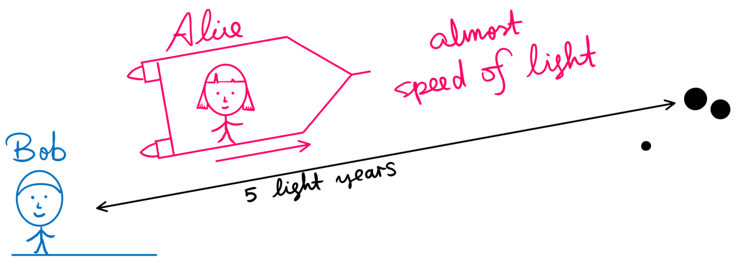
Part 1. Special Relativity

PHYS 2022, Yi Wang, Department of Physics, HKUST

Let us start our journey by a thought experiment – Alice is in a spaceship to a star, 5 light years away from us, and will return to the earth immediately after her arrival. Bob sees her off and waits her to come back. Suppose the spaceship runs almost as fast as the speed of light c, say, $v = 0.995c$.



Amazingly, at $v \sim c$, things behave dramatically different from our daily experience (Newtonian mechanics). Let's see a few surprising facts that they find:

Bob's observations about Alice's journey

- Bob wrote a letter to Alice, and incidentally Alice wrote a letter to Bob around the same time. Bob thinks that he wrote his letter earlier, but Alice insists that she wrote her letter earlier. They have considered that light needs time to travel. But they cannot resolve the dispute even after that.
- Bob finds the spaceship much heavier – when $v \sim c$, Alice needs increasingly greater amount of energy to accelerate the spaceship even by a little.
- When Alice comes back, 10 years has passed for Bob. But for Alice, her clock, her feeling, everything about Alice indicates that only 1 year has passed.

What happened? After learning this part, you will find it out, and much more – In fact, we have to think about space and time in a totally different way from we have naively thought.

Understanding spacetime

We find deeper understanding of spacetime as we understand nature better. The efforts have not come to an end so far.

- Newtonian: space and time are absolute and independent "playgrounds" for matter.
- Special relativity: space and time are unified, and depends on motion of observers (like relative orientation of an object depends on rotation of observers).
- General relativity: space and time can be curved by objects.
- Quantum gravity (conjectured): space and time may be emergent from holography, quantum entanglements, ... We have not fully understood it yet, but the hints are profound.

1 Principles of Special Relativity

1.1 Galileo's Principle of Relativity

Imagine: As in Fig. 1, Alice is moving with constant velocity $\mathbf{v}$ in a closed car with respect to (wrt for short) the ground. If Alice does not look outside the car, how can Alice find out that she is moving wrt the ground?

No. Whatever activities/experiments Alice tries, she finds no difference from if the car is not moving. To put it in the language of physics, the laws of nature she probes is identical to the laws of nature probed when she is not moving.

By "look outside", we mean connections by any means to the outside of the car, including using light, sound, gravitational waves, etc. We neglect rotation of the earth such that Alice and Bob are in inertial frames (constant velocity).

In 1632, Galileo asserts that this is true for all physical laws, and for all inertial frames. This is known as Galileo's principle of relativity. It is impressive to note that this nearly 400-year-old principle still holds now to the best of our knowledge.

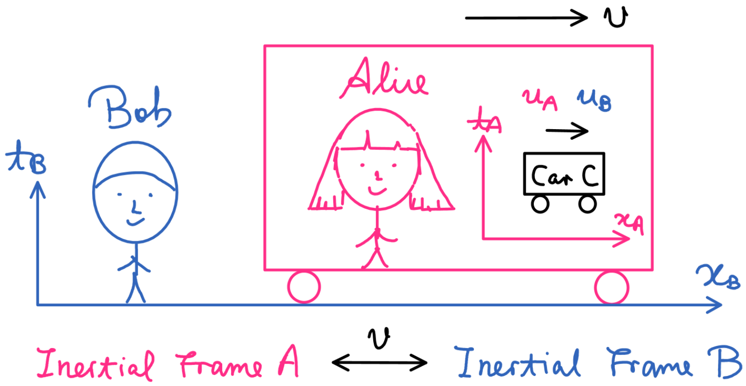


Figure 1: Alice has her time t_A and space x_A; Bob has his time x_B and coordinate x_B. Alice and Bob have relative speed v (because Alice is standing on a moving big car). A small car C has speed u_A wrt Alice and u_B wrt Bob. For simplicity, only one space direction x is shown and other space dimensions y and z are suppressed.

Galileo's principle of relativity

Laws of nature take the same form in all inertial frames.

Let us refer to Galileo's principle of relativity by ($\mathbb{R}$) as we will use it many times. There are some alternative equivalent statements of ($\mathbb{R}$):

- Motion is relative.
- There is no absolute sense of "who moves".

The change (by a relative velocity) from one inertial frame to another is called a "boost". Thus we can also say

- Laws of nature are not changed by a boost.

Isn't Galileo's relativity trivial?

Nowadays, ($\mathbb{R}$) looks pretty trivial. This is because you have already learned Newtonian mechanics.

But at Galileo's (1564-1642) time, people had no concept of acceleration; people relied on naked eyes to do astronomy. Especially, people believe in the geocentric model more than Helio-centrism (earth moves around the sun). An argument for geocentric model is that we would have fallen out of the earth if the earth is moving fast. ($\mathbb{R}$) shows that it will not happen.

Moreover, now we know that ($\mathbb{R}$) still holds in situations where Newtonian mechanics does not hold, including special relativity and beyond.

Newtonian mechanics is consistent with Galileo's relativity

In Fig. 1, does the Newtonian 2nd law (the law of nature) take the same form wrt Alice and Bob? To check that, note the relation between there reference frames are

$$t_B = t_A \ , \qquad x_B = x_A + v t_A \ . \tag{1}$$

The 2nd law in Alice's frame is $F = m a_A$. What about in Bob's frame? Bob picks up Alice's equation and transform it into that in Bob's form:

$$F = m a_A = m \ddot{x}_A = m \ddot{x}_B = m a_B \ . \tag{2}$$

We use the notation $\dot{x} \equiv dx/dt$, $\ddot{x} \equiv d^2x/dt^2$.

Can you also verify Newtonian 1st and 3rd laws?

Thus Bob indeed has the same 2nd law in his frame.

As a reminder for later reference, the velocity addition rule in Newtonian mechanics can be derived from the same frame transformation (1):

$$u_A = \dot{x}_A \ , \qquad u_B = \dot{x}_B = u_A + v \ . \tag{3}$$

From the box above, we observe: ($\mathbb{R}$) is respected by Newtonian mechanics. However, Newtonian mechanics is not the only system that satisfies ($\mathbb{R}$). To see that, and to introduce

Einstein's relativity, let us introduce another fundamental principle of nature.

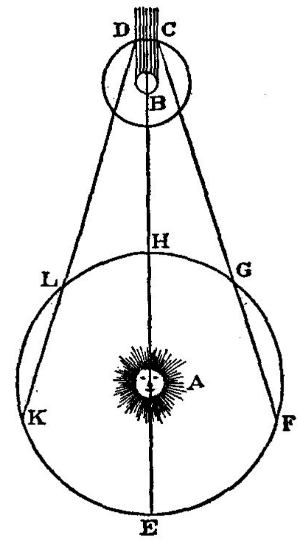
1.2 The Speed of Light

When you turn on the light, light "immediately" fills the room. In our everyday life, the speed of light is so fast that we usually ignore the propagation time needed for light. However, is the speed of light infinite, or finite?

Many delicate experiments have been designed to show that, the speed of light is actually finite at $c \simeq 3 \times 10^8$m/s. One of the first observation from astronomy is outlined in the box to the right. Nowadays, the finiteness of speed of light is not only a reality, but also has great potential for future applications. See, for example, video: seeing what's behind a wall through reflection.

(Optional) Finite c observed

As early as in 1676, Rømer noted that the actually observed eclipses of the Jupiter moon Io and the calculated time have a difference, when the eclipses happen with earth position L, K, G and F. This is interpreted as: light needs time to travel through intervals LK and GF.



The exact value of the speed of light

The speed of light c has an exact value: $c = 299,792,458$m/s.

Are you surprised by this fact? Usually, physical constants are determined by measurements. Measurements always have errors. Then, how can c have an exact value?

The modern definition of a meter (since 1983) is (distance travelled by light in 1 second)/(value of c). Indeed, historically (1889-1960) meter was defined by a real object known as "International Prototype Meter". But defining meter using a real object has many disadvantages, including

- One has to physically compare with the prototype (or its copies) to determine length.
- The accuracy of the copies made are limited by the technology of the time.
- The length of the prototype change slightly over time. And it may be damaged.

By defining meter with a constant of nature c, anyone can reproduce the meter definition. The error is only limited by his/her precision of experiments. Also, the modern definition of time and mass use the fundamental natures of quantum mechanics, through atomic frequencies and the Planck constant.

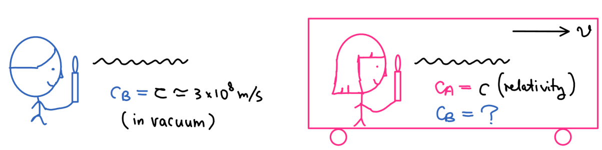
Now, let's come to the key of the section, and the corner stone of Einstein's special relativity.

We have said that the modern speed of light is a way to define length. But from now on, let's go back to the early 1900s, and still consider the speed of light as measured by distance moved per unit time, with common sense definition of distance and time duration.

What's the speed of light from a moving source?

Bob holds a candle. The speed of light from Bob's candle is $c_B = c$ wrt Bob. Alice is moving at speed v wrt Bob in a closed car. She also holds a candle. For the light from Alice's candle:

- What is the speed c_A wrt Alice?
- What is the speed c_B wrt Bob?



We can immediately get $c_A = c$ from ($\mathbb{R}$). Because otherwise Alice would know that she is moving by noticing a different value of speed of light without interacting with the outside of the car. However, what is c_B for the light from Alice's candle?

Naively, we would have expected that $c_B = c_A + v = c + v$ from the Newtonian speed addition rule (3). And this looks natural – in our everyday lives (speeds much slower than light): if

Alice is in a car moving with $v \ll c$ (wrt Bob), and throw a ball with speed u_A wrt Alice, the speed of the ball wrt Bob should be $u_B = v + u_A$.

However, the light propagation should be computed by Maxwell equations. As a result of computation, one obtains $c_B = c$! The box below derives this surprising fact for you. Multi-variable calculus is used, which is beyond the prerequisite math. Thus if you cannot follow the derivation, it's okay to skip it.

(Optional) Speed of light from Maxwell equations

Let us work in Bob's reference frame. The Maxwell's equations with vacuum permeability μ_0 and permittivity ϵ_0 are:

$$\nabla \cdot \mathbf{E} = \frac{\rho}{\epsilon_0} , \qquad (4) \qquad\qquad \nabla \cdot \mathbf{B} = 0 , \qquad (5)$$

$$\nabla \times \mathbf{E} = -\frac{\partial \mathbf{B}}{\partial t} , \qquad (6) \qquad\qquad \nabla \times \mathbf{B} = \mu_0 \left[\mathbf{J} + \epsilon_0 \frac{\partial \mathbf{E}}{\partial t} \right] . \qquad (7)$$

As we study the propagation of light, the Maxwell equations are used in the vacuum environment. Thus the charge density $\rho = 0$, and the current $\mathbf{J} = 0$. To eliminate $\mathbf{E}$ and obtain an equation for $\mathbf{B}$ only, we use the following trick: Let us apply $\nabla \times (\cdots)$ to the LHS and RHS of (7), respectively:

$$\nabla \times \text{LHS} = \nabla \times (\nabla \times \mathbf{B}) \xlongequal{\text{math identity}} \nabla(\nabla \cdot \mathbf{B}) - \nabla^2 \mathbf{B} \xlongequal{\text{using Eq. (5)}} -\nabla^2 \mathbf{B} ,$$

$$\nabla \times \text{RHS} = \mu_0 \epsilon_0 \frac{\partial}{\partial t} (\nabla \times \mathbf{E}) \xlongequal{\text{using Eq. (6)}} -\mu_0 \epsilon_0 \frac{\partial^2}{\partial t^2} \mathbf{B} .$$

Thus, LHS=RHS is a wave equation:

$$\frac{\partial^2}{\partial t^2} \mathbf{B} - \frac{1}{\mu_0 \epsilon_0} \nabla^2 \mathbf{B} = 0 . \qquad (8)$$

In a course of mathematical method of physics, one will study how to systematically solve this equation. Here we will not do so. Instead, we can always propose a solution and check it indeed solves Eq. (8) without using more math. One can check that the below E&M wave is a solution:

$$B_z = B_0 \cos[k(x - ct)] , \qquad \text{where } c \equiv \frac{1}{\sqrt{\mu_0 \epsilon_0}} . \qquad (9)$$

The take home message from the calculation is: once a beam of E&M wave is emitted, it can then propagate in the vacuum without the reference of the emitter, i.e. the motion information of the emitter is "forgotten". The speed of light is calculated by constants of nature μ_0 and ϵ_0, independent on the speed of the emitter (or the observer). Thus we conclude that the answer is $c_B = c$, the same speed that Alice observes (i.e. $c_B = c_A$)!

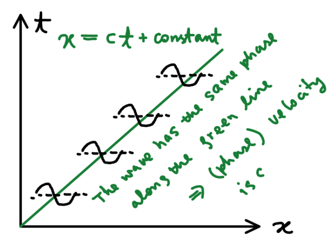
There must be something terribly wrong at this moment: The velocity addition rule (3) is inconsistent with the observer-independent speed of light (9). In such a situation, we need input from experiments to see who is right. The Michelson-Morley interferometer experiment (1887) shows that Maxwell is right. The Newtonian velocity addition does not apply to light!

Maxwell and relativity

In fact, hidden in Maxwell's equations, there is a symmetry, which already implies Einstein's special relativity. The invariant speed of light is just a consequence of that. The symmetry is known as Lorentz transformations, which is developed by Voigt, Lorentz, Larmor and Poincare during 1887-1905 before Einstein established special relativity. We will come to the Lorentz transformations in Section 7 from a mechanical point of view.

Velocity from phase

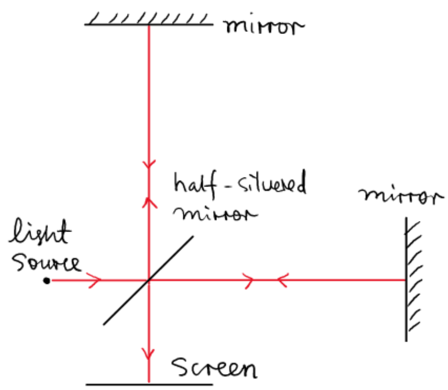
One can read off the (phase) velocity from the "phase" factor $\cos[k(x - ct)]$ in (8). This can be done by following one period of oscillation in the below figure and see how it moves.



We will encounter similar wave equations in quantum mechanics later in this course, and use complex solutions $\exp[ik(x - ct)]$. Thus, make sure to understand why (8) describes a moving wave.

If you think at a deeper level, using phase velocity above is in fact not the way to describe how fast information propagates. Then you need to solve the Maxwell equations more carefully, and find a regarded solution produced by accelerating source. It is beyond the scope of the current discussion.

The Michelson-Morley interferometer experiment



A beam of light is emitted from the source. Via a half-silvered mirror, the beam is split into two, reflected by two mirrors respectively and re-combine on the screen. Due to different length of propagation, interference fringes develop.

Note that the experiment is done on the earth. The earth moves at $v \sim 30$km/s around the sun. Thus if the velocity addition rule applies, the speed of light in the Michelson-Morley interferometer should change under rotation. Then the interference fringes should shift. But the shift is not observed, indicating no velocity addition. Maxwell wins over Newton.

Modern interferometers

The Michelson-Morley interferometer was designed to measure the variation of speed of light. Now that we know (at least assume) that the speed of light is truly a constant, we can turn the same interferometer into a precise measurement of the variation of length of the two arms. This has many applications (check Wikipedia for example), from performing Fourier transformation to measuring the diameter of stars. Last but not least, this is also how LIGO detected gravitational waves in 2015/16.

Optional historical remarks: the story of aether

In fact, what we have introduced here is *not* what researchers was thinking 100 years ago when they made the discoveries.

At the time of Maxwell, it was believed that E&M waves propagates in a media called aether, analogous to sound waves propagating in the air. Thus, the speed of light $c = 1/\sqrt{\mu_0 \epsilon_0}$ is wrt aether as well. The parameters μ_0 and ϵ_0 were considered the property of aether, instead of the property of the vacuum.

Then, the velocity addition rule had no problem – just as velocity addition of wind and sound waves propagating in the wind.

Now the question is: when the earth orbits the sun, do aether around the earth move together with the earth, or remain static?

The observation of stellar aberration (1600s - 1700s) showed that aether did not move together with the earth. And in the context of aether, the Michelson-Morley experiment shows that aether indeed moved together with the earth. This was the contradiction that puzzled physicists 100 years ago. And Einstein's contribution is to eliminate the need of aether with his theory of relativity.

As aether is abandoned, now when we learn E&M waves, we do not introduce aether any more. Thus, I choose to use the way of introduction closer to modern thinking instead of historical thinking, only leaving a remark of aether within this box.

Is Newton "wrong"?

Now that the velocity addition (in the core of Newtonian mechanics) fails for light, is Newton wrong? Should we abandon Newtonian mechanics?

In modern physics, we think of most theories as "effective" theories – they apply with certain approximations. When $v \ll c$, Newtonian mechanics is still useful, as it is pretty precise, and simpler and more intuitive than special relativity.

Thus, by learning special relativity, we do not mean to abandon Newtonian mechanics. Rather, Newtonian mechanics is still much more widely used than special relativity in the modern era (since we do not usually move at $v \sim c$ or care about the corrections suppressed by v/c).

In fact, special relativity has a similar fate – it is classical. When considering small enough particles, one has to consider quantum mechanics and we arrive at quantum field theory. And considering the quantum effect of gravity, quantum field theory is again not adequate...

1.3 Einstein's Relativity

How to resolve the contradiction between Maxwell's observer-independent c and the velocity addition rule? This is one of the two biggest problems in physics at the beginning of the 20th century.

When Einstein was 16 years old, he was already deeply puzzled by light: What if I run as fast as light? Will light be at rest? 10 years later, he found out the solution.

Here is how Einstein solved the problem in 1905: he took the observer-independency of c as a fundamental assumption of his theory. It is assumed. So problem solved :)

Seriously, an assumption is no more than an assumption. Everyone can make assumptions. But what's truly revolutionary is the implications of the assumption and how to get a consistent theory on top of that. To be clear, here we summarize the assumptions, being prepared to proceed to the profound implications.

The postulates of special relativity

- ($\mathbb{R}$) The relativity principle: Laws of nature keep the same form in all inertial frames.
- ($\mathbb{C}$) The vacuum speed of light is c in all inertial frames.

Remarks about ($\mathbb{C}$)

- The speed of light is observer-independent, but not velocity. In other words, the direction of light can be observer-dependent. We will see this explicitly later when we construct a "light clock".
- The frequency of light is observer dependent (relativistic Doppler effect). When you move towards light, the light frequency becomes higher (blue shift). When you move away from light, the light frequency becomes lower (red-shift).
- In non-vacuum situations, the speed of light will depend on the motion of the media. The speed of light in the moving frame of media can be obtained using relativistic velocity addition rules to be discussed later.

The postulates ($\mathbb{R}$) and ($\mathbb{C}$) are the original postulates in special relativity, and often referred to as the only assumptions. But to be clear, one has also to postulate observer-independency of events to complete the theoretical basis of special relativity:

The postulate of observer-independent events

($\mathbb{E}$) Let us define an event to be something that happened to a particular small object at a particular moment (i.e. *localized* at a spacetime point). Then, the occurrence (or not) of an event is observer-independent. If one observer finds that an event E happened, all observers who saw the event must agree, no matter how they move.

This postulate looks too trivial to mention. Nevertheless, let us make things clear.

For example, the followings are events that all (honest) observers must agree:

- "A beam of light is reflected by a mirror."
- "Alice and Bob met. At the time when they met, their watches both pointed to 10am."

However, the following is NOT an event:

- "Alice and Bob are 5 light years away. When Alice's watch pointed to 10am, Bob's watch also pointed to 10am."

This is not an event because Alice and Bob are at different locations. So it may be possible that an observer Charlie considers the above statement to be true; while his running Dog considers the above statement to be false.

Similarly, the following is NOT an event and may not be agreed by all observers:

- "A moving ruler has the same length with a static ruler because the two end points coincided at a moment."

Not one, but two events

In the example of 5-light-year-separated Alice and Bob, the sentence is not an event, but one can separate it into two events: When Alice looked at her watch, her watch shows 10am. And when Bob looked at his watch, his watch shows 10am. The same applies for the comparison of ruler example. It makes sense to talk about relations of events in relativity, as we will see later.

Remark: the time of events wrt a reference frame

With finite speed of light, we need a clarification about the time of events. For example, an event took place some distance away from Bob. When talking about the time of the event, it could refer to:

- The time in Bob's reference frame. Imagine Bob puts a clock at every spatial point in his frame, which "records" the time that an event happened.
- The time that Bob actually "see" the event. This needs to take the time of light propagation time between the event and Bob. This is trivial but messy.

In this course, unless otherwise emphasized, we will by default use the time in Bob's frame, instead of the time that the Bob actually see the light signal of the event reaching him, when talking about the time wrt Bob.

By "a clock at every spatial point": the clocks are static wrt Bob, and synchronized wrt Bob's frame (the coordinate time in his coordinate system). Practical synchronization can be done by using light signals to compare clocks, and deduct the time used for light propagation.
By "record": imagine that Bob has a helper at every point in his frame (i.e. coordinate system). Once any event happened, the helper can immediately write down Bob's time and spatial coordinate of the event. This will be denoted by the coordinate of the event (t_B, x_B) wrt Bob.

With the three postulates, we are now prepared to start the adventure of special relativity.

2 Time Dilation

At the beginning of this part, we promised to explain why the space traveler Alice is younger than Bob when she returns (if she initially has the same age as Bob before space travel). This is the goal of this section.

To do this, we have to touch the most fundamental concept in physics – time. Let us begin by recalling how Newton comment about time.

Time in Newton's Principia

"Absolute, true and mathematical time, of itself, and from its own nature flows equably without regard to anything external, and by another name is called duration: relative, apparent and common time, is some sensible and external (whether accurate or unequable) measure of duration by the means of motion, which is commonly used instead of true time" – Newton, Philosophiæ Naturalis Principia Mathematica

Thus, Newton believed that the "true time" exists independent of anything else. If so, can anything access and probe the true time?

- If it cannot be accessed, then why bother to define "true time"? Why shouldn't we shave this concept using Occam's razor?
- If it can be accessed by an apparatus, then why it cannot be altered? As an extrapolation of the Newtonian's third law, when there is some action from the true time to the apparatus, why the apparatus does not back-react on the true time and alter it?

What can we trust?

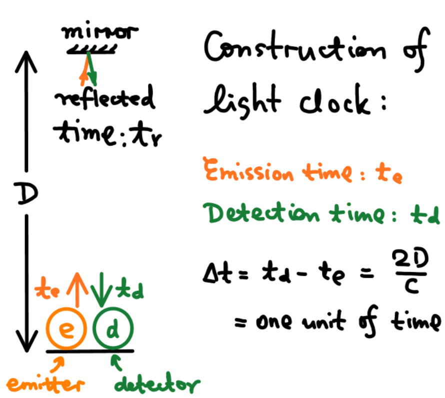
As now we are to go beyond Newton, you may wonder, what can we trust. Of course we want to stand on Newton's shoulder and make use of some of his legacy.

Almost all results for low speed motion ($v \ll c$) can be trusted (with the exception of rest energy, which we will come back to later). For high speed motion ($v \sim c$), we trust ($\mathbb{R}$), ($\mathbb{C}$), ($\mathbb{E}$) and nothing else.

For example, when discusing time: we trust that Alice's mechanical watch (static wrt Alice) defines time nicely for Alice. Because the time from mechanical watch is as a result of low speed mechanical motion. But we cannot trust Newton's conclusion about how Alice's time (as defined by the tick-tock of her mechanical watch) lapse wrt Bob, if Bob has relative motion wrt Alice with $v \sim c$. When discussing length: we trust that Alice's ruler (static wrt Alice) defines length for Alice but we cannot trust how Bob observes the length of Alice's ruler.

Fine. We are physicists. What can we do to get rid of the metaphysical thinking about time? Let us do a practical thing – define time by actually making a simplest clock: the light clock.

Shut up and construct a light clock



We define time by a tick-tock of the "light clock". The tick-tock time interval Δt is the time between the emitter emits the light (at time t_e) and the detector detects the light (at time t_d). In the middle, the light is reflected by a mirror (at time t_r). With the light clock, a "standard time interval" can be defined as

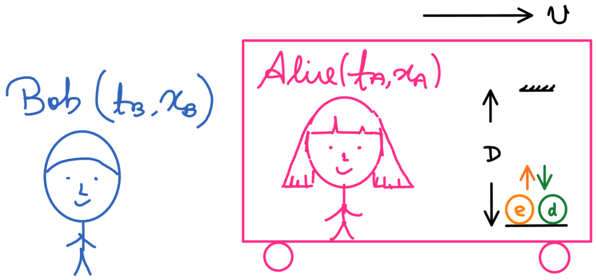
$$\Delta t \equiv t_d - t_e = \frac{2D}{c} . \tag{10}$$

Then time can be measured by number of "standard time intervals" between two events.

The emitter and detector are put extremely close together that we can consider them to be at the same point (in the figure they are drawn separately just for illustration).

To see how the time of a traveler lapse wrt a static observer, let us now load the light clock on a car with Alice.

Load the light clock into Alice's car

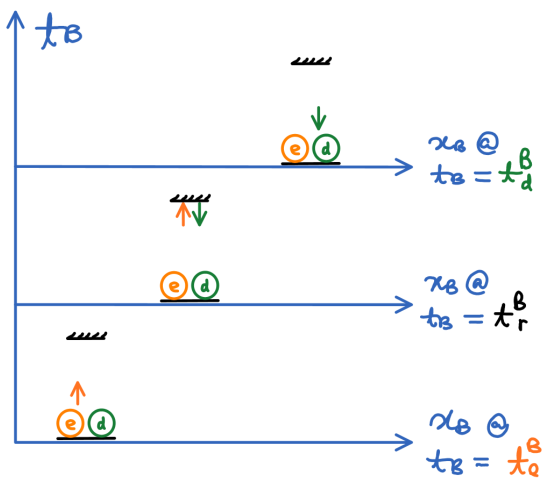


To see how the time of a traveler lapse wrt a static observer, let us now load the light clock in a car with Alice. Using ($\mathbb{R}$), Alice finds that the interval of the light clock is $\Delta t_A = 2D/c$.

Now we have to find out Δt_B and compare it with Δt_A.

Here we assume that the light clock is close enough to Alice, such that at Alice's reference, the space coordinates of both herself and the clock are $x_A = 0$.

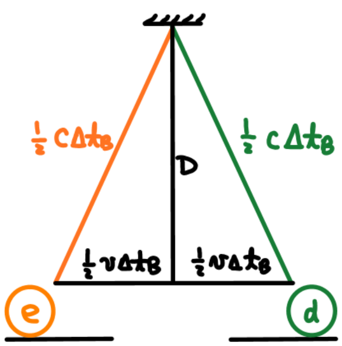
Alice's time interval of the light clock measured by Bob



The left figure is Bob's view on the moving clock. Here x_B and t_B are the space and time wrt Bob, respectively.

We take three snapshots of time: the emission of the signal at $t_B = t_d^B$, the time that light is reflected by the mirror $t_B = t_r^B$, and the time that the light is received by the receiver $t_B = t_d^B$.

The position and state of the light clock is plotted on the diagram.



In the left figure, we plot the light trajectory in the moving light clock wrt Bob. Bob must agree that the light has hit the mirror ($\mathbb{E}$). Thus the direction of light must be no longer vertical. And from ($\mathbb{C}$), the speed of light from the emitter to the mirror is still c.

From the Pythagorean theorem, we have

$$\left(\frac{1}{2}c\Delta t_B\right)^2 = \left(\frac{1}{2}v\Delta t_B\right)^2 + D^2 , \qquad (11)$$

We can thus solve Δt_B – Alice's time interval measured by Bob:

$$\Delta t_B = \frac{2D}{c}\frac{1}{\sqrt{1-\frac{v^2}{c^2}}} = \gamma \Delta t_A , \qquad \gamma \equiv \frac{1}{\sqrt{1-\frac{v^2}{c^2}}} > 1 . \qquad (12)$$

Thus in Bob's frame, Alice's light clock slows down. This is also known as the "time dilation" of moving clocks.

Same perpendicular lengths

You may have an excellent question at this moment: when the light clock is moving, how do we know that the length of the light clock is still D wrt Bob?

Consider a thought experiment: train on rail. When the train is at rest, its wheels has the same interval as the rail width. Now the train moves fast. Does its wheels have larger or smaller width compared to the rail? Neither. Otherwise there would be different type of accidents happening (inconsistent events) that contradicts ($\mathbb{E}$).

Not only Alice's light clock – but the whole Alice's frame slows down wrt Bob

At this moment, it is not convincing enough to conclude that everything about Alice slows down wrt Bob. Because we have only shown that her light clock slows down.

"I don't care about your light clock. I care about you." – Bob

What about her other clocks, her phone, her heart rate, ... ?

In fact, according to ($\mathbb{R}$), everything above slows down.

For example, Alice has a phone which can define the unit time interval. She puts the light clock and phone very close to each other. If the phone and the light clock have the same Δt in a static frame, they must have the same Δt_A wrt Alice ($\mathbb{R}$). Their agreement is an event, which Bob must agree ($\mathbb{E}$). Thus Bob must agree that the moving phone defines the same Δt_B as the light clock.

Goodbye to absolute time

By "moving clock slows down", one notes that the concept of "absolute time" by Newton is gone with the speed. Different movers have different times; there is no absolute mover and thus there is no absolute time.

We will come back to the fundamental understanding of time later this part, and later again in general relativity.

Summary and the 4-step reasoning in special relativity

I hope now you have understood why the space traveler Alice is younger than Bob when she returns – as a moving observer, she slowed down wrt Bob.

The 4 key steps of reasoning is important, and we will use similar methods repeatedly. Thus let us summarize it here:

1. Construct an apparatus in a static frame. The apparatus should be as simple as possible so we can calculate what's actually happening inside the apparatus. Here the apparatus is the light clock with standard time interval $\Delta t = 2D/c$.
2. Load the apparatus into Alice's moving car. From ($\mathbb{R}$), Alice must find the apparatus work the same way as it is at rest. Here Alice finds $\Delta t_A = \Delta t$.
3. Calculate what Bob gets. Compare the result with what Alice gets. We have calculated that $\Delta t_B = \gamma \Delta t_A$. Thus the moving light clock slows down.
4. Although the result is obtained by one particular apparatus, the comparison between Bob's frame and Alice's frame applies for all possible apparatuses. This is because otherwise, one can use the difference to identify an absolute mover ($\mathbb{R}$).

Meaning of Δt_B

Be reminded that Δt_B is NOT the time duration of Bob's light clock. Rather, it's the time duration of Alice's light clock (i.e. moving with Alice) in Bob's frame. So wrt Bob (static observer), Alice's time duration slows down.

Frame time and what Bob actually sees

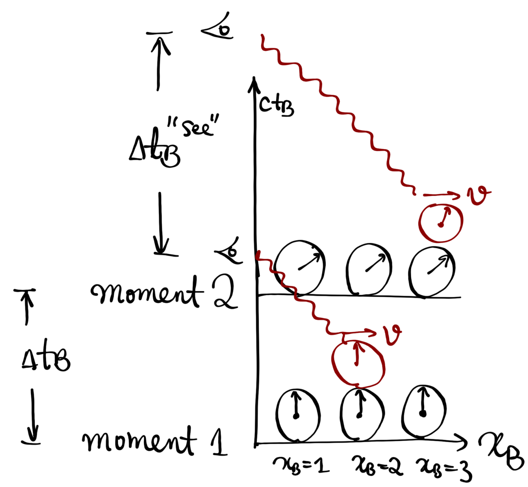
We have mentioned the concept "frame time": If not otherwise stated, when we mention time, we mean time in one's frame, instead of the time that light comes to the observer's eye. Now let us make this point clearer.

Let us put two things at each point in space (i.e. the x-direction):

- A space label. For example, imagine there is a ruler along the x-direction and the label of the ruler is the space label.
- A clock. The clocks are synchronized (see the section about synchronization for how to synchonize clocks) at all points in space.

Time dilation (Δt_B) means that, when Alice's clock (the red clock marked with a $\vec{v}$) is moving, at a next moment, it is compared with another clock in Bob's frame, and it slowed down in this comparison. The time dilation does not depend on if Alice's car moves towards or away from Bob.

On the other hand, what Bob actually "sees", denoted by $\Delta t_B^{\text{"see"}}$: since the speed of light is finite, there will be a delay for the light ray to travel from Alice's clock to Bob's eyes (the red wavy lines). $\Delta t_B^{\text{"see"}}$ depends on whether Alice is moving towards or away from Bob.



Now that you have hopefully understood time dilation and why the space traveler Alice is younger. So now it is a good time to confuse you again, with the infamous "twin paradox".

Who is moving? Who is younger? – The twin paradox

We have shown: Bob find that Alice is younger (according to the clocks in his frame). However, motion is relative ($\mathbb{R}$). Thus shouldn't Alice also find Bob younger?

Let us answer the question in two different setups:

1. Alice and Bob are both inertial frames, i.e. they move wrt each other forever. Then from ($\mathbb{R}$), both of the statements are correct: "Alice finds Bob younger" and "Bob finds Alice younger".

   At first sight, this contradicts with ($\mathbb{E}$). But in fact there is no contradiction. As inertial movers, Alice and Bob can meet only once. This is the time that Alice starts off with the same age as Bob. Afterwards they never meet again. Thus there is no local events to compare their ages (to compare Δt_B and Δt_A using *events*, they have to meet at least twice).

2. Alice first moves and later returns to meet Bob. This is the same setup as we discussed at the very beginning of this part.

   Then we can only trust Bob at the moment. This is because for Alice to return, there exist a period of time when Alice is not in an inertial frame. The principle of relativity (we used a lot of times when deriving the conclusion) does not apply for non-inertial observers. Thus we can only trust Bob's statement that indeed Alice is younger.

Time travel?

The twin paradox is an example of time travel to the future. Take a spaceship trip and back, you will be in the future. If the spacetrip is fast enough, within 1 year round trip, you can see the earth of the next century, or even later. In general relativity you will see more ways to travel to the future such as stay close to a black hole.

What about time travel to the past? Later in this part you will find difficulties within the framework of special relativity to prevent time travel to the past. There are also more general problems, such as what if you travel to the past, and prevent your father and mother to meet? What if you travel to the past and meet another version of yourself? And so on. It is probably impossible to travel to the past, though no decisive conclusion can be made at this moment.

3 Physical Picture and Physical Intuition

This section is not about any detailed rule of physics. Nevertheless, considering that modern physics is so different from classical physics, it is important to pause and discuss about method of learning. This is because modern physics is very different from classical physics. We choose to put the section here instead of at Part 0 (which is logically more suitable), because before having any real experience in modern physics, the talk is empty.

Why modern physics is different from classical physics (i.e. general physics)?

From time dilation of special relativity, you have already tasted a tinny little bite of modern physics. If you have not learned it already, I am sure that you feel it is counter-intuitive and more confusing than most parts of general physics.

And much more is coming – you will learn more about relativity, gravity, quanta, information, complexity ... There is a whole new world in front. The common features of these subjects are not only that they were discovered in the past about 100 years, but also, they are farther away from our daily experience. This is the difference from classical physics.

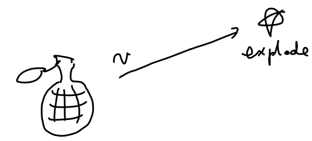
Do not be scared by that! This is why we have this section.

Incidentally, there are similar transitions in mathematics, literature and art – they also evolve from classical eras to modern eras. The modernization revolutions in literature, art, mathematics and physics are of course different, but share interesting similarities.

Let us use an exercise to talk about the methods of thinking, and the recommended way to learn modern physics.

An exercise

A "hand grenade" is a kind of bomb that explode a certain time after you trigger it. Let the time interval between trigger and explosion be Δt when the hand grenade is at rest. Now you throw the hand grenade with speed v. Calculate the distance the hand grenade travels before explosion in the framework of special relativity. In the calculation, both the earth gravity and the friction from air can be neglected.



Instead of solving the excise immediately, let us discuss 3 ways to "think about" how to solve it (*not* three ways to solve it).

- *Pattern matching.* Match (transform) the problem to a solved one: "Bob" $\leftrightarrow$ "you"; "Alice" $\leftrightarrow$ "grenade"; $\Delta t_A \leftrightarrow \Delta t$ and thus $v\Delta t_B = \gamma v \Delta t \to s$.
- *Inverse search.* The distance is asked. How to calculate distance, the speed is given, so $s = vt$. Now how to get t? Which equation has t? It is a problem about relativity so should be $\Delta t_B = \gamma \Delta t_A$ (bless you get γ instead of $1/\gamma$, or use a bit pattern matching). Thus, $s = \gamma v \Delta t$.
- *Physical picture guided.* Imagine the grenade is flying a distance and then explode. You see the words but a movie is played in your mind. The time of the grenade is slowing down. So it must be running a longer time than usual, and thus a longer distance. The flight time is longer by a γ factor, and thus the distance travelled is longer by a γ factor: $s = \gamma v \Delta t$.

When preparing exams, we may have trained a lot using mainly the 1st and 2nd methods. However, the 3rd way of thinking is the major direction to go if you'd like to become an innovative physicist, or at least think as a physicist. Because:

Why physical intuition and physical picture are important

- A physical picture connects what you learn to the real world. The other two methods does not. In research you need to know the possible applications of the theory, or how to do approximations to simplify the analysis. The physical picture guides you to do that. The other two do not. Some comments from Feynman may be nice to read at this point.
- A physical picture tells you if your answer makes sense. If you made a mistake somewhere, you have a good chance to catch it early if you have a physical picture in your mind.
- Research means new problems that pattern matching does not give you good results. (But you indeed have to use pattern matching to make sure your idea is new.)
- Researchers has to define problems before solving them (unlike exam problems which are already sharply defined). A clear physical picture tells you how to define the problem. Inverse search or pattern matching cannot.
- As a researcher, one needs to talk to people. In most discussions (especially those without a blackboard around), you talk by physical pictures.
- Computers are good at pattern matching and inverse search. The rise of AI is more likely to reduce the value of your pattern matching and inverse search works, but less likely the ones based on intuition.

I hope you are convinced now: Even physics becomes less intuitive in the modern era, you should try to make it as intuitive as possible by building up physical pictures. And here are some suggestions on how to achieve that.

How to build up physical picture/intuition in modern physics?

- Try to "play a movie" in your mind about the problem. Include as many relevant physical details in the movie as possible. When you have new understandings on something, add it to the "movie".
- When you find something not intuitive, think about it again and again. You will eventually feel happier with it.
- Pay attention to paradoxes in physics, and how they get resolved.
- Think about the same problem in different ways/angles.
- Compare with our everyday life experience. Find similarities and/or key differences.
- Simplify and modularize the problem. Build up the intuition of the simplest problems as building blocks of more complicated ones.
- If there are still parts which you cannot make it intuitive, use some (as little as possible) math derivation for the moment. Try to replace it by real intuition later.

4 Length Contraction

We have just witnessed the revolution of the concept of time – time duration is relative depending on the motion of observers. Now is the turn of space. Is space interval still absolute, or it is similarly relative depending on observers?

In this section, we will arrive at the same conclusion of length contraction in 2 ways. You are expected to understand the first well. The other is optional but recommended as to understanding the same problem in different aspects.

In fact, length contraction was hypothesized way before special relativity, by FitzGerald (1889) and Lorentz (1892) to explain the Michelson-Morley experiment. For this reason it is also known as Lorentz contraction. But it is special relativity that put length contraction into a solid and general context.

The grenade revisited

Let's think about the grenade problem from a different angle. Recall wrt the ground, the grenade has flown a distance $s = \gamma v \Delta t$. Now, what the grenade "thinks"?

It still has lifetime Δt (wrt itself). Though it would think that it has flown a distance $v\Delta t = s/\gamma$. But the observer standing on the ground thinks that it has flown a longer distance (wrt a static observer). How can it be?

Grenade sliding on a ruler:
Triggered at one end and exploded on the other.
Rest length of ruler: $s = \gamma v \Delta t$
Wrt grenade: Length of ruler is $v\Delta t = s/\gamma$.

To make the situation clearer, let us put a ruler on the ground, and let the grenade slide over the ruler. The rest length of the ruler is $s = \gamma v \Delta t$.

Wrt the grenade, the ruler (and ground) is moving with v for time Δt. Thus the ruler has length s/γ wrt the grenade.

To conclude, a moving ruler is shorter by $1/\gamma$ if placed parallel to the motion direction. Also recall that if the ruler is perpendicular to the motion direction, then the length does not change.

We would have closed up the section here. But what happened during Δt time in a grenade is opaque. Let us construct a light ruler to see what happens explicitly, which can also give insight on velocity addition.

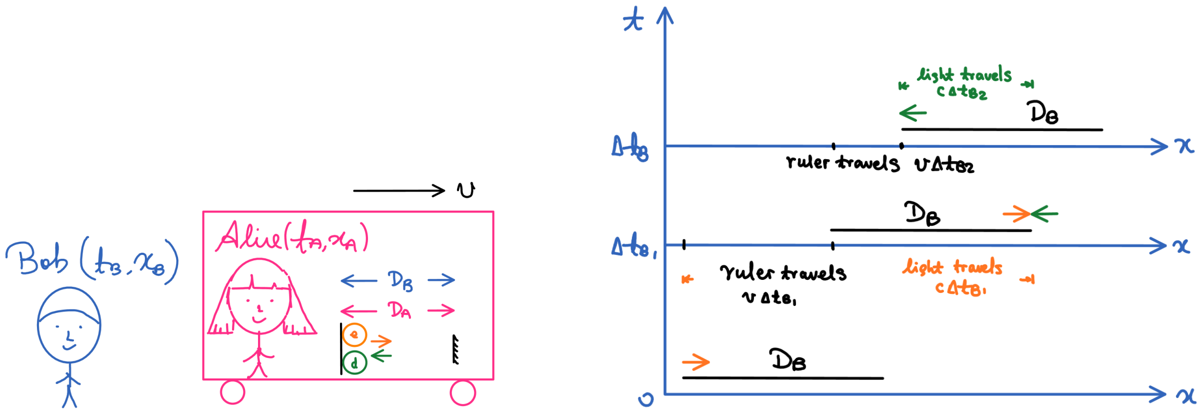
(Optional) Length contraction from the light ruler (4-step reasoning)

Although we have got the result, it is a good practice to derive it using another way. Let us construct a light ruler and study how it contracts. This will explicitly verify our promise: No matter how the ruler is defined, it should contract the same way. To see that, let us apply our familiar 4 steps:

1. In a static frame, the length of the ruler is $D = \frac{1}{2}c\Delta t$, where Δt is the time interval

between the emission and detection of light.

2 Now we load the light ruler on Alice's car. The situation and the spacetime diagram that Bob finds are in the figures below.



According to ($\mathbb{R}$), Alice consider the length of the ruler to be

$$D_A = \frac{1}{2} c\Delta t_A \ . \tag{13}$$

3 What length shall Bob get? We first divide the time interval Δt_B into two parts: Δt_{B1} and Δt_{B2} for the time interval from the emitter to the mirror and from the mirror to the detector, respectively. From the above figure (right panel),

$$c\Delta t_{B1} = D_B + v\Delta t_{B1} \ , \qquad c\Delta t_{B2} = D_B - v\Delta t_{B2} \ . \tag{14}$$

Thus

$$\Delta t_B \equiv \Delta t_{B1} + \Delta t_{B2} = \frac{D_B}{c-v} + \frac{D_B}{c+v} = \frac{2}{c}\frac{D_B}{1-\frac{v^2}{c^2}} = \frac{2}{c}\gamma^2 D_B \ , \tag{15}$$

where as always, $\gamma \equiv \frac{1}{\sqrt{1-v^2/c^2}}$. Thus

$$D_B = \frac{c}{2}\frac{\Delta t_B}{\gamma^2} = \frac{c}{2}\frac{\gamma \Delta t_A}{\gamma^2} = \frac{c}{2}\frac{\Delta t_A}{\gamma} = \frac{D_A}{\gamma} \ . \tag{16}$$

Again, we arrive at the conclusion that moving ruler contracts in the motion direction.

4 For all rulers, one should get the same conclusion as the light ruler.

Meaning of D_B

Be reminded that D_B is NOT the length of Bob's ruler. But rather, it's the length of Alice's ruler (i.e. moving with Alice) in Bob's frame. So wrt Bob (static observer), Alice's ruler (moving ruler) contracts.

(Optional) A first look at velocity addition

Let's slightly modify the light ruler: replace the light from emitter to mirror into a moving particle, with speed u_A wrt Alice. After the "mirror" gets the particle, the "mirror" still sends back a beam of light (thus it shouldn't actually be called a mirror though). Then how the light ruler experiment get modified?

- Wrt Alice, for the particle forward, $u_A \Delta t_{A1} = D_A$; and for light moving back, $c\Delta t_{A2} = D_A$. Thus,

$$\Delta t_A = \Delta t_{A1} + \Delta t_{A2} = D_A \left(\frac{1}{u_A} + \frac{1}{c} \right) \ . \tag{17}$$

- Wrt Bob, he will find the particle moving at a different velocity u_B. For the particle moving forward, $u_B \Delta t_{B1} = D_B + v\Delta t_{B1}$; and for light moving back, $c\Delta t_{B2} = D_B -$

(Optional) A speedmeter

What's the motivation to replace c with v in one way? We can ask what this apparatus can do when it's at rest. From $\Delta t = D/u + D/c$, as we have defined Δt and D, we can solve u. Thus, the apparatus is a speedmeter. No wonders that a speedmeter can tell you about speed addition.

$v\Delta t_{B2}$. Thus,

$$\Delta t_B = \Delta t_{B1} + \Delta t_{B2} = D_B \left(\frac{1}{u_B - v} + \frac{1}{c+v} \right) \ . \tag{18}$$

We already know that $\Delta t_B = \gamma \Delta t_A$ and $D_B = D_A/\gamma$. Now divide (18) by (17) (LHS and RHS, respectively), we get

$$u_B = \frac{u_A + v}{1 + u_A v / c^2} \ . \tag{19}$$

When taking $u_A \to c$ limit, we find that $u_B \to c$, consistent with ($\mathbb{C}$). In section 7, you will learn a more general version of the velocity addition formula.

(Optional) The additive rapidity

Eq. (19) looks ugly. Wouldn't nature be simpler? But who told us velocity is the best variable to describe motion? Let us define *rapidity*: $\phi(v) \equiv \text{arctanh}(v/c)$. Inserting this definition to Eq. (19), we simply get

$$\phi(u_B) = \phi(u_A) + \phi(v) \ .$$

Thus rapidity is the variable that is actually additive.

Why hyperbolic functions arises here (they originally are functions to parameterize a hyperbolic curve $x^2 - y^2 = 1$ with additive parameters, just as what trigonometric functions sin, cos, tan are functions to parameterize a circle $x^2 + y^2 = 1$ with additive angles)? Do the hyperbolic functions imply a new underlying math structure? Why motion in such a formula looks not like division of space and time, but rather hyperbolic rotation of space and time? We will return to this question in section 8.

We have got two results from the light clock now: time dilation; and after time dilation have become known, we rotate the light clock to get length contraction. Now that length contraction have become known, can we get something else?

5 Meaning of the "Same Time" (Simultaneity)

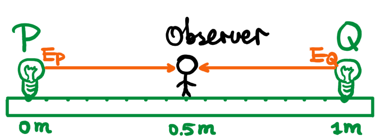
Moving ruler contracts. Now, is there a twin ruler paradox? If Alice and Bob each holds a ruler along their motion direction, and when they meet, they compare the length of their ruler, can they find each other's ruler shorter? How can that happen?

The key observation is that, to be fair, they have to compare the two ends of the ruler at the same time. Wait, do they have the same concept of the same time?

5.1 Simultaneity Depends on Which Observer

The meaning of simultaneity wrt an observer

Consider two small objects P and Q, not moving wrt each other. An observer is standing exactly at the midpoint of PQ, and not moving wrt PQ. This can be practically done with a static ruler: Let P be at the 0m point, Q be at the 1m point, and the observer at the 0.5m point.



Introduce two events: event E_P happens to object P; and event E_Q happens to object Q. For example, E_P and E_Q are the turn-on time of light bulbs at P and Q, respectively. Recall that each event happens at a particular moment.

Now we are ready to define whether E_P and E_Q happens at the same time, or one is earlier/later than the other, wrt the observer:

If the light signal from E_P and E_Q reach the observer at the same time, then E_P and E_Q happens at the same time. Otherwise, whichever reaches the observer earlier happens earlier.

Why introducing two objects?

Actual objects P, Q are needed to find an observer standing at the midpoint. Because a midpoint is defined for two points in space (P and Q), not for two events (E_P and E_Q) in spacetime.

Relating to the time of a frame

Previously, we have introduced the time of a reference frame – at different positions, time are synchronized using light signals, deducting the time used for light propagation. Here, the concept of simultaneity means that E_P and E_Q happens at the same coordinate time in this static frame.

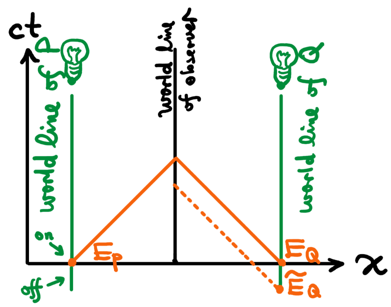
Am I wasting your time?

Oh, it looks that I am wasting your time by explaining something you already know since kindergarten. This is true. However, one little step further will

A spacetime diagram of the above scenario

Let us draw the events in a "spacetime diagram". Spacetime diagrams will turn out to be useful tools in studying relativity. On a spacetime diagram:

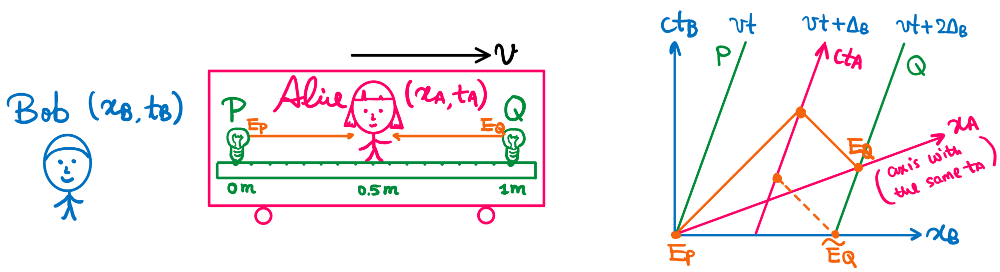
- An event is a point.
- Light travel 45° lines.
- An object (or observer, except light) is a line (called world line) with |slope| > 45° everywhere.
- An inertial observer is a straight line.
- A static object is a line parallel to the ct axis.
- Events at the same time are parallel to the x axis.



If the light rays connecting E_P and E_Q reaches the world line of the observer at the same point, they happen at the same time wrt the observer. On the contrary, $\widetilde{E}_Q$ is considered earlier than E_P wrt the observer as the light ray from $\widetilde{E}_Q$ arrives earlier at the observer.

Simultaneity is a relative concept (4-step reasoning)

1. We have defined simultaneity for a static observer with the above P-observer-Q system.
2. Let us load the system into a car and let the moving observer be Alice. Let us study the events E_P and E_Q, which happen at the same time wrt Alice. In our example, that corresponds to the light bulbs turning on at the same time at P and Q wrt Alice. See left panel of the below figure.



3. Wrt Bob, do E_P and E_Q happen at the same time? In other words, do E_P and E_Q have the same time coordinates in Bob's frame? Making use of a spacetime diagram in Bob's frame (right panel of the above figure), we immediately find that E_P is earlier than E_Q wrt Bob. On the contrary, an event $\widetilde{E}_Q$ considered to be at the same time with E_P wrt Bob, is considered earlier than E_P wrt Alice.
4. The relativity of simultaneity is not only true for the P-observer-Q system, but for all consistent definitions of simultaneity ($\mathbb{R}$).

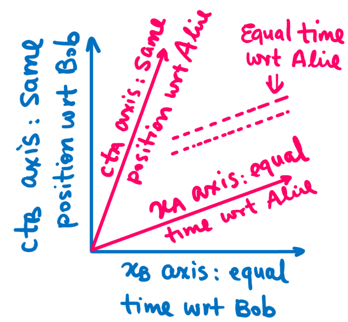
Now you should be able to resolve the puzzle of "who wrote the letter first".

Bob's spacetime diagram

- Draw Bob's axes and world lines A, P, Q.
- Draw the event that two beams of light meet Alice.
- Draw the history of these two beams of light as 45° lines.
- The intersection of light and P, Q are events E_P, E_Q.
- Compare the time of E_P and E_Q in Bob's frame.
- Draw the spacetime axes of Alice's frame in Bob's frame.

Recap: equal time slices

To condense the above analysis into one figure: When Alice is moving wrt Bob, the coordinate system of Alice drawn in Bob's coordinate system is as the figure to the right. It is similar to rotation, but both space and time axes oddly fold inwards. We will discuss this transformation (Lorentz transformation) in more details later. You should pay special attention on the equal time lines wrt Alice on this figure – different from that wrt Bob.

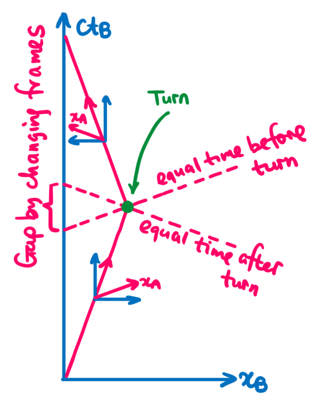


Twin paradox revisited

In our twin paradox, Alice is not in an inertial observer at the turn around time. Thus she cannot use special relativity of an inertial observer to *directly* explain her experience. However, Bob can help her to figure out what happens at the turn-around:

Before and after the turn-around, Alice is in two *different* inertial frames. Bob's age "jumps" when Alice switches from the before-turn-around frame to the after-turn-around frame. Thus in Alice's frames, there is a sudden change in Bob's age.

As mentioned, to describe what Alice sees (light arrives in her eyes), light propagation time needs to be added. I will leave the details for your exercise.



Can SR describe acceleration?

There is a common misconception saying "special relativity cannot describe the experience of an accelerating observer". This is *wrong*. The laws of special relativity (time dilation, length contraction, and more later) are expressed in inertial frames (just because the laws are mathematically simpler in inertial frames). Thus, we need to *use an inertial frame* to apply these laws to our question. However, we can *calculate in this inertial frame* what an accelerating observer sees (how light reach her eyes, for example). In this way, the experience of an accelerating observer is described *with the help of an auxiliary inertial frame.*

5.2 Causality and Types of Separations

Now we have understood: the concept of simultaneity is relative to observers. For example, Alice and Bob may consider differently on who wrote the letter first. In other words, time orders of some events (here two events: Alice writes her letter; Bob writes his letter) are reversible.

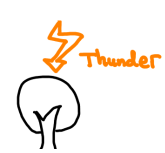
A natural question then is: Are all time orders between events reversible?

Time order associated with cause and effect

Consider an example: Lightning strikes on a tree, and then the tree dies. There are two events here:

1. Event E(strike): Lightning strikes on the tree.
2. Event E(die): The tree dies. This is an *effect*, *caused* by E(strike).

May a moving observer Ms. Bright observe that E(die) happens earlier than E(strike)?



The cause-effect relation (known as causality) is at the heart of physics. Physics is about prediction of how an initial state evolves with time. In other words, the cause-effect relation is how questions get explained in physics – "Why (effect)? Because (cause)."

As causality is so important, we hope that it is preserved in special relativity. Happily, special

Correlation

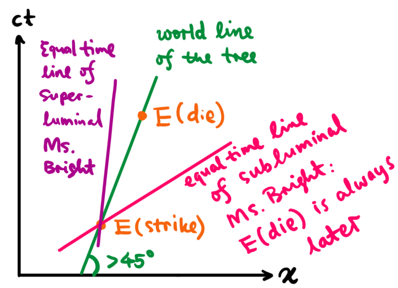
Cause-effect is a special case of correlation (conditional probability). If events E1 and E2 have nontrivial correlation, maybe

- E1 causes E2
- E2 causes E1
- Both E1 and E2 are caused by something else

relativity is indeed *consistent* with causality. Having that said, causality is an independent postulate added to special relativity. In other words, special relativity itself does not *derive* causality. Let us now see how causality and relativity work together.

Special relativity without superluminal motion is consistent with causality

Spacetime diagram of subluminal Ms. Bright and superluminal Ms. Bright. For any subluminal observer, the world line of the tree has slope greater than 45°. Then, as long as Ms. Bright moves no faster than the speed of light, she cannot flip the order of E(strike) and E(die).



Thus, if superluminal movers are forbidden, then the cause-effect order is preserved. In this course, we will assume that superluminal motion is indeed forbidden in special relativity. This is consistent with all known experiments.

More generally, no information can be sent faster than the speed of light. Because information must be sent by matter after all. And from the consistency of the theory, information can bring causal connection between events. Thus if information can be sent faster than light, then the same problems of superluminal motion can arise.

You may have an excellent question at this point: What if we push Ms. Bright so she accelerates from subluminal to superluminal? It is in fact impossible. We will come back to this point later.

In general relativity, you may hear that things can go superluminal, for example, for cosmic expansion. This is right or wrong. One has to first define velocity precisely. We will come back this point at the end of this course, in the cosmology section.

No superluminal lightning either

For causal order to be respected, not only Ms. Bright is not supposed to travel faster-than-light; but also no lightning (or any information causing the death of the tree) can be superluminal either. Otherwise causality is violated even with subluminal Ms. Bright.

Buller's poem (1923)

There was a young lady named Bright;
Whose speed was far faster than light;
She set out one day; In a relative way;
And returned on the previous night.

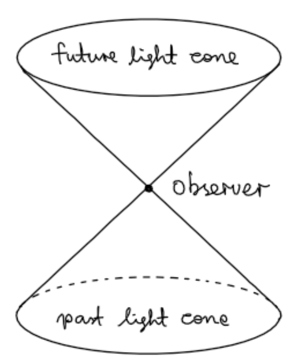
We hope this to be forbidden. Otherwise, what if Ms. Bright returned to the previous night, and locked herself in the room, how can her superluminal trip happen at all?

So far, the most convincing explanation is that, superluminal trip should be forbidden. Having that said, the possibility of a time machine is an open question for modern physics.

Causal structure of spacetime: past and future light cones

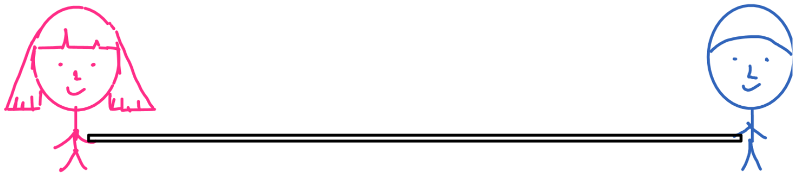
From separation of events, given a spacetime point (say, you at the present time), the spacetime is divided into three different regions according to the causal connection to the observer.

- Past light cone. This region contains all the causes of you up to now. You have not yet seen anything beyond this region.
- Future light cone. This region contains all the effects by you from now on. You can no longer change anything beyond this region.
- Outside both past and future light cones: there is no cause-effect relations between the present version of you and any point there.



No perfectly rigid body in relativity

Let's consider a thought experiment: Alice and Bob are separated by 5 light years. And they hold a 5-light-year-long rod, which is a perfect rigid body. Then can Alice and Bob send information faster than light by pushing the rod?



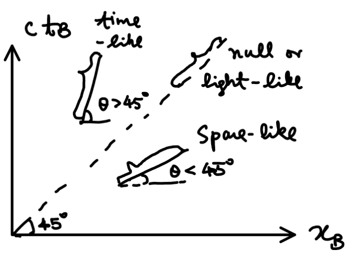
The answer is a simple straight no. Because no information can be faster than light, perfectly rigid body does not exist in special relativity.

If this answer is too brute, we can also see dynamically what happens. The rod is (usually) made of atoms and the force propagating between atoms need at least speed of light to react a push.

The speed of light limit classifies the intervals between two events into 3 classes.

Space-like, null and time-like intervals

There are three types of intervals between events.



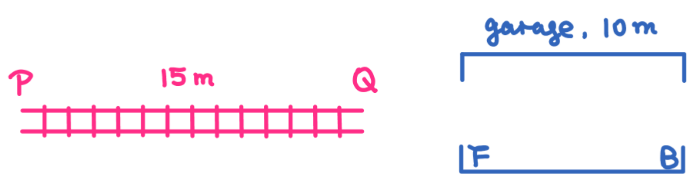
- Space-like: with slope $< 45°$. There exists observer wrt whom two space-like separated events happen at the same time. The events are thus pure space separated wrt this observer. The time order of the events can be flipped for different observers.
- Time-like: with slope $> 45°$. There exists observer wrt whom two time-like events happen at the same position. The events are thus separated only in time wrt this observer. The time order of the events is absolute and has to be agreed on for all observers.
- Light-like (or called null): the boundary separating space-like and time-like intervals, with slope 45°. Light travels with light-like lines.

6 Example: The Ladder Paradox

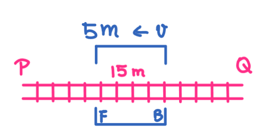
Let's apply what we have learned to get confused (or hopefully not).

The ladder and the garage

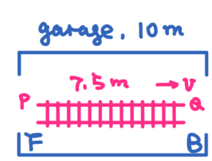
Consider a 15m long ladder (the ends are labeled P and Q). Alice holds it and moves it towards a garage, with speed 0.87c, i.e. $\gamma = 2$. The garage is 10m long, and the front door and back door are labeled F and B, respectively. Bob sits static with the garage.



What does Alice find? Alice finds that the garage moves, and thus the length of the garage is 10m/2 = 5m. Thus the ladder does not fit in the garage.



What does Bob find? Bob finds that the ladder moves, and thus the length of the ladder is 15m/2 = 7.5m. Thus the ladder fits in the garage.



Who is right?

Both are right.

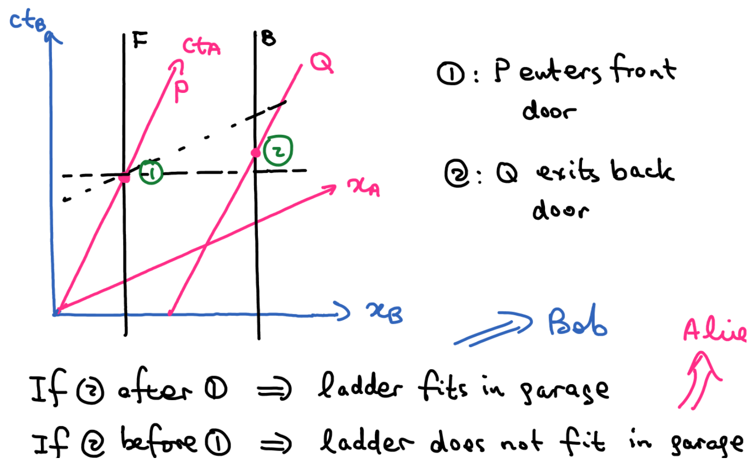
Note that "ladder fit in garage" is not an event which localized at a spacetime point. Especially, there are two important events:

1 P enters front door F.

2 Q exits back door B.

If 1 is earlier than 2, then the ladder fits in the garage. If 2 is earlier than 1, then the ladder does not fit in the garage.

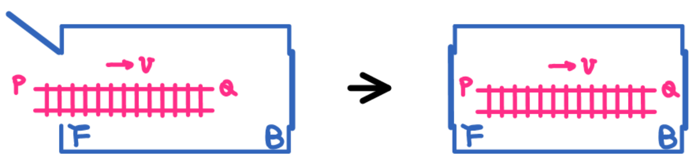
Wrt Alice, 2 is earlier than 1 (does not fit); Wrt Bob, 1 is earlier than 2 (fits).



The ladder and the garage with doors

Let's add a local event to sharpen what happens. Let the back door of the garage always be closed. Bob closes the front door once he finds the ladder moves in the garage.

Here we assume that Alice is still in an internal frame, no matter any part of the ladder may stop or not.



Now even Alice has to agree that the ladder is completely in the garage because the doors are closed. So the "paradox" becomes sharper: Alice thinks that the ladder is longer than the garage; but it is now completely in the garage!

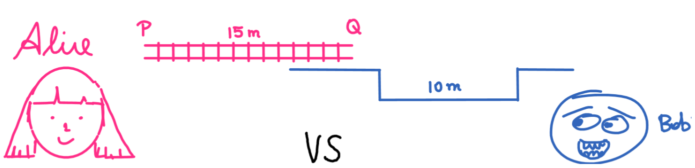
Resolution: Remember that in relativity, there is no perfectly rigid body.

Assuming that the door is relatively stronger and does not break, then the ladder must fall into parts. Wrt Alice: when Q stopped moving, the "stop" information has to be passed to P and it takes at least time 15m$/c$. During this time, the front door moved about 13m. Thus P is well inside F by this time. Of course, it fits into the 5m garage very well.

On the other hand, if the ladder is relatively stronger and does not break, then the back door is broken and there is no difference from no back door.

Replacing the garage by a trap

Let us further replace the binary doors by boundary of a trap.

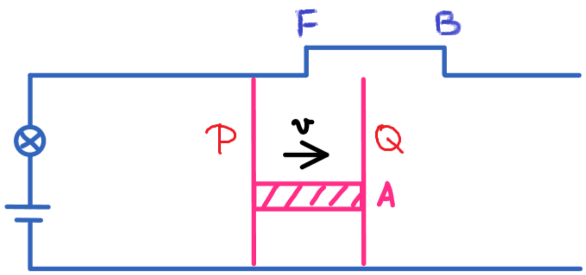


Bob' considers that Alice with the ladder start to fall in when the ladder is entirely in the trap. Thus Alice falls in. However, Alice considers herself still not yet fall-in before the Q point escapes. Who's right?

This really depends on how Alice holds the ladder. If Alice holds the ladder perfectly parallel to the ground and sliding on the ground, then Alice falls in. Because there is no perfect body. When Q falls into the trap, Q inevitably falls downwards. However, if Alice holds the ladder in the way a bit upwards, such that Q does not need ground support from the beginning, then Alice can escape.

Replacing the garage by a circuit

Let us stop destroying things and go electronic. Consider the situation below.



Alice holds two moving conducting wires (they are isolated wrt each other). Both can close the circuit. Wrt Bob, there exists a time when neither of Alice's wires closes the circuit so the bulb shuts off for a moment. Wrt Alice, at least one wire touches the circuit all the time. So shall the bulb shut off?

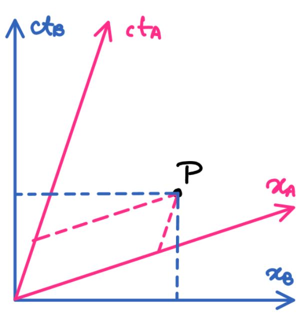
The key to resolve the problem is that, electricity needs time to conduct. The speed of electricity is comparable to c, but still less than c. Thus wrt Alice, to keep the bulb on without a off-time gap, the electric current has to reach F before Q passes F. So there is an off-time wrt Alice, too.

7 The Lorentz Transformation

Alice and Bob need a rule for their coordinate transformations

An event P happened. Both the moving Alice and the standing Bob agree on the existence of the event. However, the same event has different coordinates in Alice's and Bob's frames.

Given the coordinate of P: (t_B, x_B) in Bob's frame, how to get the coordinate (t_A, x_A) for the same event in Alice's frame?

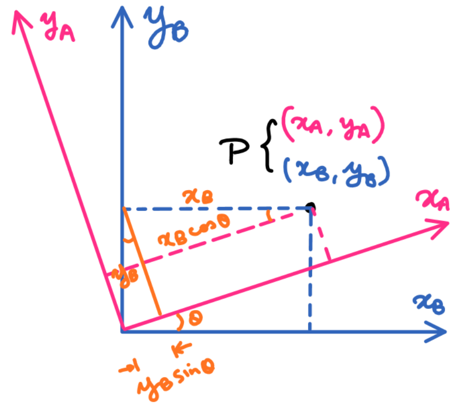


Recall that the transformation between inertial frames is called a boost. Thus the question can also be asked as: how does coordinates transform under a boost?

Before solving this problem, let us draw similarity between boost and rotation.

The case of rotation: Alice and Bob are not moving wrt each other, but facing different directions. Given Bob's coordinate of an object (x_B, y_B), how to get Alice's coordinate for the same object (x_A, y_A)? Following relations of Euclidean geometry, we get:

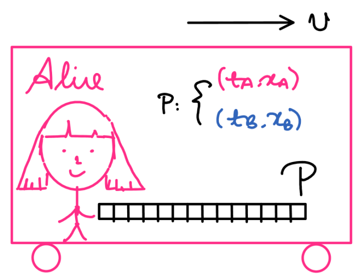
$$\begin{cases} x_A = x_B \cos\theta + y_B \sin\theta \\ y_A = -x_B \sin\theta + y_B \cos\theta \end{cases} \tag{20}$$



I hope you get a better idea about the question by this analogue. Also, in the next section we will find a surprising similarity between these space-time and space-space transformations.

Space transformation of a boost

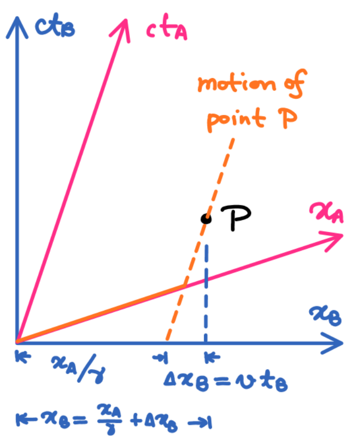
Given Bob's coordinate (t_B, x_B), now we would like to solve x_A. How to do that? An equivalent question is that, how to physically describe x_A? Note that x_A is Alice's spatial coordinate. Thus, if Alice carries a ruler (static wrt Alice) with length x_A, Alice holds one end, then the other end has coordinte x_A.



For simplicity, we choose the same origin of coordinate for Bob and Alice. Given Bob's coordinate of P: (t_B, x_B), how can we find the relation $x_A(t_B, x_B)$?

Wrt Bob, the moving ruler has length x_A/γ. Thus at Bob's time $t_B = 0$, the point P has $x_B = x_A/\gamma$. However, at $t_B \neq 0$, the end point P has also moved a distance vt_B. Thus for general t_B, $x_B = x_A/\gamma + vt_B$. In other words, we get $x_A(t_B, x_B)$ (i.e. x_A as a function of t_B and x_B) as

$$x_A = \gamma\,(x_B - vt_B) \ . \tag{21}$$



How to get $t_B(t_A, x_A)$? We can flip Alice's and Bob's role and do the calculation again. But in fact we do not have to do it. Simply following $(\mathbb{R})$, we get

$$x_B = \gamma\,(x_A + vt_A) \ . \tag{22}$$

Velocity in a general direction

Be reminded that the relative velocity v between frames is only along the x direction (and positive for Bob). If you are dealing with velocities in a general direction $\mathbf{v} = (v_x, v_y, v_z)$, you can first do a rotation to rotate it to the x direction before applying the transformation.

Lorentz vs Einstein

The Lorentz transformation is named after Lorentz for his work during 1892-1904. In other words, the Lorentz transformation is known before Einstein's special relativity (1905). It was discovered as mathematical symmetries

Time transformation of a boost

The remaining question is: How to get $t_A(t_B, x_B)$ and $t_B(t_A, x_A)$? We do not have to do more thought experiments since we can solve them from equations (21) and (22).

Inserting (21) into (22) to eliminate x_A,

$$x_B = \gamma\left[\gamma\left(x_B - vt_B\right) + vt_A\right] . \quad \to \quad t_A = \gamma\left(t_B - \frac{v}{c^2}x_B\right) . \tag{23}$$

Similarly, or simply using the principle of relativity, we get

$$t_B = \gamma\left(t_A + \frac{v}{c^2}x_A\right) . \tag{24}$$

Summary: the Lorentz transformation

Let's summarize the transformation we have got, and add back the trivial y and z directions. These rules to transform between inertial frames is known as the Lorentz transformation. Here $\beta \equiv v/c$.

$$\begin{cases} ct_A = \gamma(ct_B - \beta x_B) \\ x_A = \gamma(x_B - \beta ct_B) \\ y_A = y_B \\ z_A = z_B \end{cases} \qquad \begin{cases} ct_B = \gamma(ct_A + \beta x_A) \\ x_B = \gamma(x_A + \beta ct_A) \\ y_B = y_A \\ z_B = z_A \end{cases} \tag{25}$$

The Lorentz transformation describes the mathematical structure of special relativity. Thus our known results can be directly read-off from the formulas (25).

The Lorentz symmetry

Under the appearance of a transformation, a more physical name to refer to the equations (25) is the Lorentz symmetry: Alice and Bob are symmetric, in that though they use different coordinates to describe events, they observe the same laws of nature (for example, equations of motion for particles). Symmetry is now considered the first principle in fundamental physics. We will come back to the concept of symmetries in the part of action principle.

Time dilation, length contraction and simultaneity revisited

- Simultaneity: Consider two events E_1 and E_2, happened at the same time wrt Alice, i.e. $t_A^{E_1} = t_A^{E_2}$. Eq. (25) gives $0 = c(t_A^{E_1} - t_A^{E_2}) = \gamma[c(t_B^{E_1} - t_B^{E_2}) - \beta(x_B^{E_1} - x_B^{E_2})]$. Clearly, if the two events happens at different locations and if $\beta \neq 0$, then the two events are not at the same time wrt Bob.
- Time dilation: A clock is moving wrt Alice with coordinate $(t_A, 0)$. Note that $x_A = 0$ here since Alice carries the clock thus the clock is always located at the origin wrt Alice. From the first equation to the right panel of (25), we get $t_B = \gamma t_A$.
- Length contraction: A ruler is moving together with Alice. How is Bob supposed to measure its length? Bob should measure his coordinates of both ends of the ruler *at the same time*. The left end of the ruler is at $x_B = 0$ when $t_B = 0$. Thus the coordinate of the ruler at $t_B = 0$ is the length of the ruler. From the second equation to the left panel of (25), we get $x_B = x_A/\gamma$.

Now we are ready to solve a problem: What's wrong with the Newtonian velocity addition rule $\mathbf{v}_B = \mathbf{v} + \mathbf{v}_A$ when $v_A \sim c$? In the optional material of the length contraction section, we studied a special case. Here we will derive general formulas.

Why don't you tell us earlier?

You may feel furious here:

"I have spent great efforts in understanding time dilation, length contraction and simultaneity. But now they follow so simple equations. Why don't you start from telling us the Lorentz transformation and derive everything from there? That would have saved a lot of my time."

I am in fact pretty careful in choosing the point to introduce Lorentz transformation.

If I introduce it earlier, these effects will be just cold math formulas in your mind instead of living characters (i.e. with a physical picture).

On the other hand, I can teach velocity addition, energy and momentum first and put the Lorentz transformation to the very end. But then the physical scenarios become too complicated to be helpful. I thus choose here to be the point to introduce this powerful tool.

Addition of velocity

For simplicity, let us take the relative velocity $\mathbf{v}$ to be $(v, 0, 0)$, along the x-direction, as we have always assumed.

Now Alice throw a ball with velocity

$$\mathbf{v}_A = (v_{Ax}, v_{Ay}, v_{Az}) = \left(\frac{dx_A}{dt_A}, \frac{dy_A}{dt_A}, \frac{dz_A}{dt_A}\right) . \tag{26}$$

and Bob finds

$$\mathbf{v}_B = (v_{Bx}, v_{By}, v_{Bz}) = \left(\frac{dx_B}{dt_B}, \frac{dy_B}{dt_B}, \frac{dz_B}{dt_B}\right) \ . \tag{27}$$

Now use the rule of Lorentz transformation, one can calculate

$$v_{Bx} = \frac{dx_B}{dt_B} = \frac{\gamma d(x_A + vt_A)}{\gamma d\left(t_A + \frac{x_A v}{c^2}\right)} = \frac{dx_A + vdt_A}{dt_A + \frac{v}{c^2}dx_A} = \frac{v_{Ax} + v}{1 + \frac{v_{Ax}v}{c^2}} \ . \tag{28}$$

$$v_{By} = \frac{dy_B}{dt_B} = \frac{dy_A}{\gamma d\left(t_A + \frac{v}{c^2}x_A\right)} = \frac{v_{Ay}}{\gamma\left(1 + \frac{v_{Ax}v}{c^2}\right)} \ . \tag{29}$$

Similarly,

$$v_{Bz} = \frac{v_{Az}}{\gamma\left(1 + \frac{v_{Ax}v}{c^2}\right)} \ . \tag{30}$$

Example: velocity addition in perpendicular directions

When $\mathbf{v}_A = (0, u, 0)$, perpendicular to the relative motion direction, we get

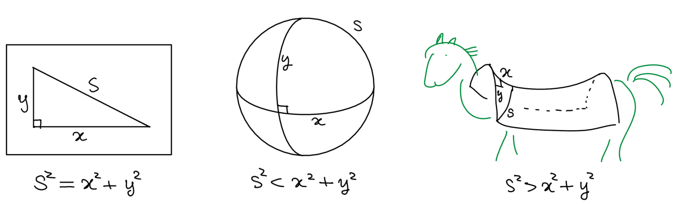
$$\mathbf{v}_B = (v, u/\gamma, 0) \ . \tag{31}$$

Note that Alice and Bob measures different velocity even in the perpendicular direction, because their time intervals are different (while the length intervals are the same).

8 The Geometry of Spacetime

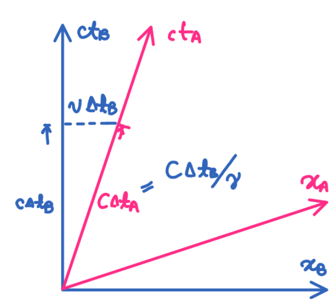
Pythagoras theorem and modern geometry

Since Gauss and Riemann, mathematicians realize that different types of geometry can be classified by how the Pythagoras theorem appears on those geometries. For example, on flat, spherical and hyperbolic surfaces, the Pythagoras theorem appears differently.



More careful studies on how the spherical and hyperbolic surfaces are curved, one can make the above expressions more precise and differentiate the relation twice to define spatial curvature.

Here we are not following this path to study pure spatial geometries in depth. Rather, we would like to ask the following question: Now that space and time are "unified" by Lorentz transformation, what does the spacetime geometry look like? Or what does the Pythagoras theorem look like for spacetime?



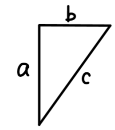
Our spacetime is NOT Euclidean

Let us reconsider time dilation. Alice carries a clock and both Alice and Bob measure the tick-tock interval.

Wrt Bob, $\Delta t_B = \gamma \Delta t_A$. What does this imply on the spacetime diagram?

Naively, from the geometry, we would have expected $c\Delta t_A > c\Delta t_B$, because $c\Delta t_A$ is the hypotenuse of the right triangle. However, from $\gamma > 1$, we see $c\Delta t_A < c\Delta t_B$. How is this possible?

As we have discussed, we cannot take the Euclidean geometry for granted. Especially, in the right figure, where a and c are time directions (of Alice or Bob), the Pythagorean theorem $a^2 + b^2 = c^2$ no longer holds. Instead, for time dilation,



$$-a^2 + b^2 = -c^2 \ . \tag{32}$$

Extra minus sign emerges in front of the square of time, but not space.

Is it a coincidence, or is it in general a new type of geometry?

Note that minus sign in time squared indicates an extra factor of i in front of time to match the Pythagoras theorem of flat space.

Our spacetime is ALMOST Euclidean

To see what has happened in another way, let us rewrite the Lorentz transformation using rapidity $\phi = \text{arctanh}\beta$. Note that $\cosh\phi = \gamma$, and $\sinh\phi = \beta\gamma$. The Lorentz transformation is then

$$\begin{cases} ct_B = ct_A \cosh\phi + x_A \sinh\phi \\ x_B = ct_A \sinh\phi + x_A \cosh\phi \end{cases} \tag{33}$$

Do you find it a bit similar to rotation?

Hyperbolic functions

$\cosh x \equiv \frac{e^x + e^{-x}}{2}$

$\sinh x \equiv \frac{e^x - e^{-x}}{2}$

$\tanh x \equiv \sinh x / \cosh x$

$\cosh^2 x - \sinh^2 x = 1$

$\cosh(i\theta) = \cos\theta$

$\sinh(i\theta) = i\sin\theta$

Now let's apply a mathematical trick. Define

$$\phi \equiv i\theta \ , \qquad ct \equiv iw \ . \tag{34}$$

We then get

$$\begin{cases} iw_B = (iw_A)\cos\theta + x_A(i\sin\theta) \\ x_B = (iw_A)(i\sin\theta) + x_A\cos\theta \end{cases} \quad \Rightarrow \quad \begin{cases} w_B = w_A\cos\theta + x_A\sin\theta \\ x_B = -w_A\sin\theta + x_A\cos\theta \end{cases} \tag{35}$$

This is *exactly* a rotation of the w-x plane, with w considered as an extra spatial dimension!

Do not ask me the "physical meaning" of the imaginarized time w at the moment. Just consider it as a trick in math. In fact, there are profound implications of imaginary time in quantum field theory (zero temperature and finite temperature). But we are far not ready to introduce them here.

To summarize: A mathematical trick of redefining variables makes Lorentz transformation identical to rotation.

The symmetry of spacetime with imaginary time

According to Pythagoras (570BC-495BC), sphere is the most beautiful and perfect object – it is symmetric under 3 dimensional rotations.

In this respect, our spacetime is even more beautiful than the most beautiful object. This is because, now w, x, y, z directions are no different from each other – they are all related by rotation. Our spacetime is symmetric (i.e. laws of nature have the same form) under rotations in planes w-x, w-y, w-z, x-y, x-z, y-z, and all kinds of combinations of them. It is indeed more symmetric than a sphere because a sphere is only invariant under rotations in x-y, x-z, y-z planes and their combinations.

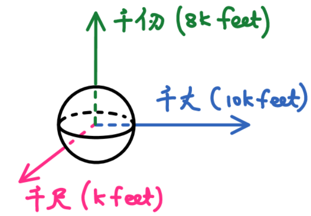
We are not saying that spacetime is a sphere. Rather sphere is curved and our spacetime is flat in special relativity. However, its symmetry includes that of a sphere, and more (there are 4 additional spacetime translations).

This further confirms that the difference between space and time is imaginary – once you use imaginary time, time becomes no different from just another spatial direction, and the 3 boost along x, y and z directions just become three additional rotations.

Natural units – the units of nature

In physics, some units are more natural than the others, because they use the same unit for the same things (or apparently different things, but with the same physical origin).

Let's consider an example of a not-so-natural unit. In ancient China, sometimes people use a different unit for height of mountains. At that time, height seems different from other directions because of gravity. But now we see that the difference is due to the presence of the earth, which spontaneously break the symmetry of empty space. Fundamentally, height has no difference from the other two directions, and they can be related by rotation. Thus, we had better to use the same unit to measure height and the rest of the spatial dimensions.



Now that we have seen space and time are not that different and can be related by Lorentz transformation, why we still give space and time different units? We don't have to.

To use the same unit for space and time, we can just set $c = 1$ and then speed is dimensionless.

This is known as the natural unit (together with $\hbar = k_B = 1$). Natural unit is widely used in theoretical high energy physics.

We shall not use natural units in this class and will always keep c explicit.

Natural vs convenient

The laws of nature are shorter in natural unit as unnecessary conversion of units are avoided. However, whether it is a convenient choice of unit depends on the physical context that we study. If we study explicit dynamics of systems with $v \ll c$, then we had better to keep c explicit because otherwise we will deal with large/small numbers everywhere. For example, for $v = 10^{-8}c \simeq 3\text{m/s}$, in natural unit the object only moves 10^{-8} unit of space in one unit of time.

The invariant interval

In Euclidean space,

$$ds^2 = dw^2 + dx^2 + dy^2 + dz^2 \ . \tag{36}$$

Recall that $w = -ict$. To bring us back to real time, we thus have

$$ds^2 = -d(ct)^2 + dx^2 + dy^2 + dz^2 \ . \tag{37}$$

In some books, the interval is defined as $ds^2 = d(ct)^2 - dx^2 - dy^2 - dz^2$. This is just a different convention without physical difference (corresponding to replacing all ds^2 here into $-ds^2$).

The importance of ds^2 is that, it is an invariant quantity under Lorentz transformation. To see that, one may first use $\{w, x, y, z\}$ and note that rotation keeps radius invariant. Or one can use the Lorentz transformation matrix to verify it explicitly. Just as (ω, x, y, z) forms a vector in 4d Euclidean space, $x^\mu = (ct, x, y, z)$ $(\mu = 0, 1, 2, 3)$ also form a 4 dimensional vector (4-vector for short), which lives in the so called *Minkowski space* – which is just the mathematical name of our spacetime.

In fact, our spacetime is more complicated than the Minkowski space – in general relativity, you will see that our spacetime is curved instead of totally flat. This is considered to be more general spacetime but still with some Minkowski signature.

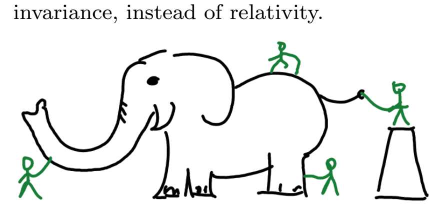
In fact, "relativity" is not a good name to the theory. By its name, it appears to emphasize that things are relative – no longer as invariant as we thought.

But in fact, the essence of relativity is the other way round: Despite of being viewed from different perspectives (wrt different observers in different inertial frames), the events and their intrinsic relations are invariant.

Relativity is just the technology of study. But the spirit of the theory is the underlying invariance.

Recall the principles ($\mathbb{R}$), ($\mathbb{C}$) and ($\mathbb{E}$). Every each principle tell you a piece of invariance, instead of relativity.

This is just like how you prove a theorem in Euclidean geometry – it is not important if the figure is placed vertically or horizontally (i.e. wrt observers with different viewing angles). What's important is the intrinsic relation between the geometrical objects.



Another analogue is the story of blind men and an elephant. The goal of physics is to understand the elephant, instead of how different men feel differently by not being able to feel the whole thing.

Time-like, space-like and null from the interval

The sign of ds^2 corresponds to different types of separations of events:

$$\begin{aligned} ds^2 < 0 &\quad\Leftrightarrow\quad \text{time-like} \\ ds^2 > 0 &\quad\Leftrightarrow\quad \text{space-like} \\ ds^2 = 0 &\quad\Leftrightarrow\quad \text{light-like, null} . \end{aligned} \tag{38}$$

And now you know why null is called null – it is indeed null.

(Optional) A metric of spacetime

A metric is a "standard" to measure coordinate distance. For our spacetime in special relativity, the metric $g_{\mu\nu}$ (as a symmetric 4×4 matrix), infinitesimal coordinate distances and the invariant interval can be related as

$$ds^2 = \sum_{\mu,\nu=0}^{3} g_{\mu\nu} dx^\mu dx^\nu \ , \tag{39}$$

where

$$x^\mu = \{ct, x, y, z\} \ , \qquad g_{\mu\nu} = \begin{pmatrix} -1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} . \tag{40}$$

This seems to be a trivial rewriting of what we have obtained. However, here we have separated two different things:

- Coordinates x^μ: What are the directions of spacetime?
- Geometry: What kind of geometry does spacetime satisfy?

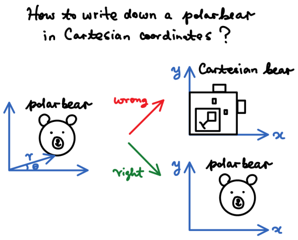
Even if we use a different coordinate system, for example, spherical coordinate for the spatial part, we can still choose a metric to keep ds^2 invariant:

$$x^\mu = \{ct, r, \theta, \phi\} \ , \qquad g_{\mu\nu} = \begin{pmatrix} -1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & r^2 & 0 \\ 0 & 0 & 0 & r^2 \sin^2\theta \end{pmatrix} . \tag{41}$$

In fact, the advantage brought about by a metric is not only the freedom to choose coordinates. Rather, it can describe the intrinsic curvature of spacetime which is not characterized by a different coordinate system. But this is beyond the scope of special relativity. You will see later that in general relativity, the metric $g_{\mu\nu}$ is the most important object to study. It can get curved by matter, and can guide matter how to move.

(Optional) Inner product of 4-vectors

The invariant internal ds^2 can be considered as the inner product of 4-vector dx^μ with itself. In general, we can use the metric to measure the inner product to measure the inner product of any 4-vectors. For two 4-vectors p^μ and k^ν, their inner product $\sum_{\mu\nu} g_{\mu\nu} p^\mu k^\nu$ is Lorentz invariant (i.e. invariant under Lorentz transformation).



9 Relativistic Momentum and Energy

Before discussing energy and momentum in relativity, let us first recall why we need them at all. In Newtonian mechanics, it is nice to have energy and momentum because they are

Transforming Slimes

Suppose there are slimes of 3 shapes: △, □ and ○. When conserved. (See the below box for why conservation laws are good after all.) Thus in relativity, we should check if the Newtonian definition of energy and momentum are still conserved; and if not, how to generalize the definition of energy and momentum, to make them still conserved.

What's good about conservation laws?

- Physically, conservation laws are the magic that we do not need to care what's happening in the middle. Given the initial state, we can immediately make a number of predictions (no greater than the number of conservation laws in the system).
- Mathematically, the equations of motion (like a set of $m\ddot{x} + V' = 0$) are second order ordinary differential equations (ODEs) in x which could be difficult. The conservation laws turns some of these into 1 or 0 order ODEs in x which are much simpler (e.g. $\frac{1}{2}m\dot{x}^2 + V = E$, or charge conservation).
- Practically, in modern physics, the objects being studied are often extremely fast ($v \sim c$), small ($\Delta x \times \Delta p \sim \hbar$), heavy ($GM/(rc^2) \sim 1$), early (less than a second after the big bang), etc. It is often hard to directly build experiments to directly monitor exactly what is happening. Instead, one often needs conserved quantities to make useful observations. For examples, on modern colliders such as the Large Hadron Collider, the trajectories of particles (they are both small and fast in the above sense) are studied at a much later time compared to the time where interaction happened. If charge, energy or momentum were not conserved, it becomes much harder to study such particle physics processes (in fact, it will become not clear how to define observables even in principle without the help of energy and momentum conservation).

That's cool. But why are we still lucky to have energy and momentum conservation in special relativity? In a later part "From Action to Laws of Nature", we will show you a positive answer – the existence of these conservation laws are not a result of luck, but a result of the fact that no time moment or spatial location is special. This applies in special relativity as well as the Newtonian mechanics. Wait for that part if you are curious about the origin of these conservation laws.

9.1 Relativistic Momentum

Can we still use Newtonian momentum conservation?

First, let us recall how the Newtonian momentum conservation are consistent with ($\mathbb{R}$) within Newtonian mechanics – forget relativity for a moment here and assume $v \ll c$.

For example, for Alice's momentum conservation for two particles:

$$m_1\mathbf{v}_1^A + m_2\mathbf{v}_2^A = m_1\mathbf{v}_1'^A + m_2\mathbf{v}_2'^A \ . \tag{42}$$

What will Bob find? In Newtonian mechanics, mass of the particles do not change. And velocity simply adds. Thus for Bob with a relative velocity $\mathbf{v}$, $\mathbf{v}_1^B = \mathbf{v}_1^A + \mathbf{v}$, etc. Thus, equation (42) can be consistently written into the same form in Bob's frame:

$$m_1\mathbf{v}_1^B + m_2\mathbf{v}_2^B = m_1\mathbf{v}_1'^B + m_2\mathbf{v}_2'^B \ . \tag{43}$$

Happily, all the terms with $\mathbf{v}$ cancel in the above equation. So it takes the same form as (42). Thus, ($\mathbb{R}$) is indeed satisfied.

Now, let us take relativity into account. For simplicity, let us assert that the mass of the particles still does not change with velocity. *Assuming* that Alice still has the momentum conservation equation (42), then what will Bob find?

Following the velocity addition rule, Bob finds (we only write the term for the first

What if mass depends on velocity

Indeed, we can assume that mass depends on velocity. We will get essentially the same relativistic particle explicitly to emphasize the key points). Let the relative velocity between Alice and Bob be along the x direction as usual. Then in the x direction, we have:

$$m_1 \frac{(v_{1Ax} + v)}{(1 + \frac{v_{1Ax} v}{c^2})} + \cdots = m_1 \frac{(v'_{1Ax} + v)}{(1 + \frac{v'_{1Ax} v}{c^2})} + \cdots \ . \tag{44}$$

In the y direction (and z direction has a similar equation), we have:

$$m_1 \frac{v_{1Ay}}{\gamma(1 + \frac{v_{1Ax} v}{c^2})} + \cdots = m_1 \frac{v'_{1Ay}}{\gamma(1 + \frac{v'_{1Ax} v}{c^2})} + \cdots \ . \tag{45}$$

Unhappily, these equations *do not* take the same form as (42), because of the red-colored factors. Thus, (42) takes different form wrt Alice and Bob. It cannot be a piece of law of nature in special relativity.

The same argument applies to the Newtonian energy.

Now that we have seen the Newtonian energy and momentum are not conserved quantities (and the unexplained belief that there are still conserved energy and momentum after all), the best we can do is to search for new quantities in special relativity, where in the $v \ll c$ limit the quantity returns to energy and momentum in the Newtonian mechanics.

In fact, we only need the energy and momentum *differences* to return to Newtonian mechanics when $v \ll c$. You will find a surprise about energy in this respect soon.

Using a proper time to fix momentum conservation

How to fix the inconsistency of (44) and (45) with ($\mathbb{R}$)?

If we can remove the red factor, we should be able to get a equation satisfying ($\mathbb{R}$). How? We recall that the red factor arises because

$$dt_B = d\left[\gamma\left(t_A + \frac{x_A v}{c^2}\right)\right] = \gamma(1 + \frac{v_{1Ax} v}{c^2}) dt_A \ . \tag{46}$$

In other words, every observer has their own time. Measuring motion with individual observer's own time ($d(\cdots)/dt_A$ and $d(\cdots)/dt_B$) is the root of evil.

How can we get rid of the observer's own time in defining momentum? Or, is there an observer-independent way to measure a time-like interval?

As I am trying to rephrase the question to approach the answer, now you should be able to figure it out: we have learned an observer-independent way to measure intervals using ds^2. To make a real number with correct time dimension, we define the *proper time*:

$$d\tau \equiv \sqrt{-ds^2/c^2} = \sqrt{dt^2 - \frac{dx^2}{c^2}} = \frac{dt}{\gamma} \ . \tag{47}$$

This is the time variable that we need. Thus, the momentum which gives a relativistic generalization of momentum conservation is

$$\mathbf{p} = m \frac{d\mathbf{x}}{d\tau} = \gamma m \mathbf{v} \ . \tag{48}$$

From the first equal sign of equation (48), one observes that $\mathbf{p}$ transforms in the same way as $\mathbf{x}$ under Lorentz transformation. This is as promised that momentum conservation takes the same form in different frames.

However, this is confusing from another aspect: we know that space ($\mathbf{x}$) and time (ct) has combined into spacetime ($ct, \mathbf{x}$) in relativity as one entity. Now that momentum transform the same as space, is there a counterpart comes together which transform

Physical meaning of proper time

Why a time variable can be defined *independently of* an observer? Haven't we said that there is no absolute time?

Because there exists a moving object (it moves, therefore it is). The moving object defines its own frame and thus defines its own time. In the own frame of the moving object, $d\mathbf{x} = 0$. Thus indeed $d\tau = dt$. This verifies that the proper time is the time measured by the moving object itself. Imagine a clock moving together with the object. The proper time is the interval read from this clock.

the same as time (ct)? What do we get if we naively replace $\mathbf{x}$ with ct in (48)? We get

$$(\text{the naive time-like counterpart of } \mathbf{p}) = m\frac{d(ct)}{d\tau} = \gamma mc \ . \tag{49}$$

What's this? The nature should not leave it unexplained! We will return to this question in the next subsection.

Caution: Here we have only showed that the definition (48) is consistent with ($\mathbb{R}$). We have not proved for you the actual conservation. In Newtonian mechanics, momentum conservation is derived from the 3rd law (which is an independent assumption from the Newtonian 1st and 2nd laws). Here we did not impose a 3rd law (although we can do so) and thus we do not have the right tool to derive momentum conservation. In fact, given an action, the relativistic momentum conservation can be derived from an action principle. We will not expand this point here.

The relativistic force

Now that there is momentum, it is straightforward to define force:

$$\mathbf{F} \equiv \frac{d\mathbf{p}}{dt} = \frac{d(\gamma m\mathbf{v})}{dt} = \gamma m\dot{\mathbf{v}} + \gamma^3 m\mathbf{v}(\mathbf{v}\cdot\dot{\mathbf{v}})/c^2 \ . \tag{50}$$

Here we use dt instead of $d\tau$. This definition is convenient, because it is intuitive to think of force as one static observer pushes an object and see its acceleration – how much additional efforts the static observer has to consume to push the rocket further.

Note that when $v \to c$, it takes $F \to \infty$ to change v. As a result, one can never accelerate a subluminal object to speed of light (or beyond). Thus c is indeed the speed limit.

9.2 Relativistic Energy

The relativistic kinetic energy

The relativistic force implies work to accelerate objects, and thus the kinetic energy. Consider an object initially at rest at location $x = 0$, and is accelerated by a force in the x-direction, into a state with position x and speed v. The work that the force has done turns into the kinetic energy of the object:

$$K = \int_0^x F\ dx = \int_0^x \frac{d(\gamma mv)}{dt}\ dx = \int_0^v v\ d(\gamma mv) = (\gamma-1)mc^2 \ . \tag{51}$$

In the second-to-last step we have converted the integration variable to v, and at the final step, we have performed integration by parts. The detail is left as an exercise.

A few comments in order before to proceed:

- When $v \ll c$ (so that we can do a Taylor expansion around $v \to 0$), the relativistic kinetic energy returns to our familiar Newtonian form as expected: $K \to \frac{1}{2}mv^2$.
- We have discussed kinetic energy following an acceleration process in the x-direction from rest. Is it general enough? Since kinetic energy is a label of a state, it should not care how it is obtained in the history of the object. Thus the kinetic energy is general enough no matter the force is not in the x-direction, or the object has an initial velocity.

Chasing the light?

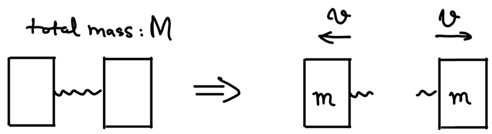
When Einstein was 16 years old, he started to dream what would happen if he can run as fast as light. Would light stop oscillating and would that contradict Maxwell's theory of E&M?

Now his dream has came to an end. No one can be accelerated to the speed of light and thus this seemingly contraditive thought experiment would never happen.

Now that we have the kinetic energy, what's the total energy? Before that, let us first study mass. In Newtonian mechanics (and chemistry), mass is conserved – the total mass before and after a reaction is the same. What about in relativity?

Mass is no longer conserved

We study a system split into two objects. Consider two objects connected by a spring. The spring is compressed and stores some potential energy. Initially the objects are at rest. The initial mass of the whole system is M. After releasing the spring, the two objects move apart, each has mass m and moving apart with speed v.



In Newtonian mechanics, there is no doubt that $M = 2m$. But is it true when $v \sim c$?

To study the property of the system, we add a small probe velocity in the y direction $v_y \ll v$ (One can comfortably take v_y to be infinitesimal). Then interestingly, momentum conservation in the y direction can calculate M for us.

- Before the split: $p_y = Mv_y$.
- After the split: $p_y = 2\gamma m v_y$ for the same p_y (momentum conservation), where $\gamma = 1/\sqrt{1 - v^2/c^2}$ accounts not only velocity in the y direction, but the the total velocity.

Thus, $M = 2\gamma m > 2m$! The good old mass conservation breaks down.

Does v_y change?

No. Because we can view the event in another frame with $v_y = 0$ at the beginning. Then in this frame $v_y = 0$ at the end. Now switch to a frame with a small $v_y \ll c$. Using the velocity addition rule, v_y does not change. (And the in x direction v_x only get a small correction of order $1/\sqrt{1 - v_y^2/c^2} - 1 = \mathcal{O}(v_y^2/c^2)$.)

The relativistic rest energy

Consider the above $M \to 2m$ split process. Let's consider the limit $\gamma \gg 1$, i.e. the final objects are flying very close to the speed of light. In this limit, the kinetic energy of the final objects are $K \simeq 2\gamma mc^2$.

From energy conservation, the kinetic energy of the final objects is contained in the initial object. Thus, the energy of the initial object is

$$E = 2\gamma mc^2 = Mc^2. \tag{52}$$

Recall that the initial object is at rest. Thus, the energy here is the rest energy of an object.

The two final objects also has their rest energies, mc^2 each. In our $\gamma \gg 1$ limit, these rest energies can be neglected. In fact, initially Einstein used light to derive the relativistic energy, corresponding to no rest mass. But the derivation needs more understanding of E&M than assumed in this course. We thus follow another route here.

The rest energy

$$E_{\text{rest}} = mc^2 \tag{53}$$

is the part of energy that an object has, even if it is not moving at all. This energy is huge in our daily standard since c is a huge number in our daily life (and we have $c \times c$).

In Newtonian mechanics, due to mass conservation, this energy is not noticed in energy conservation. However, in relativistic situations this energy can be released. For example,

- We have already encountered a non-trivial case of rest energy above – the two final objects after the split having $E = Mc^2 > 2mc^2$.
- In more realistic situations, the usage of nuclear energy makes use of the rest energy of matter. In nuclear fissions and fusions, a much greater amount of energy can be released compared to usual chemical or mechanical reactions.

Daily energy usage of Hong Kong

To see the huge rest energy of matter, for example, the daily energy consumption of Hong Kong is of order 10^{15}J. That corresponds to about 10 grams of matter – about the weight of one AAA battery. To compare, the chemical energy that an AAA battery can provide is a few thousand joules.

Why do stars shine?

Before relativity was understood, there was a big mystery that how stars can shine for longer than the human history, if the stars are powered by chemical energy. Instead, the nuclear fusion in the star can power the stars for billions of years.

- In labs, one can collide particles and anti-particles. The particles disappear after they meet and the rest energy can be completely released in the form of light.

The relativistic energy

In general, for an object with mass m and speed v, combining the kinetic energy (51) and the rest energy (53), the total energy of the object is

$$E = \gamma mc^2 \ . \tag{54}$$

This sounds a bit different from what you heard: $E = mc^2$. This more famous formula (famous because without having to explain γ in popular science) may stand for one of the two meanings:

- The rest energy of an object.
- The total energy of an object, with the mass defined as the velocity-dependent "relativistic mass" $m_{\rm rel} \equiv \gamma m$ and thus $E = m_{\rm rel}c^2$. Nowadays we understand mass as a Lorentz invariant label of a particle; the convention of $m_{\rm rel}$ is seldomly used and thus we will not use $m_{\rm rel}$ further here.

The zero-point of rest energy

The particle-anti-particle annihilation phenomenon is a very important check about the concept of rest energy. Because the zero point of energy makes no physical sense unless all the energy can be released.

The 4-dimensional momentum vector

Remember the question we asked at (49)? Happily, up to a factor of c (which is nothing more than a conversion of unit as we have explained natural unit), the energy we have got in (54) is exactly the missing "naive time-like counterpart of momentum" in (49). The 4 dimensional momentum is then $p^\mu = (E/c, \mathbf{p}) = (\gamma mc, \gamma m\mathbf{v}) = \gamma m\dot{x}^\mu$.

Recall that for the spacetime coordinate 4-vector, we have an invariant quantity $c^2\Delta t^2 - \Delta x^2$. Is there a counterpart for the momentum 4-vector? Yes. Squaring equation (54), we get

$$m^2c^4 = E^2 - \frac{v^2}{c^2}E^2 \xlongequal{\text{using ratio of (54) and (48)}} E^2 - p^2c^2 \ . \tag{55}$$

Thus, though E and $\mathbf{p}$ are dependent on observers, the combination on the RHS is independent of observers, but is just the invariant mass squared of the particle. This is not surprising. Because the momentum and coordinate 4-vectors lives in the same space and is measured by the same metric. (Just as it is not surprising that both the magnitude of 3d coordinate vector and 3d momentum are rotational invariant.)

Light is light

Applying (55) to light, we have $v = c$. Thus, $m = 0$. In other words, if we view light as particles (see the quantum mechanics part for more information), their mass has to vanish.

The 4-vector of the momentum of light has the form $p^\mu = (E/c, \mathbf{p})$, where the energy of light $E = c|\mathbf{p}|$. This 4-vector has vanishing Lorentz norm. In the notation of the metric (optional material), we have $\sum_{\mu,\nu} g_{\mu\nu}p^\mu p^\nu = 0$.

(Optional) The relativistic Doppler effect of light

Alice is moving with 3-velocity $\mathbf{v}$ wrt Bob. If Bob observes that the energy of light is E_B, what's the corresponding energy E_A of the same beam of light in Alice's frame?

We can solve this problem by the Lorentz invariance of inner products of 4-vectors. The inner product $\sum_{\mu\nu} g_{\mu\nu}k^\mu p^\nu$ is invariant under Lorentz transformations, where k^μ and p^ν are the 4-vector of Alice's momentum and the momentum of the beam of light.

In Bob's frame: $k^\mu = (\gamma mc, \gamma m\mathbf{v})$, $p^\nu = (E_B/c, \mathbf{p})$. The Lorentz invariant inner product is $\sum_{\mu\nu} g_{\mu\nu}k^\mu p^\nu = -\gamma m(E_B - \mathbf{v}\cdot\mathbf{p}) = -\gamma mE_B(1-(v/c)\cos\theta)$, where θ is the angle between $\mathbf{v}$ and $\mathbf{p}$ in Bob's frame.

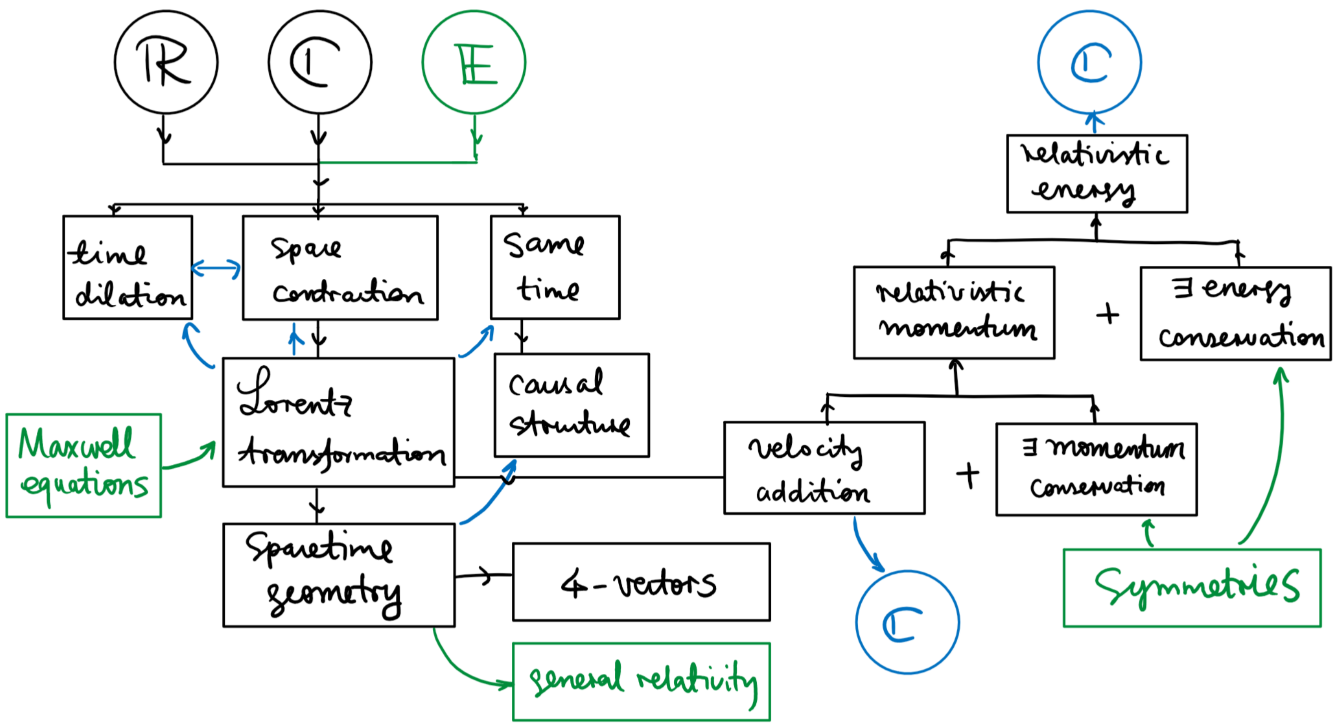
In Alice's frame: $k^\mu = (mc, \mathbf{0})$, $p^\nu = (E_A/c, \mathbf{p}')$. The Lorentz invariant inner product is $\sum_{\mu\nu} g_{\mu\nu}k^\mu p^\nu = -mE_A$. It must be equal to the Lorentz invariant in Bob's frame. Thus, $E_A = \gamma E_B(1-(v/c)\cos\theta)$.

From either classical E&M or quantum mechanics, we will find that the frequency of light is $\omega \propto E$. Thus the frequency of light wrt Alice and Bob has a similar relation $\omega_A = \gamma\omega_B(1 - (u/c)\cos\theta)$.

Here the purpose of this optional box is not only to tell you the formula of the relativistic Doppler effect, but also to show the power of Lorentz invariance for the inner product. The Doppler effect can also be derived just by considering the effect of time dilation and space contraction. But the calculation is much more complicated for a general direction θ.

10 Epilogue: Summary and What's Next

The below diagram is a recap of what we have learned. The green color denote external contents (you may ignore) and the blue denote alternative logic.



Further reading about the content

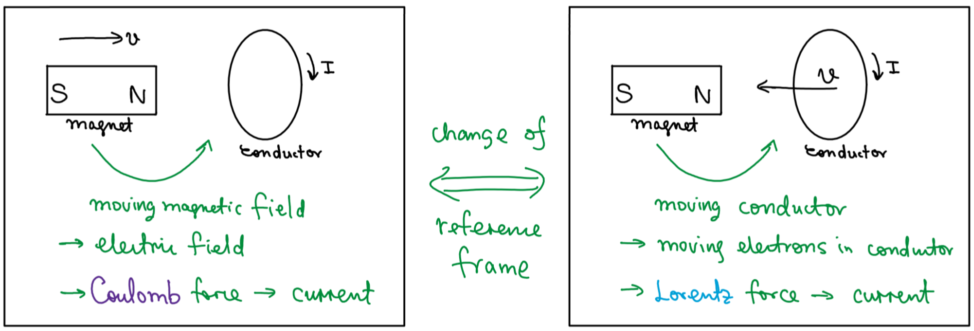
- If you are interested to explore more about the geometry of spacetime and formulate special relativity with geometrical emphasize, read "Spacetime Physics" by Taylor and Wheeler (and the a few general relativity books below).
- If you are not happy about our guess of relativistic energy and momentum, but rather want to derive them, wait and we will do it in the part "From Action to Laws of Nature". See also "Theoretical Minimum" (Video Lectures) by Susskind.
- If you consider the math used this text too sloppy, read "Relativity: Special, General, and Cosmological" by Rindler.
- If you like to learn more about the relation between electrodynamics and relativity before getting to the full details of electrodynamics, read Lecture Notes on Modern Physics by Baumann.
- Many good books on general relativity also starts with a dense and high-level introduction of special relativity, for example Gravity: An Introduction to Einstein's General Relativity by Hartle, Gravitation by Misner, Thorne and Wheeler and Gravitation And Cosmology by Weinberg.

What happens next in a university physics program?

- Electrodynamics as a deeper study of E&M.

When Einstein was young, he was deeply puzzled by two observations. Both has roots in E&M. One is light cheasing, which you have understood by now. The other was the magnet-conductor paradox, as he emphasized in the first paragraph of his 1905 paper. The paradox is shown in the below figure:

$$\vec{F} = \underbrace{q\vec{E}}_{\text{Coulomb}} + \underbrace{q\,\vec{v}\times\vec{B}}_{\text{Lorentz}}$$

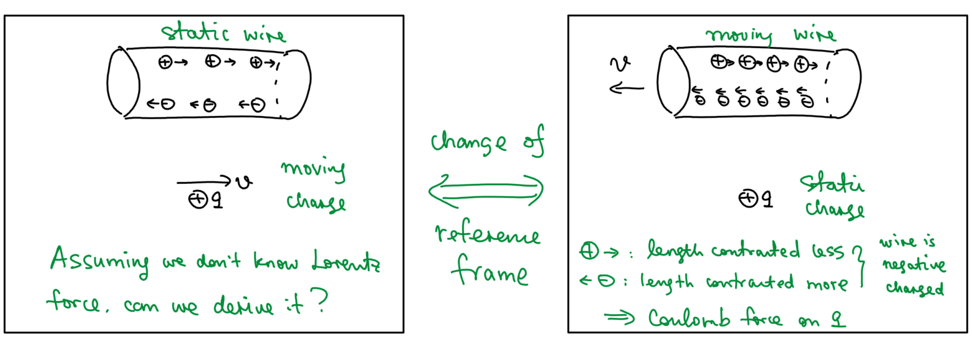


Before special relativity, the Coulomb law and the Lorentz law are considered as two independent fundamental laws of nature. However, by a change of frame, in the moving-magnet frame, the induct current is explained using the Coulomb law and in the moving-conductor frame the induced current is explained using the Lorentz law. This indicates that one should not consider them both fundamental – one same thing shouldn't be explained by two fundamental principles in physics! If the Coulomb law is more fundamental (at least it is more familiar), we should be able to derive the Lorentz law from it.

How relativity helps? Roughly speaking, 3-dimensional vectors such as **E** and **B** should be extended into 4-dimension vectors. Thus, the Coulomb and Lorentz laws, depending only on **E** and **B**, respectively, should be combined into one law using the form of 4-velocity and 4-electric-magnetic vector.

Without getting into the math, let us use a thought experiment to intuitively understand how to "derive" Lorentz force from Coulomb force by a change of frame.

Consider a wire conducing electric current, and a charge. The wire and the charge has relative motion between each other.



Here the Lorentz force is derived by the different amount of length contraction effect of moving charges.

In electrodynamics, you will explore the full connection between E&M and relativity.

- You may learn how special relativity works with gravity in "general relativity". We will also have a part to mention it briefly.
- Symmetry, transformation and group. Einstein's postulates (to be more precise, the Lorentz transformation) has the mathematical structures of a group (Lorentz group). "Group theory" is widely used in physics, including relativity, particle physics and solid state physics. And they are of their own importance in math as well.
- You may learn how special relativity works with quantum mechanics in "quantum field theory".

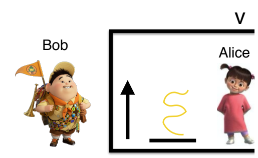
11 Exercises

E1.1 Plain waves

Consider "plain wave" $e^{ik(x-ct)}$. Interpret the real part of $e^{ik(x-ct)}$ as the amplitude of the wave.

1. For fixed t, show the plain wave indeed looks like a wave.
2. Figure out the moving direction and speed of the wave.

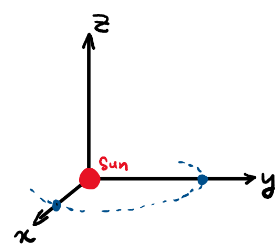
E2.1 Light travel and time dilation



Consider Alice and Bob have relative motion against each other, with velocity v. Alice carries a candle, which emits light (speed of light is c) perpendicular to the motion direction (wrt Alice).

1. What's the speed of this light ray wrt Bob?
2. What's the travel direction of this light ray wrt Bob?
3. Wrt Bob, the time of Alice slows down. Why doesn't the speed of this light ray slow down? Explain the relation between slower time and speed of light.

E5.1 Spacetime diagram of the sun-earth system



Use the frame that the sun is static, draw a spacetime diagram of the sun's and the earth's motion in x-direction (a direction in the earth's orbit plane). Show on this diagram how the sunlight (emitted in the x-direction) reaches the earth. The vertical axis is ct.

E5.2 Spacetime diagram with constant proper acceleration

Alice is moving with constant "proper acceleration" in the x-direction: $x^2 - t^2$ =constant (to simplify the discussion, let's work in one spatial dimension only). Not all light towards the observer can reach Alice. The region where light cannot reach Alice is called the "Rindler horizon". Draw a spacetime diagram of Alice's motion and find where the Rindler horizon is.

E5.3 What the twin actually sees

Draw the spacetime diagram of the twin paradox and show what the static twin actually sees (i.e. in the order that light reaches his eyes) about the aging of the moving twin (when she was moving outwards and when she was moving back).

E5.4 Distance between spaceships.

An observer A is considered at rest in the whole setup of this question. Two spaceships B and C are equidistant to A and are initially also at rest, and the distance between them is L. Now A sends a light signal. After receiving the light signal, B and C immediately start to move at a velocity v in the same direction (neglect the period of acceleration). What is the distance between B and C after they are moving? Give your answer wrt A and wrt B, respectively. Hint: draw a spacetime diagram to find out what happened.

E7.1 Use Lorentz transformation in calculations

Use Lorentz transformation to calculate:

1 Time dilation.

2 Rule contraction.

3 For two events which happened at the same time wrt Alice, calculate the time difference wrt Bob, given Bob's speed v, and the distance between the two events being L wrt Alice.

E7.2 Examples of velocity addition

Alice is moving away from Bob with velocity $\vec{v} = (v, 0, 0)$, and sending out light rays.

1 If the light ray is along x_A direction (x direction wrt Alice), calculate the velocity of the light ray $\vec{v}_B = (v_{Bx}, v_{By}, v_{Bz})$ wrt Bob.

2 If the light ray is along y_A direction (y direction wrt Alice), calculate the velocity of the light ray $\vec{v}_B = (v_{Bx}, v_{By}, v_{Bz})$ wrt Bob.

E7.3 Speed of light in the media

Consider the speed of light in static air (refractive index $n = 1.0003$). How does the speed of light in the air change wrt moving observers moving with speed v? How does the speed of light in the air change when there is a wind with speed v?

E8.1 The spacetime interval is Lorentz invariant

Show that under Lorentz transformation, the spacetime interval ds^2 is unchanged (although t and x change). For simplicity, work in two dimensions t and x only (i.e. no motion or rotation in y and z directions).

E8.2 Our motion in spacetime

Show that using proper time, everyone is moving in spacetime (not space) with the same 4-speed (i.e. the size of the 4-velocity $dx^\mu/d\tau$).

E9.1 Integrate momentum to get energy

Let us use the relativistic momentum $p = \gamma m v$ to derive the expression of the kinetic energy, in a different way from what we did in class.

Consider a ball at rest at $x = 0$ with mass m, and act a constant force F on this ball toward the x direction. The ball then accelerates because of the force.

Note that $F = dp/dt$, and the kinetic energy can be calculated from the work done by the force:

$$K = \int_0^{x_1} F\ dx = \int_0^{x_1} \left(\frac{dp}{dt}\right) dx = \int_0^{p_1} v\ dp = \int_0^{v_1} v\ d(\gamma m v)\ , \tag{56}$$

where x_1, p_1 and v_1 are the distance, momentum and velocity at a later time t_1.

Continue the calculation and derive the kinetic energy of the ball at time t_1.

Read Einstein's original papers on relativity

Nowadays Einstein's original papers can be easily found online. For example, his first paper. You will find most parts the paper accessible except that in electrodynamics he used different notations from modern convention.

Index