt{dt^2 - dx^2 - dy^2 - dz^2}$

Alternatively, you can evaluate the square root in the above equation directly.

FOUR-MOMENTUM CAN BE CONSERVED IN ALL REFERENCE FRAMES

We now have a suitable candidate for a relativistic generalization of the concept of momentum.

The final touch is to verify that a law of conservation of four-momentum is in fact consistent with the principle of relativity.

It is important to check this (this is where the law of conservation of Newtonian momentum fails!).

Consider an arbitrary collision of two objects moving along the x axis.

The law of conservation of four-momentum says that

$$\mathbf{P}_1 + \mathbf{P}_2 = \mathbf{P}_3 + \mathbf{P}_4$$

where $\mathbf{P}_1$ and $\mathbf{P}_2$ are the objects' four-momenta *before* the collision and $\mathbf{P}_3$ and $\mathbf{P}_4$ are the objects' four-momenta *after* the collision.

This equation can be fruitfully rewritten in the form

$$\mathbf{P}_1 + \mathbf{P}_2 - \mathbf{P}_3 - \mathbf{P}_4 = 0$$

which essentially says that the difference between the system's initial and final total momenta is zero.

In component form (remember these are 4-vectors), this last equation tells us that

$$\begin{bmatrix} P_{1t} \\ P_{1x} \\ P_{1y} \\ P_{1z} \end{bmatrix} + \begin{bmatrix} P_{2t} \\ P_{2x} \\ P_{2y} \\ P_{2z} \end{bmatrix} - \begin{bmatrix} P_{3t} \\ P_{3x} \\ P_{3y} \\ P_{3z} \end{bmatrix} - \begin{bmatrix} P_{4t} \\ P_{4x} \\ P_{4y} \\ P_{4z} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} \quad \text{or} \quad \begin{aligned} P_{1t} + P_{2t} - P_{3t} - P_{4t} &= 0 \\ P_{1x} + P_{2x} - P_{3x} - P_{4x} &= 0 \\ P_{1y} + P_{2y} - P_{3y} - P_{4y} &= 0 \\ P_{1z} + P_{2z} - P_{3z} - P_{4z} &= 0 \end{aligned}$$

When expressed in component form, the single equation becomes a set of four equations, one for each of the four components of the four-momentum.

Each of these equations has to be independently satisfied if four-momentum is to be conserved.

Let us assume that we have observed a collision in the Home Frame and we have determined it satisfies the law of conservation of four-momentum in that frame.

The principle of relativity requires that the same law apply in every other inertial reference frame, i.e.,

If	$\mathbf{P}_1 + \mathbf{P}_2 - \mathbf{P}_3 - \mathbf{P}_4 = 0$	in the Home Frame
Then	$\mathbf{P}'_1 + \mathbf{P}'_2 - \mathbf{P}'_3 - \mathbf{P}'_4 = 0$	in any Other Frame

If this statement is not true, our proposed relativistic generalization of the idea of momentum is not any better than Newtonian momentum.

If the statement is true, then the law of conservation of four-momentum represents at least a possible relativistic generalization of the law of conservation of ordinary momentum.

We can in fact easily show that this is true for any collision as viewed in any Other inertial reference frame.

Consider the x component of the conservation law in the Other Frame.

According to the transformation law for the components of the four-momentum given earlier,

$$P'_{1x} + P'_{2x} - P'_{3x} - P'_{4x} = \gamma(-\beta P_{1t} + P_{1x}) + \gamma(-\beta P_{2t} + P_{2x}) \\ -\gamma(-\beta P_{3t} + P_{3x}) - \gamma(-\beta P_{3t} + P_{3x})$$

Collecting the terms on the right that are multiplied by γ and those multiplied by $\gamma\beta$, we get

$$P'_{1x} + P'_{2x} - P'_{3x} - P'_{4x} = -\gamma\beta(P_{1t} + P_{2t} - P_{3t} - P_{4t}) \\ +\gamma(P_{1x} + P_{2x} - P_{3x} - P_{4x})$$

But if both the t and x components of the four-momentum are conserved in the Home Frame, then the first two lines the result derived above

$$P_{1t} + P_{2t} - P_{3t} - P_{4t} = 0$$
$$P_{1x} + P_{2x} - P_{3x} - P_{4x} = 0$$

tell us that the quantities in parentheses equal zero.

So

$$P'_{1x} + P'_{2x} - P'_{3x} - P'_{4x} = -\gamma\beta(0) + \gamma(0) = 0$$

meaning that if both the t and x components of the system's total four-momentum are conserved in the Home Frame, then the x component of the system's total four- momentum will also be observed to be conserved in the Other Frame, as hoped.

The proof that the other components of the system's total four-momentum are also conserved in the Other Frame is essentially the same.

What we have shown is that the law of conservation of four-momentum that we stated is consistent with the principle of relativity in the sense that if it holds in one frame it holds in all.

That does not make the law true: it simply makes it possible.

But now let me argue for the truth of this law.

(1) We know from a multitude of experiments at low velocities that some quantity that reduces to newtonian momentum at such velocities is conserved.

(2) Conservation of Newtonian momentum will not work: the law is inconsistent with the principle of relativity.

(3) The hypothetical law of conservation of four-momentum is compatible with the principle of relativity.

(4) The three spatial components of the four-momentum do reduce to the components of Newtonian momentum at low velocities.

(5) Therefore, if the law of conservation of four- momentum were true, both would explain the low-velocity experimental data and maintain compatibility with the principle of relativity.

In short, our observations at low velocities suggest that something like momentum is conserved, and in the absence of compelling alternatives, it only makes sense to believe that it is the total four- momentum of a system that is conserved.

Of course, no matter how suggestive a theoretical argument might be, there is no substitute for direct experimental evidence.

Since the 1950s, physicists have been using particle accelerators to create beams of subatomic particles traveling at speeds very near the speed of light and colliding these particles with stationary targets or other particle beams.

At such speeds, the distinction between Newtonian momentum and four-momentum is very clear, and analysis of a typical experiment involves applying conservation of four-momentum to anywhere from thousands to millions of particle collisions.

The result is that conservation of four-momentum is implicitly tested thousands of times daily in the course of such research.

In spite of this enormous wealth of data, no compelling evidence of a violation of the law of conservation of four-momentum has ever been seen.

THE TIME COMPONENT OF FOUR-MOMENTUM

Our derived equations make it clear that conservation of four-momentum is only consistent with the principle of relativity if all four components of the four-momentum (P_t as well as P_x, P_y, and P_z are independently conserved.

The three spatial components of an object's four-momentum correspond (at low velocities) to the three components of its Newtonian momentum.

But the law of conservation of four-momentum requires that something else (that is, P_t) be conserved as well.

What is the physical interpretation of this new conserved quantity?

Our derivation showed that the time component of an object's four-momentum is:

$$P_t = \frac{m}{\sqrt{1 - v^2}}$$

We know that this reduces to the mass m of the object at low velocities but is not exactly equal to the mass.

Let us use the binomial approximation to find out how this quantity differs from m when $v \ll 1$.

Since $(1 - x)^a \simeq 1 - ax$ if $x < 1$, we have

$$P_t = \frac{m}{\sqrt{1 - v^2}} = m(1 - v^2)^{-1/2} \approx m[1 - (-\tfrac{1}{2})v^2] = m + \tfrac{1}{2}\, mv^2$$

when $v \ll 1$.

The first term here is the mass of the particle, and when v is zero, that is what P_t becomes.

But when v is nonzero but still very small, we have an additional term in P_t that corresponds to the kinetic energy of the particle.

If P_t is conserved in a collision at low velocities, what we are saying is that the sum of the masses plus the sum of the kinetic energies of the particles involved in the collision are conserved.

If the masses of the particles remain unchanged in the collision, then conservation of P_t tells us that the kinetic energy of the particles is conserved.

Thus the statement about the conservation of P_t is (at low velocities) basically the Newtonian statement of conservation of [kinetic] energy!

As we have generalized the concept of momentum, so now we will generalize the concept of energy.

We define the component P_t of a particle's four-momentum to be the particle's relativistic energy E and assert that it is this relativistic energy that is the fourth quantity conserved in an isolated system.

The relativistic energy E of an object moving at a speed v as measured in a given reference frame is thus defined to be

$$E \equiv P_t = \frac{m}{\sqrt{1 - v^2}}$$

Note that if the speed v of the object is an appreciable fraction of the speed of light, the approximation given by

$$P_t = m + \frac{1}{2} mv^2$$

does not hold.

The relativistic kinetic energy of an object is defined (for all v) to be the difference between the object's total relativistic energy E and its mass-energy m:

$$K \equiv E - m = \frac{m}{\sqrt{1-v^2}} - m = m\left(\frac{1}{\sqrt{1-v^2}} - 1\right)$$

The relativistic kinetic energy K becomes approximately equal to $\frac{1}{2}mv^2$ only for small values of: v : in general, K > $\frac{1}{2}mv^2$.

In Newtonian mechanics, conservation of energy and momentum were thought of as separate concepts.

But just as special relativity binds space and time into a single geometry, so here it binds the laws of conservation momentum and energy into a single statement: The total four-momentum of an isolated system of particles is conserved.

Conservation of energy is impossible without conservation of momentum, and vice versa: energy and momentum are indissolvably bound together as parts of the same whole.

Note, however, that the relativistic energy of an object is not just the kinetic energy of that object (even at low velocities) but includes the mass of the object as well.

The fact that $E = P_t$ is conserved does not imply that the mass of the object and its kinetic energy are separately conserved, only that the whole (i.e., the relativistic energy) is conserved.

This implies that processes that convert mass to kinetic energy, and vice versa, do not necessarily violate the law of conservation of four-momentum and therefore might exist.

This subject will be more fully explored later.

The equation

$$P_t = m + \frac{1}{2} mv^2$$

is expressed in SR units, where velocity is unitless and both kinetic energy and mass are measured in kilograms.

If we would like to express the relativistic energy of the particle in the SI unit of joules = kg m^2/s^2, we must multiply the energy in kilograms by two powers of the conversion factor $c = 2.998 \times 10^8$ m/s to get the units to come out right.

Therefore, the equation in SI units would read

$$E \approx mc^2 + \frac{1}{2}mv^2 \qquad \text{when } v^2 << c^2 \text{ (SI units)}$$

where the energy E is measured in joules and the speed v is measured in meters per second.

In particular, when the particle is at rest, its relativistic energy in SI units is

$$E_{\text{rest}} = mc^2$$

This is the famous equation that has served as an icon representing both the essence of special relativity and also Einstein's achievement.

It should be recognized that this equation is simply a special case of the more general equation.

Nonetheless, it does focus our attention on the startling new idea implicit in the definition of relativistic energy: An object at rest has relativistic energy simply by virtue of its mass, and this mass-energy is a part of the total energy that is conserved in an interaction within an isolated system.

THE SPATIAL PART OF FOUR-MOMENTUM

We have seen that the spatial components P_x, P_y, P_z of an object's four momentum vector become approximately equal to the components of the object's Newtonian momentum at low velocities.

Just as we defined an object's relativistic energy E to be the time component of its four-momentum, so we define an object's relativistic momentum p to be the magnitude of the spatial components of its four-momentum:

$$p \equiv \sqrt{P_x^2 + P_y^2 + P_z^2}$$

The relativistic momentum is thus the relativistic generalization of the magnitude of an object's Newtonian momentum vector.

We can express an object's relativistic momentum in terms of its mass m and the speed v as follows:

$$p = \sqrt{\left(\frac{mv_x}{\sqrt{1-v^2}}\right)^2 + \left(\frac{mv_y}{\sqrt{1-v^2}}\right)^2 + \left(\frac{mv_z}{\sqrt{1-v^2}}\right)^2} = \frac{m\sqrt{v_x^2 + v_y^2 + v_z^2}}{\sqrt{1-v^2}}$$

$$= \frac{mv}{\sqrt{1-v^2}}$$

Note that p becomes approximately equal to the magnitude mv of the object's Newtonian momentum when v is very much smaller than 1.

The equation derived earlier

$$m = |\mathbf{P}| = \sqrt{P_t^2 - P_x^2 - P_y^2 - P_z^2}$$

which expresses how an object's frame-independent mass m can be computed using the frame-dependent components of its four-momentum, can be written in terms of E and m as

$$m^2 = P_t^2 - (P_x^2 + P_y^2 + P_z^2) = E^2 - p^2$$

We can also use an object's relativistic energy E and relativistic momentum p in a given frame to determine the object's speed in that frame:

$$\frac{p}{E} = \frac{mv/\sqrt{1 - v^2}}{m/\sqrt{1 - v^2}} = v$$

This relationship applies to each individual spatial component as well:

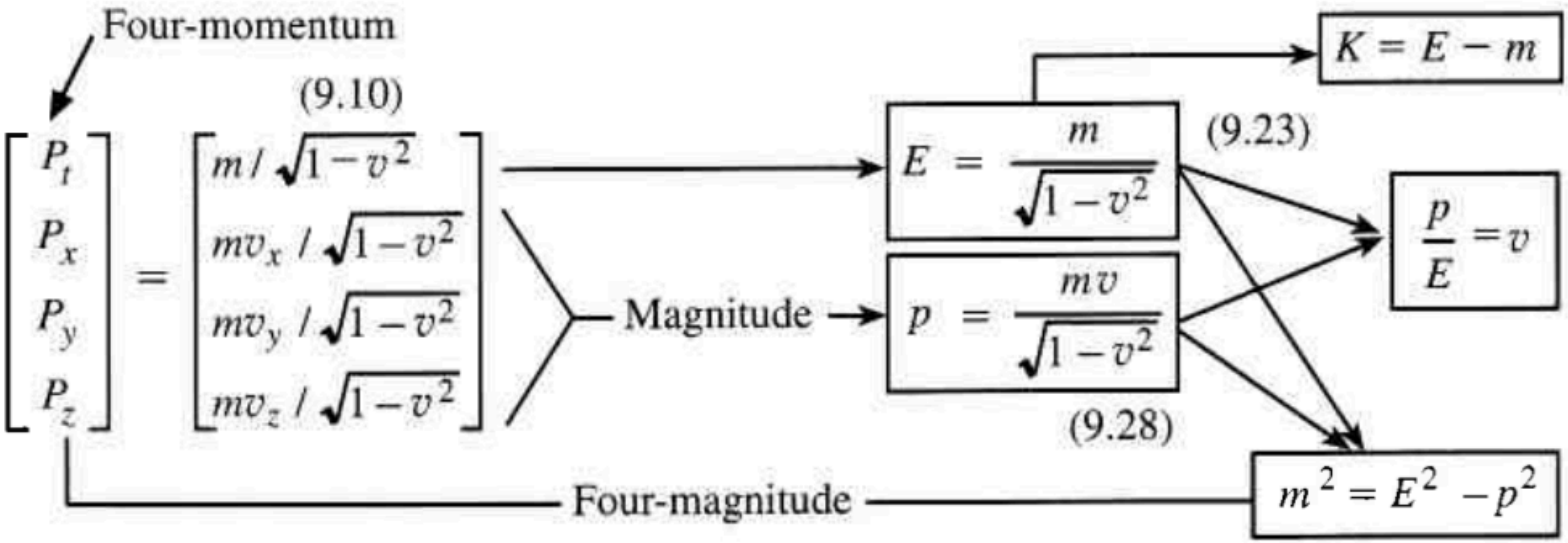
$$\frac{P_x}{E} = \frac{mv_x}{\sqrt{1-v^2}} \frac{\sqrt{1-v^2}}{m} = v_x \qquad \text{similarly} \qquad \frac{P_y}{E} = v_y \qquad \frac{P_z}{E} = v_z$$

Indeed, the spatial components of the four-momentum can be thought of as expressing the rate at which relativistic energy is transported through space:

$$\begin{bmatrix} P_t \\ P_x \\ P_y \\ P_z \end{bmatrix} = \begin{bmatrix} m/\sqrt{1-v^2} \\ mv_x/\sqrt{1-v^2} \\ mv_y/\sqrt{1-v^2} \\ mv_z/\sqrt{1-v^2} \end{bmatrix} = \begin{bmatrix} E \\ Ev_x \\ Ev_y \\ Ev_z \end{bmatrix}$$

Our derived equations express relationships between the quantities E, p, m, K, and v that are really helpful to know when working with the four-momentum.

These equations and the ways they connect to the four- momentum vector are summarized in the Figure.



(Virtually) everything you need to know about four-momentum.

CONSERVATION OF FOUR-MOMENTUM

Up to now, we saw that conservation of the total ordinary (nNewtonian) momentum of an isolated system of objects is inconsistent with the principle of relativity, because of the complicated nature of the Einstein velocity transformation.

On the other hand, we saw that conservation of the total four-momentum of an isolated system is consistent with the principle of relativity.

Moreover, the four-momentum of an object reduces to the Newtonian momentum in the low-velocity limit.

Therefore, if anything like "momentum" is to be conserved, it must be in fact four-momentum that is conserved.

Let us assume that this is so.

The purpose of the next discussion is to explore some of the surprising consequences and experimental tests of this assumption.

First of all, we will discuss how to draw energy-momentum diagrams of collision processes and what such diagrams can tell us.

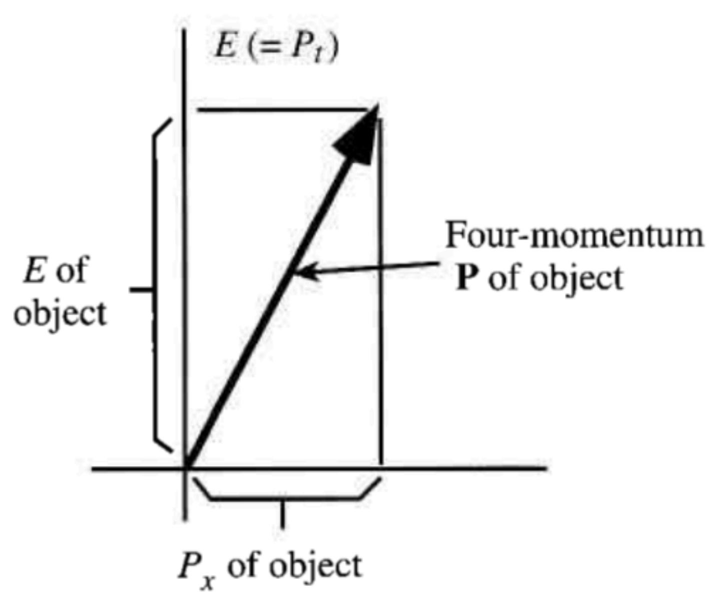
Then, using such diagrams, we will explore in some depth the assertion (hinted at in the last chapter) that mass is just another form of (relativistic) energy, and processes do exist that can convert mass to energy, and vice versa.

We will also learn how light itself can be represented by a four-momentum vector.

Finally, we will apply the law of conservation of four-momentum to a variety of examples, ranging from practical experimental tests to speculations about relativistic space travel.

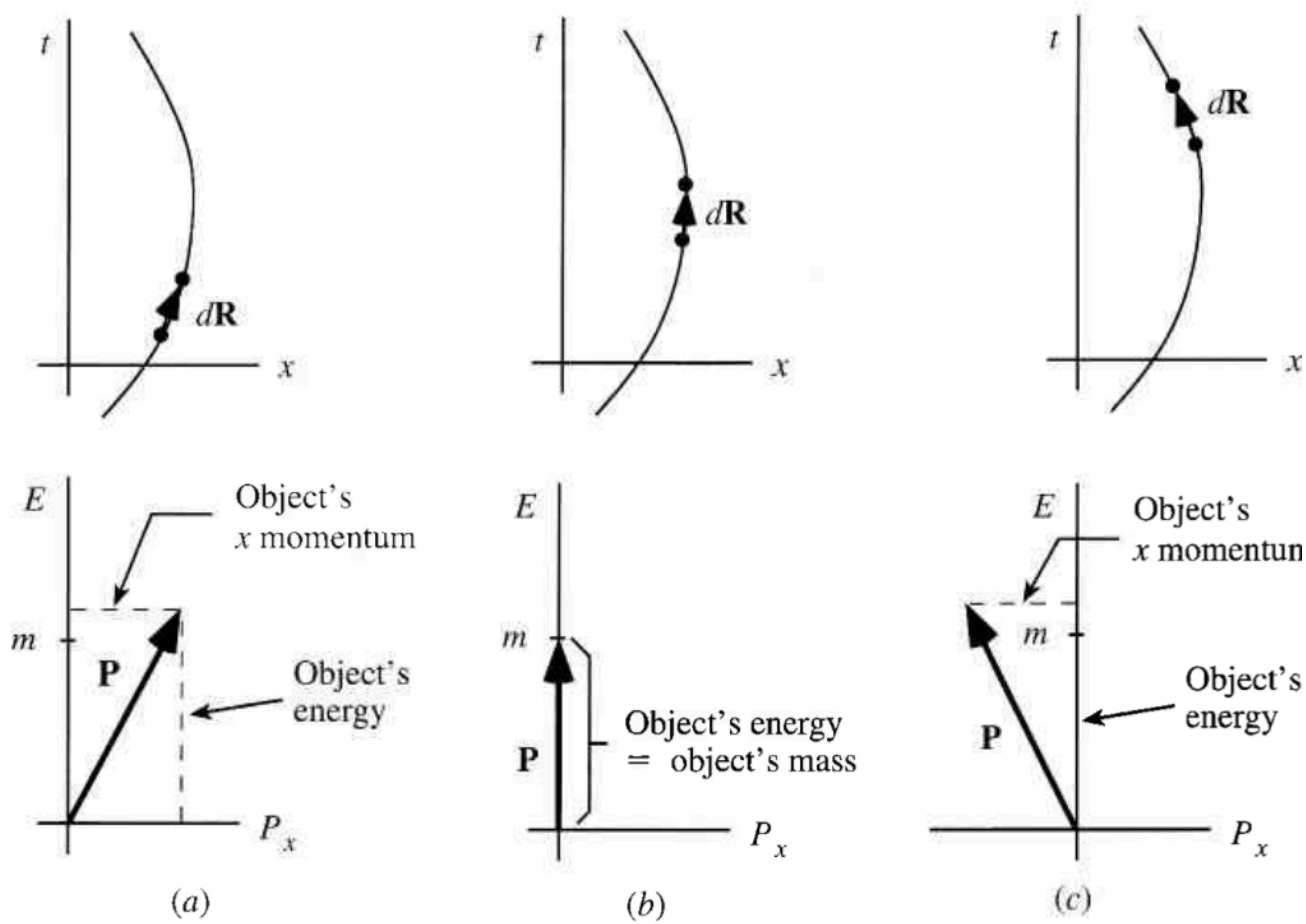
ENERGY-MOMENTUM DIAGRAMS

The four-momentum of an object moving in the spatial x direction can be visually represented as an arrow on a special kind of spacetime diagram called an energy-momentum diagram (see Figure).



Energy-momentum diagram showing the four- momentum of a certain object moving in the +x direction. The object's four-momentum is represented on the diagram by an arrow. The projections of this arrow on the vertical and horizontal axes represent the values of the object's relativistic energy E (= P_t) and its relativistic x momentum P_x. Note that the object's relativistic momentum p= $|P_x|$.

Just as the direction of the arrow representing an object's ordinary momentum is tangent to its path through space, the direction of the arrow representing an object's four-momentum is tangent to its worldline in space-time (because the object's four-momentum vector **P** = m d**R**/dt at any given point along its worldline is proportional to the object's differential displacement in space-time d**R** along that worldline around that point) (see Figure).



(a) At any given time, the arrow representing an object's four-momentum on an energy-momentum diagram points in a direction tangent to the object's worldline, because **P** is proportional to d**R**. (b) When an object is at rest (even if just at an instant), its four-momentum is vertical and its energy is equal to its mass. (c) When the object moves in the x- direction, its x momentum is negative but its energy remains positive (and indeed greater than its mass).

Since the inverse slope of an object's worldline at any instant is equal to its x velocity at that instant, the inverse slope of the object's four-momentum arrow at a given time (i.e., run/rise = P_x/E) should also be equal to its x velocity at that time if the two vectors are to be parallel.

Our equations from earlier say essentially the same thing:

$$\frac{P_x}{E} = \frac{mv_x}{\sqrt{1-v^2}} \frac{\sqrt{1-v^2}}{} = v_x$$

The four-magnitude of an object's four-momentum is its mass m: this value is frame-independent and independent of the object's motion in a given frame.

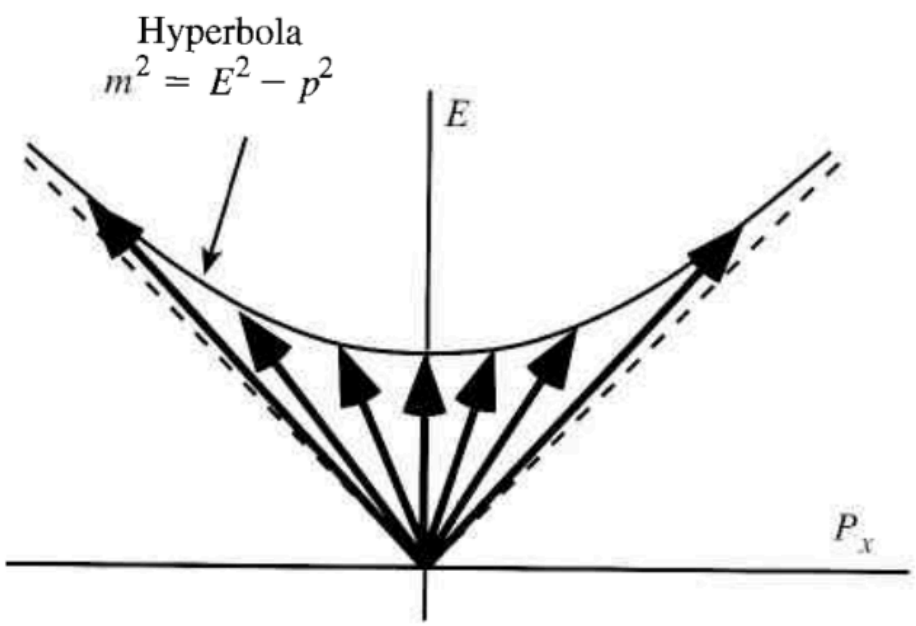
But the length of the arrow representing the object's four-momentum on an energy-momentum diagram depends on the object's velocity: the length of the arrow on the diagram is not proportional to the four-magnitude of the corresponding four- momentum.

In fact, we have (assuming that the object is moving in the x direction so that $P_y = P_z = 0$ and $p= |P_x|$

$$m^2 = E^2 - p^2 = E^2 - (P_x)^2$$

This means that the tips of the four-momentum arrows for objects of identical mass m traveling at different x velocities (or the four-momentum arrows for a single accelerating object observed at different times) lie along a curve on the diagram defined by the equation $m^2 = E^2 - p^2$.

This curve is in fact a hyperbola, as shown in the Figure.



Energy-momentum diagram showing the four-momentum arrows for a set of identical objects of mass m moving at different velocities in the Home Frame. The tips of all these arrows lie on the hyperbola defined by the equation $m^2 - E^2 = p^2$. Note that as the object's x velocity approaches 1 (and thus p/E approaches 1), both p and E have become very large if the difference of their squares is to remain fixed.

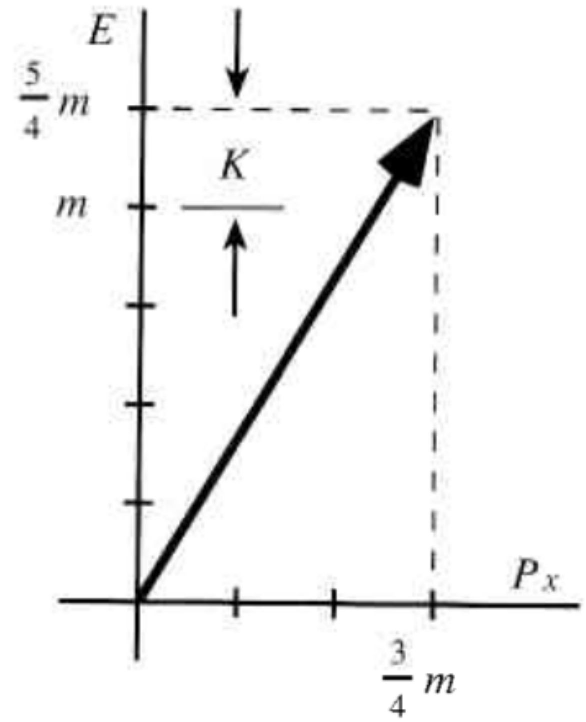
If you know an object's x velocity and its mass m, it is easy to draw an energy-momentum diagram showing its four-momentum vector.

The steps are as follows:

1. Set up your E and P_x axes.
2. Draw a line from the origin of those axes having the slope $1/v_x$.
3. Compute the value of $E = m/\sqrt{1 - v_x^2}$ for the object.
4. Draw a horizontal line from this value on the E axis until it intercepts the line that you drew in step 2.
5. The arrow representing the object's four-momentum lies along the line drawn in step 2, with its tip at the intersection found in step 4.

We can easily read an object's relativistic kinetic energy K = E - m directly from an energy-momentum diagram.

For example, K for an object of mass m moving at a speed v=3/5 is m/4, as shown in the Figure.



An energy-momentum diagram of an object of mass m traveling at an x velocity v_x =3/5. The four-momentum arrow for such an object has a slope of 5/3 and an energy of 5m/4, since

$$\sqrt{1-(3/5)^2} = \sqrt{1 = 9/25} = \sqrt{16/25} \quad = 4/5 \ ,$$

implying that E = $m\sqrt{1-v_x^2}$ = 5m/4. For the arrow to have the correct slope, we must have P_x= 3m/4. Note that the object's relativistic kinetic energy K = E- m = m/4 can be read directly from the diagram.

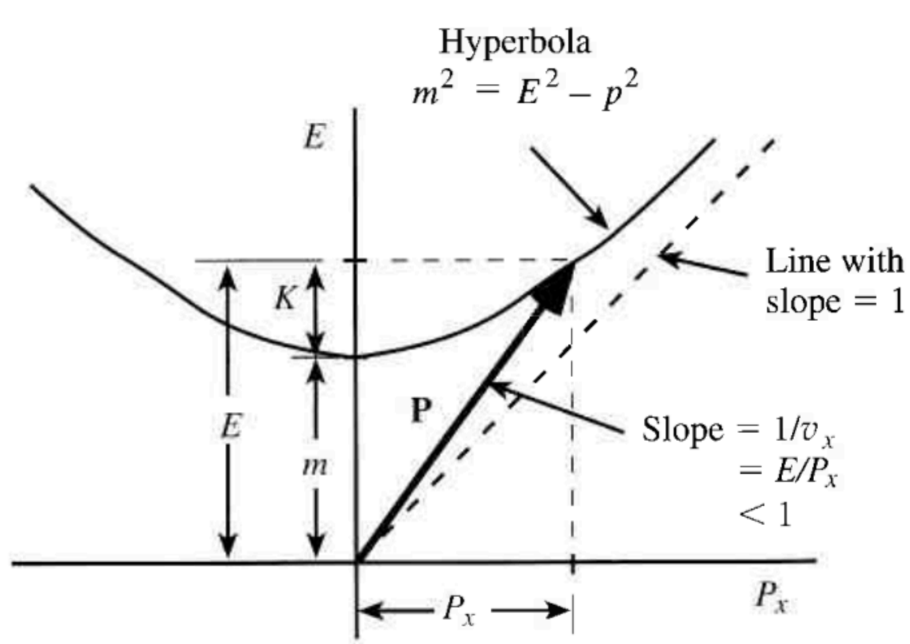
Note that $m/4 \neq \frac{1}{2}m(3/5)^2$ but in fact is substantially larger!

The last two figures together make it clear that as an object's speed v approaches 1 (the speed of light), both the object's total relativistic energy E and its kinetic energy K go to infinity.

This means that you would have to supply an infinite amount of energy to accelerate an object of nonzero mass to the speed of light.

This is the most practical reason that no object can go faster than the speed of light: al the energy in the universe could not accelerate even a mote of dust to that speed!

Virtually all that you need to know to construct and interpret an energy-momentum diagram is summarized in the Figure.



Virtually everything that you need to know about four-momentum diagrams. No matter what the x velocity of an object of mass m might be, the tip of its four-momentum arrow lies on the hyperbola $m^2 = E^2 - p^2$ for whatever that object's mass m might be. The inverse slope of the four-momentum arrow is equal to v, which always has a magnitude less than 1.

The law of conservation of four-momentum (like the law of conservation of ordinary momentum) is most useful when applied ot an isolated system of objects undergoing some type of collision process (i.e., some kind of sudden interaction between the objects in the system that may be strong and complicated but limited in time).

In such a case, the system has a clearly defined state "before" and "after" the collision, making it easy to compute the total four-momentum in the system both before and after the collision.

The law of conservation of four-momentum states that the system should have the same total four-momentum after the collision process as it had before.

What does this really mean mathematically?

Since four-momentum is a (four-dimensional) vector quantity, conservation of four-momentum means that each component of the system's total four-momentum is separately conserved.

For example, consider a system consisting of two objects, and let the objects' four-momenta before the collision be $\mathbf{P}_1$ and $\mathbf{P}_2$, and after the collision be $\mathbf{P}_3$ and $\mathbf{P}_4$.

Conservation of four- momentum requires that

$$E_1 + E_2 = E_3 + E_4$$

$$P_{1x} + P_{2x} = P_{3x} + P_{4x}$$

$$P_{1y} + P_{2y} = P_{3y} + P_{4y}$$

$$P_{1z} + P_{2z} = P_{3z} + P_{4z}$$

remembering that the time component of a four-momentum vector (i.e., the relativistic energy) is usually given the more evocative symbol E instead of P_t.

Each one of above equations has to be separately true for four-momentum to be conserved.

Now, we will focus primarily on objects moving in only one dimension, which we can take to be the x direction.

This simplifies the mathematics significantly without any substantial loss of understanding, allowing us to ignore the y and z components of the four-momenta (which are always zero) and focus on the t and x components.

Let us consider a specific example.

Imagine that somewhere in deep space a certain rock with mass $m_1 = 12$ kg is moving in the +x direction with $v_{1x} = +4/5$ in some inertial reference frame.

This rock then strikes another rock of mass $m_2 = 28$ kg at rest ($v_{2x} = 0$).

Instead of instantly vaporizing into a cloud of gas (as any real rocks colliding at this speed would), let us pretend that the first rock simply bounces of the more massive rock and is subsequently observed to have an x velocity $v_{3x} = -5/13$.

What is the x velocity v_{4x} of the larger rock after the collision?

The first step in solving this problem is to calculate the energy E_1 and the x momentum P_{1x} of the smaller rock before the collision.

Using the definitions of these four-momentum components, we find that

$$E_1 \equiv \frac{m_1}{\sqrt{1 - v_{1x}^2}} = \frac{m_1}{\sqrt{1 - (4/5)^2}} = \frac{m_1}{\sqrt{9/25}} = \frac{m_1}{3/5} = \frac{5(12\text{ kg})}{3} = 20\text{ kg}$$

$$P_{1x} \equiv \frac{m_1 v_{1x}}{\sqrt{1 - v_{1x}^2}} = \frac{m_1(+4/5)}{3/5} = \frac{4(12\text{ kg})}{3} = +16\text{ kg}$$

Similarly, the energy and x momentum of the larger rock before the collision are

$$E_2 \equiv \frac{m_2}{\sqrt{1-v_{2x}^2}} = \frac{m_2}{\sqrt{1-(0)^2}} = m_2 = 28 \text{ kg}$$

$$P_{2x} \equiv \frac{m_2 v_{2x}}{\sqrt{1-v_{2x}^2}} = \frac{m_1(0)}{\sqrt{1-0^2}} = 0 \text{ kg}$$

The energy and x momentum of the smaller rock after the collision are

$$E_3 \equiv \frac{m_1}{\sqrt{1-v_{3x}^2}} = \frac{m_1}{\sqrt{1-(-5/13)^2}} = \frac{m_1}{\sqrt{144/169}} = \frac{m_1}{12/13}$$

$$= \frac{13(12 \text{ kg})}{12} = 13 \text{ kg}$$

$$P_{3x} \equiv \frac{m_1 v_{3x}}{\sqrt{1-v_{3x}^2}} = \frac{m_1(-5/13)}{12/13} = \frac{-5(12 \text{ kg})}{12} = -5 \text{ kg}$$

Conservation of four-momentum requires that the four-momentum vectors before the collision add up to the same value after the collision:

$$t \text{ component:} \quad x \text{ component:} \qquad \begin{bmatrix} E_1 \\ P_{1x} \end{bmatrix} + \begin{bmatrix} E_2 \\ P_{2x} \end{bmatrix} = \begin{bmatrix} E_3 \\ P_{3x} \end{bmatrix} + \begin{bmatrix} E_4 \\ P_{4x} \end{bmatrix}$$

$$\begin{bmatrix} E_4 \\ P_{4x} \end{bmatrix} = \begin{bmatrix} E_1 \\ P_{1x} \end{bmatrix} + \begin{bmatrix} E_2 \\ P_{2x} \end{bmatrix} - \begin{bmatrix} E_3 \\ P_{3x} \end{bmatrix} = \begin{bmatrix} 20\text{ kg} \\ 16\text{ kg} \end{bmatrix} + \begin{bmatrix} 28\text{ kg} \\ 0 \end{bmatrix} - \begin{bmatrix} 13\text{ kg} \\ -5\text{ kg} \end{bmatrix} = \begin{bmatrix} 35\text{ kg} \\ +21\text{ kg} \end{bmatrix}$$

Knowing the energy and momentum of an object is sufficient information to determine both its mass and velocity.

Using $$m^2 = P_t^2 - (P_x^2 + P_y^2 + P_z^2) = E^2 - p^2$$

we see that the mass of the larger rock is

$$m = \sqrt{E_4^2 - P_{4x}^2} = \sqrt{(35\text{ kg})^2 - (21\text{ kg})^2} = (7\text{ kg})\sqrt{5^2 - 3^2} = (7\text{ kg})4 = 28\text{ kg}$$

after the collision (just as it was before).

According to

$$\frac{P_x}{E} = \frac{mv_x}{\sqrt{1-v^2}} \frac{\sqrt{1-v^2}}{m} = v_x$$

its final x velocity is

$$v_{4x} = \frac{P_{4x}}{E_4} = \frac{+21 \text{ kg}}{35 \text{ kg}} = \frac{3}{5}$$

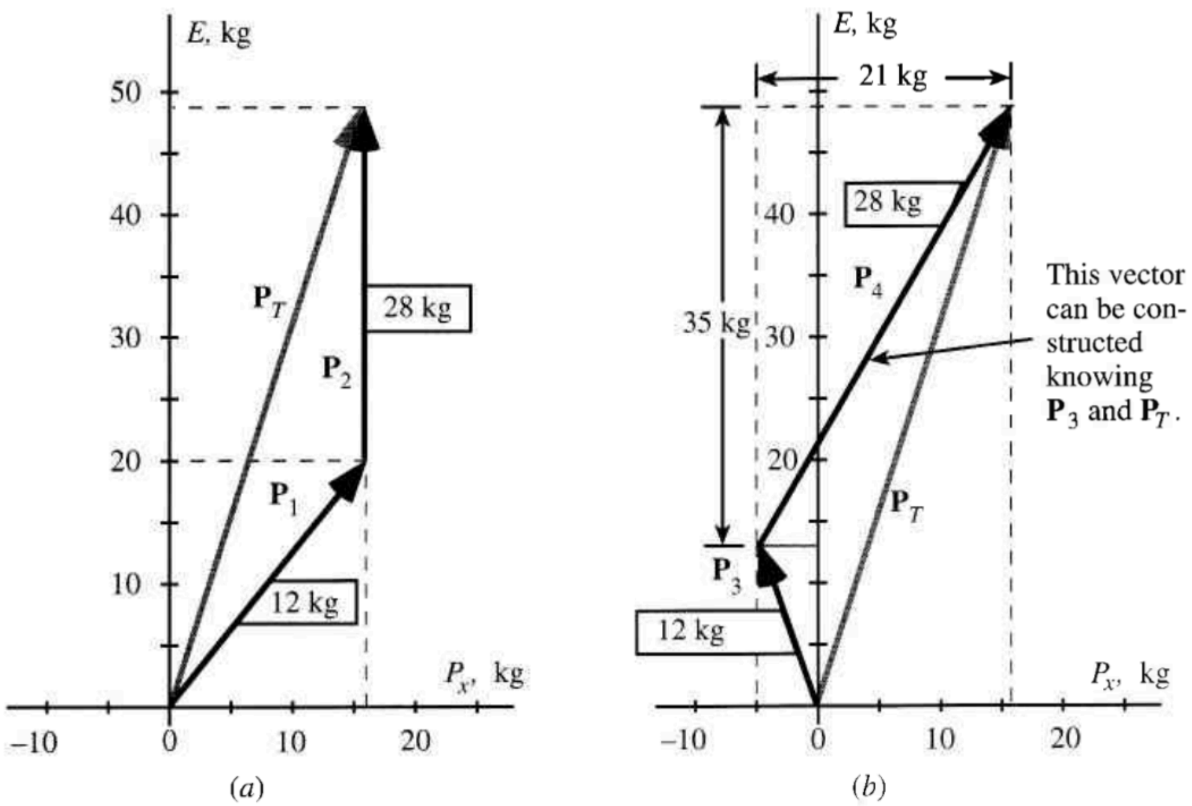
You can easily show in this case that newtonian momentum is not conserved by the collision just described.

SOLVING CONSERVATION PROBLEMS GRAPHICALLY

It is possible to solve a conservation of four-momentum problem graphically using an energy-momentum diagram.

The sum of four- momenta is defined like the sum of ordinary vectors (you simply add the components), so you can add four-momentum arrows on an energy-momentum diagram just as you would ordinary vector arrows (by putting the tail of one vector on the tip of the other while preserving their directions).

Using this technique, we see in the Figure Part (a) that in the rock example, the system's total four-momentum before the collision has components E_T = 48 kg, $P_{T,x}$ =16 kg.



a) The four-momenta of the rocks before the collision. The vector sum of these four-momenta is represented by the arrow $\mathbf{P}_T$. Since the magnitudes of the individual four-momenta arrows (which equal the masses of the corresponding rocks) cannot be read directly from the diagram, I have adopted the expedient of attaching a "flag" to each four-momentum arrow that states its magnitude. (b) The four-momenta of the rocks after the collision. The vector sum of these rocks' four- momenta is still $\mathbf{P}_T$, by four-momentum conservation. Since $\mathbf{P}_3$ is also known, it is possible to construct the unknown four-momentum $\mathbf{P}_4$, read its components from the diagram as shown, and compute its corresponding mass and x velocity.

The two rocks' four-momentum arrows after the collision have to add up to the same total four-momentum arrow, and since we know the smaller rock's four-momentum after the collision, we can construct the larger rock's final four- momentum arrow (Figure Part (b)).

We can then read the components of this arrow right off the diagram, getting the same results as in the earlier numerical calculation.

This kind of graphical approach to the problem is not usually much easier than the algebraic approach, but it does have some advantages:

(1) It provides a more visual and concrete way of dealing with the problem and may be helpful to you if you find the algebraic approach rather abstract.

(2) When used in conjunction with the algebraic method, it serves as a useful check on the algebraic results: it is more difficult to make an error using the graphical method.

(3) In some cases, as we will see, simply looking at the diagram can yield qualitative information that is very difficult to get from the algebraic equations alone.

In short, the graphical method represents an alternative method for solving problems involving conservation of four-momentum that often complements the algebraic approach.

Armed with both these techniques, we are now ready to explore some of the strange and interesting consequences of the law of conservation of four-momentum.

THE MASS OF A SYSTEM OF PARTICLES

As we have seen, the relativistic energy of an object is not simply equal to its kinetic energy (even at low velocities) but involves the mass of the object as well.

The fact that $E = P_t$ is conserved by the internal interactions of an isolated system does not imply that the mass of an object and its kinetic energy are separately conserved, only that the sum of these two things are conserved.

This implies that processes that convert mass to kinetic energy, and vice versa, do not necessarily violate the law of conservation of four-momentum and therefore may exist.

Mass and kinetic energy are seen in the theory of special relativity to be simply two parts of the same whole (the relativistic energy).

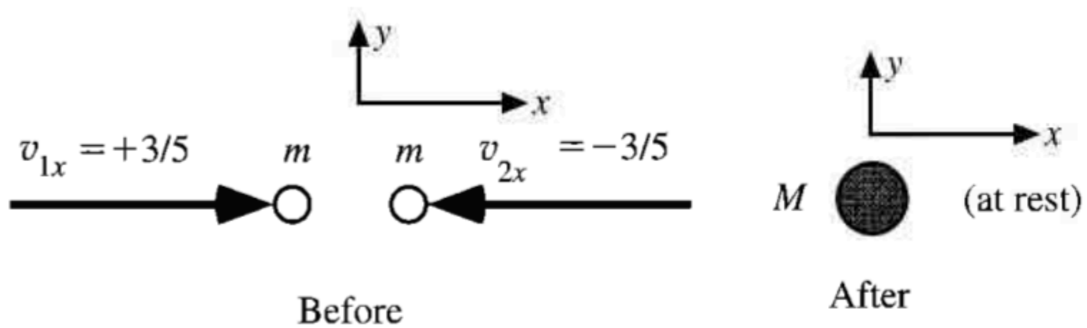
There is no reason to presuppose a barrier between these two manifestations of relativistic energy that would preclude the conversion of one into the other.

We will be considering a variety of examples of processes that do just that.

We will begin with a simple example that illustrates a crucial thing we need to understand about “mass” before we can go further: *the mass of a system is generally different from the sum of the masses of its parts*.

Consider the collision of two identical balls of putty with mass m = 4 kg which in some inertial frame are observed to have x velocities of $v_{1x} = +3/5$ and $v_{2x} = -3/5$; that is, these putty balls are approaching each other with equal speeds.

Imagine that when these putty balls collide, they stick together, as shown in the Figure.



The inelastic collision of two putty balls as seen in the frame where they initially have equal speeds but opposite directions. m = 4.0 kg.

Note that before the collision, the x component of the system's total four-momentum is zero:

$$P_{1x} + P_{2x} = \frac{m(+3/5)}{\sqrt{1-(3/5)^2}} + \frac{m(-3/5)}{\sqrt{1-(-3/5)^2}} = \frac{m(3/5 - 3/5)}{\sqrt{1-9/25}} = 0$$

so conservation of four-momentum implies that the x component of the final mass' four-momentum is zero as well, meaning it must be at rest.

What of relativistic energy conservation in this case?

A Newtonian analysis of this collision would speak of the kinetic energy being converted into thermal energy in this inelastic collision.

Such an analysis would also assert that the mass of the coalesced particle is $M = m + m = 2m$.

But we have more constraints to consider in a relativistic solution to this problem.

If the spatial components of the four-momentum are conserved in this collision, the time component must also be conserved, whether the collision is elastic or not.

But how can we think of the relativistic energy being conserved in this case, since no mention has been made of thermal energy in the definition of the relativistic energy given earlier?

The answer is direct and surprising.

Since the final object is motionless, its relativistic energy is simply equal to its mass M.

But by conservation of four-momentum, we have

$$M = E_1 + E_2 = \frac{m}{\sqrt{1-(3/5)^2}} + \frac{m}{\sqrt{1-(-3/5)^2}} = \frac{2m}{\sqrt{16/25}} = \frac{10m}{4} = 10\ \text{kg}$$

which is not equal to 2m = 8 kg!

Conservation of four-momentum thus requires that the final object have a greater mass than the sum of the masses that collided to form it!

We know from experience with collisions at low speeds that when two objects collide and stick together, their energy of motion gets converted to thermal energy: the final object is a little warmer than the original objects.

In this case, actually, the final object will be a lot warmer than the original objects, so much so that any real putty balls colliding at such speeds would vaporize instantly.

What the above equation is telling us is that the final object has to be more massive than the original objects and that the increased thermal energy is somehow correlated with this.

But where does this extra mass actually reside?

The final object has the same number of atoms as the original objects did.

Does each atom gain some extra mass somehow?

This seems absurd.

The increased thermal energy in the final object means that its atoms will jostle around more vigorously.

Can the motion of these atoms "have mass" in some sense?

This seems crazy: individual particles have the same mass no matter how they move.

So where is this extra mass?

There is only one fully self-consistent way to answer this question: *the extra mass is a property of the system as a whole* and does not reside in any of its parts.

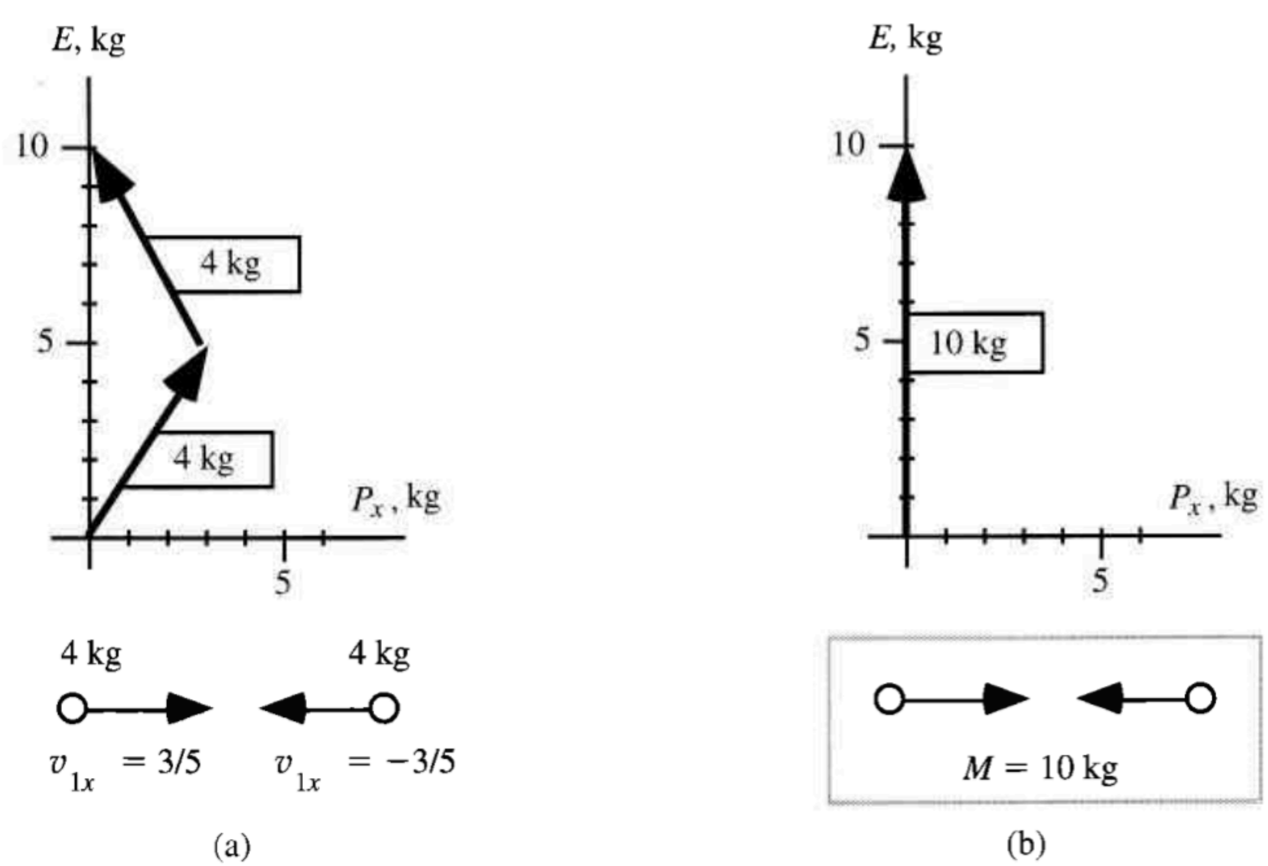
This can be vividly illustrated as follows.

Consider the "system" consisting of the two balls of putty before they collide.

If they are considered separate objects, the putty balls each have a mass m of 4 kg and a relativistic energy of 5 kg and one has an x momentum of -3 kg and the other +3 kg.

On the other hand, fi we consider the balls to constitute a system, the system has a total x momentum of zero and a total energy of 10 kg, meaning that its mass M = is equal to 10 kg.

This is illustrated by the figure.



(a) Putty balls before the collision, considered as two individual objects. Each object has its own mass, energy, and x momentum. (b) Putty balls before the collision, considered a single system. The system's x momentum is zero, meaning that its total energy of 10 kg is also equal to its mass.

So we see that the thermal energy produced by the collision is not the source of this extra mass: the extra mass was present in the "system" before the collision and remains the same after the collision.

So in some sense, mass is not "created" by the collision process at all: the collision simply manifests the mass of the system of two initial objects in the mass of a single final object.

If we focus on the masses of the individual objects ni the system before and after the collision, we think of mass being created.

But if we focus on the system before and after the collision, we see that its mass remains the same.

It is possible to get unnecessarily hung up on the difference between the mass of a system and the masses of its parts.

The reason this seems screwy is that we are used to treating mass as if it were additive: the mass of a jar of beans is the sum of the masses of the individual beans plus the mass of the container, right?

This is true enough at low velocities.

But if we had common experience with a jar full of beans that bang around inside the jar with speeds close to that of light, then we would be used to the idea that such a jar would have a different mass than the mass of the individual beans.

Mass is simply not additive in the way that energy and x momentum are.

There are actually many examples of things in the world where the whole is greater than the sum of its parts.

The meaning of a poem is not the same as the sum of the meaning of the individual letters in the words.

The "life" of an organism cannot be localized in any of its parts.

We simply need to start thinking about mass in the same way as we think about these things.

The best way to look at this is to think of the mass of a system of particles as a property of the system as a whole (i.e., the magnitude of the system's total four-momentum vector) and something that simply does not have very much to do with the masses of its parts.

The only self-consistent way to define the mass of a system is as the magnitude of the system's total four-momentum, and if this definition leads to the mass of a system being greater or less than the masses of its parts, well, that's the way it is!

THE FOUR-MOMENTUM OF LIGHT

We all have experienced the fact that light carries energy: we have felt sunlight warm our skin or seen an electric motor powered by solar cells or learned that plants convert the energy in sunlight into chemical energy.

Since we have seen that energy is the time component of four-momentum, it follows that light should have an associated four-momentum vector.

What does the four-momentum of light look like?

Previously, we have explored the four-momenta of objects (rocks or putty balls or the like) that could be considered to be particles that have a well-defined position in space and thus a well-defined worldline through spacetime.

The analogous thing in the case of light would be a "flash" or "burst" of light energy that is similarly localized in space.

We can consider a continuous beam of light to be composed of a sequence of closely spaced flashes, much as we might imagine a stream of water to be a sequence of closely spaced drops.

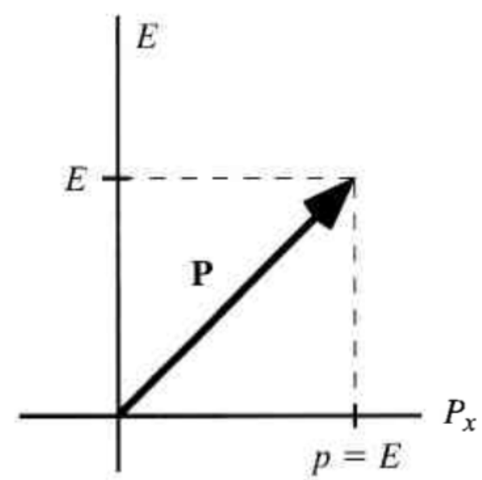
So what does the four-momentum of a flash of light look like?

Arguably, the most basic feature of any object's four-momentum is that it is parallel to that object's worldline.

If this is true for our flash of light, then it follows that the four-momentum vector for a flash of light must have a slope of ±1 on an energy-momentum diagram.

The four-momentum vector of a flash with a given energy E moving in the +x direction will thus look something like the vector shown in the Figure.

An energy-momentum diagram showing the four-momentum of a flash of light moving in the +x direction. If the four- momentum is to be parallel to the flash's worldline, then it must be drawn with slope +1 on the energy-momentum diagram. This implies that the flash's relativistic momentum p must have the same value (in SR units) as its energy: p = E



You can see from this diagram that if the flash's four-momentum vector is to have such a slope, it must have a spatial relativistic momentum p equal to its relativistic energy E.

This is in fact consistent with

$$\frac{p}{E} = \frac{mv/\sqrt{1-v^2}}{m/\sqrt{1-v^2}} = v$$

which in the case of light tells us that

$$\frac{p}{E} = v = 1 \quad \Rightarrow \quad p = E \qquad \text{(for a light flash)}$$

One immediate implication of this important formula is that light must carry momentum (as well as energy).

Light bouncing of a mirror will thus transfer momentum to the mirror (causing it to recoil) in much the same way that a ball bouncing of an object transfers momentum to the object and causes it to recoil.

This has been experimentally verified and it is now known that the pressure exerted by light due to its momentum plays an important part in the evolution of stars, the evolution of the early universe, and many other astrophysical processes.

Another immediate consequence is that a flash of light has zero mass.

We have defined the mass of an object in special relativity to be the invariant magnitude of its four-momentum.

According the above result, the mass of a flash of light is

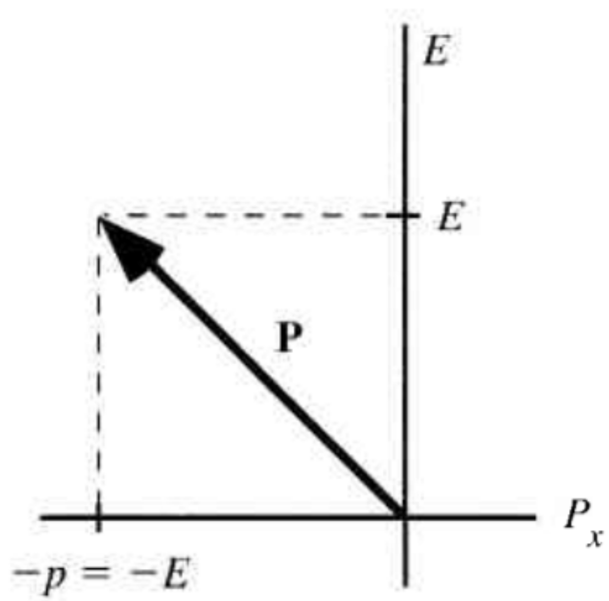
$$m^2 = E^2 - p^2 = 0$$

This is actually a good thing.

If a flash of light were to have some nonzero mass m, then its energy E = $m/\sqrt{1-v^2}$ and relativistic momentum p = $mv/\sqrt{1-v^2}$ would both have to be infinite, since v = 1 for light, which makes the denominator zero in each expression.

But since m = 0 as well, these equations instead read E= 0/0 and p = 0/0: the ratio 0/0 is "undefined" instead of being infinite, meaning that the equations E = $m/\sqrt{1-v^2}$ and p = $mv/\sqrt{1-v^2}$ simply do not tell us anything useful about the four-momentum of light.

If a light flash is moving in the -x direction instead of the +x direction, the slope of its four-momentum arrow on an energy-momentum diagram is -1 instead of +1 and its x momentum is negative: P_x = -p =-E (note that p and E are positive by definition), as shown in the Figure.



An energy-momentum diagram showing the four-momentum of a flash of light moving in the -x direction. In this case, its x momentum is negative: P_x= -p, where p is equal to the magnitude of the flash's spatial momentum (p = $|P_x|$ in this case). Note that we still have p = E: this is a general relation for light, independent of the direction the flash is traveling.

APPLICATIONS TO PARTICLE PHYSICS

High-energy particle physics is one area of physics where special relativity is not only useful but absolutely essential.

Because subatomic particles have such small masses, it is possible with current technology to accelerate them to speeds almost indistinguishably less than that of light, and this is done daily at particle accelerator facilities around the world.

Even natural processes such as the decay of radioactive nuclei can produce particles moving at relativistic speeds.

In the realm of high-energy particle physics, standard newtonian ideas are inadequate, but the ideas of special relativity, and most particularly the law of conservation of four-momentum, have been shown to be both descriptive and accurate (indeed, the frequency and precision with which the implications of special relativity are tested at the world's particle accelerators makes it one of the best-tested theories in all of physics).

This is the arena, therefore, where special relativity is applied on a daily basis to practical, real-world problems.

The purpose of the following discussion is to provide a short introduction to subatomic particle physics and to present some examples of how the law of conservation of four-momentum is applied in this context.

Part of the point is to make it clear that special relativity is not just a beautiful theoretical speculation but has important practical and testable implications and represents one of the cornerstones on which the edifice of modern physics is constructed.

UNITS FOR PARTICLE PHYSICS

Let us first review a few things about the units we use to describe the four-momentum of an object.

In the SR unit system, mass, momentum, and energy all have units of kilograms.

Thus the fundamental unit of energy in the SR system is the kilogram of energy.

To convert from this unit to joules 1(J= 1kg-m²/s²), the conventional SI unit of energy, one must multiply by two powers of the conversion factor c:

$$1\text{ kg } (\textit{of energy}) = 1\text{ kg}\left(\frac{2.998\times10^{8}\text{ m}}{1\text{ s}}\right)^{2} = 8.988\times10^{16}\text{ kg}\cdot\text{m}^{2}/\text{s}^{2}$$
$$= 8.988\times10^{16}\text{ J}$$

Thus the kilogram of energy is a lot of energy, roughly equivalent to the energy output of a very large (1000-MW) electrical power plant over a time period of 2.8 years.

This implies that a tiny amount of mass (i.e., rest energy) corresponds to a huge amount of energy in other forms.

We will see that subatomic and nuclear reactions can convert a significant fraction of the reactants' rest energy into other forms of energy, thus producing prodigious amounts of such energy from a small amount of converted mass.

The fundamental unit of momentum in the SR system is the kilogram of momentum.

To convert to the conventional SI units for momentum (kg-m/s), one must multiply by one power of the conversion factor c:

$$1 \text{ kg } (\textit{of momentum}) = 1 \text{ kg}\left(\frac{2.998 \times 10^8 \text{ m}}{1 \text{ s}}\right) = 2.998 \times 10^8 \text{ kg}\cdot\text{m/s}$$

Again, the kilogram of momentum is a substantial amount of momentum, about equal to the momentum of a 30-ton truck traveling at roughly 22,000 mi/h.

The fact that mass, energy, and momentum are all measured in kilograms in the SR system reflects the fundamental unity of these quantities: the mass, energy, and momentum of an object are simply different aspects (specifically, the magnitude, time component, and the magnitude of the spatial components, respectively) of its four- momentum vector.

But we also see that the kilogram is an inconveniently large unit of mass, energy, and momentum when we are studying the dynamics of subatomic particles.

For example, a proton moving at four-fifths the speed of light in the +x direction has a mass, energy, and x momentum of about 16.7×10^{-27} kg, 27.8×10^{-27} kg, and 2.23×10^{-27} kg, respectively; it is inconvenient, to say the least, to work with numbers of this size.

Particle physicists find it convenient to modify the SR system of units one step further and measure mass, momentum, and energy in terms of a different unit called the electronvolt, or eV, where 1eV is defined to be the energy gained by a single electron as it passes through a potential difference of 1 volt.

The conversion factor to joules is

$$1 \text{ eV} \equiv 1.602 \times 10^{-19} \text{ J}$$

This is a very small unit of energy, well suited to the study of subatomic particle interactions.

We can link this unit to the SR unit of kilograms:

$$1\ \text{eV} = 1.602 \times 10^{-19}\ \text{J}\left(\frac{1\ \text{kg}}{8.998 \times 10^{16}\ \cancel{\text{J}}}\right) = 1.782 \times 10^{-36}\ \text{kg}$$

Using this conversion factor, all masses, energies, and momenta can be expressed in electronvolts.

For example, the mass of the electron is

$$m_e = 9.11 \times 10^{-31}\ \cancel{\text{kg}}\left(\frac{1\ \text{eV}}{1.782 \times 10^{-36}\ \cancel{\text{kg}}}\right) = 5.11 \times 10^{5}\ \text{eV} = 0.511\ \text{MeV}$$

SUBATOMIC PARTICLE PROPERTIES

The table lists some subatomic particles and their properties, including masses given in units of MeV $=10^6$ eV.

The superscript attached to the particle symbol in each case indicates the sign of the charge of the particle.

SOME SUBATOMIC PARTICLE PROPERTIES*

Category	Name	Symbol	Antiparticle	Mass, MeV	Half-life, s
Field particles are carriers	Gluon	g	Same	0	Stable
of fundamental forces.	Photon	γ	Same	0	Stable
	W boson	W^+	W^-	81,000	$\approx 10^{-24}$
	Z boson	Z^0	Same	92,000	$\approx 10^{-24}$
Leptons are fairly lightweight	Neutrino[†]	ν	$\bar{\nu}$	0 (?)	Stable (?)
particles that do not part-	Electron	e^-	e^+	0.511	Stable
icipate in what is called the	Muon	μ^-	μ^+	105.66	1.524×10^{-6}
"strong" nuclear interaction.	Tau	τ^-	τ^+	1784	3.0×10^{-13}
Mesons are all construct-	Pi meson	π^0	Same	134.97	0.58×10^{-16}
ed of a quark and an anti-	(Pion)	π^+	π^-	139.57	1.804×10^{-8}
quark. There are actually	K meson	K^+	K^-	493.6	0.857×10^{-8}
two different kinds of K^0	(Kaon)	K^0	$\overline{K}^0$	497.7	$359 / 0.618 \times 10^{-10}$
particle, one with a longer	D meson	D^+	D^-	1869	7.4×10^{-13}
half-life than the other.	Psi	ψ	Same	3097	6.9×10^{-21}
	B meson	B^+	B^-	5278	9.1×10^{-13}
	Upsilon	Υ	Same	9460	9.0×10^{-21}
Baryons are all construct-	Proton	p^+	p^-	938.27	Stable? ($> 10^{38}$)
ed of triplet of quarks.	Neutron	n	$\bar{n}$	939.57	621
The baryons listed here	Lambda	Λ^0	$\bar{\Lambda}^0$	1116	1.82×10^{-10}
are all constructed of	Sigma plus	Σ^+	$\bar{\Sigma}^+$	1189	0.554×10^{-10}
the combinations of the *u*,	Sigma zero	Σ^0	$\bar{\Sigma}^0$	1193	5×10^{-20}
d, and *s* quarks (the	Sigma minus	Σ^-	$\bar{\Sigma}^-$	1197	$1.03 \times 10^{--10}$
three lowest-mass quarks).	Xi zero	Ξ^0	$\bar{\Xi}^0$	1315	2.0×10^{-10}
	Xi minus	Ξ^-	$\bar{\Xi}^-$	1321	1.14×10^{-10}
	Omega minus	Ω^-	$\bar{\Omega}^-$	1672	0.57×10^{-10}

Every particle has a corresponding antiparticle having opposite charge and the same mass and lifetime.

The conventional notation for an antiparticle involves putting a bar over the symbol for the particle (for example, $\bar{\Sigma}^+$ is the negatively charged antiparticle corresponding to the Σ^+particle), but in certain cases where there is no ambiguity, the bar is omitted and the charge superscript is reversed (for example, e^+, μ^+, π^-, K^-, p^-).

EXAMPLE: A MOVING ELECTRON

Problem

In the laboratory frame, an electron is observed to move with such a speed that its relativistic kinetic energy is K= 0.20 MeV.

How fast is the electron moving in that frame?

What is the relativistic momentum p in that frame?

If the electron is moving along the +x axis, what is its four-momentum vector in that frame?

Solution

The electron's total energy is given by

$$E = K + m = 0.20 \text{ MeV} + 0.51 \text{ MeV} = 0.71 \text{ MeV}$$

Knowing this and the mass of the electron allows us to calculate its relativistic momentum:

$$m^2 = E^2 - p^2 \qquad \text{so} \qquad p^2 = E^2 - m^2$$

Thus

$$p = [(0.71\ \text{MeV})^2 - (0.51\ \text{MeV})^2]^{1/2} \approx 0.49\ \text{MeV}$$

Then, we have

$$v = \frac{p}{E} = \frac{0.49\ \text{MeV}}{0.71\ \text{MeV}} \approx 0.69$$

Since the electron's motion is entirely along the +x axis, its four-momentum is

$$\mathbf{P} = \begin{bmatrix} P_t \\ P_x \\ P_y \\ P_z \end{bmatrix} = \begin{bmatrix} E \\ +p \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} 0.71\ \text{MeV} \\ 0.49\ \text{MeV} \\ 0 \\ 0 \end{bmatrix}$$

EXAMPLE: KAON DECAY

Problem

A K° kaon at rest decays into two π° pions.

Choose the +x direction to be the direction of motion of one of the π° pions.

In what direction does the other pion move?

What are the velocities and kinetic energies of the two pions?

Solution

In this case, the law of conservation of four-momentum implies that the sum of the product pions' four-momenta must equal the four-momentum of the original kaon.

Since the kaon is at rest, the time component of its four-momentum is equal to its mass (M = 498 MeV), and therefore the spatial components of the kaon's four-momentum are equal to zero.

Similarly, we have that $P_y = P_z = 0$ for the pion moving in the +x direction, since it has no component of velocity in the y and z directions by hypothesis.

Let us call the relativistic energy of this pion E_1 and its relativistic momentum p_1.

The other pion we know nothing about yet: let us represent its relativistic energy by E_2 and the spatial components of its four-momentum by P_{2x}, P_{2y}, and P_{2z}.

The mass of each pion produced by the decay is m= 135 MeV.

With these symbols defined, the law of conservation of four-momentum then requires that

K meson four-momentum

Four-momentum of pion moving in +*x* direction

Four-momentum of second pion

$$\begin{array}{l} t \text{ component:} \\ x \text{ component:} \\ y \text{ component:} \\ z \text{ component:} \end{array} \quad \begin{bmatrix} M \\ 0 \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} E_1 \\ +p_1 \\ 0 \\ 0 \end{bmatrix} + \begin{bmatrix} E_2 \\ P_{2x} \\ P_{2y} \\ P_{2z} \end{bmatrix}$$

This equation reads $0 = 0 + P_{2y}$, implying that $P_{2y} = 0$.

Similarly it implies that $P_{2z} = 0$.

This equation also reads $0 = p_1 + P_{2x}$, implying that $P_{2x} = -p_1$.

In short, the final three equations imply that the second pion moves in the -x direction but that the relativistic momenta of the two pions are the same: $p_1 = |P_{1x}| = |P_{2x}| = p_2$.

Now, both pions have the same mass as well, so we have

$$m^2 = E_1^2 - p_1^2 = E_2^2 - p_2^2 = E_2^2 - p_1^2$$

since $p_1 = p_2$.

This implies that $$E_1^2 = E_2^2 \Rightarrow E_1 = E_2$$

since energies are always positive.

Therefore, the time component of conservation of four-momentum reduces in this case to

$$M = E_1 + E_2 = E_1 + E_1 \Rightarrow E_1 = \tfrac{1}{2}M = \tfrac{1}{2}(498 \text{ MeV}) = 249 \text{ MeV}$$

The two pions thus have the same kinetic energy,

$$K = E_1 - m = 249 \text{ MeV} - 135 \text{ MeV} = 114 \text{ MeV}$$

the same relativistic momentum,

$$m^2 = E_1^2 - p_1^2$$

$$m^2 \Rightarrow p_1 = p_2 = \sqrt{E_1^2 - m^2} = \sqrt{(249\ \text{MeV})^2 - (135\ \text{MeV})^2} \approx 209\ \text{MeV}$$

and the same speed

$$v_1 = \frac{p_1}{E_1} = \frac{209\ \text{MeV}}{249\ \text{MeV}} = 0.839 \qquad v_2 = \frac{p_2}{E_2} = \frac{p_1}{E_1} = v_1$$

though the two pions move in opposite directions along the x axis.

THE FOUR-MOMENTUM OF A PHOTON

I have mentioned before that light can be understood as an electromagnetic wave: the wave nature of light can be easily demonstrated with diffraction experiments.

But since the early part of the last century, it has also been understood that light behaves under certain circumstances as if it were made up of tiny particles, called **photons**.

The experimental evidence for this particle-like aspect of light is beyond the scope of this class.

Suffice it to say that in subatomic particle interactions, the particle nature of light becomes its crucial aspect.

The link between these different ways of looking at light is the assertion (first made by Einstein ni 1905) that each photon has an energy E = $h\upsilon$, where υ = the frequency of the light (thought of as a wave) and h is Planck's constant.

The frequency υ has units of seconds^{-1} and in fact is equal to $1/\lambda$, where λ is the wavelength of the light in question (if λ is measured in the SR unit of seconds).

Therefore, the energy of a photon of light is given by

$$E = h\nu = \frac{h}{\lambda} \qquad \text{(in SR units)}$$

where

$$h = 7.37 \times 10^{-51}\ \text{kg}\cdot\text{s}$$

If E is expressed in energy units of electron volts instead of the standard SR unit of kilograms, then

$$h = 4.14 \times 10^{-15}\ \text{eV}\cdot\text{s} = 4.14\ \text{eV}\cdot\text{fs}$$

where 1 fs = 1 femtosecond = 10^{-15} s.

For future reference, note that 1fs of distance is

$$1\ \text{fs (distance)} = 2.998 \times 10^{-7}\ \text{m} = 299.8\ \text{nm}$$

As discussed earlier, the only way we can create a self-consistent four-momentum vector for anything that carries energy at the speed of light is to assume that the object has zero mass.

This means that photons must have zero mass, which in turn means that the photon's energy E and its relativistic momentum p must be equal:

$$0 = m^2 = E^2 - p^2 \Rightarrow E = p \qquad \text{(for a photon)}$$

We will work with both photons and neutrinos in later examples.

The conventional symbol for the photon is the Greek letter γ (gamma); the symbol for the neutrino is the Greek letter υ (nu).

Both particles have zero charge.

For the sake of this class you should assume that both these kinds of particles have exactly zero mass.

The crucial thing to know about such particles is that the energy of a massless particle is equal to it relativistic momentum.

Such particles will always travel at exactly the speed of light: if E = p, then we have

$$\upsilon = \frac{p}{E} = 1$$

We will find these bits of knowledge very useful in the examples and problems that follow.

EXAMPLE: COMPTON SCATTERING

Problem

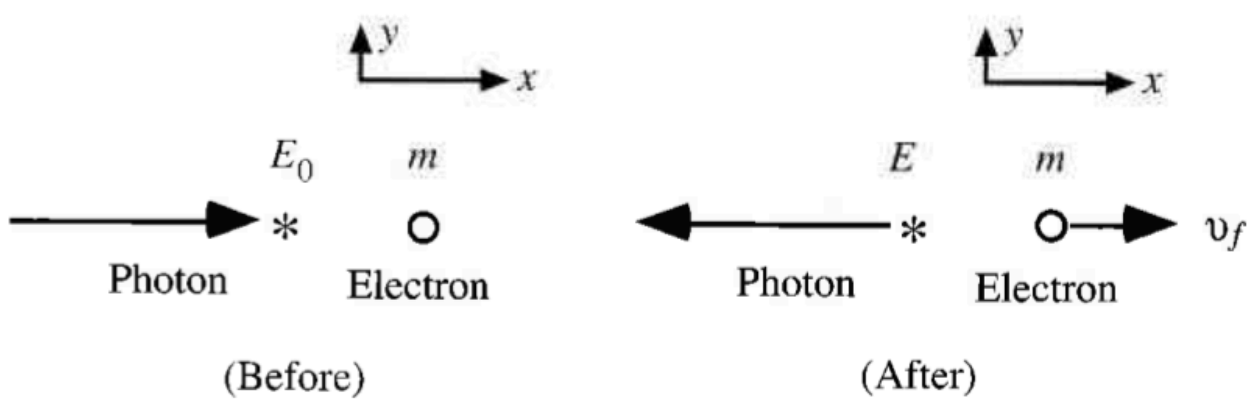
Certain radioactive substances radiate high-energy photons called gamma rays.

The energy of a single such photon can be measured with modern equipment.

Imagine that such a photon with energy E, collides with an electron of mass m at rest and rebounds from it back along the direction from which it came.

What is the energy of the rebounding photon?

See Figure.

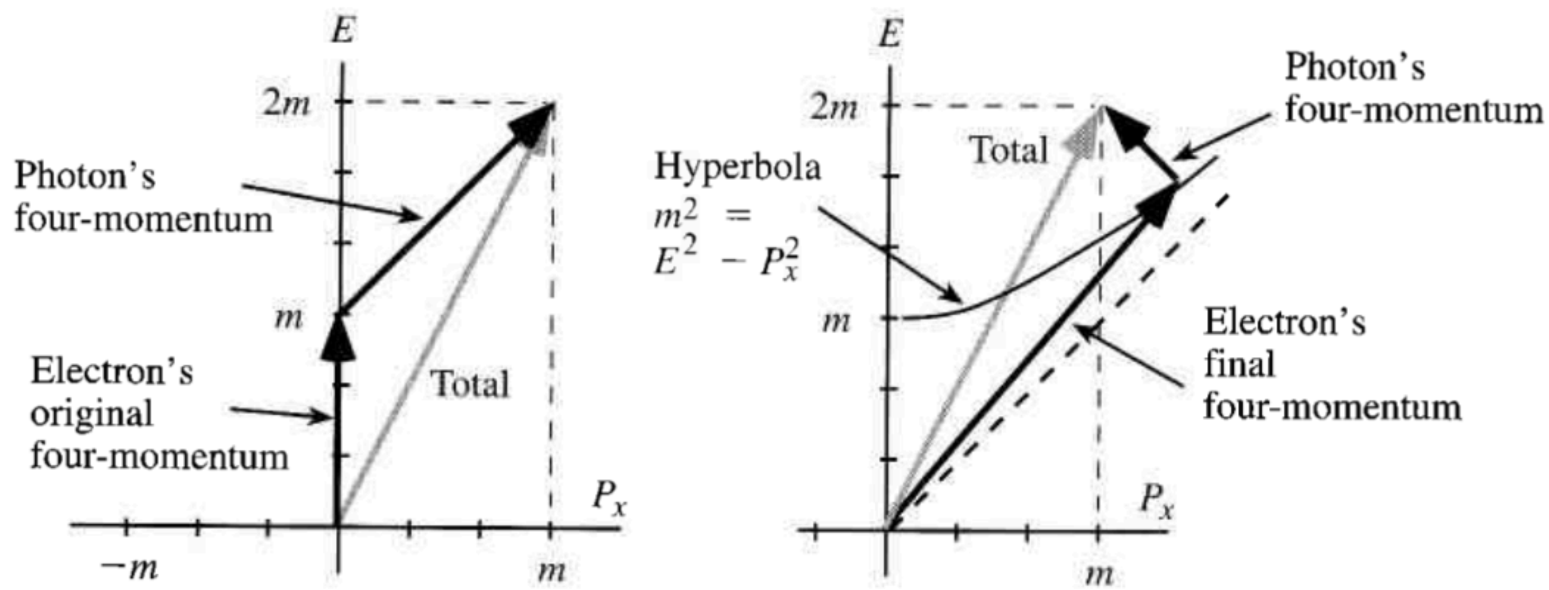


A photon with energy E, collides with an electron at rest. What is the energy E of the rebounding photon?

Solution

The next Figure shows an energy-momentum diagram of such a collision, taking the energy of the photon to be equal to the electron's mass m for the sake of concreteness.

This diagram clearly shows that E < Eo, that is, the photon's energy after hthete collision is smaller than its original energy (it actually looks that if we choose Eo = m, E turns out to be a bit larger than m/3).



(a) Energy-momentum diagram for the four-momenta of the photon and electron before the coli- sion. Note that the slope of the photon four-momentum must = 1. (b) The electron's final four- momentum will have its tip somewhere on the hyperbola shown. The final photon four-momentum must have a slope of -1 if it is moving in the -x direction. So draw a line of slope -1from the tip of the total four-momentum vector until it intersects the hyperbola: this is where the tip of the electron's final four-momentum arrow must be located.

The solution can be found algebraically as well.

Since the electron is at rest before he collision, its original x momentum is zero and its original energy is equal to its mass m.

Since p = E for photons and E is always positive, the original photon's x momentum must be +E, and the recoiling photon's x momentum must be -E.

Let E_f and P_f be the energy and x momentum of the electron after the collision.

The law of conservation of four-momentum implies that

$$
\begin{array}{l} t \text{ component:} \\ x \text{ component:} \\ y \text{ component:} \\ z \text{ component:} \end{array}
\qquad
\begin{bmatrix} E_0 \\ +E_0 \\ 0 \\ 0 \end{bmatrix} + \begin{bmatrix} m \\ 0 \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} E \\ -E \\ 0 \\ 0 \end{bmatrix} + \begin{bmatrix} E_f \\ P_{fx} \\ P_{fy} \\ P_{fz} \end{bmatrix}
$$

The bottom two equations tell us that $P_{fy} = P_{fz} = 0$, meaning that the recoiling electron moves along the x axis with relativistic momentum $p_f = |P_{fx}|$.

In this problem, E_o and m are known (we will assume that the original photon energy is given, and we can look up the electron's mass).

So these equations provide two equations in the three unknowns E, E_f and p_f.

But we also know that

$$m^2 = E_f^2 - p_f^2$$

since the magnitude of the recoiling electron's four-momentum must be equal to the electron's mass.

This provides the third equation that we need to solve for the three unknowns.

Looking at the x-component part of the equations above we have,

$$P_{fx} = E_0 + E = + p_f$$

(since E_o and E are positive, so is P_{fx}.) Using this result and solving for E_f we get

$$E_f^2 = m^2 + p_f^2 = m^2 + (E_0 + E)^2$$

Plugging this into the t-component part our equations to eliminate E_f, we get

$$E_0 + m = E + \sqrt{(E^0 + E)^2 + m^2}$$

or

$$(E_0 + m - E)^2 = (E_0 + E)^2 + m^2$$

Our task now is basically to solve for E.

Writing out the squares in the last equation, we get

$$E_0^2 + 2mE_0 - 2E_0E - 2mE + m^2 + E^2 = E_0^2 + 2E_0 E + E^2 + m^2$$

Canceling E_0^2, E^2, and m from both sides, we get

$$+ 2mE_0 - 2E_0E - 2mE = 2E_0 E \qquad \text{or} \qquad 2mE_0 = 4E_0E + 2mE$$

or

$$(2E_0 + m)E = mE_0 \qquad \text{implying that} \qquad E = \frac{mE_0}{2E_0 + m}$$

When $E_o = m$ (the case shown in the Energy-momentum graph), E = m/3, roughly equal to the result that we found graphically on the diagram.

An experiment to test this formula can be performed as follows.

Photons with a known energy are directed against a metal block.

Some of these photons collide with free electrons in the metal and bounce back.

The energy of these rebounding photons can be measured.

If a range of initial photon energies is used, the energies of the rebounding photons can be determined as a function of the original photon energy E_o.

This experiment was first performed by Arthur Compton in 1920 and has been repeated many times since. The results are in complete agreement with our results.

This is not only a vindication of the idea of conservation of four-momentum and the assumptions that we have made about how to handle photon four-momentum, but it represents compelling evidence for the idea that light can really be treated as if it were made up of photons that rebound from electrons as if they were billiard balls!

Historically, Compton's work was what finally convinced the physics community that the photon model of light was something more than a theorist's speculation.

EXAMPLE: PAIR PRODUCTION

At the level of subatomic particle physics, an inelastic collision is one that converts energy of motion into energy of mass by the creation of new particles.

Analogously, the Compton scattering process just described would be considered an elastic collision: no mass-energy was created or destroyed, implying that the kinetic energy of the system was conserved.

Fundamental particle theory predicts that when new particles are created out of kinetic energy, they must be created in pairs: every particle must be accompanied by its antiparticle.

For example, if an electron is created by an inelastic collision, an antielectron (a positron) must also be created.

Now a single gamma-ray photon can carry more energy than the mass energy of an electron-positron pair.

Nonetheless, the reaction

$$\gamma \rightarrow e^+ + e^-$$

is not observed to occur.

In fact, such a process cannot conserve four-momentum, as I will now demonstrate.

Let us take the direction of motion of the photon to define the +x direction.

Let the energies of the gamma-ray photon, the positron, and the electron be E_o, E_1 and E_2, respectively: conservation of the time component of the four-momentum thus requires

$$E_0 = E_1 + E_2$$

Our earlier work implies that for any particle with nonzero mass (such as the electron or the positron in the case under consideration),

$$m^2 = E^2 - P_x^2 - P_y^2 - P_z^2 \Rightarrow P_x^2 = E^2 - P_y^2 - P_z^2 - m^2 < E^2 \Rightarrow |P_x| < E$$

since E is always positive.

Plugging this into the first equation, we see that

$$E_0 > P_{1x} + P_{2x}$$

Now, the photon's relativistic momentum is equal to E_o, because it is massless.

Since the photon is moving in the +x direction by hypothesis, its x component of momentum is $P_x = +E_0$.

Therefore, conservation of the x component of four-momentum requires

$$E_0 = P_{1x} + P_{2x}$$

Do you see that this leads to a contradiction?

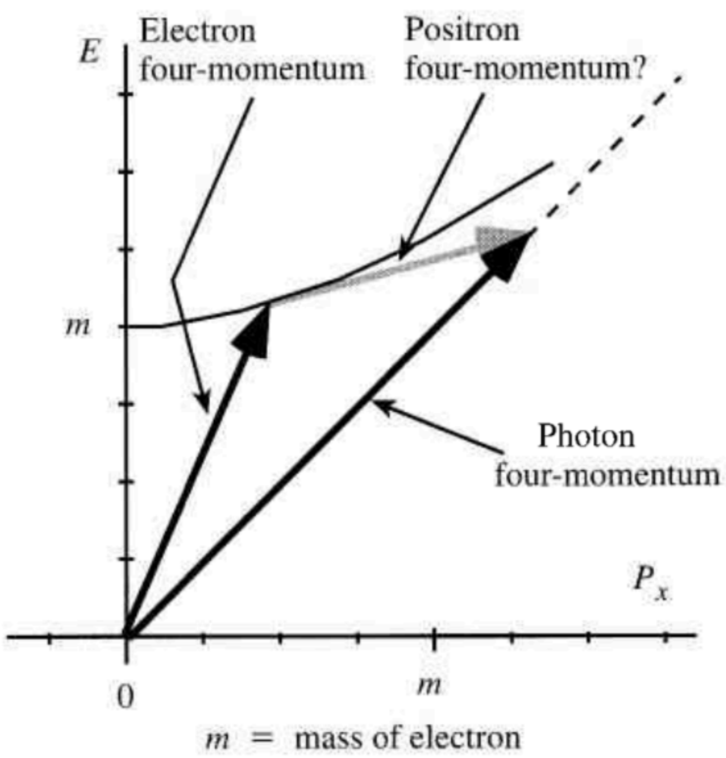
The last equation implies that E_o is strictly equal to $P_{1x} + P_{2x}$.

The previous equation implies that E_o is strictly greater than $P_{1x} + P_{2x}$.

These equations cannot be simultaneously true, and yet they must be simultaneously true for four-momentum to be conserved.

The conclusion is that four-momentum cannot be conserved for the indicated process described and thus the process cannot occur.

This assertion is even more vividly illustrated by the energy-momentum diagram shown in the Figure.



This diagram illustrates that it is impossible for a single photon to produce an electron-positron pair. The four-momentum vector of the initial photon has a slope of 1 on this diagram, since the photon moves at the speed of light. The created electron must move at a speed < 1, since it has nonzero mass, so the slope of its four-momentum will be less than 1. Conservation of four-momentum requires that the four-momenta of the electron and positron add up to that of the photon. But the only way for this to happen is if the positron's four-momentum has a slope greater than 1, implying that the positron is moving faster than light. This is impossible, so the decay process is impossible.

This diagram shows that one of the pair of particles created from the photon would have to be moving faster than the speed of light fi the total four-momentum of the created pair is to be equal to the original four-momentum of the photon.

It is possible, however, to create an electron-positron pair from a single gamma-ray photon by means of the reaction

$$\gamma + e^- \text{ (at rest)} \rightarrow e^- + e^- + e^+$$

In this interaction, the gamma-ray photon collides with an electron at rest and disappears, leaving an electron-positron pair in its place.

The presence of the extra electron in this interaction enables the four-momentum of the system to be conserved, as the worked example below demonstrates.

Problem

Consider the collision process described by the last equation.

Define the +x direction to be the direction of motion of the original gamma-ray photon.

For simplicity's sake, imagine that the product particles in this reaction move together with the same velocity in the +x direction after the collision.

Show that four-momentum can be conserved in such a process if the gamma-ray photon has a certain energy and calculate this energy.

Solution

Let the unknown energy of the gamma-ray photon be E_o.

Since it is massless and moving in the +x direction, its four-momentum will be $[P_t, P_x, P_y, P_z] = [E_o, +E_o, 0, 0]$.

The original electron at rest has a four-momentum [m, 0, 0, 0], where m is the mass of the electron.

The three product particles all have the same velocity by hypothesis, and they also have the same mass, so their total four-momentum will be the same as the four-momentum of a single particle of mass 3m traveling at their joint velocity.

Let us denote the total four-momentum vector of these three particles by $[E_3, P_{3x}, P_{3y}, P_{3z}$l and note that we must have,

$$(3m)^2 = E_3^2 - P_{3x}^2 - P_{3y}^2 - P_{3z}^2$$

The law of conservation of four-momentum then requires that

Target electron

Incoming photon — Joint four-momentum of product particles

$$
\begin{array}{l}
t \text{ component:} \\
x \text{ component:} \\
y \text{ component:} \\
z \text{ component:}
\end{array}
\quad
\begin{bmatrix} E_0 \\ E_0 \\ 0 \\ 0 \end{bmatrix} + \begin{bmatrix} m \\ 0 \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} E_3 \\ P_{3x} \\ P_{3y} \\ P_{3z} \end{bmatrix}
$$

The last two equations imply that if the product particles move together after the collision, they must move in the +x direction.

The second equation implies that $P_{3x} = E_o$.

We then find that

$$(3m)^2 = E_3^2 - E_0^2 \quad \Rightarrow \quad E_3 = \sqrt{E_0^2 + (3m)^2}$$

Plugging this into the first equation, we get

$$
\begin{aligned}
E_0 + m = \sqrt{E_0^2 + (3m)^2} \quad &\Rightarrow \quad E_0^2 + 2E_0 m + m^2 = E_0^2 + 9m^2 \\
&\Rightarrow \quad 2E_0 m = 8m^2 \\
&\Rightarrow \quad E_0 = 4m
\end{aligned}
$$

We see that four-momentum can be conserved in this process if the original photon has an energy of 4m = 4(0.511 MeV) = 2.04 MeV.

This is twice as much energy as one might think would be required to produce the rest mass of the new particles, but a certain amount of the energy of the photon must be channeled into kinetic energy of the product particles if the four-momentum of the entire system is to be conserved.

Special Relativity - A More Mathematical Approach

Although Newtonian mechanics gives an excellent description of Nature, it is not universally valid.

When we reach extreme conditions — the very small, the very heavy or the very fast — the Newtonian Universe that we're used to needs replacing.

You could say that Newtonian mechanics encapsulates our common sense view of the world.

One of the major themes of twentieth century physics is that when you look away from our everyday world, common sense is not much use.

One such extreme is when particles travel very fast.

The theory that replaces Newtonian mechanics is due to Einstein. It is called special relativity.

The effects of special relativity become apparent only when the speeds of particles become comparable to the speed of light in the vacuum.

The speed of light is $c = 299792458\ ms^{-1}$

This value of c is exact.

It may seem strange that the speed of light is an integer when measured in meters per second.

The reason is simply that this is taken to be the definition of what we mean by a meter: it is the distance travelled by light in 1/299792458 seconds.

For the purposes of this discussion, we'll be quite happy with the approximation $c \simeq 3 \times 10^8$ ms^{-1}.

The first thing to say is that the speed of light is fast.

Really fast.

The speed of sound is around 300 ms^{-1}; escape velocity from the Earth is around 10^4 ms^{-1}; the orbital speed of our solar system in the Milky Way galaxy is around 10^5 ms^{-1}.

As we shall soon see, nothing travels faster than c.

The theory of special relativity rests on two experimental facts.

We will look at the evidence for these shortly.

In fact, we have the first of these is simply the older Galilean principle of relativity.

The second postulate is more surprising:

- **Postulate 1:** The principle of relativity: the laws of physics are the same in all inertial frames
- **Postulate 2:** The speed of light in vacuum is the same in all inertial frames

On the face of it, the second postulate looks nonsensical.

How can the speed of light look the same in all inertial frames?

If light travels towards me at speed c and I run away from the light at speed v, surely I measure the speed of light as c - v.

Right? Well, no.

This common sense view is encapsulated in the older Galilean transformations.

Mathematically, we derive this "obvious" result as follows: two inertial frames, S and S′, which move relative to each with velocity v = (v, 0, 0), have Cartesian coordinates related by

$$x' = x - vt \quad , \quad y' = y \quad , \quad z' = z \quad , \quad t' = t \qquad (1)$$

If a ray of light travels in the x direction in frame S with speed c, then it traces out the trajectory x/t = c.

The transformations above then tell us that in frame S′ the trajectory if the light ray is x'/t' = c - v.

This is the result we claimed above: the speed of light should clearly be c - v.

If this is wrong (and it is) something must be wrong with the Galilean transformations.

But what?

Our immediate goal is to find a transformation law that obeys both postulates above.

As we will see, the only way to achieve this goal is to allow for a radical departure in our understanding of time.

In particular, we will be forced to abandon the assumption of absolute time, enshrined in the equation $t' = t$ above.

We will see that time ticks at different rates for observers sitting in different inertial frames.

Lorentz Transformations

We stick with the idea of two inertial frames, S and S', moving with relative speed v.

For simplicity, we'll start by ignoring the directions y and z which are perpendicular to the direction of motion.

Both inertial frames come with Cartesian coordinates: (x, t) for S and (x', t') for S'.

We want to know how these are related.

The most general possible relationship takes the form

$$x' = f(x,t) \quad , \quad t' = g(x,t)$$

for some function f and g.

However, there are a couple of facts that we can use to immediately restrict the form of these functions.

The first is that the law of inertia holds; left alone in an inertial frame, a particle will travel at constant velocity.

Drawn in the (x, t) plane, the trajectory of such a particle is a straight line.

Since both S and S' are inertial frames, the map $(x, t) \rightarrow (x', t')$ must map straight lines to straight lines; such maps are, by definition, linear.

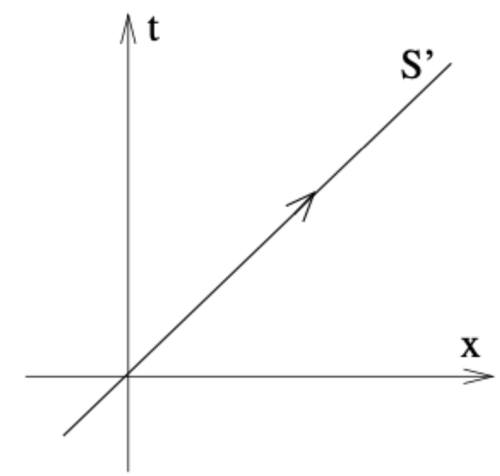
The functions f and g must therefore be of the form

$$x' = \alpha_1 x + \alpha_2 t \quad , \quad t' = \alpha_3 x + \alpha_4 t$$

where α_i, $i = 1, 2, 3, 4$ can each be a function of v.

Secondly, we use the fact that S' is travelling at speed v relative to S.

This means that an observer sitting at the origin, $x' = 0$, of S' moves along the trajectory $x = vt$ in S shown in the figure.



Or, in other words, the points $x = vt$ must map to $x' = 0$.

There is actually one further assumption implicit in this statement: that the origin $x' = 0$ coincides with $x = 0$ when $t = 0$.

Together with the requirement that the transformation is linear, this restricts the coefficients α_1 and α_2 above to be of the form,

$$x' = \gamma(x - vt) \tag{2}$$

for some coefficient .

Once again, the overall coefficient can be a function of the velocity: $\gamma = \gamma_v$.

We've used subscript notation γ_v rather than the more standard $\gamma(v)$ to denote that depends on v.

This avoids confusion with the factors of (x - vt) which aren't arguments of but will frequently appear after γ like in the last equation.

There is actually a small, but important, restriction on the form of γ_v: it must be an even function, so that $\gamma_v = \gamma_{-v}$.

There are a couple of ways to see this.

with The first is by using rotational invariance, which states that γ can depend only on the direction of the relative velocity $\mathbf{v}$, but only on the magnitude $v^2 = \mathbf{v} \cdot \mathbf{v}$.

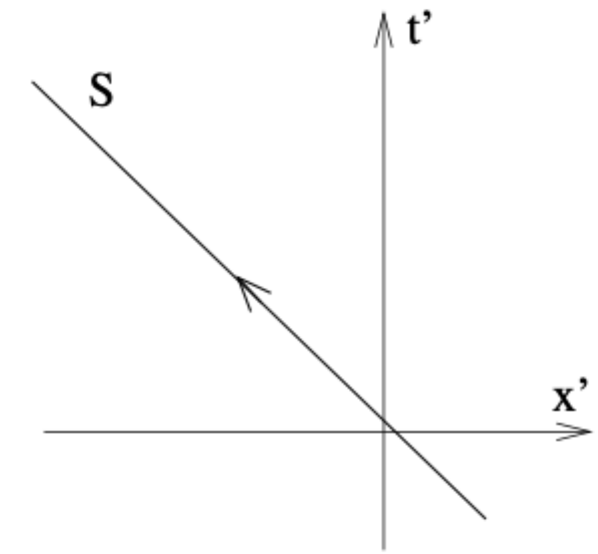
Alternatively, if this is a little slick, we can reach the same conclusion by considering inertial frames $\tilde{S}$ and $\tilde{S}'$ which are identical to S and S$'$ except that we measure the x-coordinate in the opposite direction, meaning $\tilde{x}$ = -x and $\tilde{x}'$ = -x$'$.

While S is moving with velocity +v relative to S$'$, $\tilde{S}$ is moving with velocity -v with respect to $\tilde{S}'$ simply because we measure things in the opposite direction.

That means that

$$\tilde{x}' = \gamma_{-v} \left(\tilde{x} + v\tilde{t}\right)$$

Comparing to (2), we see that we must have $\gamma_v = \gamma_{-v}$ as claimed.



We can also look at things from the perspective of S′, relative to which the frame S moves backwards with velocity -v.

The same argument as before now tells us that

$$x = \gamma(x' + vt') \tag{3}$$

Now the function $\gamma = \gamma_{-v}$.

But by the argument above, we know that $\gamma_v = \gamma_{-v}$.

In other words, the coefficient appearing in (3) is the same as that appearing in (2).

At this point, things don't look too different from what we've seen before.

Indeed, if we now insisted on absolute time, so t = t′, we're forced to have $\gamma = 1$ and we get back to the Galilean transformations.

However, as we've seen, this is not compatible with the second postulate of special relativity.

So let's push forward and insist instead that the speed of light is equal to c in both S and S$'$.

In S, a light ray has trajectory

$$x = ct$$

While, in S$'$, we demand that the same light ray has trajectory

$$x' = ct'$$

Substituting these trajectories into (2) and (3), we have two equations relating t and t$'$,

$$ct' = \gamma(c - v)t \quad \text{and} \quad ct = \gamma(c + v)t'$$

A little algebra shows that these two equations are compatible only if is given by

$$\gamma = \sqrt{\frac{1}{1 - v^2/c^2}} \tag{4}$$

We'll be seeing a lot of this coefficient in what follows.

Notice that for v << c, we have $\gamma \approx 1$ and the transformation law (2) is approximately the same as the Galilean transformation.

However, as v → c we have $\gamma \to \infty$.

Furthermore, becomes γ imaginary for v > c which means that we're unable to make sense of inertial frames with relative speed v > c.

Equations (2) and (4) give us the transformation law for the spatial coordinate.

But what about for time?

In fact, the temporal transformation law is already lurking in our analysis above.

Substituting the expression for x′ in (2) into (3) and rearranging, we get

$$t' = \gamma \left(t - \frac{v}{c^2} x \right) \tag{5}$$

We shall soon see that this equation has dramatic consequences.

For now, however, we merely note that when $v << c$, we recover the trivial Galilean transformation law $t' \approx t$.

Equations (2) and (5) are the *Lorentz transformations*.

Lorentz Transformations in Three Spatial Dimensions

In the above derivation, we ignored the transformation of the coordinates y and z perpendicular to the relative motion.

In fact, these transformations are trivial.

Using the above arguments for linearity and the fact that the origins coincide at $t = 0$, the most general form of the transformation is

$$y' = \kappa y$$

But, by symmetry, we must also have $y = \kappa y'$.

Clearly, we require $\varkappa = 1$.

The other possibility $\varkappa = -1$ does not give the identity transformation when $v = 0$.

Instead, it is a reflection.

With this we can write down the final form of the Lorentz transformations.

Note that they look more symmetric between x and t if we write them using the combination ct,

$$\begin{aligned} x' &= \gamma\left(x - \frac{v}{c}\,ct\right) \\ y' &= y \\ z' &= z \\ ct' &= \gamma\left(ct - \frac{v}{c}x\right) \end{aligned} \tag{6}$$

where γ is given by (4).

These are also known as *Lorentz boosts*.

It's also worth stressing again the special properties of these transformations.

To be compatible with the first postulate, the transformations must take the same form if we invert them to express x and t in terms of x' and t', except with v replaced by -v.

And, after a little bit of algebraic magic, they do.

Secondly, we want the speed of light to be the same in all inertial frames.

For light travelling in the x direction, we already imposed this in our derivation of the Lorentz transformations.

But it's simple to check again: in frame S, the trajectory of an object travelling at the speed of light obeys $x = ct$.

In S', the same object will follow the trajectory $x' = \gamma(x - vt) = (ct - vx/c) = ct'$.

What about an object travelling in the y direction at the speed of light?

Its trajectory in S is y = ct.

From (6), its trajectory in S′ is y′ = ct′/ and x′ = vt′.

Its speed in S′ is therefore $v'^2 = v_x{}^2 + v_y{}^2$, o r $$v'^2 = \left(\frac{x'}{t'}\right)^2 + \left(\frac{y'}{t'}\right)^2 = v^2 + \frac{c^2}{\gamma^2} = c^2$$

Spacetime Diagrams

We'll find it very useful to introduce a simple spacetime diagram to illustrate the physics of relativity.

In a fixed inertial frame, S, we draw one direction of space — say x — along the horizontal axis and time on the vertical axis.

But things look much nicer if we rescale time and plot ct on the vertical instead.

In the context of special relativity, space and time is called Minkowski space.

Although the true definition of Minkowski space requires some extra structure on space and time which we will meet later).

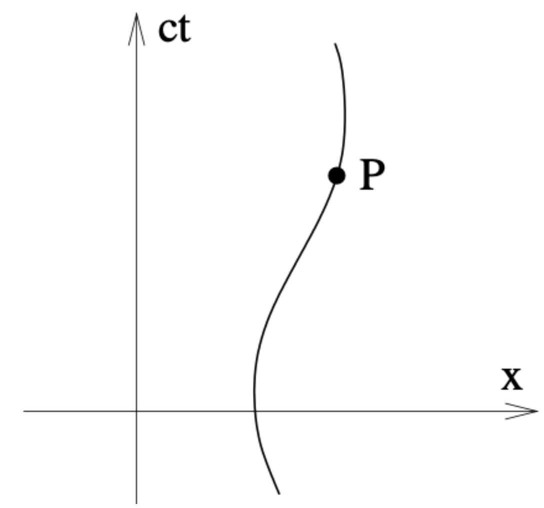
This is a spacetime diagram.

Each point, P, represents an *event*.

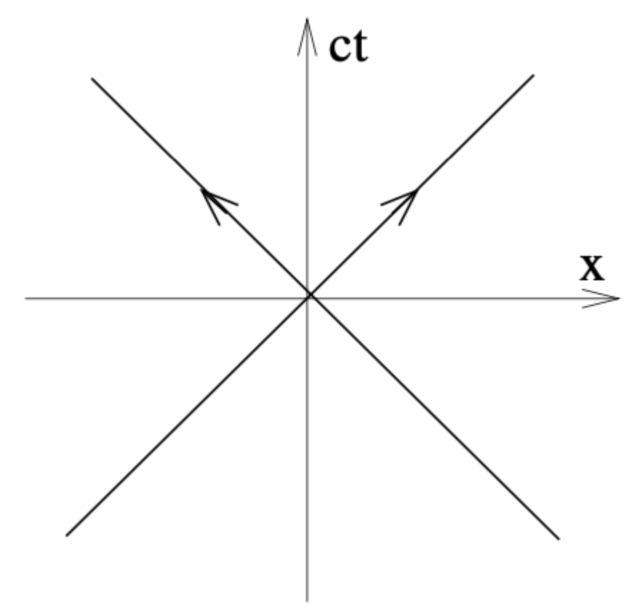
In the following, we'll label points on the spacetime diagram as coordinates (ct, x) i.e. giving the coordinate along the vertical axis first.

This is backwards from the usual way coordinates but is chosen so that it is consistent with a later, standard, convention that we will meet.

A particle moving in spacetime traces out a curve called a worldline as shown in the figures.



Because we've rescaled the time axis, a light ray moving in the x direction moves at 45°.



We'll later see that no object can move faster than the speed of light which means that the worldlines of particles must always move upwards at an angle steeper than 45°.

The horizontal and vertical axis in the spacetime diagram are the coordinates of the inertial frame S.

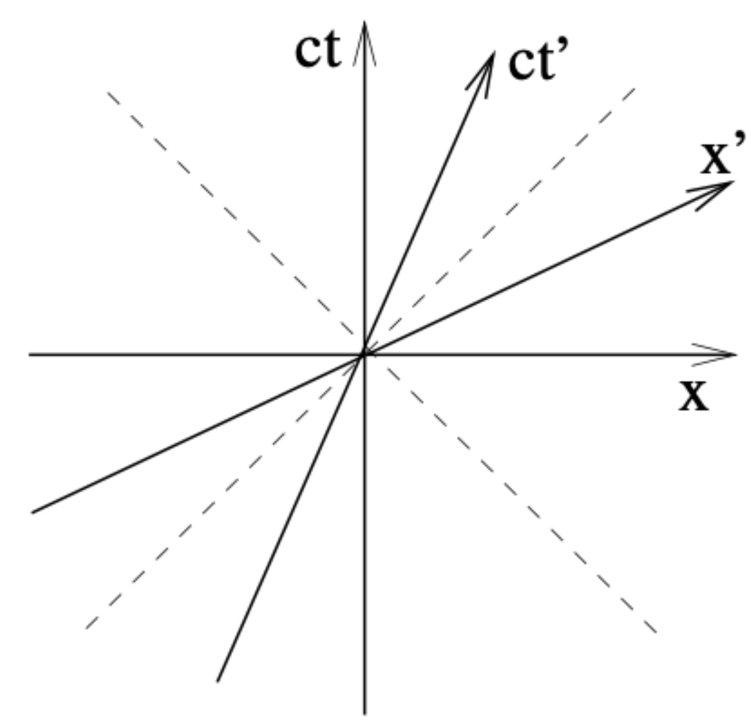
But we could also draw the axes corresponding to an inertial frame S′ moving with relative velocity v = (v, 0, 0).

The t′ axis sits at x′ = 0 and is given by

$$x = vt$$

Meanwhile, the x′ axis is determined by t′ = 0 which, from the Lorentz transformation (6), is given by the equation

$$ct = \frac{v}{c}x$$



These two axes are drawn on the figure to the right.

They can be the thought of as the x and ct axes, rotated by an equal amount towards the diagonal light ray.

The fact the axes are symmetric about the light ray reflects the fact that the speed of light is equal to c in both frames.

Relativistic Physics

Now we will explore some of the more interesting and surprising consequences of the Lorentz transformations.

Simultaneity

We start with a simple question: how can we be sure that things happen at the same time?

In Newtonian physics, this is a simple question to answer.

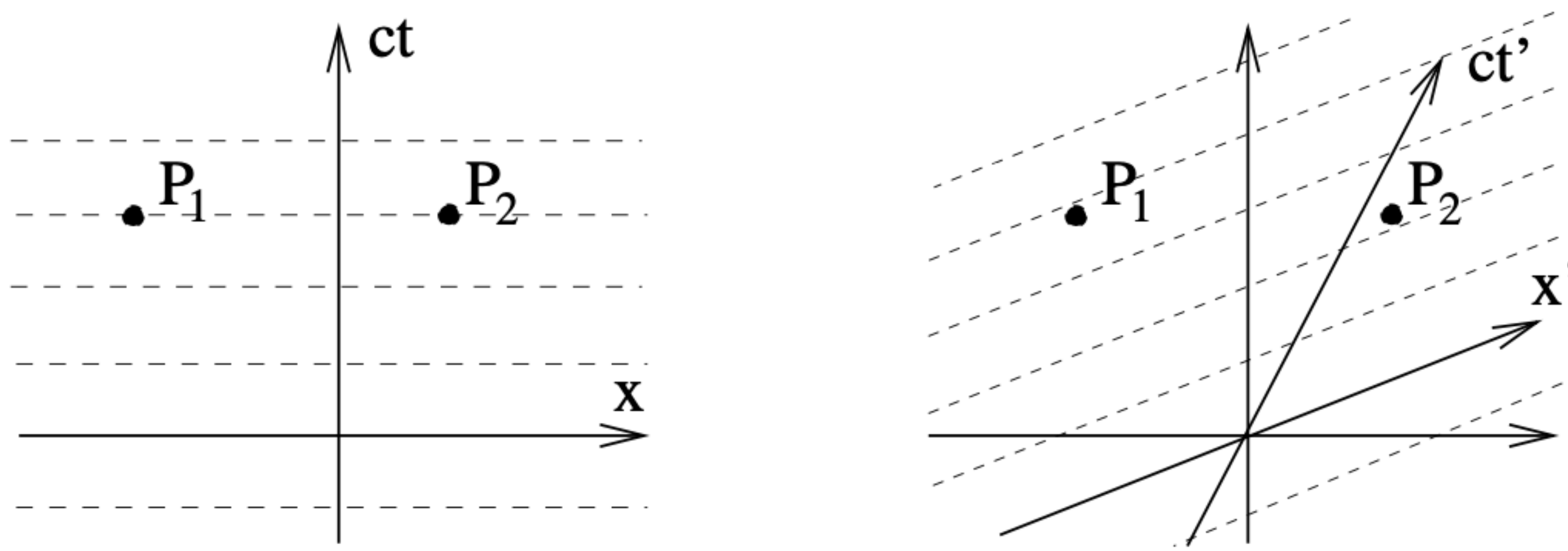
In that case, we have an absolute time t and two events, P_1 and P_2, happen at the same time if $t_1 = t_2$.

However, in the relativistic world, things are not so easy.

We start with an observer in inertial frame S, with time coordinate t.

This observer sensibly decides that two events, P_1 and P_2, occur simultaneously if $t_1 = t_2$.

In the spacetime diagram on the left of the Figure below we have drawn lines of simultaneity for this observer.



But for an observer in the inertial frame S', simultaneity of events occurs for equal t'.

Using the Lorentz transformation, lines of constant t' become lines described by the equation
$t - vx/c^2 = \text{constant}$.

These lines are drawn on the spacetime diagram on the right of the above Figure.

The upshot of this is that two events simultaneous in one inertial frame are not simultaneous in another.

An observer in S thinks that events P_1 and P_2 happen at the same time.

All other observers disagree.

A Potential Confusion: What the Observer Observes

We'll pause briefly to press home a point that may lead to confusion.

You might think that the question of simultaneity has something to do with the finite speed of propagation.

You don't see something until the light has travelled to you, just as you don't hear something until the sound has travelled to you.

This is not what's going on here!

A look at the spacetime diagram in the last Figure shows that we've already taken this into account when deciding whether two events occur simultaneously.

The lack of simultaneity between moving observers is a much deeper issue, not due to the finiteness of the speed of light but rather due to the constancy of the speed of light.

The confusion about the time of flight of the signal is sometimes compounded by the common use of the word observer to mean “inertial frame”.

This brings to mind some guy sitting at the origin, surveying all around him.

Instead, you should think of the observer more as a Big Brother figure: a sea of clocks and rulers throughout the inertial frame which can faithfully record and store the position and time of any event, to be studied at some time in the future.

Of course, this means that there is a second question we can ask which is: what does the guy sitting at the origin actually see?

Now we have to take into account both the relative nature of simultaneity and the issues related with the finite speed of propagation.

This adds an extra layer of complexity which we will discuss later.

Causality

We've seen that different observers disagree on the temporal ordering of two events.

But where does that leave the idea of causality?

Surely it's important that we can say that one event definitely occurred before another.

Thankfully, all is not lost: there are only some events which observers can disagree about.

To see this, note that because Lorentz boosts are only possible for $v < c$, the lines of simultaneity cannot be steeper than 45.

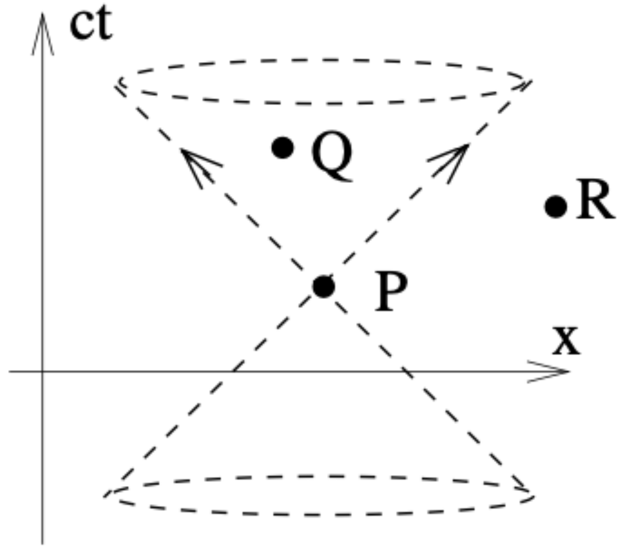
Take a point P and draw the 45 light rays that emerge from P.

This is called the light cone.

For once, in the figure, I've drawn this with an extra spatial dimension present to illustrate how this works in spatial dimensions bigger than one.

The light cone is really two cones, touching at the point P .

They are known as the future light cone and past light cone.



For events inside the light cone of P, there is no difficulty deciding on the temporal ordering of events.

All observers will agree that Q occurred after P .

However, for events outside the light cone, the matter is up for grabs: some observers will see R as happening after P; some before.

This tells us that the events which all observers agree can be causally influenced by P are those inside the future light cone.

Similarly, the events which can plausibly influence P are those inside the past light cone.

This means that we can sleep comfortably at night, happy in the knowledge that causality is preserved, only if nothing can propagate outside the light cone.

But that's the same thing as travelling faster than the speed of light.

The converse to this is that if we do ever see particles that travel faster than the speed of light, we're in trouble.

We could use them to transmit information faster than light.

But another observer would view this as transmitting information backwards in time.

All our ideas of cause and effect will be turned on their head.

You will therefore be relieved to learn that we will show in later why it is impossible to accelerate particles past the light speed barrier.

There is a corollary to the statement that events outside the lightcone cannot influence each other: there are no perfectly rigid objects.

Suppose that you push on one end of a rod.

The other end cannot move immediately since that would allow us to communicate faster than the speed of light.

Of course, for real rods, the other end does not move instantaneously.

Instead, pushing on one end of the rod initiates a sound wave which propagates through the rod, telling the other parts to move.

The statement that there is no rigid object is simply the statement that this sound wave must travel slower than the speed of light.

Finally, let me mention that when we're talking about waves, as opposed to point particles, there is a slight subtlety in exactly what must travel slower than light.

There are at least two velocities associated to a wave: the group velocity is (usually) the speed at which information can be communicated.

This is less than c.

In contrast, the phase velocity is the speed at which the peaks of the wave travel.

This can be greater than c, but transmits no information.

Time Dilation

We'll now turn to one of the more dramatic results of special relativity.

Consider a clock sitting stationary in the frame S' which ticks at intervals of T'.

This means that the tick events in frame S' occur at $(ct'_1, 0)$ then $(ct'_1 + cT', 0)$ and so on.

What are the intervals between ticks in frame S?

We can answer immediately from the Lorentz transformations (6).

Inverting this gives

$$t = \gamma \left(t' + \frac{vx'}{c^2} \right)$$

The clock sits at $x' = 0$, so we immediately learn that in frame S, the interval between ticks is

$$T = \gamma T'$$

This means that the gap between ticks is longer in the stationary frame.

A moving clock runs more slowly.

But the same argument holds for any process, be it clocks, elementary particles or human hearts.

The correct interpretation is that time itself runs more slowly in moving frames.

On Muons and Planes

Away from the world of gedankenexperiments, there are a couple of real experimental consequences of time dilation.

Certainly the place that this phenomenon is tested most accurately is in particle accelerators where elementary particles routinely reach speeds close to c.

The protons spinning around the LHC have $\gamma \approx 3500$.

The previous collider in CERN, called LEP, accelerated electrons and positrons to $\gamma \approx 2 \times 10^5$.

Although the electrons in LEP were travelling faster than the protons in LHC, the greater mass of the protons means that there is substantially more energy in the LHC collisions).

The effect of time dilation is particularly vivid on unstable particles which live much longer in the lab frame than in their own rest frame.

An early demonstration was seen in muons in 1941.

These are heavier, unstable, versions of the electron.

They decay into an electron, together with a couple of neutrinos, with a half-life of $\tau \approx 2 \times 10^{-6}$ s.

Muons are created when cosmic rays hit the atmosphere, and subsequently rain down on Earth.

Yet to make it down to sea level, it takes about $t = 7 \times 10^{-6}$ s, somewhat longer than their lifetime.

Given this, why are there any muons detected on Earth at all?

Surely they should have decayed.

The reason that they do not is because the muons are travelling at a speed $v \approx 0.99c$, giving $\gamma \approx 10$.

From the muon's perspective, the journey only takes $t^0 = t / \gamma \approx 7 \times 10^{-7}$ s, somewhat less than their lifetime.

Note that elementary particles are, by definition, structureless.

They're certainly not some clock with an internal machinery.

The reason that they live longer can't be explained because of some mechanical device which slows down: it is time itself which is running slower.

A more direct test of time dilation was performed in 1971 by Hafele and Keating.

They flew two atomic clocks around the world on commercial airliners; two more were left at home.

When they were subsequently brought together, their times differed by about 10^{-7}s.

There are actually two contributions to this effect: the time dilation of special relativity that we've seen above, together with a related effect in general relativity due to the gravity of the Earth.

Twin Paradox

Two twins, Luke and Leia, decide to spend some time apart. Leia stays at home while Luke jumps in a spaceship and heads at some speed v to the planet Tatooine.

With sadness, Leia watches Luke leave but is relieved to see — only a time T later from her perspective — him safely reach the planet.

However, upon arrival, Luke finds that he doesn't like Tatooine so much.

It is a dusty, violent place with little to do.

So he turns around and heads back to Leia at the same speed v as before.

When he returns, he finds that Leia has aged by $T_{Leia} = 2T$.

And yet, fresh faced Luke has only aged by $T_{Luke} = 2T/\gamma$.

We see, that after the journey, Luke is younger than Leia.

In fact, for large enough values of , Luke could return to find Leia long dead.

This is nothing more than the usual time dilation story.

So why is it a paradox?

Well, things seem puzzling from Luke's perspective.

He's sitting happily in his inertial spaceship, watching Leia and the whole planet flying off into space at speed v.

From his perspective, it should be Leia who is younger.

Surely things should be symmetric between the two?

The resolution to this “paradox” is that there is no symmetry between Luke’s journey and Leia’s.

Leia remained in an inertial frame for all time.

Luke, however, does not.

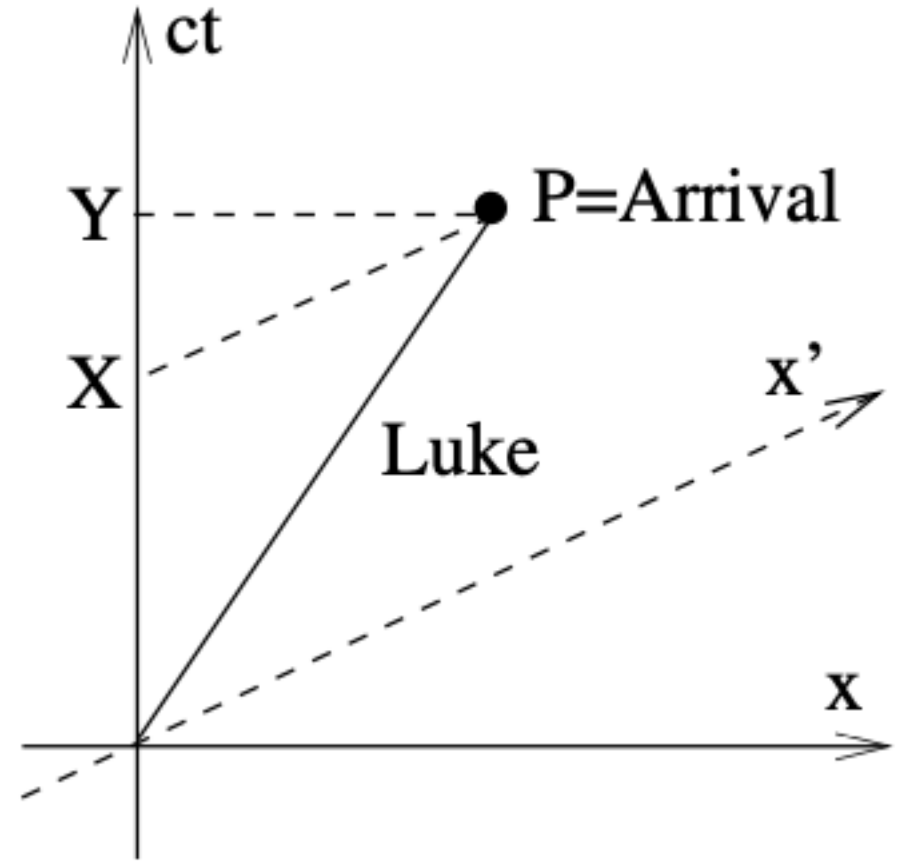
When he reaches Tatooine, he has to turn around and this event means that he has to accelerate.

This is what breaks the symmetry.

We can look at this in some detail.

We draw the space-time diagram in Leia’s frame.

Luke sits at $x = vt$, or $x' = 0$.



Leia sits at $x = 0$.

Luke reaches Tatooine at point P.

We've also drawn two lines of simultaneity.

The point Y is when Leia thinks that Luke has arrived on Tatooine.

The point X is where Luke thinks Leia was when he arrived at Tatooine.

As we've already seen, it's quite ok for Luke and Leia to disagree on the simultaneity of these points.

Let's figure out the coordinates for X and Y .

Event Y sits at coordinate (cT,0) in Leia's frame, while P is at (cT,vT).

The time elapsed in Luke's frame is just the usual time dilation calculation,

$$T' = \gamma\left(T - \frac{v^2T}{c^2}\right) = \frac{T}{\gamma}$$

We can also work out the coordinates of the event X.

Clearly this takes place at x = 0 in Leia's frame.

In Luke's frame, this is simultaneous with his arrival at Tatooine, so occurs at $t' = T' = T/\gamma$.

We can again use the Lorentz transformation

$$t' = \gamma\left(t - \frac{v^2x}{c^2}\right)$$

now viewed as an equation for t given x and t'.

This gives us

$$t = \frac{T'}{\gamma} = \frac{T}{\gamma^2}$$

So at this point, we see that everything is indeed symmetric.

When Luke reaches Tatooine, he thinks that Leia is younger than him by a factor of γ.

Meanwhile, Leia thinks that Luke is younger than her by the same factor .

Things change when Luke turns around.

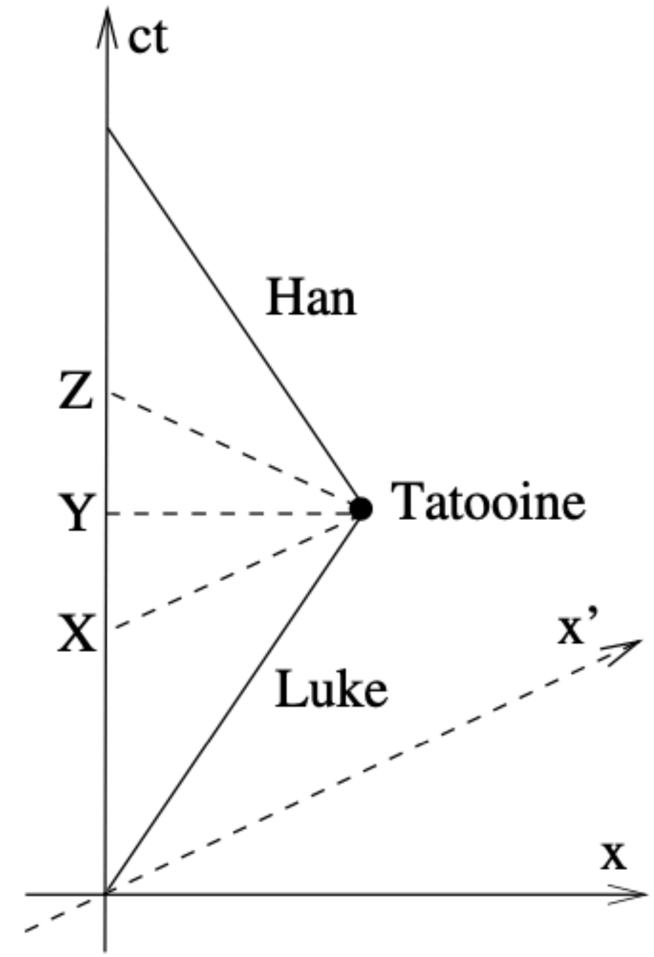
To illustrate this, let’s first consider a different scenario where he doesn’t return from Tatooine.

Instead, as soon as he arrives, he synchronizes his clock with a friend – let’s call him Han – who is on his way to meet Leia.

Now things are still symmetric.

Luke thinks that Leia has aged by T/γ^2 on the outward journey;

Han also thinks that Leia has aged by T/γ^2 on the inward journey.



So where did the missing time go?

We can see this by looking at the spacetime diagram of Han's journey.

We've again drawn lines of simultaneity.

From Han's perspective, he thinks that Leia is sitting at point Z when he still convinced that she's sitting at point X.

It's not hard to check that at point Z, Leia's clock reads $t = 2T - T/\gamma^2$.

From this perspective, we can also see what happens if Luke X does return home.

When he arrives at Tatooine, he thinks Leia is at point X.

Yet, in the time he takes to turn around and head home, the acceleration makes her appear to rapidly age, from point X to point Z.

Length Contraction

We've seen that moving clocks run slow.

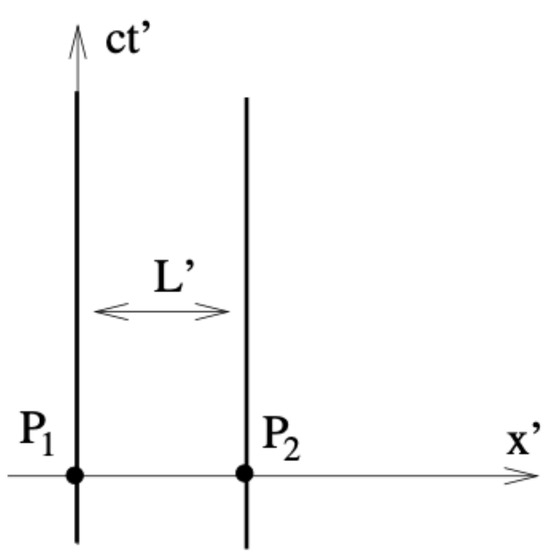
We will now show that moving rods are shortened.

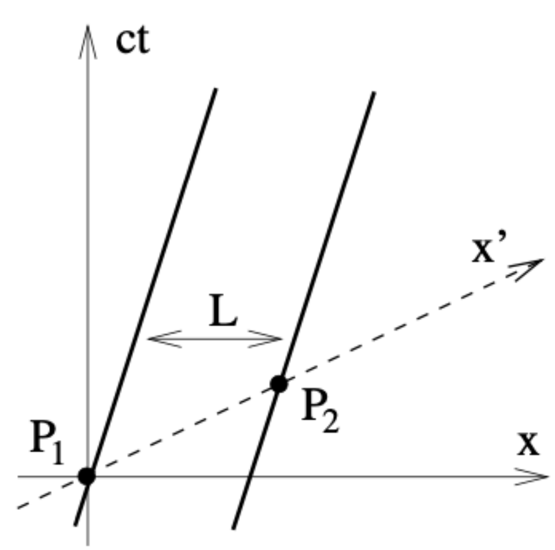
Consider a rod of length L' sitting stationary in the frame S'.

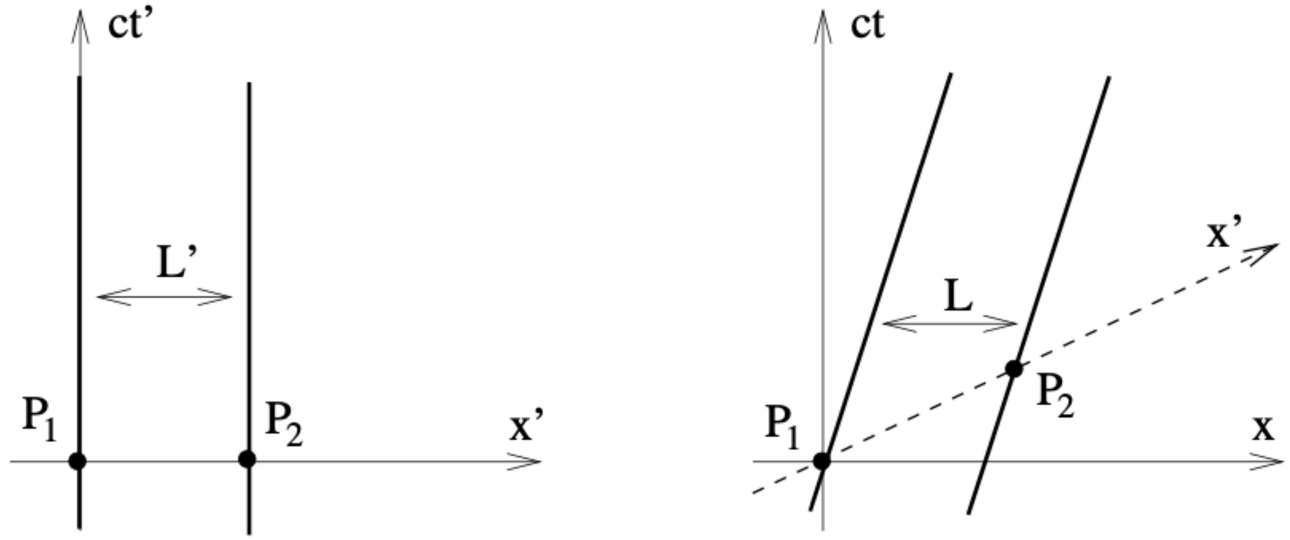
What is its length in frame S?

To begin, we should state more carefully something which seems obvious: when we say that a rod has length L', it means that the distance between the two end points at equal times is L'.

So, drawing the axes for the frame S', the situation looks like the picture on the left.







The two, simultaneous, end points in S' are P_1 and P_2.

Their coordinates in S' are $(ct', x') = (0, 0)$ and $(0, L')$ respectively.

Now let's look at this in frame S.

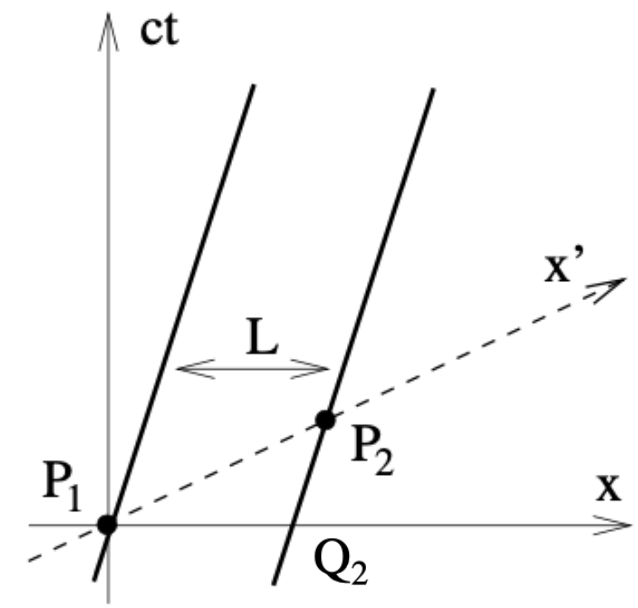
This is drawn in right-hand picture.

Clearly P_1 sits at $(ct, x) = (0, 0)$.

Meanwhile, the Lorentz transformation gives us the coordinate for P_2

$$x = \gamma L' \quad \text{and} \quad t = \frac{\gamma v L'}{c^2}$$

But to measure the rod in frame S, we want both ends to be at the same time



And the points P_1 and P_2 are not simultaneous in S.

We can follow the point P_2 backwards along the trajectory of the end point t
Q_2, which sits at

$$x = \gamma L' - vt$$

We want Q_2 to be simultaneous with P_1 in frame S.

This means we must move back a time $t = \gamma v L'/c^2$, giving

$$x = \gamma L' - \frac{\gamma v^2 L'}{c^2} = \frac{L'}{\gamma}$$

This is telling us that the length L measured in frame S is $$L = \frac{L'}{\gamma}$$

It is shorter than the length of the rod in its rest frame by a factor of γ.

This phenomenon is known as *Lorentz contraction*.

Putting Ladders in Barns

Take a ladder of length 2L and try to put it in a barn of length L.

If you run fast enough, can you squeeze it?

Here are two arguments, each giving the opposite conclusion

- From the perspective of the barn, the ladder contracts to a length $2L/\gamma$. This shows that it can happily fit inside as long as you run fast enough, with $\gamma \geq 2$
- From the perspective of the ladder, the barn has contracted to length L/γ. This means there's no way you're going to get the ladder inside the barn. Running faster will only make things worse

What's going on?

As usual, to reconcile these two points of view we need to think more carefully about the question we're asking.

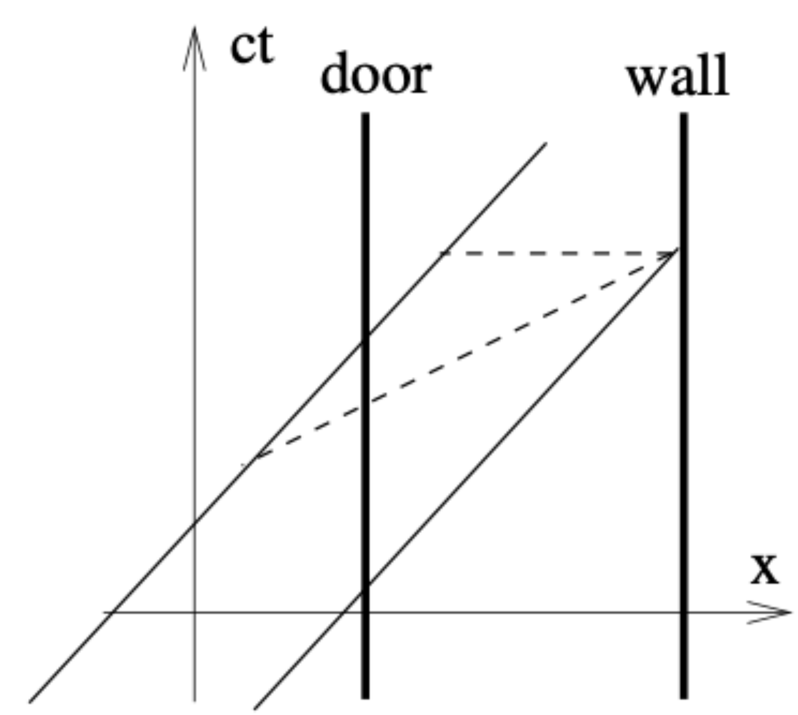
What does it mean to "fit a ladder inside a barn"?

Any observer will agree that we've achieved this if the back end gets in the door before the front end hits the far wall.

But we know that simultaneity of events is not fixed, so the word "before" in this definition suggests that it may be something different observers will disagree on.

Let's see how this works.

The spacetime diagram in the frame of the barn is drawn in the figure with $\gamma > 2$.



We see that, from the barn's perspective, both back and front ends of the ladder are happily inside the barn at the same time.

We've also drawn the line of simultaneity for the ladder's frame.

This shows that when the front of the ladder hits the far wall, the back end of the ladder has not yet got in the door.

Is the ladder in the barn? Well, it all depends who you ask.

Addition of Velocities

A particle moves with constant velocity u' in frame S' which, in turn, moves with constant velocity v with respect to frame S.

What is the velocity u of the particle as seen in S?

The Newtonian answer is just $u = u' + v$.

But we know that this can't be correct because it doesn't give the right answer when $u' = c$.

So what is the right answer?

The worldline of the particle in S' is

$$x' = u't' \tag{8}$$

So the velocity of the particle in frame S is given by

$$u = \frac{x}{t} = \frac{\gamma(x' + vt')}{\gamma(t' + vx'/c^2)}$$

which follows from the Lorentz transformations (6).

Actually, we've used the inverse Lorentz transformations since we want S coordinates in terms of S′ coordinates, but these differ only changing -v to v).

Substituting (8) into the expression above, and performing a little algebra, gives us the result we want:

$$u = \frac{u' + v}{1 + u'v/c^2}$$

Note that when u′ = c, this gives us u = c as expected.

We can also show that if |u′| < c and |v| < c then we necessarily have -c < u < c.

The proof is simple algebra, if a little fiddly

$$c - u = c - \frac{u' + v}{1 + u'v/c^2} = \frac{c(c - u')(c - v)}{c^2 + u'v} > 0$$

where the last equality follows because, by our initial assumptions, each factor in the final expression is positive.

An identical calculation will show you that $c < u$ as well.

We learn that if a particle is travelling slower than the speed of light in one inertial frame, it will also be travelling slower than light in all others.

The Geometry of Spacetime

> *The views of space and time which I wish to lay before you have sprung from the soil of experimental physics, and therein lies their strength. They are radical. 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.*
>
> Hermann Minkowski, 1908

We have seen that time is relative, length is relative, simultaneity is relative.

Is nothing sacred anymore?

Well, the answer is yes: there is one measurement that all observers will agree on.

The Invariant Interval

Let’s start by considering a spacetime with just a single spatial coordinate, x.

In frame S, two events P_1 and P_2 have coordinates (ct_1, x_1) and (ct_2, x_2).

The events are separated by $\Delta t = t_1 - t_2$ in time and $\Delta x = x_1 - x_2$ in space.

We define the *invariant interval* Δs^2 as a measure of the distance between these two points:

$$\Delta s^2 = c^2 \Delta t^2 - \Delta x^2$$

The advantage of the invariant interval is that it is something all observers agree upon.

In frame S', we have

$$\begin{aligned} \Delta s^2 &= \gamma^2 \left(c\Delta t' + \frac{v \Delta x'}{c} \right)^2 - \gamma^2 \left(\Delta x' + v \Delta t' \right)^2 \\ &= \gamma^2 (c^2 - v^2) \Delta t'^2 - \gamma^2 \left(1 - \frac{v^2}{c^2} \right) \Delta x'^2 \\ &= c^2 \Delta t'^2 - \Delta x'^2 \end{aligned} \tag{10}$$

where, in going from the first line to the second, we see that the cross-terms $t'x'$ cancel out.

Including all three spatial dimensions, the definition of the invariant interval is

$$\Delta s^2 = c^2\Delta t^2 - \Delta x^2 - \Delta y^2 - \Delta z^2 \tag{11}$$

which, again, is the same in all frames.

The only non-trivial part of the calculation is (10) above since y and z are invariant under a boost in the x direction).

The spacetime of special relativity is topologically R^4.

When endowed with the measure of distance (11), this spacetime is referred to as Minkowski space.

Although topologically equivalent to Euclidean space, distances are measured differently.

To stress the difference between the time and spatial directions, Minkowski space is sometimes said to have dimension d = 1 + 3. (For once, it's important that you don't do this sum!).

The invariant interval provides an observer-independent characterization of the distance between any two events.

However, it has a strange property: it is not positive definite.

Two events whose separation is $\Delta s^2 > 0$ are said to be timelike separated.

They are closer together in space than they are in time.

Pictorially, such events sit within each others light cone.

In contrast, events with $\Delta s^2 < 0$ are said to be spacelike separated.

They sit outside each others light cone.

From our earlier discussion, we know that two observers can disagree about the temporal ordering of spacelike separated events.

However, they agree on the ordering of timelike separated events.

Note that since $\Delta s^2 < 0$ for spacelike separated events, if you insist on talking about Δs itself then it must be purely imaginary.

However, usually it will be perfectly fine if we just talk about Δs^2.

Finally, two events with $\Delta s^2 = 0$ are said to be lightlike separated.

Notice that this is an important difference between the invariant interval and most measures of distance that you're used to.

Usually, if two points are separated by zero distance, then they are the same point.

This is not true in Minkowski spacetime: if two points are separated by zero distance, it means that they can be connected by a light ray.

A Rotational Analogy

There's a simple analogy to understand the meaning of the invariant interval.

Let's go back to consider three dimensional Euclidean space with coordinates $\mathbf{x} = (x, y, z)$.

An observer measures the position of a stationary object — let's say a helicopter — and proudly announces the x and y and z coordinates of the helicopter.

Meanwhile, a second observer shares the same origin as the first, but has rotated his axes to use coordinates $\mathbf{x}' = (x', y', z')$ where $\mathbf{x}' = \mathbf{R}\mathbf{x}$ for some rotation matrix $\mathbf{R}$ - tells us how to transform one frame into the other , i.e., to perform the rotation of coordinates.

He too sees the helicopter, and declares that it sits at coordinates x', y' and z'.

Of course, there's no reason why the coordinates of the two observers should agree with each other.

However, there is one quantity that should be invariant: the distance from the origin (which is shared by both observers) to the helicopter.

In other words, we should find that

$$s^2_{\text{Euclidean}} = x^2 + y^2 + z^2 = x'^2 + y'^2 + z'^2 \tag{12}$$

This leads to a property of the rotation matrix - it obeys $R^T R = 1$, i.e, it a *unitary* matrix.

The essence of special relativity is nothing more than an extrapolation of the discussion above.

The Lorentz boosts can be should be thought of as a rotation between space and time.

The individual spatial and temporal coordinates are different for the two observers, but there remains an invariant distance.

The only thing that's different is that the time and space directions in this invariant distance (11) come with different minus signs.

We'll now explore this relationship between Lorentz boosts and rotations in some detail.

The Lorentz Group

We have defined the interval (11) as the measure of distance which is invariant under Lorentz transformations.

However, it is actually better to look at things the other way: the invariant interval is the *primary* object.

This is a property of spacetime which defines the Lorentz transformations.

Let's see how the argument runs this way around.

If we sit at the origin in a fixed frame S, the coordinates of an event can be written as a *four vector* X.

We won't denote that this is a vector by bold font or squiggly underlines (which we're really saving for three-dimensional spatial vectors).

We're just getting sophisticated now and just the capital letter will have to suffice.

However, we will sometimes use index notation, in which the components of the 4-vector are

$$X^\mu = (ct, x, y, z) \qquad \mu = 0, 1, 2, 3$$

Note that we write the indices running from $\mu = 0$ to $\mu = 3$ rather than starting at 1.

The zeroth component of the vector is time (multiplied by c).

The invariant distance between the origin and the point P can be written as an inner product, defined as

$$X \cdot X \equiv X^T \eta X = X^\mu \eta_{\mu\nu} X^\nu \tag{13}$$

In the first expression above we are using matrix-vector notation and in the second we have resorted to index notation, with the summation convention(repeated indices are summed) for both indices μ and υ.

The matrix η is given by

$$\eta = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix}$$

This matrix is called the *Minkowski metric*.

With this expression for the Minkowski metric, the inner product becomes

$$X \cdot X = c^2t^2 - x^2 - y^2 - z^2$$

which is indeed the invariant distance (11) between the origin and the point X as promised.

Following our characterization of distances using the invariant interval, a four vector obeying $X \cdot X > 0$ is said to be *timelike*; one with $X \cdot X < 0$ is said to be *spacelike*; and one with $X \cdot X = 0$ is said to be *lightlike* or, alternatively, *null*.

The Lorentz transformation can also be be thought of as a 4 x 4 matrix Λ, rotating the coordinates in frame S to coordinates in frame S', such that the four vector becomes

$$X' = \Lambda X$$

This can also be written index notation as $X'^{\mu} = \Lambda^{\mu}{}_{\upsilon} X^{\upsilon}$.

The Lorentz transformations are defined to be those matrices which leave the inner product invariant.

This means that

$$X' \cdot X' = X \cdot X$$

From our definition (13), we see that this is true only if Λ obeys the matrix equation

$$\Lambda^T \eta \Lambda = \eta \tag{14}$$

Let's try to understand the solutions to this.

We can start by counting how many we expect.

The matrix Λ has 4 x 4 = 16 components.

Both sides of equation (14) are symmetric matrices (identical on both sides of the diagonal), which means that the equation only provides 10 constraints on the coefficients of Λ.

We therefore expect to find 16 - 10 = 6 independent solutions.

The solutions to (14) fall into two classes.

The first class is very familiar.

Let's look at solutions of the form

$$\Lambda = \left(\begin{array}{c|ccc} 1 & 0 & 0 & 0 \\ \hline 0 & & & \\ 0 & & R & \\ 0 & & & \end{array}\right) \tag{15}$$

where R is a 3 x 3 matrix.

These transformations change space, but leave time intact.

The condition (14) reduces to a condition for the matrix R,

$$R^T R = 1$$

where the right-hand side is understood to be the 3 x 3 unit matrix.

But this is something that we've seen before: it is the requirement for R to be a rotation matrix.

There are three such independent matrices, corresponding to rotations about the three different spatial axes.

The remaining three solutions to (14) are the Lorentz boosts that have preoccupied us for much of this discussion.

The boost along the x axis is given

$$\Lambda = \left(\begin{array}{cc|cc} \gamma & -\gamma v/c & 0 & 0 \\ -\gamma v/c & \gamma & 0 & 0 \\ \hline 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{array}\right) \tag{17}$$

These are precisely the Lorentz transformations (6).

Two further solutions to (14) come from boosting along the y and z directions.

The set of all matrices Λ obeying (14) form the Lorentz group, denoted O(1, 3).

You can easily check that they indeed obey all axioms of a group.

Taking the determinant of both sides of (14), we see that $\det\Lambda^2 = 1$, so the Lorentz group splits into two pieces with $\det\Lambda = \pm 1$.

The subgroup with $\det\Lambda = 1$ is called the proper Lorentz group and is denoted SO(1, 3).

Rapidity

We previously derived the velocity addition law (9).

Let's see how we get this from the matrix approach above.

We can focus on the 2 →ı 2 upper-left hand part of the matrix in (16). We'll write this as

$$\Lambda[v] = \begin{pmatrix} \gamma & -\gamma v/c \\ -\gamma v/c & \gamma \end{pmatrix}$$

If we combine two boosts, both in the x direction, the resulting Lorentz transformation is

$$\Lambda[v_1]\Lambda[v_2] = \begin{pmatrix} \gamma_1 & -\gamma_1 v_1/c \\ -\gamma_1 v_1/c & \gamma_1 \end{pmatrix} \begin{pmatrix} \gamma_2 & -\gamma_2 v_2/c \\ -\gamma_2 v_2/c & \gamma_2 \end{pmatrix}$$

It takes a little bit of algebra, but multiplying out these matrices you can show that

$$\Lambda[v_1]\,\Lambda[v_2] = \Lambda\left[\frac{v_1 + v_2}{1 + v_1 v_2/c^2}\right]$$

which is again the velocity addition rule (9), now for the composition of boosts.

The algebra involved in the above calculation is somewhat tedious; the result somewhat ugly.

Is there a better way to see how this works?

We can get a clue from the rotation matrices R.

Now the 2 x 2 matrix which rotates a plane by angle θ is

$$R[\theta] = \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix}$$

If we perform two rotations in succession, we have

$$R[\theta_1]R[\theta_2] = R[\theta_1 + \theta_2]$$

But the nice addition rule only worked because we were clever in parameterizing our rotation by an angle.

In the case of Lorentz boosts, there is a similarly clever parameterization.

Instead of using the velocity v, we define the *rapidity* φ by

$$\gamma = \cosh\varphi$$

We can see one of the nice things about this definition if we look at

$$\sinh\varphi = \sqrt{\cosh^2\varphi - 1} = \sqrt{\gamma^2 - 1} = \frac{v\gamma}{c}$$

This is the other component of the Lorentz boost matrix.

We can therefore write

$$\Lambda[\varphi] = \begin{pmatrix} \cosh\varphi & -\sinh\varphi \\ -\sinh\varphi & \cosh\varphi \end{pmatrix}$$

Looking again at the composition of two Lorentz boosts, we see that the rapidities add, just like the angles of rotation

$$\Lambda[\varphi_1]\,\Lambda[\varphi_2] = \Lambda[\varphi_1 + \varphi_2]$$

The matrix description of the Lorentz boost (17) shows most clearly the close relationship between rotations and boosts.

Relativistic Kinematics

So far, our discussion has been focussed on what the world looks like to different observers.

Let's now return to the main theme of these lectures: the motion of particles.

Remember that our ultimate goal is to construct laws of physics which look the same to all inertial observers.

For this reason, we will start by defining some of the basic elements that go into the laws of physics: velocity, momentum and acceleration.

We want to define these in such a way that they have nice transformation properties when viewed from different inertial frames.

Proper Time

One can describe the trajectory of particle in an inertial frame in terms of a curve $x(t)$ and velocity $u = dx/dt$.

There's nothing incorrect with this description in special relativity but, as we will see, there's a much better way to parameterize the trajectory of a particle.

Let's start by considering a particle at rest at the origin of frame S' with $x' = 0$.

The invariant interval between two different points on the worldline of the particle is

$$\Delta s^2 = c^2 \Delta t'^2$$

We see that the invariant interval between two points on the worldline is proportional to the time experienced by the particle.

But this must be true in all frames.

The time experienced by the particle is called the *proper time*, τ.

In all frames, it is given by

$$\Delta \tau = \frac{\Delta s}{c}$$

where Δs is real as long as the particle doesn't travel faster than the speed of light, so it sits on a timelike trajectory.

We keep promising to prove that a particle is unable to travel faster than light...we are almost there!

Proper time provides a way to parameterize the trajectory of a particle in a manner that all inertial observers will agree on.

Consider the trajectory of a general particle, not necessarily travelling in a straight line.

Viewed from an inertial frame S, the worldline can be parameterised by x(τ) and t(τ).

This has several advantages.

For example, we can use this formulation to determine the time experienced by a particle moving along a general trajectory.

Along a small segment of its trajectory, a particle experiences proper time

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

from which we have

$$\frac{dt}{d\tau} = \gamma \tag{18}$$

Note that here is a function of the speed, u, of the particle seen by the observer in S.

From this, the total time T experienced by a particle as it travels along its worldline is simply the sum of the proper times associated to each small segment,

$$T = \int d\tau = \int \frac{dt}{\gamma} \tag{19}$$

4-Velocity

We'll now explain why it's useful to parameterize the trajectory of a particle in terms of proper time τ.

We can write a general trajectory in spacetime using the 4-vector:

$$X(\tau) = \begin{pmatrix} ct(\tau) \\ \mathbf{x}(\tau) \end{pmatrix}$$

From this, we can define the 4-velocity,

$$U = \frac{dX}{d\tau} = \begin{pmatrix} c\,dt/d\tau \\ d\mathbf{x}/d\tau \end{pmatrix}$$

Using the relationship (18) between the proper time of the particle τ and the observer's time t we can write this as

$$U = \frac{dt}{d\tau}\begin{pmatrix} c \\ \mathbf{u} \end{pmatrix} = \gamma \begin{pmatrix} c \\ \mathbf{u} \end{pmatrix} \tag{20}$$

where $\mathbf{u} = d\mathbf{x}/dt$.

This definition of the 4-velocity has a nice property: if an observer in frame S measures a particle's 4-velocity as U, then an observer in frame S' with $X' = \Lambda X$ will measure the 4-velocity

$$U' = \Lambda U \tag{21}$$

This transformation holds only because $d\tau$ is invariant, meaning that it is the same for all observers.

In contrast, if we had tried to define a 4-velocity by, say, $V = dX/dt$ then both X and t would change under a Lorentz transformation and we would be left with a messy, complicated expression for V in frame S'.

Our definition of U differs from V by the extra factor of in (20).

This is all important!

We now have two objects which transform nicely under Lorentz transformations: the coordinates X $\rightarrow \Lambda$X and the 4-velocity U $\rightarrow \Lambda$U.

Quantities like this are called 4-vectors.

It's a name that we've already used to label points in spacetime.

More generally, the *definition* of a 4-vector is any 4-component object A which transforms as A $\rightarrow \Lambda$A under a Lorentz transformation.

Because of the simple transformation law (21), we can immediately import some of the things that we learned from our previous discussion of Lorentz groups.

In particular, from the definition of Λ given in (14), we know that the inner product

$$U \cdot U = U^T \eta U$$

is invariant.

It is the same for all observers: $U \cdot U = U' \cdot U'$.

Let's look at a simple example.

A particle which is stationary in frame S has 4-velocity $$U^\mu = (c, 0, 0, 0)$$

and so $U \cdot U = c^2$.

But this must be true in all frames.

We can check this explicitly from (20) (we'll take the middle equation to illustrate the point) which gives us

$$U \cdot U = \left(\frac{dt}{d\tau}\right)^2 (c^2 - u^2) = \left(\frac{dt}{d\tau}\right)^2 \frac{c^2}{\gamma^2} = c^2$$

This result also helps answer a puzzle.

In Newtonian mechanics, if we want to specify the velocity, we only have to give three numbers u.

But in special relativity, the velocity is a 4-vector U.

Nonetheless, we still only need specify three variables because U is not any 4-vector: it is constrained to obey $U \cdot U = c^2$.

Addition of Velocities Revisited

Earlier, we derived the rule for the addition of velocities in one-dimension.

But what if the velocity of a particle is not aligned with the relative velocity between S and S$'$?

The addition of velocities in this case is simple to compute using 4-vectors.

We start with a particle in frame S travelling with 4-velocity

$$U = \begin{pmatrix} \gamma_u c \\ u\gamma_u \cos\alpha \\ u\gamma_u \sin\alpha \\ 0 \end{pmatrix}$$

Here we've added the subscript to $\gamma_u = (1 - u^2/c^2)^{1/2}$ to distinguish it from the γ factor arising between the two frames.

Frame S$'$ moves in the x-direction with speed v relative to S.

The Lorentz boost is given in (16).

In frame S$'$, the 4-velocity is then

$$U' = \Lambda U = \gamma_u \begin{pmatrix} \left(1 - (uv/c^2)\cos\alpha\right)\gamma_v c \\ (u\cos\alpha - v)\gamma_v \\ u\sin\alpha \\ 0 \end{pmatrix} \equiv \begin{pmatrix} \gamma_{u'} c \\ u'\gamma_{u'}\cos\alpha' \\ u'\gamma_{u'}\sin\alpha' \\ 0 \end{pmatrix} \tag{22}$$

Dividing the t and x components of this 4-vector, we recover the velocity transformation law (9) for the speed in the x-direction, namely

$$u'\cos\alpha' = \frac{u\cos\alpha - v}{1 - uv\cos\alpha/c^2}$$

Meanwhile, dividing the y component by the x component gives us a formula for the angle α' that the particles trajectory makes with the x$'$-axis,

$$\tan\alpha' = \frac{u\sin\alpha}{\gamma_v(u\cos\alpha - v)} \tag{23}$$

4-Momentum

The *4-momentum* is defined by

$$P = mU = \begin{pmatrix} mc\gamma \\ m\gamma\mathbf{u} \end{pmatrix} \tag{24}$$

where m is the mass of the particle, usually referred to as the *rest mass*.

Importantly, it will turn out that P is the quantity that is conserved in the relativistic context.

The spatial components give us the relativistic generalization of the 3-momentum,

$$\mathbf{p} = m\gamma\mathbf{u} \tag{25}$$

Notice that as the particle approaches the speed of light, $u \rightarrow c$, the momentum diverges $p \rightarrow \infty$.

Since momentum is conserved in all processes, this is really telling us that massive particles cannot break the speed of light barrier.

Here the word “massive” doesn’t mean “really really big”: it just means “not massless”, or $m \neq 0$).

This is sometimes interpreted by viewing the quantity m as a velocity-dependent relativistic mass.

In these terms, the relativistic mass of the particle diverges $m\gamma \to \infty$ as the particle approaches the speed of light.

The words may be different, but the mathematics (and underlying physics) is the same: particles are bound by Nature's speed limit.

Nothing can travel faster than the speed of light.

What is the interpretation of the time-component of the momentum 4-vector, P^0.

We can get a hint of this by Taylor expanding the γ factor,

$$P^0 = \frac{mc}{\sqrt{1-u^2/c^2}} = \frac{1}{c}\left(mc^2 + \frac{1}{2}mu^2 + \dots\right) \tag{26}$$

The first term is just a constant.

But the second term is something familiar: it is the non-relativistic kinetic energy of the particle.

This, coupled with the fact that all four components of P are conserved, strongly suggests that the right interpretation of P^0 is the energy of the particle (divided by c), so

$$P = \begin{pmatrix} E/c \\ \mathbf{p} \end{pmatrix} \tag{27}$$

To show that P^0 is indeed related to the energy in this way requires a few more techniques than we will develop in this course.

The expansion of (26) shows that both the mass and the kinetic energy contribute to the energy of a particle.

These combine to give

$$E = m\gamma c^2 \tag{28}$$

Notice that as the particle approaches the speed of light, its energy diverges.

Yet again, we see a barrier to breaking the speed limit: as we approach the speed of light, the energy required to make a particle go just a little faster gets bigger and bigger.

For a stationary particle, all its energy is contained in its rest mass, giving us the famous slogan

$$E = mc^2$$

There's a nice way to rearrange (28), to replace the u in the factor with p defined in (25).

But the algebra is laborious.

Instead there's a cute trick that gives the result much more quickly: we look at the inner product $P \cdot P$.

In the rest frame of the particle, $P = (mc,0,0,0)$ and we have

$$P \cdot P = m^2c^2 \quad (29)$$

But the inner product is an invariant, holding in any frame.

From (27), we have

$$P \cdot P = \frac{E^2}{c^2} - \mathbf{p}^2$$

Equating these two expressions gives

$$E^2 = \mathbf{p}^2 c^2 + m^2 c^4 \tag{30}$$

This is the generalization of $E = mc^2$ to include the kinetic energy.

This equation can also be derived the hard way by playing around with (28) and (25).

The identification $P^0 = E/c$ has dramatic consequences.

In Newtonian mechanics, we boasted about the conservation of energy, but implicit in everything we did was the more elementary fact that mass is conserved.

Relativity teaches us that the conservation of mass is subsumed into the conservation of energy.

There is nothing that guarantees that they are individually conserved.

Just as potential energy can be converted into kinetic energy, so too can mass be converted into kinetic energy.

In Japan, in 1945, this fact was vividly demonstrated.

Massless Particles

Until now, we built our discussion of particle trajectories on proper time.

But, looking back earlier work, proper time is only defined for time-like trajectories.

This is fine for massive particles.

But what about for massless particles?

We can sidestep the need for proper time by looking at the invariant of the 4-momentum (29) which, for particles with m = 0, tells us that the 4-momentum must be null,

$$P \cdot P = 0$$

This means that the 4-momentum of a massless particle necessarily lies along a light ray.

This fact also allows us to clarify one of our original postulates of special relativity: that the speed of light is the same for all inertial frames.

You may wonder why the propagation of light, an electromagnetic phenomenon, is singled out for special treatment.

The answer is: because the photon – the particle of light – is massless.

In fact, a better way of stating the postulate is to say that there is an upper speed limit in the Universe, which is the same for all inertial observers.

Any massless particle must travel at this speed limit.

All massive particles must go slower.

We know of only two types of massless particles in the Universe: the photon and the graviton.

Both of these owe their particle-like nature to quantum mechanics (actually, this is true of all particles) and have a classical analog as light waves and gravity waves respectively.

You've all seen light waves (literally!) and individual photons have been routinely measured in experiments for more than a century.

Gravitational waves were observed for the first time in 2015, although compelling indirect evidence had existed for decades.

There appears to be no hope at all of detecting an individual graviton, at least within our lifetimes.

Until the late 1990s, it was thought that neutrinos were also massless. It is now known that they have a small, but finite mass.

Actually, there's a caveat here: there are three different types of neutrino: an electron neutrino, a muon neutrino and a tau neutrino.

The differences between their masses are known to be of order of 0.01 - 0.1 eV and there are constraints which limit the sum of their masses to be no greater than 0.3 eV or so.

But the absolute scale of their masses has not yet been determined.

In principle, one of the three neutrinos may be massless).

From (30), the energy and momentum of a massless particle are related by $E^2 = p^2c^2$.

The four momentum takes the form

$$P = \frac{E}{c}\begin{pmatrix} 1 \\ \hat{\mathbf{p}} \end{pmatrix}$$

where $\hat{\mathbf{p}}$ is a unit vector in the direction of the particle's motion.

To get an expression for the energy, we need a result from quantum mechanics which relates the energy to the wavelength of the photon or, equivalently, to the angular frequency $\omega = 2\pi c/\lambda$,

$$E = \hbar\omega = \frac{2\pi\hbar c}{\lambda}$$

There's something rather nice about how this equation ties in with special relativity.

Suppose that in your frame, the photon has energy E.

But a different observer moves towards the light with velocity v.

By the Lorentz transformation, he will measure the 4-momentum of the photon to be $P' = \Lambda P$ and, correspondingly, will see a bigger energy $E' > E$.

From the above equation, this implies that he will see a smaller wavelength.

But this is nothing other than Lorentz contraction.

The phenomenon of different observers observing different wavelengths of light is called the *Doppler effect*.

Tachyons and Why They're Nonsense

It is sometimes stated that a particle which has imaginary mass, so that $m^2 < 0$, will have $P \cdot P < 0$ and so travel consistently at speeds $u > c$.

Such particles are called *tachyons*.

They too would be unable to cross Nature's barrier at $u = c$ and are consigned to always travel on spacelike trajectories.

Although, consistent within the framework of classical relativistic particle mechanics, the possibility of tachyons does not survive the leap to more sophisticated theories of physics.

All our current best theories of physics are written in the framework of quantum field theory.

Here particles emerge as ripples of fields, tied into small lumps of energy by quantum mechanics.

But in quantum field theory, it is not unusual to have fields with imaginary mass $m^2 < 0$.

The resulting particles do not travel faster than the speed of light.

Instead, imaginary mass signals an instability of the vacuum.

Particle Physics

Our goal in this section is to describe various relativistic phenomena that arise in particle physics.

All these processes occur in the absence of external forces, so $F = 0$ and we will rely only on conservation of 4-momentum, meaning

$$\frac{dP}{d\tau} = 0$$

Of course, conservation of 4-momentum includes both conservation of 3-momentum and conservation of energy.

Firstly, we need to choose a frame of reference in which to calculate: the smart frame to choose is nearly always the center of mass of the system.

It should more correctly be called the center of momentum frame, for it is the one with vanishing spatial 3-momentum.

Second, you will often be presented with a situation where there is one particle with momentum P about which you know nothing.

A good way to eliminate this is often to rearrange your equation so it takes the form P = . . . and then square it to get the right-hand side to be $P \cdot P = m^2c^2$.

Let's now see how this works in a few examples.

Particle Decay

Consider a single particle with rest mass m_1 which decays into two particles with rest masses m_2 and m_3.

Conservation of 4-momentum tells us

$$P_1 = P_2 + P_3$$

or, equivalently,

$$E_1 = E_2 + E_3 \quad \text{and} \quad \mathbf{p}_1 = \mathbf{p}_2 + \mathbf{p}_3$$

In the rest frame of the decaying particle, we can write (using (30)),

$$E_1 = m_1 c^2 = \sqrt{p_2^2 c^2 + m_2^2 c^4} + \sqrt{p_3^2 c^2 + m_3^2 c^4} \geq m_2 c^2 + m_3 c^2$$

which tells us the unsurprising result that a particle can only decay if its mass is greater than that of its decay products.

Some algebra gives the velocities v_2 and v_3 of the decay products in the center of mass frame

$$\gamma_2 = \frac{m_1^2 + m_2^2 - m_3^2}{2 m_1 m_2} \quad \text{and} \quad \gamma_3 = \frac{m_1^2 + m_3^2 - m_2^2}{2 m_1 m_3}$$

Now we will look at some slightly different problems.

An Example: Higgs Decay

The LHC has taught us that the Higgs boson has mass $m_h c^2 \approx 125$ GeV .

It mostly decays into two photons.

In particle physics, photons are always denoted by γ.

Do not confuse them with the the factor in the Lorentz transformations!

The “equations” in which the photon γ’s appear are more like chemical reactions than true equations.

The decay of the Higgs into two photons is written as

$$h \to \gamma\gamma$$

Similar decays occur for other particles, most notably the neutral pion, a meson (meaning that it is made of a quark and anti-quark) with mass $m_\pi c^2 \approx 140$ M eV .

This too decays as $\pi^0 \to \gamma\gamma$.

To be concrete (and more relevant!) we'll focus on the Higgs.

Conservation of 4-momentum tells us (in, hopefully, obvious notation) that

$$P_h = P_\gamma + P'_\gamma$$

If we sit in the rest frame of the Higgs(or the center of 3-momentum frame because the total 3-momentum = 0), so $P_h^\mu = (m_h c, 0)$, the photons must have equal and opposite(must add up to 0) 3-momentum, and therefore equal energy $E_\gamma = \frac{1}{2} m_h c^2$.

The photons must be emitted back-to-back but, because the problem is rotationally symmetric, can be emitted at any angle.

What if we're not sitting in the rest frame of the Higgs?

Suppose that the Higgs has energy E_h and the energy of one of the photons is measured to be E_γ.

What is the angle θ that this photon makes with the path of the Higgs?

We have no information about the second photon, with 4-momentum P'_γ.

So we rearrange the conservation of momentum to read $P'_\gamma = P_h - P'_\gamma$.

Upon squaring this, we have $P'_\gamma \cdot P'_\gamma = 0$, so

$$
\begin{aligned}
0 = (P_h - P_\gamma) \cdot (P_h - P_\gamma) &= P_h \cdot P_h + P_\gamma \cdot P_\gamma - 2P_h \cdot P_\gamma \\
&= m_h^2 c^2 - \frac{2E_h E_\gamma}{c^2} + 2\mathbf{p}_h \cdot \mathbf{p}_\gamma \\
&= m_h^2 c^2 - \frac{2E_h E_\gamma}{c^2} + \frac{2E_\gamma}{c} \cos\theta \sqrt{E_h^2/c^2 - m_h^2 c^2}
\end{aligned}
$$

where, in the last equation, we have used $E^2 = p^2c^2 + m^2c^4$ (which is just $E = pc$ for the photon).

This can now be rearranged to give the answer for θ.

Particle Collisions

Let's now look at the physics of relativistic collisions.

We'll collide two particles together, both of mass m.

They will interact in some manner, preserving both energy and 3-momentum, and scatter at an angle θ.

$$P_1 + P_2 = P_3 + P_4$$

As we mentioned previously, it's easiest to see what happens in the center of mass frame.

Without loss of generality, we'll take the initial momenta to be in the x-direction.

After the collision, the particles must have equal and opposite momenta, which means they must also have equal energy.

This, in turn, ensures that in the centre of mass frame, the speed v after the collision is the same as before.

We can choose our axes so that the initial and final momenta are given by

$$P_1^\mu = (mc\gamma_v, mv\gamma_v, 0, 0) \quad , \quad P_2^\mu = (mc\gamma_v, -mv\gamma_v, 0, 0)$$
$$P_3^\mu = (mc\gamma_v, mv\gamma_v\cos\theta, mv\gamma_v\sin\theta, 0) \quad , \quad P_4^\mu = (mc\gamma_v, -mv\gamma_v\cos\theta, -mv\gamma_v\sin\theta, 0)$$

where we've put the subscript on γ_v to denote its argument.

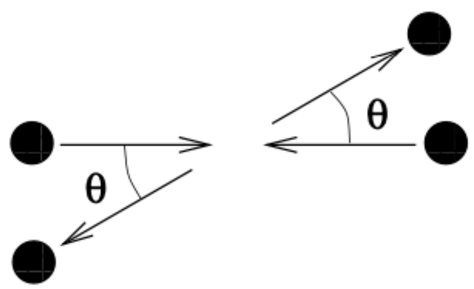
We can also look at the same collision in the lab frame.

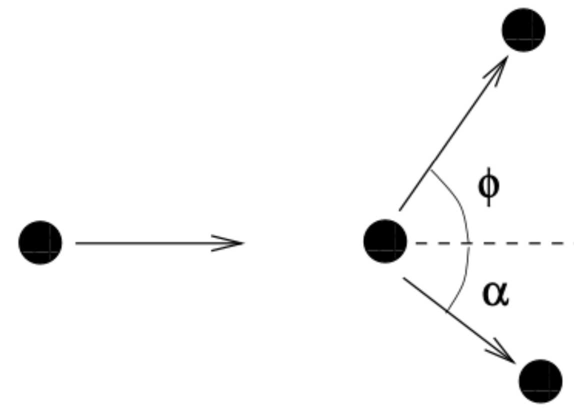
This refers to the situation where one of the particles is initially at rest (Presumably in your lab).

By the velocity addition formula, the other particle must start with speed

$$u = \frac{2v}{1 + v^2/c^2}$$

You can also derive this result by writing down the momenta P'_1 and P'_2 in the lab frame and equating $(P_1 + P_2)^2 = (P'_1 + P'_2)^2$





Collisions in the center of mass frame on the left and the lab frame on the right

In the lab frame, the angles ϕ and α at which the particles scatter are not equal.

They can be easily determined using the addition of 4-velocities that we saw earlier .

Set u = -v in equation (23) and use the identity tan(x/2) = sin x/(1 + cos x) to get

$$\tan\phi = \frac{1}{\gamma_v}\tan\theta/2 \quad \text{and} \qquad \tan\alpha = \frac{1}{\gamma_v}\tan(\theta/2+\pi/2)$$

One of the more interesting examples of collisions is Compton Scattering, in which the color of light changes after scattering off an electron (because it changes its energy and therefore its frequency).

Particle Creation

Just as mass can be converted into kinetic energy, so kinetic energy can be converted into mass through the creation of new particles.

Roughly speaking, this is the way we discover new particles of Nature.

Suppose we collide two particles, each of mass m.

After the collision, we hope to be left with these two particles, together with a third of mass M.

How fast must the original two particles collide?

Conservation of momentum gives us

$$P_1 + P_2 = P_3 + P_4 + P_5$$

where $P^2_1 = P^2_2 = P^2_3 = P^2_4 = m^2c^2$, while $P^2_5 = M^2c^2$.

Let's work in the center of mass frame of the colliding particles, each of which has speed v.

In this case, we have

$$(P_1 + P_2)^2 = 4m^2\gamma_v^2 c^2 = (P_3 + P_4 + P_5)^2 \tag{31}$$

Since we are in the center of mass frame, the final momenta must take the form $P_3 + P_4 + P_5 = ((E_1 + E_2 + E_3)/c, 0)$ so that

$$(P_3 + P_4 + P_5)^2 = \frac{1}{c^2}(E_1 + E_2 + E_3)^2 \geq \frac{1}{c^2}(2mc^2 + Mc^2)^2$$

where, for each particle, we've used the fact that $E = \sqrt{m^2c^4 + p^2c^2} \geq mc^2$.

Substituting this into (31) gives

$$4m^2\gamma_v^2 c^2 \geq 4m^2c^2 + M^2c^2 + 4Mmc^2 \quad \Rightarrow \quad \gamma_v \geq 1 + \frac{M}{2m} \tag{32}$$

This makes sense.

The amount of minimum amount of kinetic energy per particle is $T = \gamma_v mc^2 - mc^2 = \frac{1}{2}Mc^2$.

With this minimum amount, the two colliding particles can combine their kinetic energies to form the new particle.

After the collision, all three particles are then at rest.

What if we re-do this experiment in the lab frame, in which of the original particles is at rest and the other has speed u?

Now we have $P_1 = (m\gamma_u c, m\gamma_u u)$ and $P_2 = (mc, 0)$, so

$$(P_1 + P_2)^2 = P_1^2 + P_2^2 + 2P_1 \cdot P_2 = 2m^2c^2 + 2m^2\gamma_u c^2$$

But we don't have to compute $(P_3 + P_4 + P_5)^2$ again since the beauty of taking the square of the 4-momenta is that the result is frame independent.

We have

$$2m^2c^2 + 2m^2\gamma_u c^2 \geq 4m^2c^2 + M^2c^2 + 4Mmc^2 \quad \Rightarrow \quad \gamma_u \geq 1 + \frac{2M}{m} + \frac{M^2}{2m^2}$$

It's certainly not enough to give the incoming particle kinetic energy $T = ½Mc^2$ as one might intuitively expect.

Instead, if you want to create very heavy particles, $M >> m$, you need to give your initial particle a kinetic energy of order $T \approx ½M^2c^2/2m$.

This scales quadratically with M, rather than the linear scaling that we saw in the center of mass frame.

The reason for this simple: there's no way that the end products can be at rest.

The need to conserve momentum means that much of the kinetic energy of the incoming particle goes into producing kinetic energy of the outgoing particles.

This is the reason that most particle accelerators have two colliding beams rather than a single beam and a stationary target.

The LHC primarily collides protons in its search to discover new elementary particles.

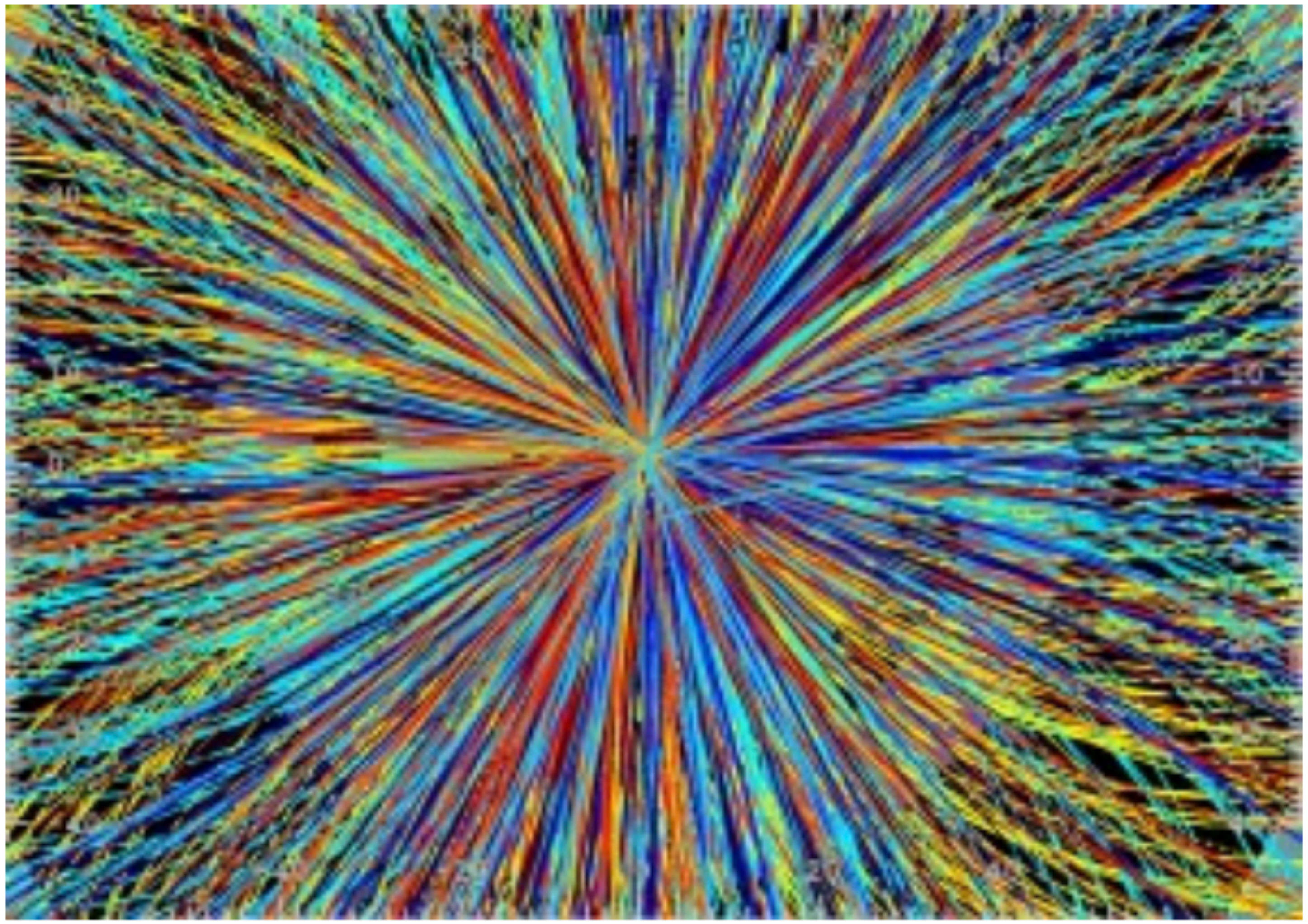
However, for one month a year, it switches to collisions of lead nuclei in an attempt to understand a new form of matter known as the quark-gluon plasma.

Each lead nuclei contains around 200 protons and neutrons.

The collision results in a dramatic demonstration of particle creation, with the the production of many thousands of particles – protons, neutrons, mesons and baryons.

Here's a very pretty picture.

It's one of the first collisions of lead nuclei at LHC in 2010, shown here in all its glory by the ALICE detector.



To end the class, I am going to discuss several potentially interesting relativistic phenomena.

Photon Rockets

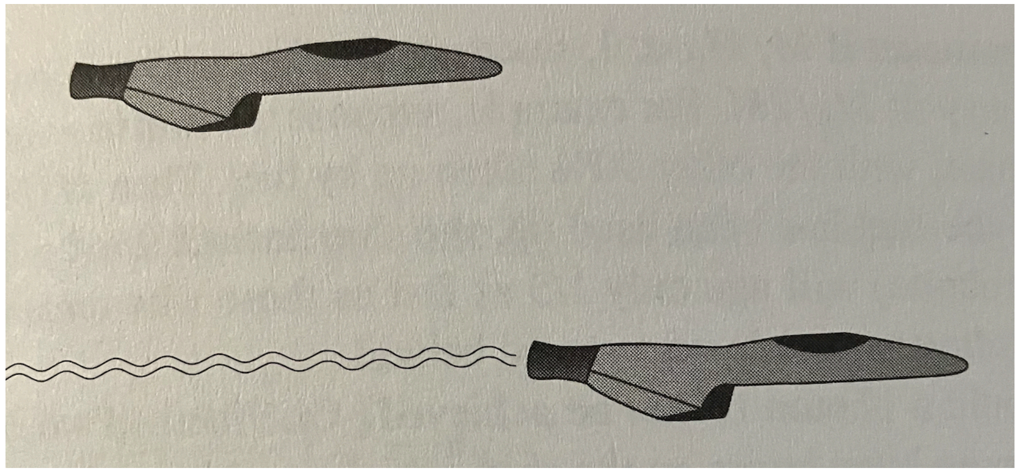
A spaceship of mass M_0 is at rest in some initial frame in outer space.

It suddenly starts shooting out photons from an exhaust nozzle at the rear of the ship.

The photons have momentum, so the spaceship recoils in the forward direction, becoming a *photon rocket*.

It is a particularly efficient rocket, because the photon propellent moves as fast as any propellant can.

After some period of time, the spaceship has ejected a stream of photons with total energy Ephotons in the spaceship's initial rest frame, and the spaceship's mass has been reduced to M, as shown in the Figure.



Conserving energy and momentum for the spaceship and thr stream of photons, we have

$$M_0c^2 = E_{photons} + E_{ship}$$

$$0 = \mathbf{p}_{photons} + \mathbf{p}_{ship}$$

These are identical to the equations we had we we discussed particle decay earlier.

The big difference is that now we must include the entire stream of photons.

We thus have

$$E = \frac{(M_0^2 + M^2)c^2}{2M_0}$$

for the energy of the ship any time after the ignition and therefore we have

$$\gamma = \frac{1}{\sqrt{1 - v^2/c^2}} = \frac{E}{Mc^2} = \frac{(M_0^2 + M^2)c^2}{2M_0Mc^2} = \frac{1 + (M/M_0)^2}{2(M/M_0)}$$

for the ship's γ-factor (Note that M/M_0 at ignition corresponds to $\gamma = 1$, since the ship is then at rest.)

When we design an interstellar spaceship, we will want it to be highly relativistic, meaning that reach large values of γ.

Based on the last equation, this will happen for a photon is $M/M_0 << 1$, since we can then neglect the term $(M/M_0)^2$ in the numerator, so $\gamma \simeq M_0/2M$.

How could a photon rocket be achieved?

One method would be to use onboard fuel to pump up huge lasers on the ship, which would emit beams of photons out the exhaust nozzle.

It is likely that such a rocket could be built, but it is much less likely that it could achieve relativistic speeds, because it is not terribly efficient.

Another method would be to carry separate(!) matter and antimatter onboard fuel chambers.

The matter and antimatter particles would be allowed to annihilate one another at a controlled rate, and since the annihilation would be largely into photons, the photons could be made to emerge from the exhaust.

Needless to say , such a rocket would be quite difficult to build and use.

You would not want to be on the ship if there was any onboard fuel confinement failure!

The Appearance of Moving Objects

This discussion concerns a so-far unobserved optical illusion predicted by special relativity.

Let us consider the type of measurement that is being made when a single observer watches a moving object, or when a snapshot is taken.

A snapshot of a moving train won't necessarily record its "true" length, since photons that arrive at the camera when the picture is taken will generally have left different parts of the train at different times.

Thus a snapshot made of a train traveling on a track running past us will make the train appear longer during its approach and shorter during its departure.

This effect is, however, just a consequence of the finite speed of light, and should not be blamed on the theory of relativity.

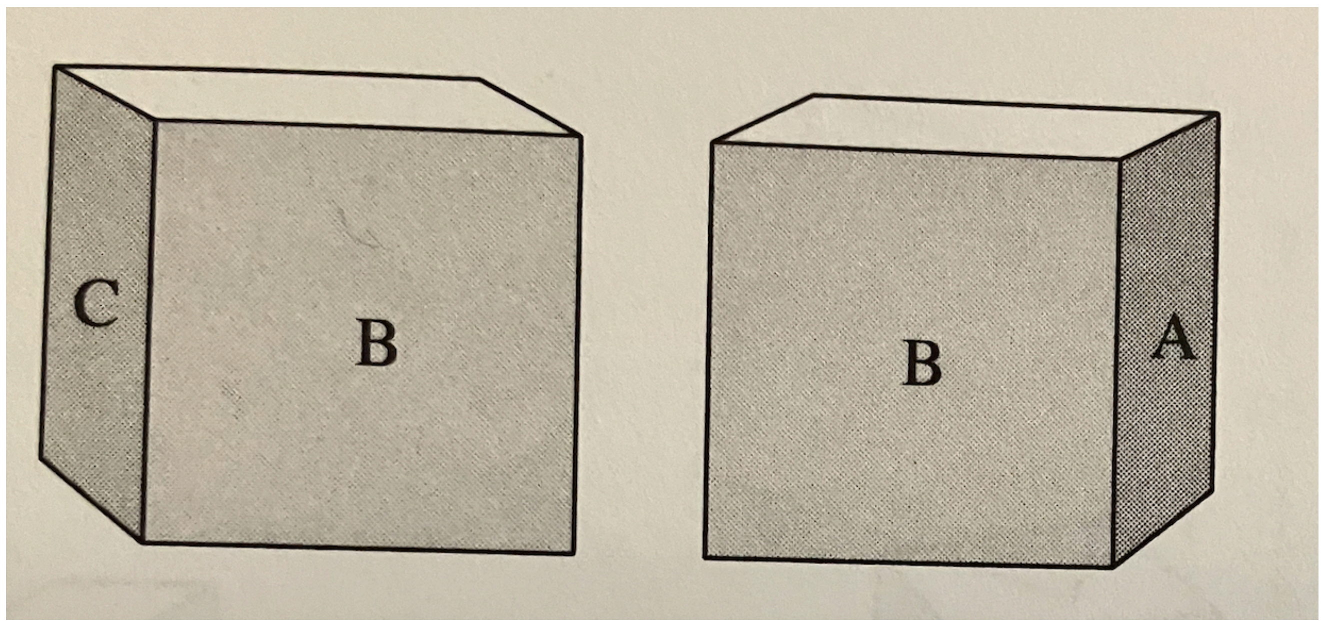
A detailed study of how a moving 3-dimensional object would appear to a single observer (or camera) was carried out by J.L. Terrell.

We have thought of a train as a 1-dimensional object having an apparent length depending on its speed and position, but the visual appearance of 3-dimensional objects is more subtle.

The result is somewhat surprising.

It turns out if an observer watches a distant moving object, that object will appear to be simple rotated.

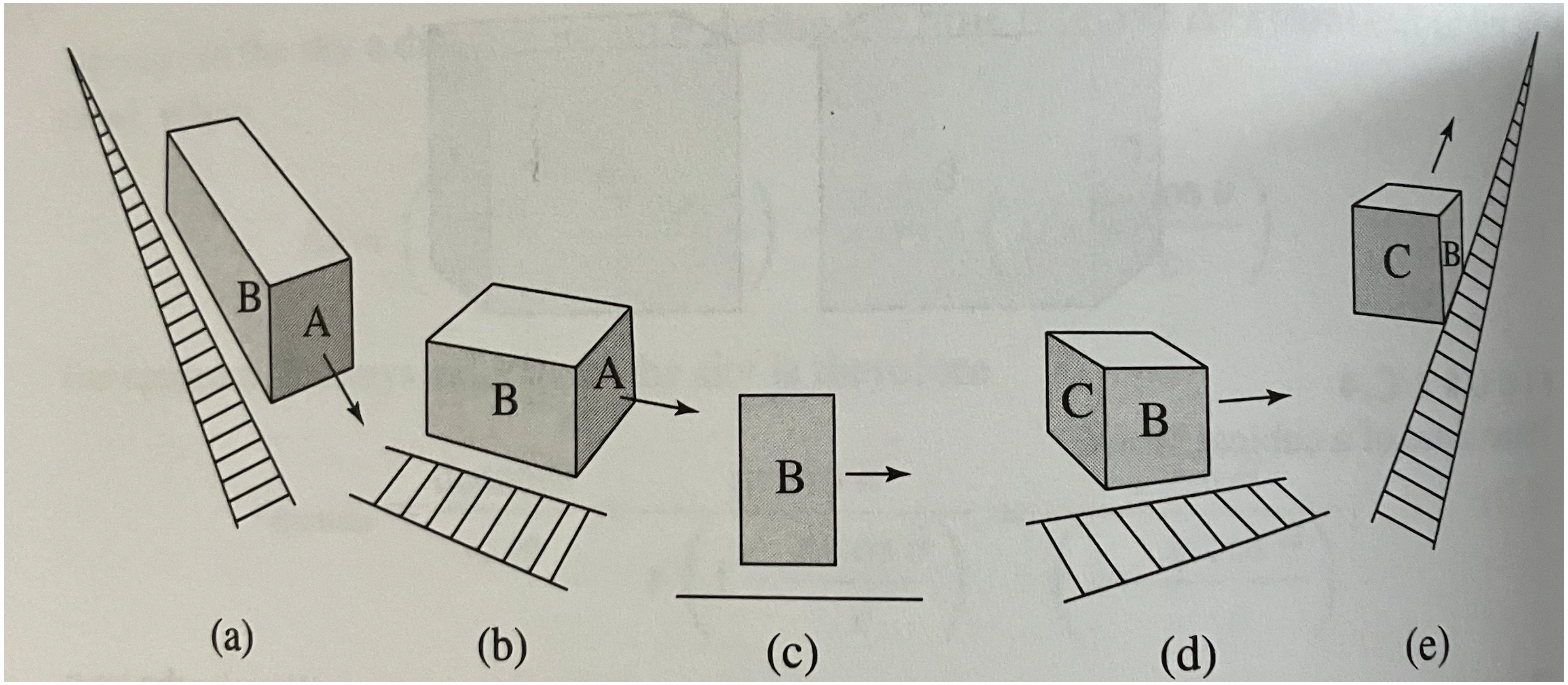
A conventional 3-dimensional object to look at is a cubical block of wood with lettered sides for the purpose of identification, as shown in the Figure.



Two views of the block are shown.

We suppose this block has unit edge length, and that it moves past us left to right at high speed, so that side B is parallel to the track and faces us as the block goes by.

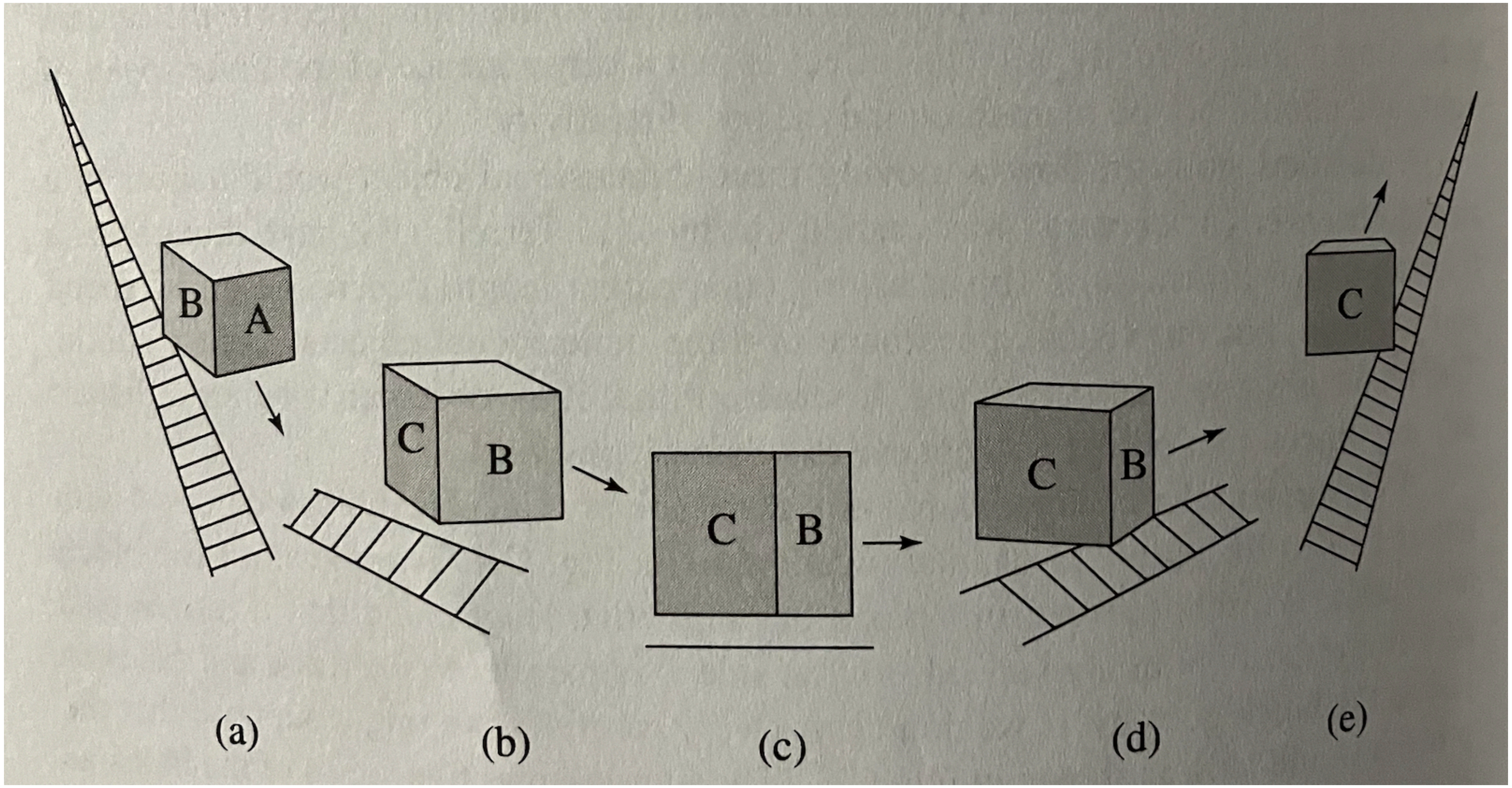
If we don't think very carefully, we might suppose that process would look as shown in the next Figure.



The figure shows 5 views of the block as it approaches, passes and recedes.

The block is shown stretched out as it approaches, Lorentz-contracted as it passes by, and compressed as it leaves.

In fact, the block would *actually* look as shown in the next Figure.



It does *not* appear to be distorted, but rather it looks like an ordinary cubical block that has been rotated about a vertical axis.

As it approaches, it suffers an apparent partial twist, so that somewhere between the first and second views we see on side B.

Then in the second view we have entirely lost side A, which is supposed to be toward us, and are beginning to see side C, which is supposed to be away from us.

This twist persists as the block passes by, but becomes less pronounced as the block departs.

Why does the block appear to be twisted?

The basic idea is that if the block is moving very fast, the block can get out of the way of the light from side C as long as the light moves at a sufficiently large angle with respect to the block's direction of motion.

Similarly, the block tends to run into the light from side A so that light never escapes from the block by the time the block reaches the second position in the last Figure.

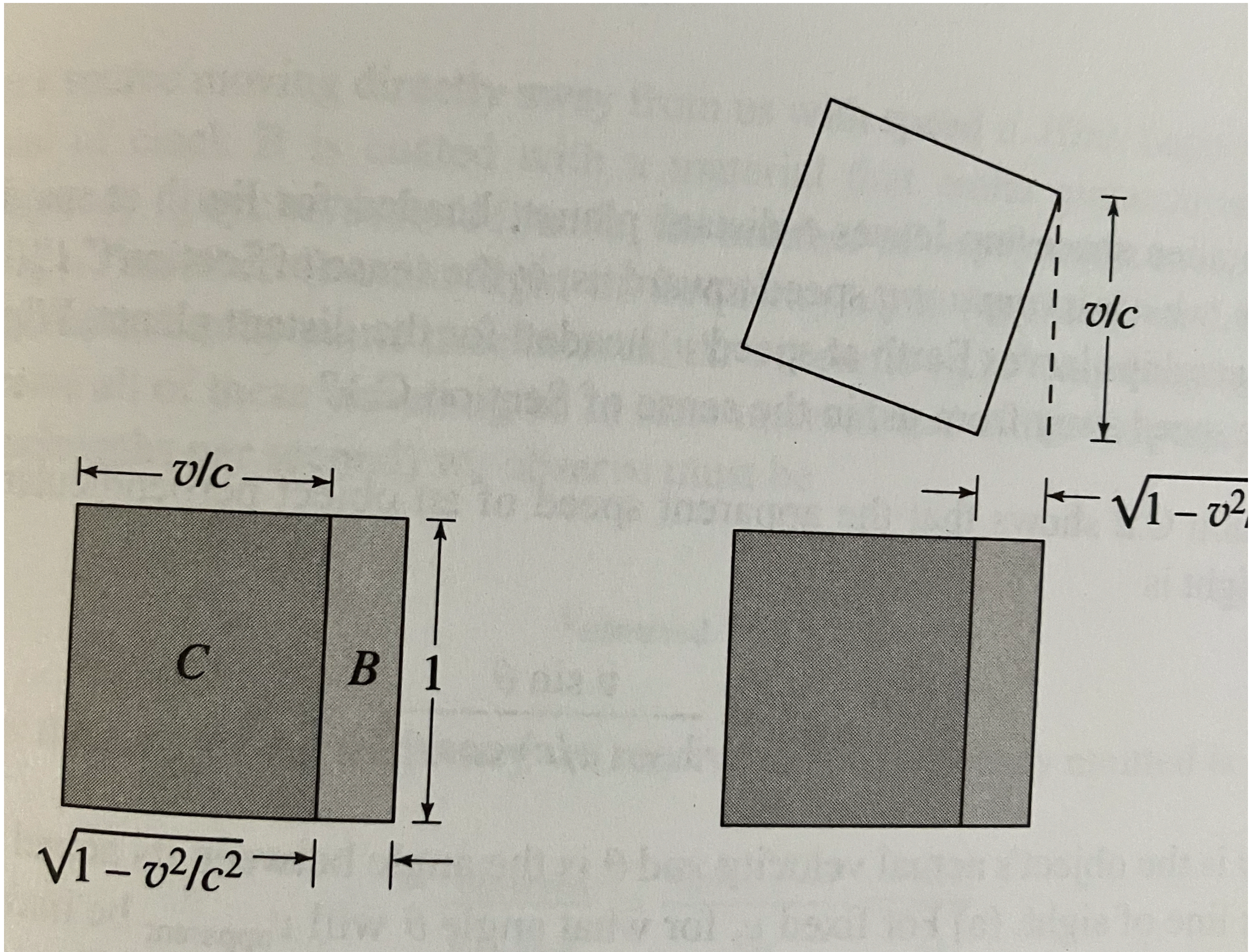
The reason why the visual appearance of a moving block is identical to that of a rotated block is mostly understood when the block is at its closest position to us.

Normally in this position we would expect to see only side B, contracted by the factor $\sqrt{1-v^2/c^2}$

In fact, side C can also be seen in this position because the block moves out of the way of the light being emitted from that side, allowing us to see it.

The light from the back edge of the side C had to be emitted earlier than the light from the front edge in order for it to reach our eye at the same time.

The next Figure shows the apparent width of side C is v/c.



(a)
Moving block ($V = 0.9c$)

(b)
Stationary rotated block

Therefore the block will appear as in Part (a) of the above Figure.

Part (b) of the above Figure shows a similar block, stationary, but physically rotated.

A snapshot of the object is the same as that of a fast-moving but unrotated block.

Other objects also look rotated, so that for example, a sphere will always look spherical, and will not appear to be squashed in its direction of motion.

To be precise, we should say that the exactly rotated appearance holds true only when the object subtends a small angle at our eye.

Large objects will be somewhat distorted, and the analysis is more complicated.

It is interested that since the Sun and Moon move with respect to us, the twist effect makes them appear very slightly rotated from their actual orientations.

It can be shown that the rotation angle $\Delta\theta$ satisfies

$$\cos \Delta\theta = \sqrt{1 - v^2/c^2}$$

where v is the relative sideways velocity of the object.

The rotation angle can be used to find out how much of the object is hidden and how much is uncovered due to this “Terrell Twist” effect.

The Cosmic Speed Limit

It is often said that “nothing can go faster than the speed of light”.

We are naturally suspicious!

People used to say that no piloted aircraft could ever travel faster than the speed of sound.

That limitation was hard to overcome, but it was overcome,

Who says we can’t travel faster than light?

Some Difficulties

Needless to say, there are difficulties.

It can be shown, using standard physics and mathematics, that if a constant force is applied to to an object, it keeps speeding up, but its acceleration decreases so that it never quite reaches the speed of light, no matter how long tthe force is applied.

This is shown now.

Rectilinear Motion(1-dim)

$$F = m_0 \frac{dv}{dt}\left(\frac{1}{\sqrt{1-\frac{v^2}{c^2}}}\right) + m_0 v \frac{-\frac{1}{2}\left(-2\frac{v}{c^2}\right)\frac{dv}{dt}}{\left(1-\frac{v^2}{c^2}\right)^{3/2}}$$

$$= m_0 \frac{dv}{dt}\frac{1}{\left(1-\frac{v^2}{c^2}\right)^{3/2}} = m_0 \gamma^3 \frac{dv}{dt}$$

Newton's law is **modified** by the factor γ^3 which has a dramatic effect as $v \rightarrow c$

Now suppose that we have a constant force F = constant.

We can then integrate the equation as follows:

$$F\,dt = m_0 \gamma^3(v)\,dv$$

$$Ft = m_0 \int_0^v \gamma^3(v)\,dv$$

Now

$$\frac{d}{dv}(\gamma v) = \gamma + v\frac{d\gamma}{dv}$$

$$\frac{d\gamma}{dv} = \frac{d}{dv}\left(1-\frac{v^2}{c^2}\right)^{1/2} = \frac{\frac{v}{c^2}}{\left(1-\frac{v^2}{c^2}\right)^{3/2}} = \gamma^3 \frac{v}{c^2}$$

$$\frac{d}{dv}(\gamma v) = \gamma + \gamma^3 \frac{v^2}{c^2} = \gamma^3\left(\frac{1}{\gamma^2} + \frac{v^2}{c^2}\right) = \gamma^3$$

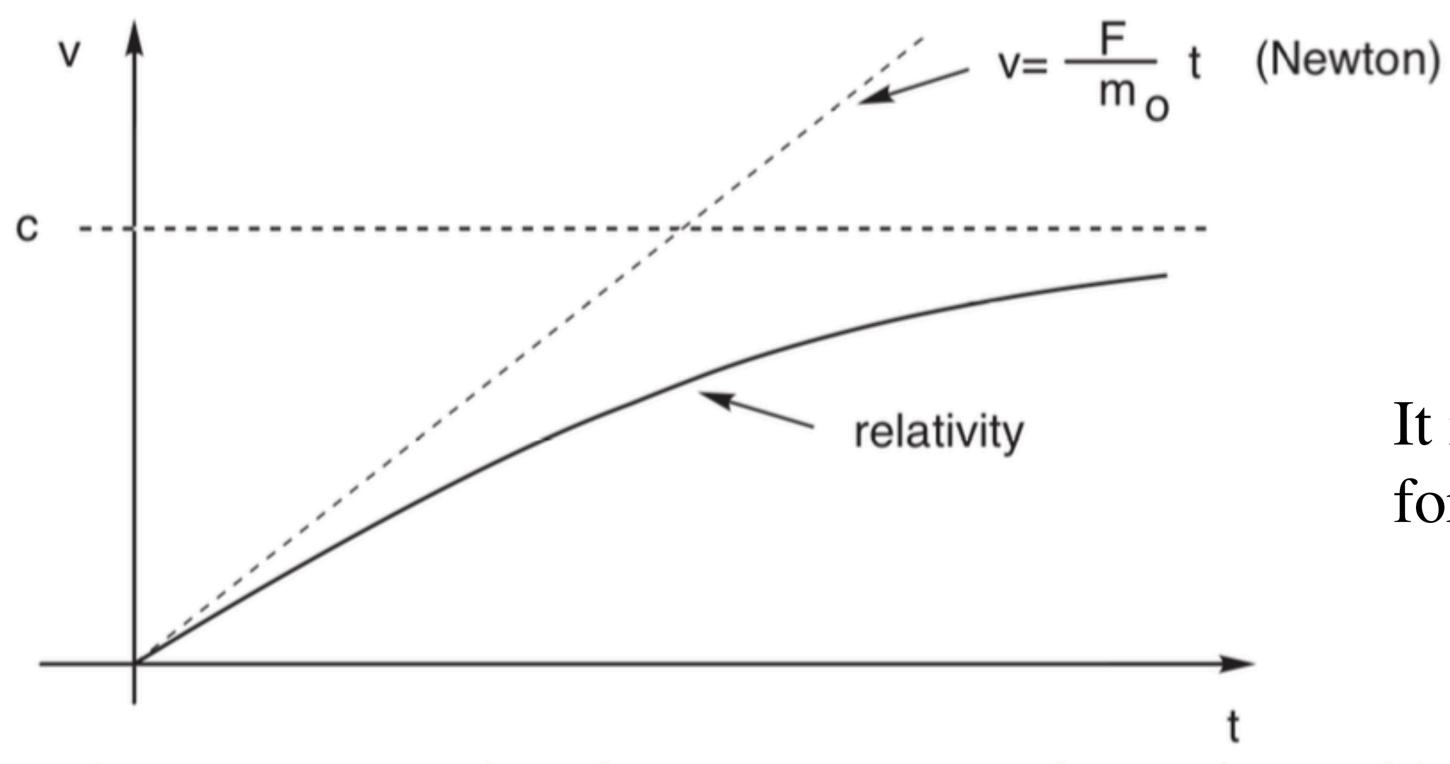
Therefore

$$Ft = m_0 \int_0^v d(\gamma v) = m_0 \gamma v = m_0 \frac{v}{\sqrt{1 - \frac{v^2}{c^2}}}$$

$$F^2 t^2 = m_0^2 \frac{v^2}{1 - \frac{v^2}{c^2}} \rightarrow v^2 = \frac{\left(\frac{Ft}{m_0}\right)^2}{1 + \left(\frac{Ft}{m_0 c}\right)^2}$$

$$\rightarrow v = \frac{dx}{dt} = \frac{F}{m_0 c} \frac{ct}{\sqrt{1 + \left(\frac{F}{m_0 c}\right)^2 t^2}}$$

A plot of v versus t is shown.



It is clear that no matter how long a constant force is applied we still have v < c.

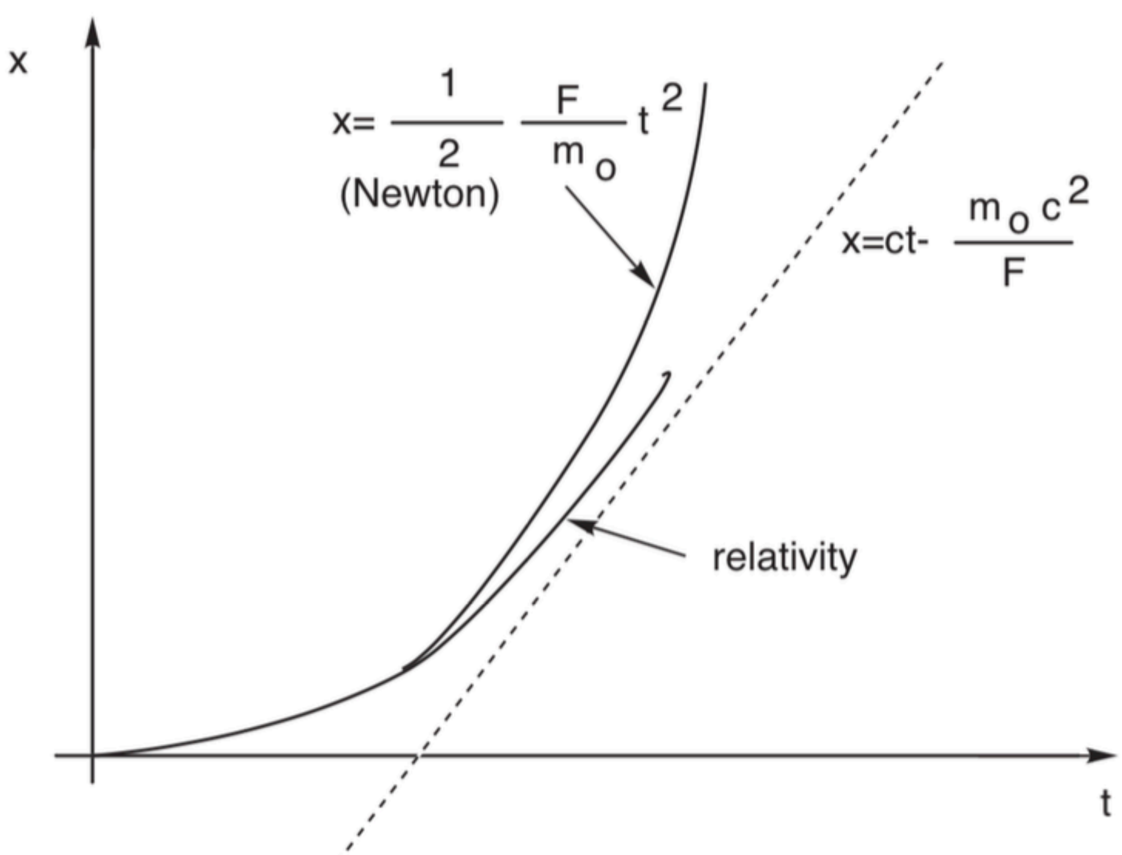
Continuing the integration

$$dx = \frac{F}{m_0 c}\frac{ct}{\sqrt{1+\left(\frac{F}{m_0 c}\right)^2 t^2}}dt$$

$$x = \frac{F}{m_0}\int_0^t \frac{F}{m_0}\frac{t}{\sqrt{1+\left(\frac{F}{m_0 c}\right)^2 t^2}}dt = \frac{F}{m_0}\left(\frac{m_0 c}{F}\right)^2 \times \int_0^{Ft/m_0 c}\frac{u}{\sqrt{1+u^2}}du$$

$$= \frac{m_0 c^2}{F} \times \int_0^{Ft/m_0 c} d(\sqrt{1+u^2}) = \frac{m_0 c^2}{F}\left(\sqrt{1+\left(\frac{F}{m_0 c}\right)^2 t^2} - 1\right)$$

A plot of x versus t is shown below:



Whenever, one does complex calculation you should check your results by calculating limits where the answer is known.

Letting t → 0 we get

$$\sqrt{1+\left(\frac{F}{m_0 c}\right)^2 t^2} \approx 1 + \frac{1}{2}\left(\frac{F}{m_0 c}\right)^2 t^2$$

$$v \approx \frac{F}{m_0} t \text{ as expected}$$

$$a \approx \frac{F}{m_0} \text{ as expected}$$

$$x \approx \frac{1}{2} \frac{F}{m_0} t^2 \text{ as expected}$$

Letting t → ∞ we get

$$v \to c \quad \text{and} \quad x \to ct \text{ as expected}$$

Suppose, however, there were some more exotic way to give an object a velocity v > c.

The the quantity 1 - v^2/c^2 would become negative, so the famous factor $\sqrt{1 - v^2/c^2}$ would become imaginary.

If the "superluminal" (faster-than-light) object happened to be a clock, for example, it would read imaginary time according to the time dilation formula and have an imaginary length according to the Lorentz-contraction formula.

Any moving particle of mass m would also have imaginary momentum and imaginary energy.

None of this makes sense.

Nevertheless, people have speculated about the existence of single particles called tachyons, as we mentioned earlier, which always travel faster than the speed of light.

These hypothetical particles are pointlike and without any kind of clock, so time dilation and length contraction don't matter.

Also tachyons have *real* momentum $\mathbf{p} = \gamma m\mathbf{v}$ and real energy $E = \gamma mc^2$, because they are assigned an imaginary mass m, so that the imaginary number i in m cancels out with the i in

$$\gamma = 1/\sqrt{1 - v^2/c^2} = 1/i\sqrt{v^2/c^2 - 1}$$

If tachyons exist, and if they interact with ordinary matter (including massless photon or particles with real mass like electrons and protons), there are ways to infer their presence.

High-energy physicists have carried out experiments to look for them in these indirect wyas, but so far none has been found.

If tachyons exist, then the world of particles is divided into three parts: (i) ordinary particles with real mass and velocities $v < c$, (ii) photons (and possibly other massless particles) with velocity $v = c$, and (iii) tachyobsm with imaginary mass, that always travel with $v > c$.

A pretty picture in some ways, but up to now there is no evidence that the third category exists.

What about signals of any kind?

Might we someday be able to send a signal faster than the speed of light?

As we have seen, the agent carrying the message cannot be an ordinary material object and cannot be light itself.

Nevertheless, suppose somehow that a message could be sent faster than the speed of light, using tachyons, “wormholes”, quantum entanglement, mental telepathy, or any other hypothetical means.

Would such a signal cause any problems?

Causality Paradoxes

It turns out that if we could send signals faster than light we could construct a very different kind of paradox.

Most paradoxes, as we have seen, are *apparent* paradoxes; they are not self-contradictions, but only counterintuitive.

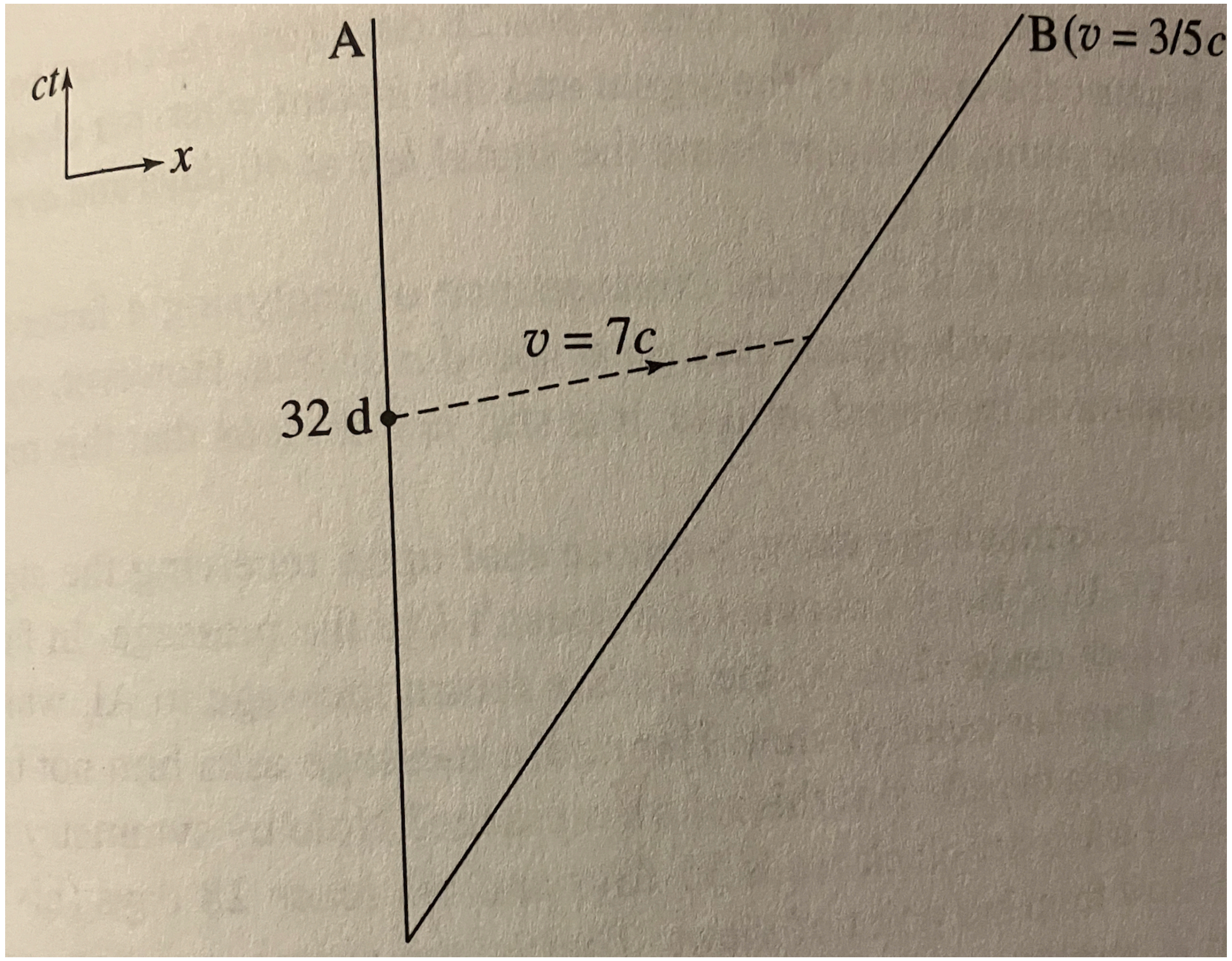
We could explain all of them by just being careful about using special relativity consistently.

As an example of what can happen with superluminal signals, consider the following story about the twins (from the earlier "paradox"), call them Al and Bertha.

One day Bertha departs from Earth, moving at constant speed (3/5)c.

Thirty two days after she leaves, Al decides to send her a signal that moves at speed 7c.

A spacetime diagram of the events is shown below in the Figure.



How far away is Bertha when she receives the message, and what does her clock read at that instant?

At time t in Al's frame, she has moved away a distance (3/5)ct.

This is also the distance the signal moves at speed 7c during the time interval Δt = t-32 days(unit=d); that is

$$x = \frac{3}{5}ct = 7c(t - 32d)$$

Solving this equation for t gives t = 35d.

That is, in Al's frame she receives the signal 3 days after it was sent; at that time she is a distance (7c)(3d) =21cd away.

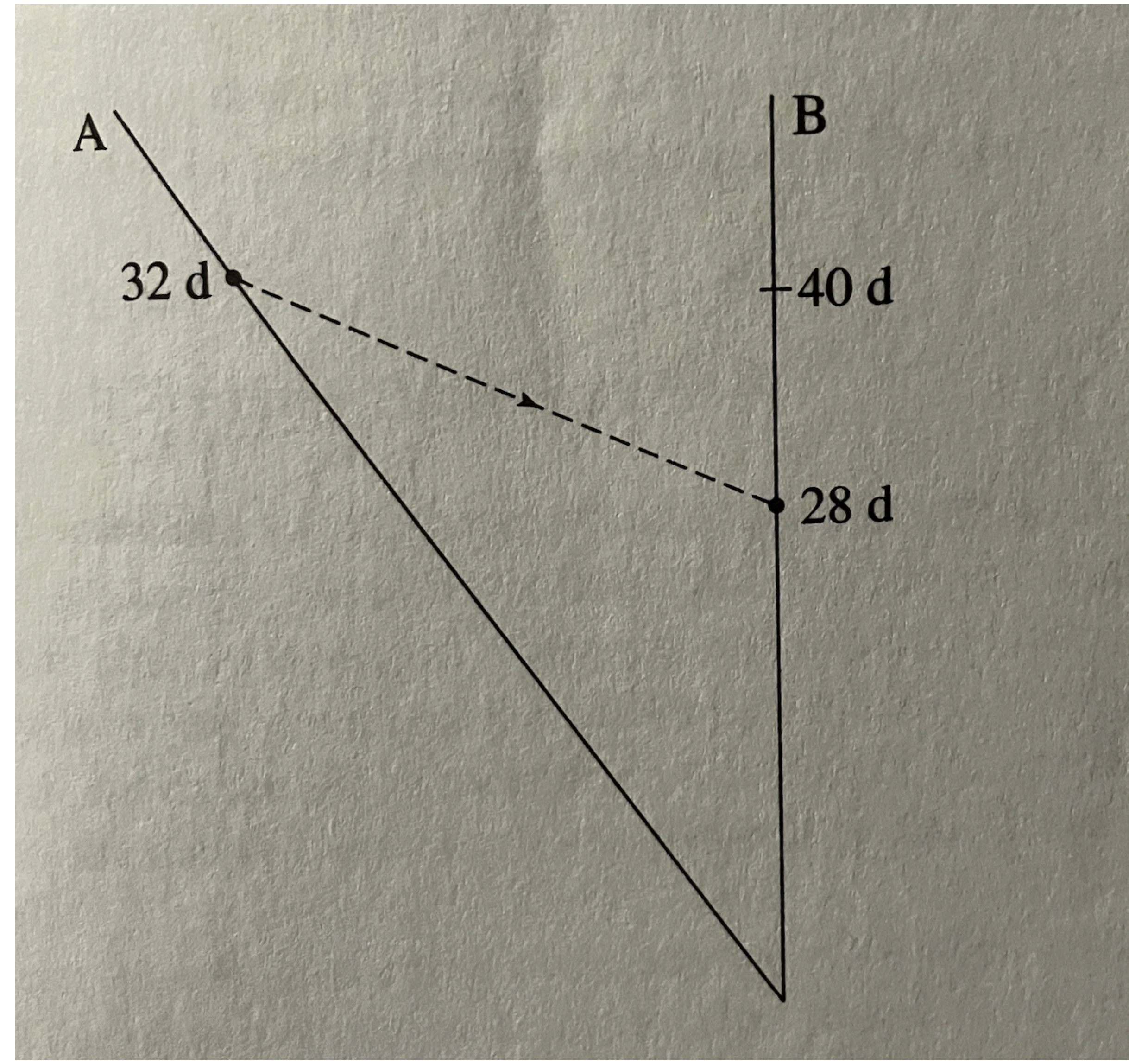
In Al's frame her clock has been running slow, so although his clock reads 35 days, hers reads only

$$35d\sqrt{1-(3/5)^2} = 28d$$

when she receives the signal.

There is nothing particularly strange in this: the communication is very fast covering 21cd in only 3 days, but it does not seem paradoxical in any way.

Now in the figure below we draw the same events from Bertha's point of view.



To her, Al moves off to the left at (3/5)c, as shown.

When his clock reads 32 days he sends the signal.

However, since his clock has been running slow from her point of view, her clocks reads

$$32d\sqrt{1-(3/5)^2} = 40d$$

when he sends the signal.

And when the signal arrives, her clock reads 28 days (this fact must be the same in all frames because the arrival of the signal and the instant her clock reads 28 days are at the same point), so in her frame the signal left at 40 days and arrived at 28 days.

It travelled backwards in time!

The result is weird.

It is a special consequence of analyzing a faster-than-light signal.

It cannot happen with signals that go at speeds c or less.

However, as strange as it is to have a signal travel backwards in time, it is still not obvious that this represents a paradox.

However, let’s continue with the story.

Suppose that upon receiving the signal when her clock reads 28 days, Bertha decides she doesn’t like the message.

In fact, 4 days later, when her clock reads 32 days, she sends a return message to Al, which travels at speed $v = 7c$ from her point of view.

Her return message asks him not to send the first message!

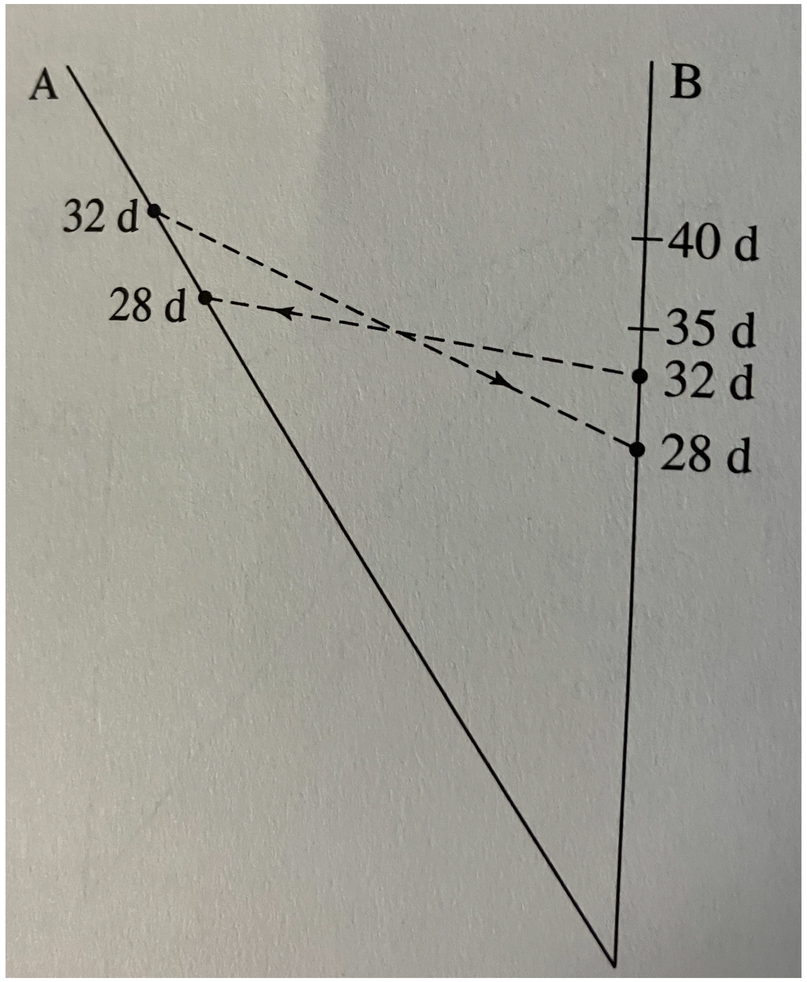
What is the effect of this return message?

Note by symmetry that it will arrive back at Al when her clock reads 35 days and his reads 28 days (his clock has been running slow from her point of view).

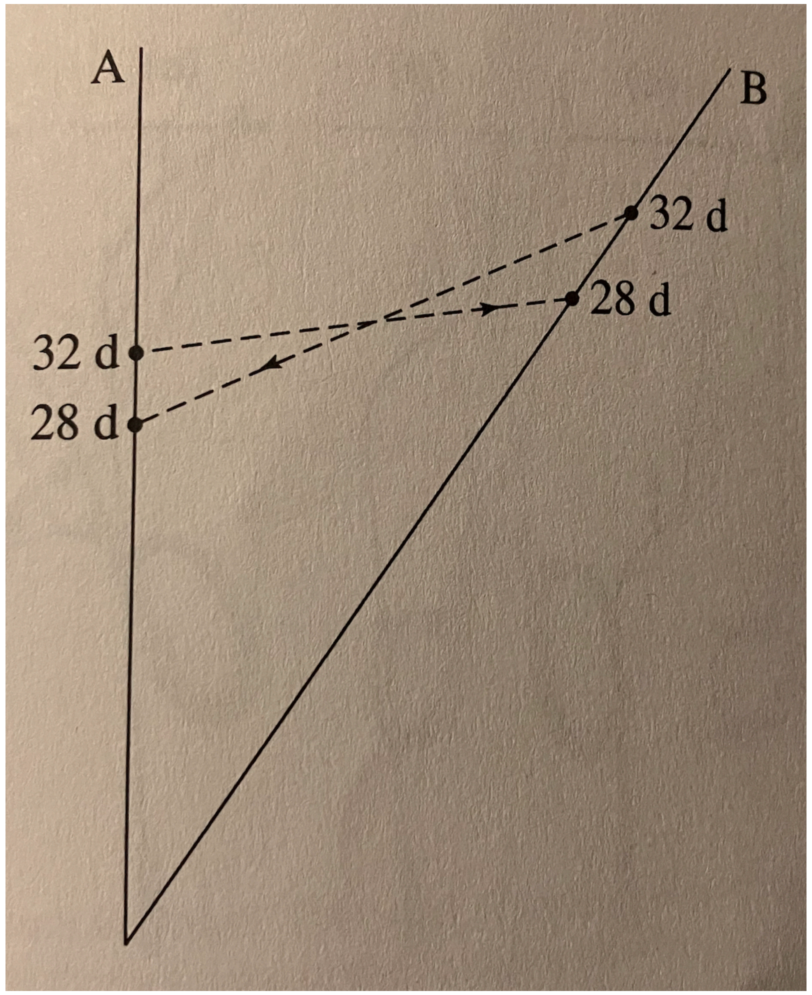
That is, the return message arrives at Al 4 days before he sent the original message!

The spacetime diagram in her rest frame is shown below on the left and the same complete scenario from the point of view of Al's rest frame is shown below on the right.

Bertha



Al



This is a paradox of a very different sort.

To make it even more clear, suppose that Al had a combined transmitter/receiver that is programmed as follows: it will send a message to Bertha when its clock reads 32 days unless it has received a prior message that indicates it should not send the message.

Bertha’s transmitter/receiver, on the other hand, is programmed to send a message to Al when its clock reads 32 days only if it receives a prior message from Al.

So if Al does send a message, it must be that his transmitter/receiver has not received a prior message from Bertha.

Contrariwise, if Al does not send the original message, his transmitter/receiver must have received a message from Bertha.

But Bertha does not send a message unless she receives one from Al.

Both scenarios appear to be contradictory.

Such a paradox is called a causality paradox.

Causality paradoxes can arise if message travel faster than light.

“Things” That Go Faster Than Light

In spite of all of our discussion so far, there certainly are phenomena that travel faster than the speed of light, or could in principle travel faster than light.

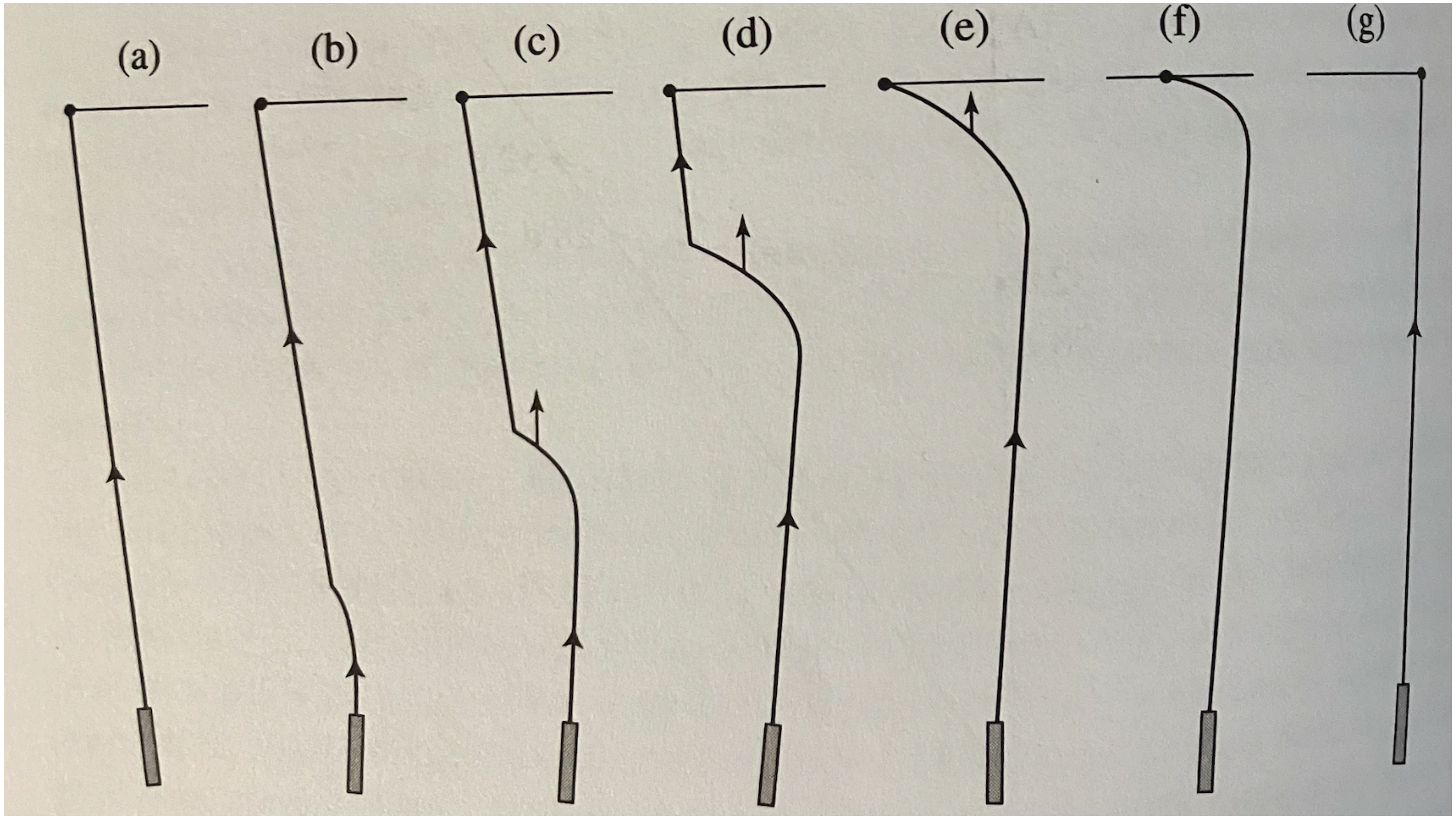
Suppose for example we aim a laser at the Moon and observe through a telescope the bright dot where the laser beam is partially reflected from the Moon’s surface.

If we turn the laser from left to right, we observe the reflected dot on the Moon surface move from left to right.

Is it possible to turn the laser so quickly that the dot will move across the Moon’s surface faster than the speed of light?

Picture the laser as a hose sending out a stream of photons toward the Moon.

The Figure below shows the stream at discrete times.



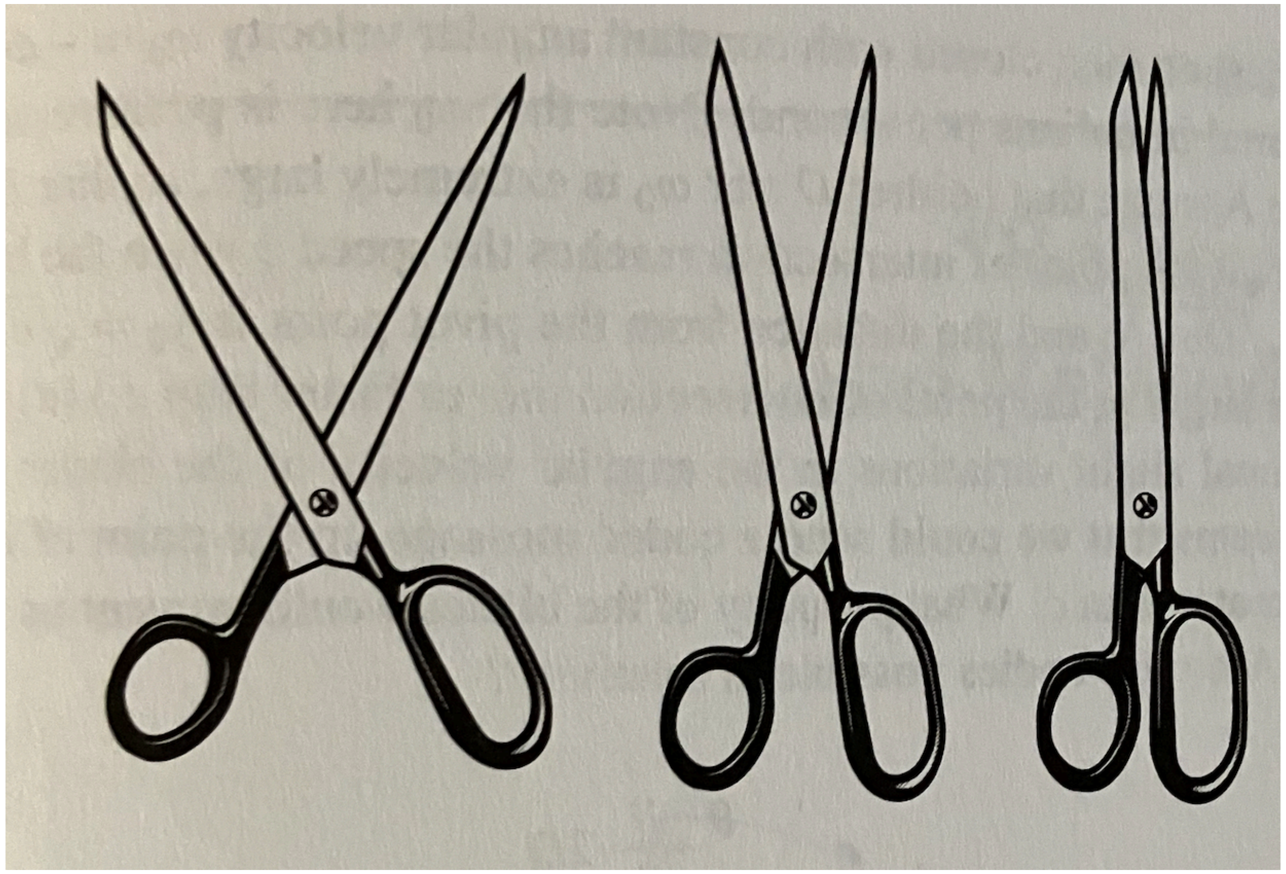
The dot stays at the left edge of the Moon until Part (e), when it starts to move to the right, and finally reaches the right side of the Moon in Part (g).

Once it starts to move, the dot actually can travel across the Moon's surface faster than the speed of light

Note that no single photon travels with the dot, and note also that no one living on the left edge of the Moon can send a signal to someone living on the right edge - all potential signals are controlled by the person turning the laser on the Earth.

Therefore such a faster-than-light dot violates no principle of relativity.

Another "thing" that could travel faster than light is the point of intersection of the two blades of a very long pair of scissors as the scissor blades close.



A sequence of picture is shown in the Figure.

While the atoms in each blade move towards each other at speeds less than c, thew abstract point of intersection can move faster than c!

Its is essential that the blades close so that each of them remains straight.

This, in general, cannot be done.

To remain straight the upper parts of the blades would have to realize that it time for them to move to retain the structure of the scissors.

However, for very long scissors, signals at less than c, cannot do it in time and the blades cannot remain straight - there are no such things a rigid bodies in relativity.

Earlier I showed you what happens if we apply a constant force forever.

Let me now continue with these ideas so we can develop the parameters of a realistic interstellar spaceflight.

The derivation is mathematic intensive.

So, if mathematic is not your thing, just follow along until the end whenI will draw conclusions from all the mathematics.

Acceleration in Special relativity

The following is a derivation of a relationship between the time that passes with a rest frame (Earth) and the time that passes within a moving frame (a starship), that starting from rest, moves with a predetermined constant acceleration to a predetermined destination and back.

More complicated than simple time dilation

We begin with the Lorentz Transformations for a point P in S and moving frame S',

where S' moves with velocity v in the positive x direction.

The Lorentz Transformations from the Earth frame S to ship frame S' is stated first,

followed by the Inverse Lorentz Transformations ($S' \rightarrow S$).

The differentials of these transformations are found

and rearranged to find the time transformation equations between frames

where the point accelerates from rest at a constant rate.

We will get same equations as in on earlier slides(rectilinear motion).

The Lorentz Transformations for point P (x', y', z', t') are given below:

where S' moves with speed v in the positive x direction,

t is the rest frame time, c is the speed of light, and $\gamma = \sqrt{1 - v^2/c^2}$

$$x' = \frac{x - vt}{\sqrt{1 - v^2/c^2}} = \gamma(x - vt) \tag{1a}$$

$$y' = y \tag{1b}$$

$$z' = z \tag{1c}$$

$$t' = \frac{t - xv/c^2}{\sqrt{1 - v^2/c^2}} = \gamma(t - xv/c^2) \tag{1d}$$

The Inverse Lorentz Transformations for point P (x, y, z, t),

$$x = \frac{x' + vt'}{\sqrt{1 - v^2/c^2}} = \gamma(x' + vt') \tag{2a}$$

$$y = y' \tag{2b}$$

$$z = z' \tag{2c}$$

$$t = \frac{t' + x'v/c^2}{\sqrt{1 - v^2/c^2}} = \gamma(t' + x'v/c^2) \tag{2d}$$

Taking the differentials of (2a) and (2d),

$$dx = \frac{dx' + vdt'}{\sqrt{1 - v^2/c^2}} = \gamma(dx' + vdt') = \gamma(v'_x + v)dt' \tag{3a}$$

$$dt = \frac{dt' + (v/c^2)dx'}{\sqrt{1 - v^2/c^2}} = \gamma(1 + v'_x v/c^2)dt' \tag{3b}$$

where $v'_x = dx'/d't$ is the speed of P in S.

Dividing (3a) by (3b) to find the speed v_x of the point in S,

$$\frac{dx}{dt} = \frac{= \gamma(v'_x + v)dt'}{\gamma(1 + v'_x v/c^2)dt'} = \frac{v'_x + v}{1 + v'_x v/c^2}$$

$$v_x = \frac{v'_x + v}{1 + v'_x v/c^2} \tag{4}$$

Differentiating v_X to find the acceleration a_X of the point in S,

$$a_x \equiv \frac{d}{dt}v_x = \frac{d}{dt}\left(\frac{v'_x+v}{1+v'_xv/c^2}\right) = \frac{dv'_x/dt}{1+v'_xv/c^2} - \frac{(v'_x+v)(v/c^2)}{(1+v'_xv/c^2)^2}\left(\frac{dv'_x}{dt}\right)$$
$$= \frac{dv'_x/dt}{\gamma^2(1+v'_xv/c^2)^2}$$

Substituting $dt = \gamma(1+v'_xv/c^2)dt'$ in the above,

$$a_x = \frac{dv'_x/dt}{\gamma^2(1+v'_xv/c^2)^2} = \frac{dv'_x/dt'}{\gamma^3(1+v'_xv/c^2)^3}$$

$$a_x = \frac{a'_x}{\gamma^3(1+v'_xv/c^2)^3} \qquad (5a)$$

The acceleration of point P in S', $a'_X = dv'/dt'$ is held constant;

to emphasize this and to simplify the writing,

the acceleration of the point in the moving frame S' will be represented with "a".

The ship and its contents should thought of as the point P accelerating in S',

while P and S' momentarily have the same instantaneous speed v.

Summarizing(It might help to recall that a point moving in simple harmonic motion can have a non-zero instantaneous acceleration and instantaneous speed of zero),

$$a'_x \equiv a \;\;,\;\; v'_x = 0 \;\text{ and }\; v_x = v \;\text{ from (4)}$$

Therefore (5a) can be written,

$$a_x = \frac{a}{\gamma^3} \quad , \quad \frac{dv_x}{dt} = (1 - v^2/c^2)^{3/2} a \tag{5b}$$

Integrating (5b),

$$\int_0^v \frac{dv}{(\sqrt{1 - v^2/c^2})^3} = a \int_0^t dt \to \frac{v}{\sqrt{1 - v^2/c^2}} = at \tag{5c}$$

Rearranging (5c)

$$\frac{v^2}{1 - v^2/c^2} = a^2t^2 \to v^2(1 + a^2t^2/c^2) = a^2t^2$$

$$v = \frac{at}{\sqrt{1 + a^2t^2/c^2}} \tag{5d}$$

which is the same result as my earlier derivation.

Rearranging again to find another useful form of (5c)

$$at = \frac{v}{\sqrt{1-v^2/c^2}} \rightarrow \left(\frac{1}{at}\right)(v) = \sqrt{1-v^2/c^2}$$
$$\rightarrow \left(\frac{1}{at}\right)\left(\frac{at}{\sqrt{1+a^2t^2/c^2}}\right) = \sqrt{1-v^2/c^2}$$

$$\sqrt{1-v^2/c^2} = \frac{1}{\sqrt{1+a^2t^2/c^2}}$$

Integrating (5e) to find the time transform equations with constant acceleration,

$$\tau = \int_0^t dt' = \int_0^t (1-v^2/c^2)^{1/3} dt = \int_0^t \frac{dt}{\sqrt{1+a^2t^2/c^2}}$$
$$\frac{c}{a}\ln\left(\frac{a}{c}t + \sqrt{1+a^2t^2/c^2}\right) = \frac{c}{a}\sinh^{-1}\left(\frac{a}{c}t\right)$$

Time Transform with Constant Acceleration

$$\tau = \frac{c}{a}\sinh^{-1}\left(\frac{a}{c}t\right) \tag{5f}$$

Finding the inverse of (5f),

$$t = \frac{c(e^{at/c} - e^{-at/c})}{2a} = \sinh\left(\frac{a}{c}\tau\right) \tag{5g}$$

This the **Inverse Time Transform with Constant Acceleration**,

where t is the time within S (the rest frame),

a is the acceleration of P in the moving frame,

c is the speed of light,

and τ is time within the moving frame S'.

we now have all the equations we need in most useful form to work out some non-trivial examples

Journey to the Stars

NASA's newly discovered Earth-like planet Kepler-452b is about 1,400 light-years away.

Though it's is over a thousand light-years away, equation (6a)(see later)

can be used to determine the time it would take to reach Kepler-452b

travelling at a constant acceleration of 1g in both the frame of the Earth (obviously over 1000 years)

and in the frame of starship (amazingly, within an average life time).

The journey is divided up into four parts, accelerating to the midpoint of the outward destination, turning the engines around to decelerate to rest at the destination, accelerating back to the midpoint, turning the engines again at the midpoint to return to rest on an Earth that is over 2000 years older that when the ship left it.

To find the rest frame time (time on Earth) it would take to reach a given distance,

in a ship travelling with constant acceleration,

equation (5d) is integrated to find the time needed to reach the ship's turnaround

(midway to the destination).

This time is then doubled to find the total rest frame time to reach its destination.

Since the engine turn-around time is assumed negligible and ship decelerates at the same rate

as it accelerates to the midpoint,

by symmetry the time of the ship's constant deceleration to rest is the same as the outbound time.

Rearranging (5d), integrating to find the distance traveled x with constant acceleration,

staring from rest after time t,

$$v = \frac{dx}{dt} = \frac{at}{\sqrt{1 + a^2t^2/c^2}} \rightarrow dx = \frac{atdt}{\sqrt{1 + a^2t^2/c^2}}$$

$$\rightarrow \int_0^x dx = \int_0^t \frac{atdt}{\sqrt{1 + a^2t^2/c^2}}$$

$$x = \left(\frac{c^2}{a}\right)(\sqrt{1 + a^2t^2/c^2} - 1) \tag{6a}$$

which is the same result as my earlier derivation(rectilinear motion).

Let X represent the outward bound distance, then multiplying (6a) by 2 and solving for time t,

$$X = 2\left(\frac{c^2}{a}\right)(\sqrt{1 + a^2t^2/c^2} - 1) \tag{6b}$$

$$\mathbf{X} = \left(\frac{c^2}{a}\right)(\sqrt{1 + a^2t^2/c^2} - 1) \rightarrow \frac{a\mathbf{X}}{c^2} + 1 = \sqrt{1 + a^2t^2/c^2} \rightarrow t^2 = \frac{\mathbf{X}^2}{c^2} + 2\frac{\mathbf{X}}{a}$$

$$t = \sqrt{\frac{\mathbf{X}^2}{c^2} + 2\frac{\mathbf{X}}{a}} \tag{7a}$$

If the destination distance X (the distance to the midpoint engine turnaround is X/2),

then total time T_X to the destination is twice the time t_X to the midpoint (twice t_X):

$$T_X = 2t_X = 2\sqrt{\frac{1}{c^2}\left(\frac{X}{2}\right)^2 + \frac{2}{a}\left(\frac{X}{2}\right)} = 2\sqrt{\frac{X^2}{4c^2} + \frac{X}{2}}$$

$$T_X = 2\sqrt{\frac{X^2}{4c^2} + \frac{X}{2}} \tag{7b}$$

$$t = \sqrt{\frac{x^2}{c^2} + \frac{2x}{a}}$$

$$t' = \frac{c}{a}\sinh^{-1}\left[\frac{a}{c}t\right]$$

$$t' = \frac{c}{a}\sinh^{-1}\left[\frac{a}{c}\sqrt{\frac{x^2}{c^2} + \frac{2x}{a}}\right]$$

Substituting half of (5f) into (7a) and multiplying by 2,

$$T' = 2\frac{c}{a}\sinh^{-1}\left[\frac{a}{c}\sqrt{\frac{X^2}{4c^2} + \frac{X}{2}}\right]$$

If (5f) and (5g) are used to find the rest frame round-trip T and moving frame round-trip time T$'$, respectively, then the total of each time has to be understood as the sum of four equal blocks of time - by symmetry, two equal blocks of outward trip time and two equal blocks of return time,

$$T' = 4\frac{c}{a}\sinh^{-1}\frac{a}{c}\frac{T}{4} \tag{8a}$$

$$T = \frac{4c}{a}\sinh\frac{aT'}{4c} \tag{8b}$$

Summarizing

STARSHIP TIME TO EARTH TIME	$T' = 4\frac{c}{a}\sinh^{-1}\left(\frac{a}{c}\frac{T}{4}\right)$
EARTH TIME TO STARSHIP TIME	$T = \frac{4c}{a}\sinh\frac{aT'}{4c}$
EARTH TIME - STARSHIP TIME	$T - T' = T - 4\frac{c}{a}\sinh^{-1}\left(\frac{a}{4c}T\right) = \cdot\ T = \frac{4c}{a}\sinh\frac{aT'}{4c} \quad - T'$
MAXIMUM SPEED	$v = \frac{at}{\sqrt{1+a^2t^2/c^2}}$
FARTHEST DISTANCE	$X = 2\left(\frac{c^2}{a}\right)(\sqrt{1 + a^2t^2/c^2} - 1)$
EARTH TIME TO DESTINATION	$T_X = 2\sqrt{\frac{X^2}{4c^2} + \frac{X}{2}}$
SHIP TIME TO DESTINATION	$T'_X = 2\frac{c}{a}\sinh^{-1}\left(\frac{a}{c}\sqrt{\frac{X^2}{4c^2} + \frac{X}{2}}\right)$

Here are some typical values of the various parameters for a = 1g

T (years)	t (years)	d (ly)	v/c	γ
1	1.19	0.56	0.77	1.58
2	3.75	2.90	0.97	3.99
5	83.7	82.7	0.99993	86.2
8	1840	1839	0.9999998	1895
12	113,243	113,242	0.99999999996	116,641

Here are some of the times you will age when journeying to a few well known space marks, arriving at low speed:

d (ly)	Stopping at:	T (years)
4.3	Nearest star	3.6
27	Vega	6.6
30,000	Centre of our galaxy	20
2,000,000	Andromeda Galaxy	28

I came across this neat plot done with these equations.