Special theory of relativity

Electrodynamics and Michelson-Morley experiment

- At the end of the 19th century physics was solidly shaped by two great theories: Newton's mechanics and Maxwell's electrodynamics. Maxwell's equation provided a complete description of all phenomena known at the time associated with electricity and magnetism.

- Maxwell’s equations in vacuum (Gaussian units) are:

\[ \nabla E = 0 \]
\[ \nabla B = 0 \]
\[ \nabla \times E = -\frac{1}{c} \frac{\partial B}{\partial t} \]
\[ \nabla \times B = \frac{1}{c} \frac{\partial E}{\partial t} \]

They determine the time evolution of electric \( E(r,t) \) and magnetic \( B(r,t) \) fields. By taking curls of the last two equations one easily finds:

\[ \nabla^2 E - \frac{1}{c^2} \frac{\partial^2}{\partial t^2} E = 0 \]
\[ \nabla^2 B - \frac{1}{c^2} \frac{\partial^2}{\partial t^2} B = 0 \]

where \( \nabla^2 \) is Laplacian:

\[ \nabla^2 f = \frac{\partial^2 f}{\partial x^2} + \frac{\partial^2 f}{\partial y^2} + \frac{\partial^2 f}{\partial z^2} \]

These equations describe the propagation of waves. The plane-wave solution is:

\[ E(r,t) = E_0 \cos(kr - \omega t) \]
\[ B(r,t) = B_0 \cos(kr - \omega t) \]

where the vectors \( E_0 \) and \( B_0 \) are perpendicular to each other and to the wavevector \( k \) which determines the direction of propagation and the wavelength \( \lambda = 2\pi/|k| \). The wave frequency is \( \nu = \omega/2\pi \). By substituting the plane-wave solutions in the wave equations one finds \( \omega^2 - c^2k^2 = 0 \), which implies \( \omega = ck \) or \( \lambda \nu = c \). In order to determine the velocity of wave propagation (phase velocity), we ask how fast must \( r \) change in time to keep the phase fixed, \( kr - \omega t = \text{const} \). We find

\[ \frac{dr}{dt} = \frac{\omega}{k} = c \]

Hence, all electromagnetic waves propagate with the velocity \( c \approx 3 \times 10^8 \) m/s, the speed of light.

- Maxwell’s equations have been extremely successful in describing electromagnetic phenomena, and were spectacularly successful in predicting the existence of electromagnetic waves subsequently discovered by Hertz. Therefore, their validity can be hardly doubted.

- However, Maxwell’s equations are written without regard to the frame of reference. Consequently, they predict that the speed of light is the same in any reference frame, regardless of how fast one moves with respect to the light source, or the source of any other electromagnetic radiation. Material objects behave differently. If a particle was kicked with relative velocity \( v \) from a source, and the source is moving with velocity \( u \) relative to an observer, then the observer finds that the particle moves with velocity \( v + u \). In other words, velocities of objects are relative and depend on the reference frame in which motion is measured.
• In order to explain the constancy of the speed of light without contradicting Newtonian mechanics, a proposal was put forward that an invisible substance called ether fills space and acts as a medium for electromagnetic waves, in a similar fashion to air being a medium for sound waves. Then, the speed of light predicted by Maxwell’s equation would be the velocity of radiation measured with respect to ether, that is the velocity measured in the reference frame in which ether is at rest (doesn’t flow). If we were to think of ether as a fluid filling the universe, it would have to be at rest everywhere in space in some reference frame, otherwise we would see various optical distortions of distant stars, possibly time dependent (just like looking through non-uniformly flowing water). Therefore, there should be a preferred reference frame in the universe, as far as electromagnetic radiation is concerned.

Michelson-Morley experiment

• Michelson and Morley attempted to detect motion through ether based on the described physical picture. They constructed the first interferometer. A monochromatic collimated light beam was split by semi-transparent mirrors into two beams which propagated at right angles. These two beams were reflected back by regular mirrors positioned at the same distance from the splitter. The reflected beams would fall back on the semi-transparent splitter and interfere. The observed interference fringes, specifically their locations, allowed measuring distances (or velocities) traveled by the two split light beams with accuracy set by the light wavelength (which can be nanometers).

• Since the two split beams propagated at right angles, they could have different velocities according to the ether hypothesis. For example, one could rotate the equipment to align the path of one beam with the velocity vector of Earth relative to ether. The other beam would be perpendicular to the relative motion of Earth through ether, so it would propagate with different velocity than the first beam. One could then measure interference fringes while gradually rotating the equipment and look for changes.

• Michelson and Morley found that interference fringes did not meet the expectations from the ether hypothesis. Their experiment was more sensitive than required to see a clear signal if ether existed. Many other subsequent experiments, up to modern times, confirmed this finding.

• By now, it is overwhelmingly clear that there is no ether.

Einstein’s theory

• After the Michelson-Morley experiment, Einstein attempted to find a different resolution to the apparent conflict between Maxwell’s electromagnetism and Newtonian mechanics. He made a bold proposition that the Newtonian rule of adding relative velocities, known as Galilean transformations, applied only when velocities were much smaller than the speed of light. This did not contradict any observation made to date on the kinematics of material objects, because all tests of Newtonian mechanics had been indeed conducted only at small velocities in comparison to c. However, it left open the question of how the laws of kinematics should be modified at extremely large velocities in order to accommodate the properties of electromagnetic radiation. Einstein constructed the minimal possible generalization of Newtonian kinematics which accomplished this goal, and it turned out to be a spectacular success despite some of its strange predictions.

• Einstein put forward two postulates:

1. All reference frames are physically equivalent. There is no preferred frame of reference in the universe which could be distinguished by some experiment.

2. The speed of light c is the same in all reference frames, irrespective of the relative velocity of the source.

• The first postulate is very natural and holds true even in Newtonian mechanics. It is an implicit statement that ether does not exist. The second postulate is a reaffirmation of Maxwell equations’ consistency with the first postulate. The speed of light does not add up to the velocity of source. Therefore, Galilean transformations cannot be absolutely correct.
Galilean transformations

- A reference frame is defined by a coordinate system used to specify positions of objects, and a clock used to measure time. Coordinates of an object express the distance in $x$, $y$ and $z$ directions of the object from the coordinate system origin. The clock measures elapsed time since some initial moment labeled $t = 0$. The position of the origin, orientation of the coordinate system axes and the moment $t = 0$ are arbitrary.

- Any event can be specified by a point $(x, y, z, t)$ in the space-time coordinate system of any reference frame. The same event will have different coordinates in different reference frames. Since there is no physically preferred reference frame, we need transformation rules for expressing the coordinates of the same event between different reference frames.

- Galilean transformations are such rules defined in Newtonian mechanics.

- Consider two reference frames, A and B. In order to avoid mathematical complexity let us assume that $x, y, z$ axes of A and B are always parallel. Also, let us assume that at $t = 0$ measured in both reference frames the two frames coincide. Otherwise, the origin of the frame A moves with constant velocity $u$ along $x$ axis as observed in the frame B.

- Now, consider an event $(x_A, y_A, z_A, t_A)$ observed in the frame A. When and where does this event occur in the frame B? First, time is absolute in Newtonian mechanics and “flows equally fast” for all possible observers. Therefore, $t_B = t_A$. Second, the two frames move relatively to each other only in $x$ direction. The $y$ and $z$ axes coincide all the time, so $y_B = y_A$ and $z_B = z_A$. Only the $x$ coordinates need translation. The same event has the same $x$ coordinates in the two reference frames only at $t = 0$ when the two frames coincide. An object at rest in the frame A moves at velocity $u$ in the frame B, so its $x_B$ coordinate increases at the rate $u$ measured in B: $x_B = x_A + ut_B$. (note $t_B$). Now, we use the finding $t_A = t_B$ and summarize Galilean transformations:

\[
\begin{align*}
t_B &= t_A \\
x_B &= x_A + ut_A \\
y_B &= y_A \\
z_B &= z_A
\end{align*}
\]

- Galilean transformations imply the rule for adding velocities. Velocity of an object $v = (v_x, v_y, v_z)$ is a vector with components $v_x = dx/dt$, $v_y = dy/dt$ and $v_z = dz/dt$. If an object has velocity $v_A$ in the reference frame A, then its velocity in the frame B is:

\[
\begin{align*}
v_{Bx} &= \frac{dx_B}{dt_B} = \frac{d}{dt_A} (x_A + ut_A) = v_{Ax} + u \\
v_{By} &= \frac{dy_B}{dt_B} = \frac{dy_A}{dt_A} = v_{Ay} \\
v_{Bz} &= \frac{dz_B}{dt_B} = \frac{dz_A}{dt_A} = v_{Az}
\end{align*}
\]

- Despite all of its intuitive appeal and accuracy in describing the kinematics of ordinary objects around us, we see that these velocity addition rules do not work for light because $c$ must be always the same constant given by Maxwell’s equations.

Lorentz transformations

- How can we generalize Galilean transformations to obtain something consistent with Maxwell’s equations and Einstein’s second postulate? In order to answer this question we will set up the reference frames A and B exactly the same way as in the above discussion. In the spirit of the Einstein’s first postulate, the transformation rules must be the same for any relative velocity $u$ of the two frames. However, the modified rules must reduce to Galilean transformations in the limit of small velocities, that is $u \ll c$ (Galilean transformations are obviously correct in this limit).
• Next, we note that Galilean transformations are linear. We expect the modified transformations to also be linear. Otherwise uniform motion of an object in one reference frame could look like accelerated motion in the other frame. This is explicitly forbidden by the first Einstein’s postulate. The laws of physics must be the same in all reference frames: if a force acts on an object, then its effects (acceleration) must be observed in all reference frames.

• Note that we made an implicit assumption that all reference frames we consider are inertial. A reference frame is inertial if we can apply the first Newton’s law in it (if the net sum of all forces on an object exerted by other bodies is zero, then the object will sit at rest or move along a straight line with a constant velocity). Any reference frame moving uniformly with respect to an inertial frame is also inertial. However, a frame which accelerates with respect to an inertial frame is not inertial, and one can find that objects accelerate in it without being affected by other objects.

• Let us then begin by writing generic linear transformation laws. Since there is no relative motion of the frames A and B along y and z axis, there is no reason to convert these coordinates: \( y_B = y_A \), \( z_B = z_A \). The linear transformation \( A \rightarrow B \) of \( x \) and \( t \) coordinates is:

\[
\begin{align*}
t_B &= a_{11} x_A + a_{12} t_A \\
x_B &= a_{21} x_A + a_{22} t_A
\end{align*}
\]

The new thing is that we allow the time coordinate to be converted in a non-trivial way. We will soon see why this is necessary.

• Consider an object sitting at the origin in the frame A: \( x_A = 0 \). In the frame B this must be perceived as motion at velocity \( u \):

\[
\frac{dx_B}{dt_B} = \frac{a_{22} dt_A}{dt_B} = \frac{a_{22} dt_A}{a_{12} dt_A} = u
\]

where we first used \( x_B = a_{22} t_A \) and then \( t_B = a_{12} t_A \) for \( x_A = 0 \). Similarly, an object sitting at the origin in the frame B must be perceived as moving with the velocity \(-u\) in the frame A:

\[
\frac{dx_A}{dt_A} = -\frac{a_{22}}{a_{21}} = -u
\]

where we have used just the second transformation rule with \( x_B = 0 \). By comparing these two equations we immediately find \( a_{22} = ua_{12} \) and \( a_{12} = a_{21} \).

• Now consider an object moving with velocity \( c \) in the frame A. According to the second postulate, this must be perceived as motion at velocity \( c \) in the frame B as well. Therefore:

\[
\frac{dx_B}{dt_B} = \frac{a_{21} dx_A + a_{22} dt_A}{a_{11} dx_A + a_{12} dt_A} = \frac{a_{21} (dx_A/dt_A) + a_{22}}{a_{11} (dx_A/dt_A) + a_{12}} = \frac{a_{21} c + a_{22}}{a_{11} c + a_{12}} = c
\]

This was calculated by considering how small increments of coordinates and times are related in the two reference frames. It follows that

\[
a_{21} c + a_{22} = a_{12} c + a_{11} c^2 \implies a_{22} = a_{11} c^2
\]

since we found earlier that \( a_{12} = a_{21} \). One important observation is that we used to have \( a_{22} = u \) in the case of Galilean transformations, but now this requires \( a_{11} \) to be non-zero so that time itself becomes dependent on the reference frame. Such non-Newtonian relativity of time is necessary in order to formulate coordinate transformations which do not contradict Maxwell’s electrodynamics.

• Let us summarize the findings so far and rewrite the transformation rules in a more compact form. First we introduce a single symbol \( \gamma \) to capture the equality between \( a_{12} \) and \( a_{21} \): \( \gamma = a_{12} = a_{21} \). Similarly, the relationship between \( a_{22} = a_{11} c^2 \) can be expressed by introducing a symbol \( \beta \) and writing \( a_{11} = \beta \gamma / c \) and \( a_{22} = \beta \gamma c \). This particular form is convenient because the remaining relationship \( a_{22} = ua_{12} \) we found implies \( \beta \gamma c = u \gamma \), that is \( \beta = u/c \).
• It will be convenient to express time variables as products \(ct\) because these products have the same physical dimension as the \(x\) coordinates. The transformation rules for \(A \rightarrow B\) can now be written as:

\[
\begin{align*}
ct_B &= \gamma(\beta x_A + ct_A) \\
x_B &= \gamma(x_A + \beta ct_A)
\end{align*}
\]

• The only remaining unknown is \(\gamma\) and in order to find it we need to invert the above equations and find the transformation rules \(B \rightarrow A\). According to the first Einstein’s postulate, there are no preferred reference frames so that the \(B \rightarrow A\) transformations must have the same mathematical form as the \(A \rightarrow B\) transformations. The only difference is the sign of \(u\): if the frame \(A\) moves with velocity \(u\) relative to \(B\), then \(B\) moves with velocity \(-u\) relative to \(A\). Let us obtain the inverted transformation by eliminating \(x_A\) from the second equation, \(x_A = \frac{1}{\gamma} x_B - \beta ct_A\) and substituting this into the first equation:

\[
ct_B = \gamma \beta \gamma^{-1} x_B - \gamma \beta^2 ct_A + \gamma ct_A \quad \implies \quad ct_A = \frac{-\beta x_B + ct_B}{\gamma(1 - \beta^2)}
\]

We see that the same mathematical form as in the first \(A \rightarrow B\) equation is obtained if \((1 - \beta^2) = \frac{1}{\gamma^2}\). The sign of the \(\beta\) term in the numerator is changed, but this is just as expected because \(\beta = \frac{u}{c}\) and \(u\) must appear with the opposite sign in the \(B \rightarrow A\) transformation.

• The full set of transformations is:

\[
\begin{align*}
ct_B &= \gamma(\beta x_A + ct_A) \\
x_B &= \gamma(x_A + \beta ct_A) \\
y_A &= y_B \\
z_A &= z_B
\end{align*}
\]

where

\[
\begin{align*}
\beta &= \frac{u}{c} \\
\gamma &= \frac{1}{\sqrt{1 - \beta^2}}
\end{align*}
\]

• These equations are called Lorentz transformations. Clearly, they reduce to Galilean ones whenever we can neglect \(\beta\) as a small number.

Relativistic phenomena

• Lorentz transformations predict a few unusual phenomena. The strangeness comes from the fundamental relativity of time which we are not used to in our everyday experience. For example, two events which occur simultaneously in one reference frame may appear non-simultaneous in other reference frames. Furthermore, time intervals and spatial distances between events can be perceived differently in different reference frames.

• Consider two events which occur at the same location in the frame \(A\), say \(x_{A1} = x_{A2} = 0\), but at different times \(t_{A1} = 0\) and \(t_{A2}\). The perceived time interval between these events in the frame \(A\) is \(\Delta t_A = t_{A2} - t_{A1} = t_{A2}\). According to Lorentz transformations, the events are seen to occur in the frame \(B\) at:

\[
\begin{align*}
x_{B1} &= 0, \quad t_{B1} = 0 \\
x_{B2} &= u\gamma t_{A2}, \quad t_{B2} = \gamma t_{A2}
\end{align*}
\]

Therefore, the interval between the two events is perceived in the frame \(B\) as:

\[
\Delta t_B = t_{B2} - t_{B1} = \gamma \Delta t_A
\]

This is known as time dilatation, because \(\gamma > 1\) for any finite \(u\). If we put a clock in front of a camera in a fast space ship, and broadcast the captured video to an observer on Earth, the observer will see a slower passage of time on the clock in the space ship than on his/her own clock. Interestingly, if the pilot of the space ship received a similar video transmission of the clock on Earth, he/she would also have an illusion that time on Earth goes more slowly than inside his vessel.
• Now, consider two events which occur simultaneously in the frame A, say at \( t_{A1} = t_{A2} = 0 \), but at different locations \( x_{A1} = 0 \) and \( x_{A2} = \Delta x_A \). This setup can represent a measurement of the length of some object in the frame A. Using Lorentz transformations we get:

\[
\begin{align*}
    x_{B1} &= 0 , \quad t_{B1} = 0 \\
    x_{B2} &= \gamma x_{A2} , \quad t_{B2} = \frac{u}{c^2} \gamma x_{A2}
\end{align*}
\]

The distance between events (the object length) is perceived in the frame B as:

\[
\Delta x_B = x_{B2} - x_{B1} = \gamma \Delta x_A
\]

This is called the contraction of lengths based on the premise that an observer ideally measures the length of a moving object in one instant of time. We have placed such an observer in the frame A in this analysis, so \( \Delta x_A = \gamma^{-1} \Delta x_B < \Delta x_B \). If we took an instant photograph of the fast space ship from the previous example, it would look shortened in the direction of its motion. The pilot could also take a photograph of Earth, and while he wouldn’t notice anything unusual about his own appearance, Earth would look to him like a Rugby ball.

• In our everyday experience we do not see such strange things, but then \( \gamma \to 1 \) because \( u \ll c \). If the velocity \( u \) becomes comparable with the speed of light, time dilatation and length contraction become noticeable. We have indirect evidence of these phenomena. For example, certain unstable elementary particles created in accelerator experiments have a very short lifetime before decaying into more stable particles. Yet, we find such particles in cosmic radiation, created far away and traveling toward Earth for much longer than their proper lifetime. Relativity reconciles this because the short lifetime in a particle’s own frame of reference appears significantly longer to us when the particle moves close to the speed of light.

• The theory of relativity implies that nothing can move faster than the speed of light. In order for Lorentz transformations to be physical, \( u \) must be strictly smaller than \( c \), otherwise \( \gamma \) is not a finite real number. The correct interpretation of this is that no physical influence of any kind can propagate faster than light. One event cannot be a cause of another if the time interval between them is shorter than it would take light to travel the distance between them. It is causality which is limited by the speed of light. One can reflect light from a powerful collimated lamp on clouds in the sky and make the reflected spot move faster than light by rotating the lamp very fast. This does not violate the relativistic causality.

**Velocity transformation**

• Here we derive the relativistic rules for adding velocities. Let an object move with velocity \( \mathbf{v}_A = (v_{Ax}, v_{Ay}, v_{Az}) \) in the reference frame A. Its velocity in the frame B appears to be:

\[
\begin{align*}
    v_{Bx} &= \frac{dx_B}{dt_B} = \frac{\gamma dx_A + ud_A}{\gamma \gamma dx_A + \gamma dt_A} = \frac{v_{Ax} + u}{\gamma \gamma dx_A + \gamma dt_A + 1} = \gamma \gamma^{-1} \frac{v_{Ax} + u}{1 + u v_{Ax}/c^2} \\
    v_{By} &= \frac{dy_B}{dt_B} = \frac{dy_A}{\gamma \gamma dx_A + \gamma dt_A} = \gamma^{-1} \frac{\gamma dy_A}{\gamma \gamma dx_A + \gamma dt_A + 1} = \gamma^{-1} \gamma \gamma^{-1} \frac{v_{Ay}}{1 + u v_{Ax}/c^2} \\
    v_{Bz} &= \frac{dz_B}{dt_B} = \frac{dz_A}{\gamma \gamma dx_A + \gamma dt_A} = \gamma^{-1} \frac{\gamma dz_A}{\gamma \gamma dx_A + \gamma dt_A + 1} = \gamma^{-1} \gamma \gamma^{-1} \frac{v_{Az}}{1 + u v_{Ax}/c^2}
\end{align*}
\]