Physics 5A, Introductory Mechanics and Relativity at UC Berkeley

Special Relativity

Fall 2017

1 Einstein's Principles of Relativity

1. The laws of physics are the "same" in all reference frames.
2. The speed of light in a vacuum is equal to c in all reference frames.

2 Galilean Transformations

Consider the system of two observers, A and B, undergoing relative motion (say in A's reference frame he is standing still and B is moving away from him, but B could make the equivalent statement that she is still and A is moving away with the same velocity in the opposite direction) and observing the positions of an object over time. Both of them tabulate a sequence of positions $r_A\vec{(t_i)}$ and $r_B\vec{(t_i)}$, for which we can set up a relation by vector addition:

$$\vec{r_A} = \vec{r_{B/A}} + \vec{r_B}$$

We differentiate this twice,

$$\vec{a_A} = \vec{a_{B/A}} + \vec{a_B}$$

We now make the observation that for Newton's laws to work correctly in both A's reference frame and B's, their relative *acceleration* must be zero. We can use $F = ma$ by multiplying both sides of the above relation by m, the mass of the object, but this gives different values for the force on the object, which is clearly independent of the observer. This apparent discrepancy is resolved if we recognise that Newton's laws only work correctly in *inertial* reference frames, i.e. we have to set $\vec{a_{B/A}} = 0$. Relative velocity, not relative acceleration, can exist between A and B.

We take the specific case in which B is moving only along the $x-$axis. We can do this without loss of generality, since in a specific case it is easy to set the coordinate system so that relative motion of the observers is only along the $x-$axis. Then, we can convert the observations from one frame to those of the other using the following:

$$t_A = t_B$$

$$x_A = x_B + vt_B$$

$$y_A = y_B$$

$$z_A = z_B$$

We can take the time derivative,

$$v_{x_B} = v_{x_A} - v$$

$$v_{y_B} = v_{y_A}$$

$$v_{z_B} = v_{z_A}$$

where v is the relative velocity between A and B.

Suppose the observed object is a beam of light, which has a known speed $v = c$. Then, from the above relations, we get

$$v_{x_A} = c_B + v$$

$$v_{x_B} = c_A - v$$

This seems to violate Einstein's second principle of relativity, that the speed of light should be a constant. Maxwell's equations suggest that the speed of light, or that of any electromagnetic wave, is dependent on certain natural constants:

$$c = \frac{1}{\sqrt{\varepsilon_o \mu_o}}$$

This suggests that something needs to be corrected in either Newton's theory that velocities and positions can be linearly compared from reference frames as a consequence of $F = ma$, or Maxwell's theory that the speed of light is invariant and dependent on natural constants.

Nor is this a problem that can be resolved with a specific choice of reference frame. In any real-world system, relative motion will exist due to the motion of the Earth, galaxies, galactic clusters, etc. and this can often happen at speeds that make this effect significant. The theory of special relativity is aimed at resolving this apparent contradiction.

3 Trains get struck by lightning twice all the time

Consider a train in motion with velocity $\vec{v} = v\hat{i}$, and length L. Observer A is located outside the train at $x = 0$, and Observer B is located inside the train also at $x = 0$. At $t = 0$, the train is struck by lightning twice: at $x = \frac{L}{2}$ and at $x = -\frac{L}{2}$.

From A's frame of reference, the light ray from the front travels a distance $\frac{L}{2}$ in a time t_F, which we can calculate easily:

$$x_A(t) = \frac{L}{2} - ct$$

$$x_A(t_F) = 0 \implies \frac{L}{2} = ct_F \implies t_F = \frac{L}{2c}$$

And similarly for the light ray from the back,

$$0 = -\frac{L}{2} + ct_B \implies t_B = \frac{L}{2c}$$

Therefore, we can define the time difference between these two observations,

$$\Delta t = t_F - t_B = 0$$

i.e. *simultaneity* is preserved for A.

In order to convert between A's measurements and B's, while still preserving the condition that c must be constant, we introduce an unknown even function $\gamma(v)$, so that the transformation between reference frames becomes

$$x_A = \gamma(v)(x_B + vt_B)$$

In general, for $v << c$, $\gamma \to 1$. We can do this in reverse,

$$x_B = \gamma(v)(x_A - vt)$$

Therefore, we can substitute this into the expression for x_A,

$$x_A = \gamma(v)(\gamma(v)(x_A - vt_A) + vt_B)$$

$$\therefore x_A - \gamma^2 x_A + \gamma v t_A = \gamma v t_B$$

$$\therefore t_B = \gamma\Big(t_A - \frac{1-\gamma^2}{v\gamma} x_A\Big)$$

$$t_A = \gamma\Big(t_B + \frac{1-\gamma^2}{\gamma v} x_B\Big)$$

4 Finding γ (Pixar, 2003) and the Lorentz Transformation

In order to find γ, we use the property that c is constant in all reference frames. Therefore, for a light wave moving outwards,

$$r_A = ct_A$$

$$x_A^2 + y_A^2 + z_A^2 = c^2 t_A^2$$

and

$$x_B^2 + y_B^2 + z_B^2 = c^2 t_B^2$$

As we are taking (without loss of generality) motion only in the $x-$direction,

$$x_A^2 - x_B^2 = c(t_A^2 - t_B^2)$$

Then, we use the transformation formula,

$$x_A^2 - \gamma^2(x_A - vt)^2 = c^2\Big(t_A^2 - \gamma^2(t_A - \frac{1-\gamma^2}{v\gamma}x_A)^2\Big)$$

$$x_A^2 - \gamma^2(x_A^2 - 2vx_At_A + v^2t_A^2) = c^2\Big(t_A^2 - \gamma^2(t_A^2 - \frac{2(1-\gamma^2)}{\gamma v}t_Ax_A + \gamma\frac{(1-\gamma^2)^2}{\gamma^2 v^2}x_A^2)\Big)$$

We can then compare coefficients of t_A^2:

$$-\gamma^2 v^2 t_A^2 = c^2(1-\gamma^2)t_A^2$$

$$-\gamma^2\frac{v^2}{c^2} = 1 - \gamma^2$$

$$\therefore \gamma = \frac{1}{\sqrt{1-\frac{v^2}{c^2}}}$$

Now that we have found γ, we substitute this into the previously-obtained time relation to get:

$$t_A = \gamma\Big(t_B + \frac{vx_B}{c^2}\Big),\ t_B = \gamma\Big(t_A - \frac{vx_A}{c^2}\Big)$$

This is called a *Lorentz transformation.*

5 Applying Lorentz Transformations

We are now in a position to analyse what B sees in the train-lightning example.

From the front,

$$x_B^{(F)} = \gamma\Big(\frac{L}{2} - 0\Big)$$

$$t_B^{(F)} = \gamma\Big(0 - \frac{v}{c^2}\frac{L}{2}\Big)$$

and from the back,

$$x_B^{(B)} = \gamma\Big(-\frac{L}{2} - v \cdot 0\Big)$$

$$t_B^{(B)} = \gamma\Big(0 + \frac{v}{c^2}\frac{L}{2}\Big)$$

6 Length Contraction

Our Newtonian conception of length as being invariant (absent any external forces on it that might stretch or compress the object) is only valid when simultaneity holds true, i.e. lengths are measured simultaneously in the frame doing the measuring. For relativistic velocities, this no longer holds true, and we have to apply the Lorentz transformation.

Say the two ends of the object are defined by positions x_{B_1} and x_{B_2} as measured in frame B, which, relative to frame A, is moving at a velocity v. Then,

$$x_{B_1} = \gamma(x_{A_1} - vt_{A_1}),\ x_{B_2} = \gamma(x_{A_2} - vt_{A_2})$$

Therefore

$$L_p = x_{B_2} - x_{B_1} = \gamma(x_{A_2} - x_{A_1} - v(t_{A_2} - t_{A_1})$$

where $L_p = x_{B_2} - x_{B_1}$ refers to the proper length as measured in a frame in which there is no relative motion. When we apply the condition for simultaneity, that $t_{A_1} = t_{A_2}$, this reduces to

$$L = \frac{L_p}{\gamma}$$

where $L = x_{A_2} - x_{A_1}$ is the contracted length as measured in the relatively moving frame.

7 Time Dilation

Consider, as before, two frames of reference A and B, in which B's frame is moving away from A's with a velocity v. B then throws a ball upward and catches it after some time. We characterise positions in time as (t, x); then, When B throws the ball: (t_{B_1}, x_{B_1})

When B catches the ball again: (t_{B_2}, x_{B_1})

No net displacement has occurred in B's frame of reference, over a time interval $\Delta t_B = t_{B_2} - t_{B_1}$.

Now, we apply the Lorentz transformation and analyse this situation in A's frame of reference:

$$t_{A_1} = \gamma(t_{B_1} + \frac{vx_{B_1}}{c^2})$$

$$t_{A_2} = \gamma(t_{B_2} + \frac{vx_{B_2}}{c^2})$$

$$\therefore \Delta t_A = \gamma(t_{B_2} - t_{B_1} + \frac{v}{c^2}(x_{B_2} - x_{B_1}))$$

$$= \gamma(\Delta t_B + \frac{v}{c^2}\Delta x_B)$$

Because there is no net displacement in B's frame of reference, we have the time dilation factor

$$\Delta t_A = \gamma \Delta t_B$$

8 Representing 4D Space

We can represent a point in four-dimensional space uniquely with four coordinates (x, y, z, t), but this has a unit mismatch. Therefore we instead use *ct* as a coordinate, and shift its position in the set, to obtain

$$(ct, x, y, z)$$

We can then change known equations into this form, and we see some simplification:

$$ct_A = \gamma(ct_B + \frac{v}{c}x_B),\ x_A = \gamma(x_B + \frac{v}{c}(t_B c))$$

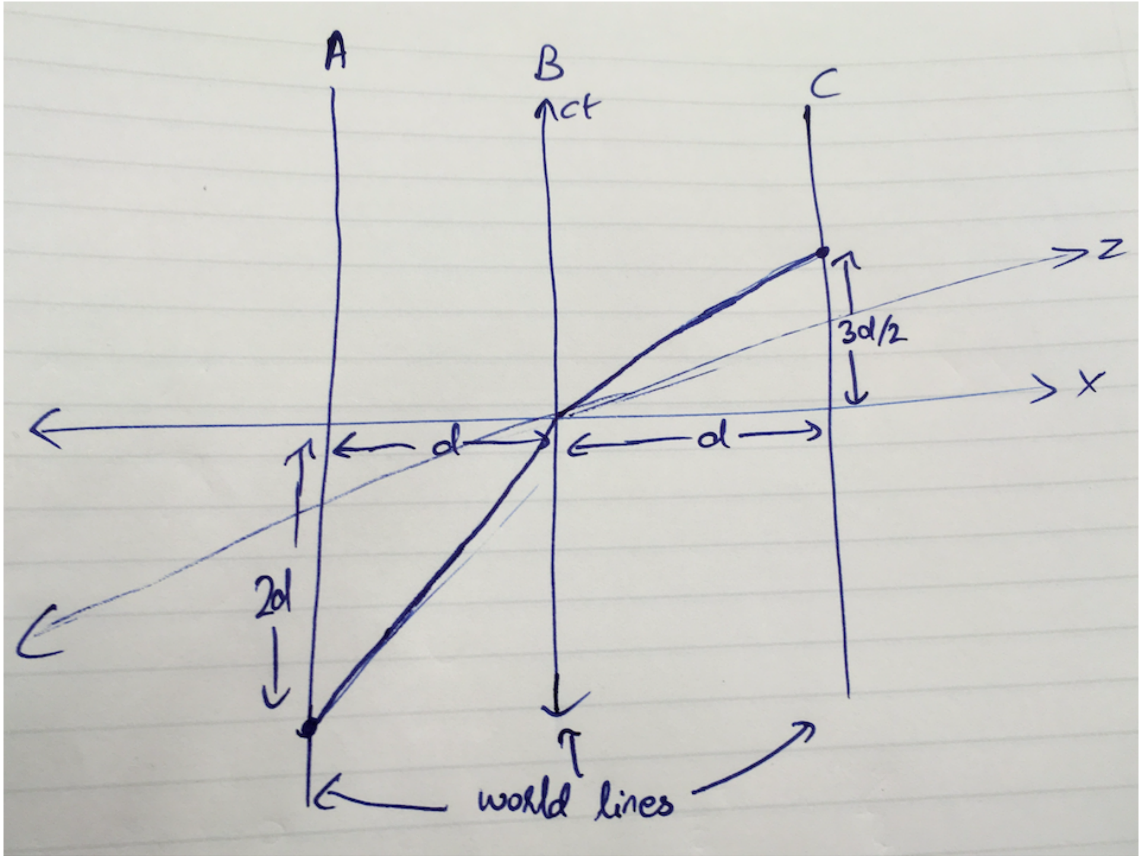
$$ct_B = \gamma(ct_A - \frac{v}{c}x_A),\ x_B = \gamma(x_A - \frac{v}{c}(t_A c))$$

We can introduce the parameter $\beta := \frac{v}{c}$ to simplify this further.

9 Spacetime Diagrams and Light Cones

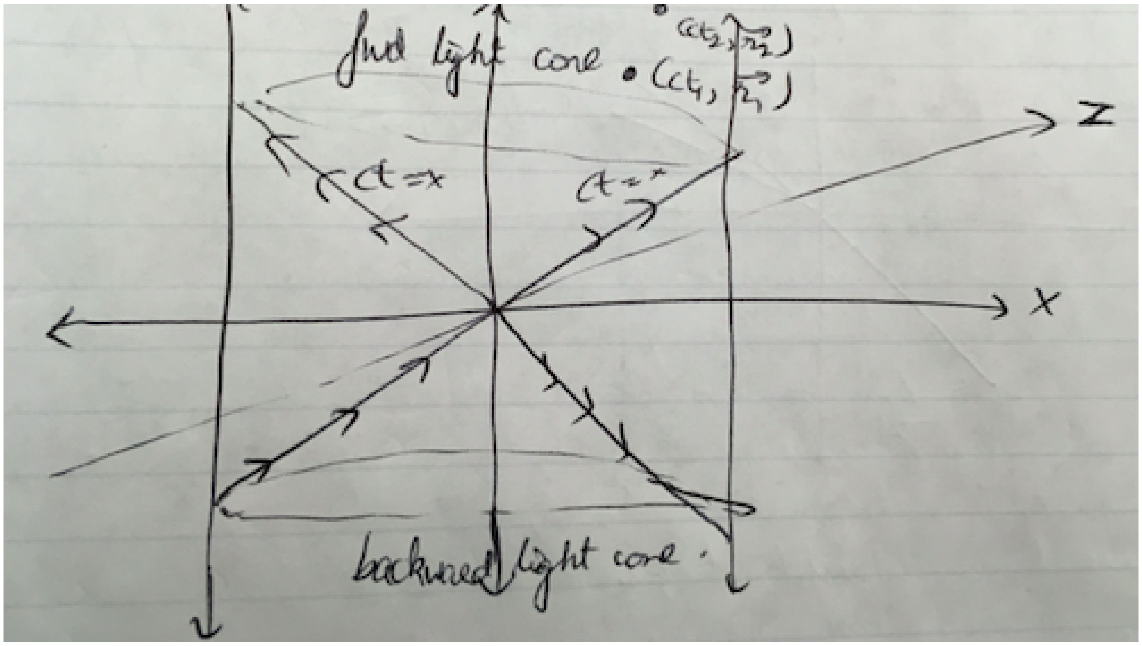
If we first define an *event* as a specific combination of the regular three-dimensional position vector and an instant in time, i.e. a unique set of coordinates (ct, x, y, z), then we can represent an event on a plot of ct against x (or against any other spatial coordinate - we restrict it to x for simplicity). This is referred to as a *spacetime diagram.*

For instance, consider a setup that involves three people, A, B and C. We set B at the origin, with A at $x = -d$ and C at $x = +d$. A has a ball that he throws to B at $u_1 = \frac{c}{2}$ and which B then throws to C at $u_2 = \frac{2c}{3}$. The presence of the ball at a point in space at each time in this interval is an *event* that shows up on the spacetime diagrams. The positions of A, B, and C are also events, but ones that never move in space, i.e. they are constant relative to the ct axis.



The slope of the spacetime diagram involving the motion of a particular object is the inverse of its β parameter, or $\frac{c}{v}$.

Light is therefore characterised by a slope of 1 on the spacetime diagram ($v = c$). Since nothing can go faster than light, we can use this to define a section of the diagram within which events can affect one another. This region is referred to as the *light cone*, and its edges are defined by the lines $ct = x$ and $ct = -x$.



For every point within the light cone shown, centered at $(0,0,0,0)$ which is B's position at $ct = 0$, B can have some causal effect on an event in that region. If it is in the *forward* light cone, i.e. $ct > 0$, then B can affect or cause an event within the cone. If it is in the *backward* light cone, i.e. $ct < 0$, then B can be affected by an event within the cone. Therefore the light cone constitutes what is called a *causally connected space.*

For points lying outside the area defined by the light cone, there can be no causal connection to the light cone's center at that particular moment in time and space. For example, consider C at $t = 0$. It lies outside the light cone of B, therefore it cannot have a causal connection to B. If C decides to do something, such as shine a laser at B, there will be a delay. Until B's position in space comes into C's light cone, or vice versa, B cannot experience any effect from C's actions at $t = 0$. Put another way, B cannot know what C is doing at $t = 0$.

10 The Spacetime Interval Δs

When translating between reference frames, we would like to be able to define a quantity analogous to length in 3D space that is invariant between reference frames. This turns out to be the *spacetime interval* defined as follows:

$$\Delta\mathbf{s}^2 = (\Delta x)^2 + (\Delta y)^2 + (\Delta z)^2 - (c\Delta t)^2$$

We can prove that this is invariant under a Lorentz transformation.

Proof.

$$\Delta\mathbf{s}^2 = (\Delta x_A)^2 + (\Delta y_A)^2 + (\Delta z_A)^2 - (c\Delta t_A)^2$$

$$= \gamma^2(\Delta x_B + \beta c\Delta t_B)^2 + (\Delta y_B)^2 + (\Delta z_B)^2 - \gamma^2(c\Delta t_B + \beta\Delta x_B)^2$$

$$= \gamma^2 \Big((\Delta x_B)^2 + 2\beta c \Delta x_B \Delta t_B + \beta^2 (c\Delta t_B)^2 \Big) - \gamma^2 \Big((c\Delta t_B)^2 + 2\beta c \Delta x_B \Delta t_B + (\beta \Delta x_B)^2 \Big) + (\Delta y_B)^2 + (\Delta z_B)^2$$

At this point we can group terms,

$$= (\gamma^2 - \beta^2\gamma^2)(\Delta x_B)^2 - \gamma^2(1-\beta^2)(c\Delta t_B) + (\Delta y_B)^2 + (\Delta z_B)^2$$

We can simplify the coefficients:

$$(\gamma^2 - \beta^2\gamma^2) = \frac{1}{1-\frac{v^2}{c^2}} - \frac{v^2}{c^2}\frac{1}{1-\frac{v^2}{c^2}} = \frac{c^2}{c^2-v^2} - \frac{v^2}{c^2-v^2} = 1$$

$$\gamma^2(1-\beta^2) = \frac{1}{1-\frac{v^2}{c^2}} \cdot \left(1 - \frac{v^2}{c^2}\right) = \frac{c^2-v^2}{c^2-v^2} = 1$$

Therefore, the expression simplifies to

$$\Delta \mathbf{s}^2 = (\Delta x_B)^2 + (\Delta y_B)^2 + (\Delta z_B)^2 - (c\Delta t_B)^2$$

□

The spacetime interval allows us to characterise three different kinds of separation between events, depending on their relative causality.

Case 1: $\Delta \mathbf{s}^2 > 0$

From the definition of $\Delta \mathbf{s}$, this suggests that

$$\left(\frac{\Delta \vec{r}}{\Delta t}\right)^2 > c^2$$

This implies that the events cannot be causally connected (the signal required for causality would have to travel faster than light), or that there is a *spacelike interval* between the two events.

Case 2: $\Delta \mathbf{s}^2 < 0$

$$\left(\frac{\Delta \vec{r}}{\Delta t}\right)^2 < c^2$$

This implies that the events can have a causal connection; there is a *timelike interval* between them.

Case 3: $\Delta \mathbf{s}^2 = 0$

$$\left(\frac{\Delta \vec{r}}{\Delta t}\right)^2 = c^2$$

This is *lightlike separation* an edge case of timelike separation, in which a causal connection is possible but only if the signal is light, or travels at exactly the speed of light.

Because the type of separation is characterised by the spacetime interval, it is independent of the reference frame. If two events are timelike/spacelike/lightlike separated in one reference frame, then that is the case in all reference frames.

11 Velocity Transformations

We define the velocity of an object in a reference frame A in a certain direction simply as follows:

$$u_{A_x} = \lim \Delta t_A \to 0 \frac{\Delta x_A}{\Delta t_A}$$

and similarly for y and z, only replacing the coordinates. This allows us to obtain a relation between velocities in the same directions in different reference frames:

$$u_{A_x} = \lim_{\Delta t_A \to 0} \frac{\gamma(\Delta x_B + \beta c \Delta t_B)/\Delta t_B}{\frac{\gamma}{c}(c\Delta t_B + \beta \Delta x_B)/\Delta t_B}$$

We get this by applying the Lorentz transformation on both Δx_A and Δt_A, and dividing the numerator and denominator by Δt_B. We can simplify this to

$$u_{A_x} = \lim_{\Delta t_B \to 0} \frac{u_{B_x} + \beta c}{1 + \frac{\beta}{c} u_{B_x}} = \frac{u_{B_x} + v}{1 + \frac{v u_{B_x}}{c^2}}$$

We can set up similar relations for the other spatial dimensions, with the exception that the βc term will not make an appearance (we are taking the case in which the frames are only moving away from one another in the $x-$direction). These relations in reverse become:

$$u_{B_x} = \frac{u_{A_x} - v}{1 - \frac{v u_{A_x}}{c^2}}$$

$$u_{B_y} = \frac{u_{A_y}}{\gamma(1 - \frac{v u_{A_y}}{c^2})}$$

$$u_{B_z} = \frac{u_{A_z}}{\gamma(1 - \frac{v u_{A_z}}{c^2})}$$

We note that the first relation, for the case $v << c$, reduces to $u_{B_x} = u_{A_x} - v$, which is exactly what we would expect from Newtonian physics.

This causes an apparent incongruity with how positions change. u_B is not a linear function of u_A, like the spatial coordinates are. We would prefer a setup in which this property is preserved.

12 Arc Length Characterisation

It is possible to characterise the motion of a particle in multiple ways. For regular motion in two or three dimensions, this is often restricted to parameterising equations of motion in time, as follows:

$$x(t) = v_{0x} t, \; z(t) = v_{0z} t - \frac{1}{2} g t^2$$

However, this can also be expressed in terms of one of the spatial coordinates:

$$z(x) = \frac{v_{0z}}{v_{0x}}x - \frac{1}{2}\frac{gx^2}{v_{0x}^2}$$

This is the difference between specifying a position in terms of the coordinates $(x(t), z(t))$ and $(x, z(x))$. We can characterise motion in a third way: by the distance that the particle has travelled along its path s, with coordinates $(x(s), z(s))$.

We can apply this to motion in four dimensions, to characterise the path without preferring one coordinate over another. The incongruity associated with the non-linearity of u_B relative to u_A can be resolved by using this characterisation of the path, rather than characterising by the parameter ct.

We define as the characterising coordinate $\Delta\tau^2 = -\frac{\Delta \mathbf{s}^2}{c^2} = \Delta t^2 - \left(\frac{\Delta\vec{r}}{c}\right)^2$, the ratio of the proper time to the distance travelled. This is also independent of the frame of reference.

We can define a **four-velocity** $u = (u_t, u_x, u_y, u_z)$ in which the individual components in a particular frame of reference are defined as

$$u = \lim_{\Delta\tau\to 0}\frac{\Delta\mathbf{R}}{\Delta\tau}$$

where R is also a four-vector with the components $(ct, \vec{r})$ in which $\vec{r}$ can be split further.

We can define this four-velocity in two different frames of reference A and B, and carry out a Lorentz transformation using this characterisation:

$$u_{A_t} = \lim_{\Delta\tau\to 0}\frac{c\Delta t}{\Delta\tau} = \lim_{\Delta\tau\to 0}\frac{\gamma_v(c\Delta t_B + \beta\Delta x_B)}{\Delta\tau}$$

$$= \gamma_v(cu_{B_t} + \beta u_{B_x})$$

Therefore the transformation with this characterisation is linear.

$$u_{A_x} = \lim_{\Delta\tau\to 0}\frac{\Delta x_A}{\Delta\tau} = \lim_{\Delta\tau\to 0}\frac{\gamma_v(\Delta x_B + \beta c\Delta t_B)}{\Delta\tau} = \gamma_v(u_{B_x} + \beta u_{B_t})$$

Instead of carrying out a Lorentz transformation, we can simplify the above expression as follows:

$$u_{A_x} = \lim_{\Delta\tau\to 0}\frac{\Delta x}{\Delta\tau} = \lim_{\Delta\tau\to 0}\frac{\Delta x}{\sqrt{(\Delta t)^2 - \left(\frac{\Delta\vec{r}}{c}\right)^2}}$$

$$= \lim_{\Delta t\to 0}\frac{\Delta x/\Delta t}{\sqrt{1 - \left(\frac{\vec{r}}{\Delta t}\right)^2\frac{1}{c^2}}}$$

$$= \frac{u_{A_x}}{\sqrt{1 - \frac{v_x^2}{c^2}}}$$

where v_x is the relative velocity between the two reference frames. We have an inconsistency in notation here, as u_{A_x} is being used to denote either the component of the 3-velocity or the 4-velocity. Therefore we set the notation for the 4-velocity as

$$\mathbf{u} = (\tilde{u_t}, \tilde{u_x}, \tilde{u_y}, \tilde{u_z})$$

Then, we can simplify the above relation to

$$\tilde{u_{A_x}} = \gamma u_{A_x}$$

13 4-momentum

We can define a 4-momentum from the definition of the 4-velocity:

$$\mathbf{P} = m\mathbf{u} = \left(\frac{E}{c}, \vec{p}\right) = m\Big(\gamma c, \gamma \vec{u}\Big)$$

Therefore we can define the energy and momentum of a relativistic particle,

$$E = m\gamma c^2, \vec{p} = m\gamma\vec{u}$$

in which m is the rest mass, which does not imply that $m\gamma$ should be called the *moving mass* because a particle does not gain mass as it moves.

We use the relation defining the invariant quantity (analogous to length) of the 4-velocity covered in the definition of the spacetime interval:

$$\mathbf{P}^2 = \vec{P}^2 - \Big(\frac{E}{c}\Big)^2 = m^2\gamma^2\vec{u}^2 - m^2\gamma^2c^2 = m^2\gamma^2c^2\Big(\frac{\vec{u}^2}{c^2} - 1\Big)$$

In a rest frame, this reduces to

$$\mathbf{P}^2 = -m^2c^2 = \vec{P}^2 - \Big(\frac{E}{c}\Big)^2$$

which we can rearrange to get

$$E^2 = (\vec{p}c)^2 + m^2c^4$$

When $\vec{u} = 0$ (set a frame in which $\vec{p} = 0$), this reduces further to

$$E = mc^2$$

14 Energy and Momentum Transformations

Because of the arc-length characterisation, we can transform energy and momentum just like positions:

$$\frac{E_A}{c} = \gamma_v\Big(\frac{E_B}{c} + \beta P_{B_x}\Big) \implies \frac{E_B}{c} = \gamma_v\Big(\frac{E_A}{c} - \beta P_{A_x}\Big)$$

$$P_{A_x} = \gamma_v\Big(P_{B_x} + \beta\frac{E_B}{c}\Big) \implies P_{B_x} = \gamma_v\Big(P_{A_x} - \beta\frac{E_A}{c}\Big)$$

This suggests that energy is not invariant between reference frames, which has strange implications for how basic laws of mechanics work relativistically.

15 Developing Relativistic Mechanics

We expect that Newton's second law $\sum \vec{F} = m\vec{a}$ should still be valid, but we cannot use this without a rigorous definition of relativistic acceleration. Therefore we use the definition $\sum \vec{F} = \frac{d\vec{p}}{dt}$. Then, the work over a path characterised by position vectors a and b is:

$$W = \int_a^b \vec{F} \cdot d\vec{r} = \int_a^b \frac{d\vec{p}}{dt} d\vec{r}$$

$$= \int_a^b \frac{d\vec{p}}{dt} \vec{u} dt = \int_a^b \vec{u} d\vec{p} = \int_a^b \Big(d(\vec{u} \cdot \vec{p}) - d\vec{u} \cdot \vec{p} \Big)$$

$$= \vec{u} \cdot \vec{p} \Big|_a^b - \int_a^b \vec{p} d\vec{u} = m\gamma\vec{u}^2 \Big|_a^b - \int_a^b m\gamma\vec{u}d\vec{u}$$

$$= m\gamma\vec{u}^2 \Big|_a^b - m\int_a^b \frac{\vec{u}d\vec{u}}{1-\frac{\vec{u}^2}{c^2}} = \left(m\gamma\vec{u}^2 + mc^2\sqrt{1-\frac{u^2}{c^2}} \right) \Bigg|_a^b$$

$$= mc^2\gamma \left(\frac{u^2}{c^2} + \frac{1}{\gamma^2} \right) \Bigg|_a^b = \gamma mc^2 \Bigg|_a^b$$

Therefore the work done is path-independent, and the answer suggests the form of the work-energy theorem $W = \Delta K$. This in turn suggests that we should define $K = \gamma mc^2$, which we shift by $-mc^2$ so that when a relativistic body is at rest it does not have a positive KE.

$$K = mc^2(\gamma - 1)$$

which suggests $E = K + mc^2$.

We can confirm this by taking $v << c$:

$$K = mc^2 \left(\frac{1}{\sqrt{1-\frac{\vec{u}^2}{c^2}}} - 1 \right)$$

whereupon we use a Taylor approximation $(1+x)^p \approx 1 + px$, which suggests

$$\left(1 - \frac{\vec{u}^2}{c^2}\right)^{-1/2} = 1 + \frac{1}{2}\frac{\vec{u}^2}{c^2}$$

$$\therefore K = mc^2 \left(\frac{1}{2}\frac{\vec{u}^2}{c^2} \right) = \frac{1}{2} m\vec{u}^2$$

16 Velocity/Momentum as Independent Variable

Under the velocity-as-independent-variable formalism (Lagrangian mechanics), we characterise energy and momentum as follows:

$$E(\vec{u}) = m\gamma(\vec{u})c^2, \vec{p} = m\gamma\vec{u}$$

The formalism in which E and $\vec{p}$ are functions of $\vec{u}$ turns out to be limited. So, we use the Hamiltonian formalism in which momentum is the independent variable, and we take:

$$E = \sqrt{(\vec{p}c)^2 + m^2c^4} \implies \vec{u}c = \frac{\vec{p}c}{E}$$

The advantage of this is that it is independent of the rest mass, i.e. it allows for a massless particle such as a photon having a nonzero energy or momentum. In this formalism, if we take $\lim_{m\to 0} E$, we get

$$E = \vec{p}c \implies \frac{\vec{u}}{c} = \frac{\vec{p}}{p} = 1$$

which suggests that the velocity of a particle without mass is $u = c$.

Lagrangian mechanics cannot give us these conclusions.