The Mathematics of Special Relativity

Jared Ruiz
Advised by Dr. Steven Kent
May 7, 2009
1 Origins of Relativity

When hearing the words “theory of relativity,” most immediately think of the equation \( E = mc^2 \), or Albert Einstein. While this is not a bad thing, ample appreciation is oftentimes not given where it is due. The road to the development, understanding, and actual application of the theory of relativity is a long and twisted one, and credit cannot be given to one man alone. Much like how Newton stood “…on the shoulders of giants…” when he developed the calculus, so too did Albert Einstein when finally realizing the theory of relativity.

Today the the necessary tools and knowledge are available that the special theory of relativity can be understood with little work. But before delving into the mathematics which lay before us, we should look at those who helped develop this theory; to understand their thoughts, their inspirations, and their struggles.

1.1 The Absolute

In 1632, Galileo Galilei published *Dialogue concerning the two chief world systems*. In it, two characters, Salviati and Simplicio, debate over whether the sun or the earth is the center of the universe (we know now that the sun is the center of our solar system). While this brought about great controversy regarding the Roman Catholic Church, it marked a monumental shift in perspective. Up until this point, the state of physics had been at a standstill for over a millenium. Science at the time consisted of reading the works of Aristotle, not hypothesizing and making new theories. Galileo was a visionary; he studied the heavens, made calculations, and although he could not prove his ideas, he supported them with solid observational evidence. So in a sense, Galileo’s work did more than persuade others to believe the Copernican view of the solar system. It was the birth of modern physics [?].

Isaac Newton was born on Christmas day 1642, the same year that Galileo passed away. In Newton’s lifetime, Galileo’s would become generally accepted, and Newton himself based his work upon Galileo’s. Again, by his own observations and study, Newton successfully developed the calculus in *Philosophiae Naturalis Principia Mathematica*. More importantly, in this work he defined a set of laws which all objects must follow in motion, and gave the law of universal gravitation. His work was incomplete, yet the *Principia*, as it is widely called, gave Newton great fame. His and Galileo’s work, oftentimes referred to as Galilean relativity or classical mechanics, would remain the foundation of physics for nearly 200 years [?].

Today, Newton’s Laws of Motion and the heliocentric theory are taught to children in grade school. So why the need for relativity? Was not classical mechanics enough? Was it possible that these great masters of old were incorrect in their assumptions? Yes, yet none realized how incorrect they were. Why then is Newton (along with Archimedes and Gauss), considered one of the greatest mathematicians of all time? It is important to understand the subtlety of their mistakes. The
mathematics which Newton and Galileo used was correct. What was incorrect was the physics. Newton and Galileo simply thought the speed of light was not constant [?].

### 1.2 Electrodynamics

In 1865, the main turning point in the history of relativity came about. James Clerk Maxwell, a Scottish physicist, unified the theories of electricity and magnetism. Up until this point, the laws and equations which governed electrical and magnetic fields were separate. All known interactions between electrical and magnetic fields with moving charges were described by Maxwell’s system of partial differential equations [?]. These are known as Maxwell’s equations, which are given for completeness.

\[
\begin{align*}
\nabla \cdot D &= 4\pi \rho \\
\nabla \cdot B &= 0 \\
\nabla \times H &= \frac{4\pi}{c} J + \frac{1}{c} \frac{\partial D}{\partial t} \\
\nabla \times E + \frac{1}{c} \frac{\partial B}{\partial t} &= 0
\end{align*}
\]

The release of Maxwell’s equations began a crisis in physics. If what Maxwell said was correct, then light traveled in waves just like sound. Then there obviously had to be some sort of medium for which light waves had to travel. Sound traveled through air, but what did light travel through? This medium was called the ether. Unfortunately, the idea of an ether conflicted with the laws of Galilean relativity. If the ether existed, then the laws of electromagnetism did not correctly transform under Galilean relativity. This was a huge problem. Either Maxwell was wrong (and experimental evidence said otherwise), or the work done by Galileo and Newton, work that had been unchanged for the last two hundred years, had to be corrected [?].

### 1.3 The Ether

The ether was thought to exist everywhere at once, since it had to follow Maxwell’s equations. The belief was that the ether had little density, and its interaction with matter so small it could barely be noticed. The ether’s sole purpose was to propagate electromagnetic waves. This ether was thought of as fixed, and in it the velocity of light was \( c = 299,792,458 \text{m/s} \). But for other observers moving relative to the ether, the speed of light ought to be different [?].

To make the idea of the ether fit with Galilean relativity, partial solutions were given. One was that the ether is the only absolute standard of rest, and that all motion was relative to it. Another was to assume the velocity of light was \( c \) when measured in a way that the source of the light is
at rest. These were disproved; the first by the aberration of starlight, the second by the Fizeau experiment (see [?] for more detail). One last avenue was that motion through the ether could be detected by bodies with extremely high mass (or that these bodies could “carry” the ether) [?].

This last argument was disproven in 1887 by two Americans, Albert Michelson (who would later go on to win the Nobel prize in physics), and Edward Morley. The thought was that as the earth moved through the ether, the velocity of light should be higher when the earth approached the source of light, and less when measured at right angles. To test this, the now famous Michelson-Morley experiment was devised. Here, a beam of light was split in two, each new beam travelling different paths of equal length. The distance traveled and time taken for the beams to run their course were carefully measured. This entire apparatus floated on mercury, so that it could be rotated (much like the rotation of the earth). To their surprise there was no considerable difference in results from when the device was spun or not. The result was inconclusive: motion through the ether could not be detected. Even more so, it appeared that the velocity of light was independent of the motion of the source (in other words, it was a constant) [?].

Partial solutions were given to explain the null-result of the Michelson-Morley experiment. Lorentz and FitzGerald argued that the lengths of moving bodies would contract in the direction they were travelling. They were somewhat correct: this contraction appears in special relativity, but in a much more general sense. However, most agreed by this point that the idea of the ether was dead. More work was done, but with little success. With one of the greatest problems ever in physics, the stage was set for Albert Einstein to make his grand appearance [?].

1.4 The Relative

Einstein realized what none other had: that the idea of an ether did not just conflict with Galilean relativity, but classical mechanics gave no room for an ether to even exist! Einstein concluded it was pointless in defining a physical quantity unless there was some means of measuring it. Furthermore, motion could only be measured physically with respect to a material body, so we have to ignore all other motion. Einstein was viewing motion as a relative idea, so motion with respect to absolute space was incorrect. This became his first postulate of relativity, the same one which Galileo and Newton had presented years earlier [?].

Postulate 1 There is no absolute standard of rest; only relative motion is observable [?].

Postulate ?? is consistent with the results of the Michelson-Morley experiment, as it makes the question of detecting motion through the ether meaningless. Next, to compensate for the lack of an ether, Einstein realized what Maxwell himself did not: that the velocity of light did not depend on the motion of observers. Light required no medium to travel through, as its velocity was constant. This was his second postulate.
Postulate 2 The velocity of light $c$ is independent of the motion of the source [?].

With these two assumptions, Einstein was able to recreate the mathematics which governed the physical universe. In 1905 he released his crowning achievement, the *Theory of Special Relativity*. At first, most felt that what this new “theory” said was impossible to grasp. Yet as time progresses, we see that relativity has slowly creeped into the general consciousness of humanity. What Einstein predicted in his work has been shown to be true by literally millions of experiments [?].

At first, we will only assume Einstein’s first postulate. This gives us the Galilean transformation. Then we will show how the innocence of the second postulate changes everything. From there, we will shift the mathematics to adhere to these guidelines, and then see what they predict. We start by looking at the world as Galileo did; through the eyes of observers moving with respect to one another in different reference frames.

## 2 Essentials of Relativity

### 2.1 Galilean Reference Frames

When talking about space and time, one imagines a set of axes on which to plot points. However as Galileo made clear, it is necessary to use different axes at different times. In classical mechanics, motion is described in a frame of reference. These frames all contain an origin, a set of Cartesian axes (usually everything is restricted to the first quadrant or octant), and a fourth axis which measures time. This four-dimensional set of axes is known as spacetime [?].
An example can illustrate what is meant by a reference frame. Imagine a traveller who is sitting in a train which is moving in a straight line with a constant velocity relative to the earth (recall Postulate 1). Inside the train, the traveller is at rest. Yet the traveller is not really at rest. Taking a step back and looking at the frame of reference of the earth, the traveller is in a completely different place at one moment in time than another, because the train is moving along its tracks. Further still, the earth is revolving about the sun, which is slowly revolving about the center of the galaxy, etc. In special relativity, we assume that all motion is uniform (there is no acceleration or deceleration).

**Definition 1.** A reference frame is *inertial* if every test particle initially at rest remains at rest and every particle in motion remains in motion without changing speed or direction.

In the train, the traveller can throw balls into the air and catch them as if the train were not moving. He can jump forward or backward, and would not unexpectedly crash into the walls of the train. Even though the train is moving along its track, the motion inside the train remains the same. “The laws of motion are the same in a ‘fixed’ frame and in a frame moving...” with the train, so long as the motion is uniform. Thus to look at motion happening in the train, we can simply look at a reference frame which is moving with the train, as opposed to looking at one that is fixed with respect to the earth. We say the person in the train moves *relative* to the reference frame of the train.

**Definition 2.** An *event* is a point \((t, x, y, z)\) in spacetime.

**Definition 3.** Any line which joins different events associated with a given object will be called a *world-line*.

In the train, the person jumping or throwing balls would all be different events. We can also think of the world-line as the history of the object.
**Definition 4.** An *observer* is a series of clocks in an inertial reference frame which measures the time at which different events take place [?].

An observer must reside inside a particular reference frame, yet an event does not have to. If we were to take an event in the train, say a person jumping, the event is taking place somewhere in spacetime. This event can be observed from inside the reference frame of the train. Additionally it can be viewed from the reference frame of the earth. Because different observers are viewing the same event in different ways, there needs to be some way to link the coordinates of the event in one reference frame to the coordinates of the same event in another. To do this, classical mechanics uses what is now known as the Galilean transformation.

**Theorem 1.** Given two observers, \( O \) and \( O' \), events in the inertial frames of reference set up by the observers are related by:

\[
\begin{align*}
  t' &= t \\
  x' &= x + vt \\
  y' &= y \\
  z' &= z
\end{align*}
\]

where \( v \) is the relative velocity which \( O' \) is moving with respect to \( O \) [?].

*Proof.* Let \( O \) and \( O' \) be two reference frames and let there be an event which occurred at some point \((t, x, y, z)\) in \( O \). Assume \( O' \) is moving with uniform velocity \( v \) in the \( x' \) direction relative to \( O \), but not moving in the \( y' \) or \( z' \) direction. Also, we will assume that the two frames coincided at \( t = t' = 0 \). Thus an event which happened at a given time \( t \) and position \( x \) in \( O \)s coordinate system would happen at some position \( x' = x + vt \) in \( O' \)s. Neither the \( y \) nor \( z \) coordinates changed. \( \square \)

Unfortunately, this theorem is incorrect. The mathematics is fine; assuming what we did (that the speed of light is *not* constant) this would be the logical outcome. It can be generalized to motion in three directions (although the mathematics would be more difficult) in a similar manner. The problem is that it assumes that space and time are absolute quantities.

### 2.2 Time Dilation

But how can time *not* be absolute? Is not a minute for one observer a minute for every other observer, regardless of where they are or how fast their motion?

**Definition 5.** The *proper time* \( \tau \) along the worldline of a particle in constant motion is the time measured in an inertial coordinate system in which the particle is at rest [?].
Imagine a set of mirrors which reflect light rays. If the mirrors are set at a fixed distance from one another, we can measure time by how often the light rays bounce off each mirror, thus effectively creating a “light-clock.” Now assuming Einstein’s second postulate (that the velocity of light is constant), if the mirrors were to move with a constant velocity to the right or left, the distance traveled by the light is farther (by the Pythagorean theorem) than when the mirrors were stationary. Furthermore, the time between ticks has slowed down. In general, observers moving with constant velocity measure time slower than observers at rest. This is true of all clocks, not just our special light-clock. This particular phenomenon is known as time dilation [?].

Definition 6. “Time Dilation is the name given to the phenomenon that happens when a nonproper-time interval is larger than a proper time interval” [?].

Consider two observers, $O$ and $O'$, where $O'$ is moving with constant velocity $v$ with respect to $O$. Suppose a photon (a bundle of electromagnetic energy) is emitted by $O$ toward $O'$ at time $t$ (according to $O$), and is received by $O'$ at time $t'$ according to his clock. Since the velocity of light is $c$ with respect to both observers (regardless if they are moving or not), and one is traveling with a constant velocity, we can assume that according to $O$, $t = kt'$, where $k$ is a constant. But there is nothing special about $O$, or the reference frame in which $O$ is stationary. We can change reference frames so that $O'$ is at rest, and $O$ is now moving with velocity $v$ (this follows from Postulate ??). So according to $O'$, $t' = kt$. This constant $k$ is known as Bondi’s $k$-factor [?].

In Galilean relativity, $k = 1$. But $c$ is constant, so solving for $k$, we can see how the time in one inertial coordinate system is related to the time in another. Consider the above diagram. Because $c$ is a constant, the convention is to set $c = 1$ geometrically, but to write $c$ (as in $ct$) algebraically. It has been set to 1 above for clarity, but it is necessary to consider it in our calculations. The time coordinate $t_B$ and space coordinate $d_B$ can be expressed as:

$$d_B = \frac{1}{2}c(k^2 - 1)t \quad \text{and} \quad t_B = \frac{1}{2}(k^2 + 1)t.$$
Now if $O'$ is moving with constant velocity $v$ according to $O$,

$$v = \frac{d_B}{t_B} = \frac{\frac{1}{2}c(k^2 - 1)t}{\frac{1}{2}(k^2 + 1)t} = \frac{c(k^2 - 1)}{(k^2 + 1)}.$$ 

Now solving for $k$ (notice how important keeping $c$ in our calculations is) gives

\[ vk^2 + v = ck^2 - c \]
\[ c + v = ck^2 - vk^2 \]
\[ c + v = k^2(c - v) \]
\[ k = \sqrt{\frac{c + v}{c - v}} > 1. \]

It follows that

\[ \text{time } E \text{ to } B \text{ measured by } O = \frac{t_B}{kt} = \frac{(k^2 + 1)t}{2kt} = \gamma(v) \]

where substituting for $k$ we find that the gamma factor $\gamma(v)$ is

\[ \gamma(v) = \frac{(k^2 + 1)t}{2kt} = \frac{c + v + 1}{2\sqrt{\frac{c + v}{c - v}}} \]
\[ = \frac{(c + v) + (c - v)}{2\sqrt{\frac{c + v}{c - v}}(c - v)} \]
\[ = \frac{2c}{2\sqrt{(c + v)(c - v)}} \]
\[ = \frac{c}{\sqrt{c^2 - v^2}} = \frac{c}{c\sqrt{1 - v^2/c^2}} = \frac{1}{\sqrt{1 - v^2/c^2}}. \]

With the gamma factor, we can easily see the effects of time dilation. One such example is known as the “Paradox of the twins,” which was originally introduced by Einstein. Suppose a set of twins are born on earth, and that one remains on earth, while the other is blasted off to space, travelling in a straight line at a constant velocity. After some time, the ship stops, turns around, and comes back to earth, following the same path and travelling at the same constant velocity. Then according to the gamma factor, if the twin on earth aged $T$ years, the twin in the rocket ship aged $T/\gamma(v) = T\sqrt{1 - v^2/c^2}$ years!

Yet, if we were to look at a frame in which the twin in the rocket was stationary and the twin
on earth moved away, the opposite would happen. The twin on earth would now be younger! Something is wrong. The problem is that the twin in the rocket ship is not in an inertial frame of reference. The rocket must have accelerated at some point to reach a constant velocity, decelerated, stopped, and then accelerated once again. Not remaining at a constant velocity destroyed the link between the twins, thus we cannot view the scenario where the earth is moving away from the rocket. Since special relativity ignored acceleration, general relativity is required to correctly solve problems of this type [?].

2.3 The Lorentz Transformation

With the gamma factor, we can correctly see how the coordinates of an event in one inertial coordinate system are related to another observer. This is known as the Lorentz transformation.

Theorem 2. The inertial coordinate systems set up by two observers are related by:

\[
\begin{align*}
    t &= t' + \frac{vx'/c^2}{\sqrt{1 - (v/c)^2}} \\
    x &= \frac{x' + vt'}{\sqrt{1 - (v/c)^2}}
\end{align*}
\]

We can write this more concisely as

\[
\begin{pmatrix}
    ct \\
    x
\end{pmatrix} = \gamma(v) \begin{pmatrix}
    1 & v/c \\
    v/c & 1
\end{pmatrix} \begin{pmatrix}
    ct' \\
    x'
\end{pmatrix} \tag{1}
\]

where \( v \) is the relative velocity.

Proof. Look at the diagram. According to \( O \), the event \( B \) has coordinates

\[
t = \frac{1}{2}(t_2 + t_1) = \frac{1}{2}(kt_2' + t_1) \quad x = \frac{1}{2}c(t_2 - t_1) = \frac{1}{2}c(kt_2' - t_1)
\]
where $k$ is Bondi’s k-factor. Now since $t = kt'$, the coordinates of the event $B$ according to $O'$ is:

$$t' = kt = k\left(\frac{1}{2}(t_2 + t_1)\right) = \frac{1}{2}(t'_2 + kt_1)$$

and since $t'$ happened between $t'_1$ and $t'_2$, we have

$$x' = \frac{1}{2}c(t'_2 - t'_1) = \frac{1}{2}c(t'_2 - kt_1)$$

This can be condensed as:

$$\begin{pmatrix} ct \\ x \end{pmatrix} = \frac{c}{2} \begin{pmatrix} 1 & k \\ -1 & k \end{pmatrix} \begin{pmatrix} t_1 \\ t'_2 \end{pmatrix}$$

$$\begin{pmatrix} ct' \\ x' \end{pmatrix} = \frac{c}{2} \begin{pmatrix} k & 1 \\ -k & 1 \end{pmatrix} \begin{pmatrix} t_1 \\ t'_2 \end{pmatrix}$$

Thus we see that

$$\begin{pmatrix} t_1 \\ t'_2 \end{pmatrix} = 2\left(\frac{1}{k} - 1\right)^{-1} \begin{pmatrix} ct \\ x \end{pmatrix} = \frac{2}{c} \begin{pmatrix} k & 1 \\ -k & 1 \end{pmatrix}^{-1} \begin{pmatrix} ct' \\ x' \end{pmatrix}.$$ 

So

$$\begin{pmatrix} ct \\ x \end{pmatrix} = \begin{pmatrix} 1 & k \\ -1 & k \end{pmatrix}^{-1} \begin{pmatrix} ct' \\ x' \end{pmatrix} = \frac{1}{2k} \begin{pmatrix} 1 & k \\ -1 & k \end{pmatrix} \begin{pmatrix} 1 & -1 \\ k & k \end{pmatrix} \begin{pmatrix} ct' \\ x' \end{pmatrix}$$

$$= \frac{1}{2} \begin{pmatrix} k + k^{-1} & k - k^{-1} \\ k - k^{-1} & k + k^{-1} \end{pmatrix} \begin{pmatrix} ct' \\ x' \end{pmatrix}$$

Now

$$k + k^{-1} = \frac{2c}{\sqrt{c^2 - v^2}} \quad \text{and} \quad k - k^{-1} = \frac{2v}{\sqrt{c^2 - v^2}}$$

so

$$\frac{1}{2} \cdot \frac{2c}{\sqrt{c^2 - v^2}} = \gamma(v) \quad \text{and} \quad \frac{1}{2} \cdot \frac{2v}{\sqrt{c^2 - v^2}} = \frac{v}{c} \gamma(v).$$

Therefore,

$$\begin{pmatrix} ct \\ x \end{pmatrix} = \gamma(v) \begin{pmatrix} 1 & v/c \\ v/c & 1 \end{pmatrix} \begin{pmatrix} ct' \\ x' \end{pmatrix}$$

Geometrically, the coordinates of the event are being transformed.
Transformation of Coordinates

Since the $2 \times 2$ matrix in (2) is based upon $v$, we can think of it as adding the velocity $v$ to the particle. From the previous theorem, we have in four dimensions

$$
\begin{pmatrix}
 ct \\ x \\
 y \\
 z
\end{pmatrix} =
\begin{pmatrix}
 \gamma & \gamma v/c & 0 & 0 \\
 \gamma v/c & \gamma & 0 & 0 \\
 0 & 0 & 1 & 0 \\
 0 & 0 & 0 & 1
\end{pmatrix}
\begin{pmatrix}
 ct' \\
 x' \\
 y' \\
 z'
\end{pmatrix}
$$

(2)

where $v$ is the relative velocity and $\gamma = \gamma(v)$. Thus we arrive at a definition for this matrix.

**Definition 7.** The $4 \times 4$ matrix in (2) is known as the *boost*, denoted $L_v$.

With the properties of the boost, we can look at how a moving observer would judge a moving particle’s velocity.

**Lemma 1.** A particle cannot travel faster than $c$, the velocity of light.

**Proof.** Suppose an observer is moving in the $x$ direction with velocity $-v$, and that a particle is travelling relative to the observer with constant velocity $u$. Then by Theorem ?? $x' = -vt + b$ for some constant $b$. In classical mechanics, the new velocity $w$ would be an addition of the old velocities, $w = u + v$. But in relativity, using the boost (in two dimensions)

$$
\begin{pmatrix}
 ct' \\
 x'
\end{pmatrix} = \gamma(u)
\begin{pmatrix}
 1 & -u/c \\
 -u/c & 1
\end{pmatrix}
\begin{pmatrix}
 ct \\
 -vt + b
\end{pmatrix}.
$$

So according to $O'$,

$$
t' = \gamma(u) \left(1 + \frac{uv}{c^2}\right) t - \frac{bu}{c^2} \quad \text{and} \quad x' = \gamma(u) (- (u + v) t + b).
$$
Then the new velocity is

\[ w = \frac{dx'}{dt'} = \frac{v + u}{1 + uv/c^2}. \]

Now, if \( v = c \), we have

\[ \frac{c + u}{1 + (cu)/c^2} = \frac{c^2 + uc}{c + u} = c. \]

Finally, if \( |u| < c \) and \( |v| < c \), \( |w| < c \) since

\[
(c - u)(c - v) = c^2 - (u + v)c + uv > 0 \iff -(u + v)c > -c^2 - uv
\]
\[
u + v < c \left(1 + \frac{uv}{c^2}\right)
\]
\[
w < c
\]

and

\[
(c + u)(c + v) = c^2 + (u + v)c + uv > 0 \iff (u + v)c > -c^2 - uv
\]
\[
u + v > -c \left(1 + \frac{uv}{c^2}\right)
\]
\[
w > -c
\]

So, we cannot combine any number of boosts together to break or even reach the speed of light. As Michio Kaku says, “The speed of light is the ultimate speed limit in the universe” [?]. Although this fact will be more important later, we use the boost to help define a general Lorentz transformation.

In four-dimensions, we can say that

\[
\begin{pmatrix}
ct \\
x \\
y \\
z
\end{pmatrix}
= L
\begin{pmatrix}
ct' \\
x' \\
y' \\
z'
\end{pmatrix}
\]

where \( L = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & H \\ 0 & 0 & K^T \end{pmatrix} \) and \( H, K \) are \( 3 \times 3 \) orthogonal matrices. It is not important to know every term of the matrix \( L \), only that it correctly represents the change in coordinate systems (unlike the Galilean transformation).

**Definition 8.** We say that \( L \) in (??) is a Lorentz transformation if \( L^{-1} = gL^Tg \), where \( g =
We will see shortly where the entries in $g$ come from. With the Lorentz transformation and the concept of time dilation, we have the necessary tools to look deeper at the mathematics behind special relativity. The gamma factor will continue to be used heavily. As for the Lorentz transformation, it will be used for both definitions and calculations. Again, the inspiration for the mathematics to be used comes from observations and experiments which confirm the speed of light is constant. Thus some of the mathematics must be forced to fit where necessary.

3 Space of Relativity

Since the universe we live in has one time coordinate and three space coordinates, we can think of all motion being described in $\mathbb{R}^4 = \mathbb{R} \times \mathbb{R}^3$.

Definition 9. In special relativity, the space we live in is called Minkowski Space, denoted by $\mathbb{M} = \mathbb{R} \times \mathbb{R}^3$ where $\mathbb{R} = \mathbb{R} \times 0$ is the time axes, and $\mathbb{R}^3 = 0 \times \mathbb{R}^3$ the space axes [5].

Minkowski space is not entirely correct when representing our universe (again, because we are in special relativity). In general relativity, acceleration and forces of gravity are viewed as curving spacetime. Minkowski space assumes that the universe is flat, not curved. However, Minkowski space is correct locally, much like measuring distances on a flat map, rather than a globe.

3.1 The Interval

The distance $D$ from any point to the origin is

$$D = \sqrt{x^2 + y^2 + z^2}.$$ 

If a light-pulse is emitted at $t = x = y = z = 0$, it will arrive at $(ct, x, y, z)$ if $ct = \sqrt{x^2 + y^2 + z^2} = D$. Thus $c^2t^2 = D^2$, and $c^2t^2 - x^2 - y^2 - z^2 = 0$ [5].

Definition 10. In Minkowski space, the interval between any two events $x = (t_1, x_1, y_1, z_1)$ and $y = (t_2, x_2, y_2, z_2)$ is defined to be

$$c^2(t_2 - t_1)^2 - (x_2 - x_1)^2 - (y_2 - y_1)^2 - (z_2 - z_1)^2$$ [5].
In an inertial coordinate system, assume that a light-pulse is transmitted when \( ct_1 = \sqrt{x_1^2 + y_1^2 + z_1^2} \), and arrives at \((ct_2, x_2, y_2, z_2)\) when \( ct_2 = \sqrt{x_2^2 + y_2^2 + z_2^2} \). Then
\[
c^2(t_2 - t_1)^2 - (x_2 - x_1)^2 - (y_2 - y_1)^2 - (z_2 - z_1)^2 = 0.
\]

If we consider an observer \( O' \), moving with constant velocity relative to this first frame, then by the Lorentz transformation
\[
c^2(t'_2 - t'_1)^2 - (x'_2 - x'_1)^2 - (y'_2 - y'_1)^2 - (z'_2 - z'_1)^2 = 0
\]
in \( O' \)'s reference frame. This property is known as the invariance of the interval, and simply expresses that the velocity of light is the constant \( c \) for both observers [?].

**Definition 11.** In \( \mathbb{M} \), two objects \( X = (X_0, X_1, X_2, X_3) \) and \( X' = (X'_0, X'_1, X'_2, X'_3) \) are called four-vectors if
\[
X = LX'
\]
where \( L \) is the general Lorentz transformation [?].

Addition and scalar multiplication of four-vectors in Minkowski space works similar to \( \mathbb{R}^4 \). In fact, \( \mathbb{M} \) forms a vector space over \( \mathbb{R} \). Every event in Minkowski space is a four-vector. If we have two events \( x = (ct_1, x_1, y_1, z_1), y = (ct_2, x_2, y_2, z_2) \in \mathbb{M} \), then the displacement four-vector between them is \( X = y - x = c(t_2 - t_1) + (x_2 - x_1) + (y_2 - y_1) + (z_2 - z_1) \). Thinking in terms of vectors rather than individual points, we can redefine how we are to look at the interval between two events.

**Definition 12.** The inner product between two four-vectors \( X, Y \in \mathbb{M} \) is:
\[
g(X, Y) = X_0Y_0 - X_1Y_1 - X_2Y_2 - X_3Y_3 \quad [?].
\]

This definition is slightly misleading, since \( g \) is not exactly an inner product. We can see that for any three four-vectors \( X, Y, Z \in \mathbb{M} \):
\[
g(X + Y, Z) = g((X_0 + Y_0, X_1 + Y_1, X_2 + Y_2, X_3 + Y_3), Z)
= (X_0 + Y_0)Z_0 - (X_1 + Y_1)Z_1 - (X_2 + Y_2)Z_2 - (X_3 + Y_3)Z_3
= X_0Z_0 + Y_0Z_0 - X_1Z_1 - Y_1Z_1 - X_2Z_2 - Y_2Z_2 - X_3Z_3 - Y_3Z_3
= g(X, Z) + g(Y, Z)
\]
Also, for a scalar \( \alpha \in \mathbb{R} \),
\[
g(\alpha X, Y) = \alpha X_0Y_0 - \alpha X_1Y_1 - \alpha X_2Y_2 - \alpha X_3Y_3 = \alpha(X_0Y_0 - X_1Y_1 - X_2Y_2 - X_3Y_3) = \alpha g(X, Y).
\]
But $g$ can be positive, negative, or zero. Thus $g$ is technically an *indefinite* inner product (however we will continue to refer to it as an inner product for short). Because of this property, $g(X, X)$ (or the interval) can be used to split Minkowski space into cones.

![Cone Diagram](image)

Events on the cone are called *lightlike*; they can only be reached from the event $x$ by travelling at the speed of light. Events inside the cone are called *timelike*; they can be reached from $x$ by travelling at speeds less than $c$. All events outside the cone are *spacelike*; they can only be reached from $x$ by travelling faster than the speed of light, and are unreachable by Lemma 1. We can think of the points on, in, and outside this cone as sets.

**Definition 13.** For an event $x \in \mathbb{M}$, the interval through $x$ is broken up into three cones:

- **Light-Cone** $C^L(x) = \{y : g(X, X) = 0\}$
- **Time-Cone** $C^T(x) = \{y : g(X, X) > 0\}$
- **Space-Cone** $C^S(x) = \{y : g(X, X) < 0\}$

where $y$ is any event in $\mathbb{M}$, and $X$ is the displacement vector from $x$ to $y$.

As the picture shows, only the light cone is an actual cone, but we will call the spacelike and timelike events cones as well, simply because it makes set notation more uniform. These cones can further be split up, based upon whether the temporal part $(t_2 - t_1)$ is positive or negative. If positive (the event $y$ takes place *after* $x$) we have the *future light-cone* and *future time-cone*, and if negative (the event $y$ takes place *before* $x$) we have the *past light-cone* and *past time-cone*. The space-cone is not split up, so we ignore the case where $t_2 = t_1$.
Definition 14. For events $x, y \in \mathbb{M}$ we define a partial ordering $<$ on $\mathbb{M}$ by $x < y$ if the displacement vector $X$ from $x$ to $y$ lies in the future light-cone $[?]$.

With the inner product, we can give a more rigorous definition to the idea of Lorentz transformations.

Definition 15. Define the Lorentz group $\mathcal{L} = \{ \lambda : \mathbb{M} \to \mathbb{M} : \forall X, Y \in \mathbb{M}, g(\lambda X, \lambda Y) = g(X, Y) \}$ $[?]$.

The Lorentz group is acting on the vector space $\mathbb{M}$. Here, each $\lambda L$ can be represented by $L$, a general Lorentz transformation. Thus

$$g(X', Y') = g(\lambda X, \lambda Y) = g(LX, LY) = g(X, Y)$$

so the Lorentz group is simply those mappings which leave the interval invariant. In Definition ?? we gave a pseudo-orthogonality condition for $L$. Now we can see where the matrix $g$ in Definition ?? comes from; it is the matrix representation of our inner product (or the interval). It is crucial in defining both the Lorentz transformation and Lorentz group. The set of orthochronous Lorentz transformations also defined in Definition ?? define another group (notice the similarities between this new group and the orthochronous Lorentz transformation).

Definition 16. The orthochronous Lorentz group $\mathcal{L}_+$ is the subgroup of $\mathcal{L}$ whose elements preserve the partial ordering $<$ on $\mathbb{M}$.

3.2 Topology on Minkowski Space

When looking deeper at the vector space $\mathbb{M}$, one may be inclined to think of the topology on $\mathbb{M}$ as the usual Euclidean topology on $\mathbb{R}^4$ ($\mathbb{M}$ essentially is $\mathbb{R}^4$ after all). This topology is:

$$\mathcal{T}^E = \{ U \in P(\mathbb{R}) : N^E_\epsilon(x) \subset U \text{ for all } x \in \mathbb{R} \}.$$  

Here

$$N^E_\epsilon(x) = \{ y : d(x, y) < \epsilon \}$$

where $d$ is the usual Euclidean metric between two points, $x, y \in \mathbb{R}^4$, and $\epsilon > 0$. Physically, using this topology is incorrect for two reasons:

1. The 4-dimensional Euclidean topology is locally homogeneous, yet $\mathbb{M}$ is not (the light cone separates timelike and spacelike events).

2. The group of all homeomorphisms of $\mathbb{R}^4$ include mappings of no physical significance.
The correct topology to use is known as the fine topology, $\mathcal{T}^{F}$. Unlike $T$, it will have the two properties:

1*. $\mathcal{T}^{F}$ is not locally homogenous, and the light cone through any point can be deduced from $\mathcal{T}^{F}$.

2*. The group of all homeomorphisms of the fine topology is generated by the inhomogeneous Lorentz group and dilatations [?].

**Definition 17.** Define the group $G$ to be the group consisting of:

1. $\mathcal{L} (X' = LX)$.

2. Translations ($X' = X + K$ where $K \in \mathbb{R}$ is a constant column vector).

3. Multiplication by a scalar, or dilatations ($X' = \alpha X$ where $\alpha \in \mathbb{R}$).

Every element in $G$ either preserves or reverses the partial ordering $<$ on $\mathbb{M}$. Specifically, those elements which reverse the partial ordering are all part of $\mathcal{L}$, since neither a translation nor dilatation has any effect on the time. From this we can define a new group.

**Definition 18.** Define the group $G_0$ to be the group consisting of:

1. $\mathcal{L}_+$.

2. Translations.

3. Dilatations.

**Definition 19.** The fine topology $\mathcal{T}^{F}$ of $M$ is defined to be the finest topology on $\mathbb{M}$ which induces the 1-dimensional Euclidean topology on every time-axes $g\mathbb{R}$ and the 3-dimensional topology on every space axes $g\mathbb{R}^3$, where $g \in G$. A set $U \in M$ is open in $\mathcal{T}^{F}$ if and only if $U \cap g\mathbb{R}$ is open in $g\mathbb{R}$, and $U \cap g\mathbb{R}^3$ is open in $g\mathbb{R}^3$ [?].

Here, we use $g\mathbb{R}$ and $g\mathbb{R}^3$ to signify that these axes can be acted upon by any element of $G$, yet the sets are no different. The neighborhoods in $\mathcal{T}^{F}$ are defined as:

$$N^F_\epsilon(x) = N^E_\epsilon(x) \cap (C^T(x) \cup C^S(x))$$

where $x \in \mathbb{M}$. Essentially, the epsilon-neighborhoods in $\mathcal{T}^{F}$ are the usual Euclidean neighborhoods with the light cone removed and the point $x$ replaced. It is necessary to show these actually fit into the definition of open sets of the fine topology.

**Theorem 3.** $N^F_\epsilon(x)$ is open in $\mathcal{T}^{F}$.
Proof. Let $A$ be either the time axes or space axes. Then

$$N^F_\epsilon(x) \cap A = \begin{cases} 
N^E_\epsilon(x) \cap A & \text{if } x \in A \\
(N^E_\epsilon(x) \setminus C^L(x)) \cap A & \text{if } x \notin A 
\end{cases}$$

If $x$ is on the $A$ axes, consider the usual Euclidean neighborhoods of $x$ intersect the $A$ axes. We are therefore left with an open interval.

If the point $x$ is not on the $A$ axes, remove the light cone from the Euclidean neighborhood, replace the point $x$, and again intersect with the $A$ axes. Here we are left with disjoint open sets, the union of which are open.

Knowing how this topology works, we are more equipped to prove statements 1* (Theorem ??) and 2* (Corollary ??) above. We first state two lemmas.

**Lemma 2.** Let $h : \mathcal{T}^F \to \mathcal{T}^F$ be a homeomorphism. Then $h$ either preserves the partial ordering or reverses it.

*Proof. See [?].

**Lemma 3.** The group of automorphisms of the set $\mathcal{M}$ preserving the partial ordering is $G_0$. 

19
Proof. See [?].

Theorem 4. The group of all homeomorphisms of $\mathcal{T}^F$ is $G$.[?]

Proof. $\mathcal{T}^F$ is defined invariantly under $G$, so every $g \in G$ is a homeomorphism. Now we have to show every homeomorphism $h : \mathcal{T}^F \to \mathcal{T}^F$ is in $G$.

Now let $g \in G$ be the element corresponding to time reflection. By the previous lemma, either $h$ or $gh$ preserves the partial ordering, and is an element of $G_0$. The group of all homeomorphisms is thus generated by $g$ and $G_0$, which is $G$.

Corollary 1. The light, time, and space cones through a point can be deduced from $\mathcal{T}^F$.[?]

Proof. For $x \in M$, let $G_x$ be the group of homeomorphisms which fix $x$. By the previous theorem, $G_x$ is generated by the Lorentz group and dilatations. Therefore there are exactly four orbits under $G_x$, the point $x$, $C^T(x) \setminus x$, $C^L(x) \setminus x$, and $C^S(x) \setminus x$.

These results are quite exciting. The topological space $(M, \mathcal{T}^F)$ correctly gives the light cone, time cone, and space cone. Even more interesting, the topology which correctly represents our universe, and therefore the one we want to use, only uses Lorentz transformations, translations, and dilatations. Translations and dilatations are far from exciting, so when working in $M$, the only maps we need to consider are Lorentz transformations, which are all invariant under $g$.

4 Predictions of Relativity

We can now see what predictions are made by the mathematics which have so far been developed. Recall that we have only assumed Postulates 1 and 2. From these, Bondi’s $k$-factor led us to develop the gamma factor, which in turn gave us the Lorentz transformation. The topology $\mathcal{T}^F$ on $M$ ensures us that the only mappings which need be considered are in the Lorentz group. All of these predictions (no matter how strange) have been verified by experiments.

If a particle is at rest in $(ct', x', y', z')$, then $x', y', z'$ are all constant. If $(ct, x, y, z)$ is another inertial coordinate system related to the first by a Lorentz transformation and moving with constant velocity $v$, then

$$t = \gamma(v) \tau + k$$

where $k$ is a constant. Thus

$$\frac{dt}{d\tau} = \gamma(v).$$

Definition 20. The four-velocity of a particle, $(V_0, V_1, V_2, V_3)$, is given by:

$$V_0 = c \frac{dt}{d\tau}, \quad V_1 = \frac{dx}{d\tau}, \quad V_2 = \frac{dy}{d\tau}, \quad V_3 = \frac{dz}{d\tau}.$$
It can be shown that a particle’s four-velocity satisfies the properties of a four-vector. If \( X = (X_0, X_1, X_2, X_3) \) then
\[
\frac{dX}{d\tau} = V.
\]
Hence by integration we have that
\[
X = \tau V + K
\]
where \( K \) is a constant column vector.

### 4.1 Effects of \( c \)

**Theorem 5.** Let an observer \( O \) be moving with constant velocity \( V \). Then \( O \) sees two events \( A \) and \( B \) as being simultaneous if and only if the displacement vector \( X \) from \( A \) to \( B \) is orthogonal to \( V \).

**Proof.** Choose a frame in which \( O \) is at rest. Then the displacement vector \( X = (X_0, a_1, a_2, a_3) \), where \( a_1, a_2, a_3 \in \mathbb{R} \) are constants. Then
\[
V = \frac{dX}{d\tau} = \left( \frac{dX_0}{d\tau}, 0, 0, 0 \right) = c\gamma(v).
\]
So \( g(V, X) = c\gamma(v)X_0 \), and \( O \) thinks the events \( A \) and \( B \) are simultaneous \( \iff X_0 = 0 \iff g(X, V) = 0 \). \( \square \)

**Corollary 2.** Two events which are simultaneous for one observer may not be simultaneous for another observer moving at constant velocity with the respect to the first observer.

**Proof.** Let \( O \) be an observer moving with velocity \( V \), and let \( A \) and \( B \) be two simultaneous events in \( O \)’s inertial coordinate system. Now let \( O' \) be an observer moving with velocity \( U \) relative to \( O \). In \( O \)’s coordinate system, \( X = (X_0, a_1, a_2, a_3) \), where \( a_1, a_2, a_3 \) are constants. But since \( O' \) is moving with a constant velocity (according to \( O \)), then \( X' = (X_0, X_1, X_2, X_3) \), \( U = \frac{dU}{d\tau} = \left( \frac{dX_0}{d\tau}, \frac{dX_1}{d\tau}, \frac{dX_2}{d\tau}, \frac{dX_3}{d\tau} \right) \) and if we assume \( g(X', X') > 0 \), then
\[
g(X', U) \geq 0.
\]
\( \square \)

**Lemma 4.** A particle with four-velocity \( V \) has \( g(V, V) = c^2 \).

**Proof.** In an inertial coordinate system where the particle is at rest, \( V = (c, 0, 0, 0) \), hence \( g(V, V) = c^2 \). But \( g \) is invariant, thus this result holds in any inertial coordinate system. \( \square \)

**Theorem 6.** If a rod has length \( R_0 \) in a rest frame, then in an inertial coordinate system oriented in the direction of the unit vector \( \mathbf{e} \) and moving with respect to the rod with velocity \( \mathbf{v} \), the length...
of the rod is:

\[ R = \frac{R_0 \sqrt{c^2 - v^2}}{\sqrt{c^2 - v^2 \sin^2(\theta)}} \]

where \( \theta \) is the angle between \( \mathbf{e} \) and \( \mathbf{v} \).

**Proof.** Let \( V \) be the four-velocity of the rod. If \( X \) is the displacement vector whose length is \( R_0 \), then the displacement vector \( X' \) is

\[ X' = X + \tau V. \]

So

\[ R^2 = -g(X, X) \quad \text{and} \quad R_0^2 = -g(X', X'). \]

Now, \( g(V, Y) = 0 \), since the two ends of the rod are simultaneous in this frame. In the moving frame, \( X = (0, R \mathbf{e}) \) (where \( \mathbf{e} \) is the unit vector) and \( V = \gamma(v)(c, \mathbf{v}) \). So

\[ g(X, V) = -R\gamma(v) \mathbf{e} \cdot \mathbf{v} = -R\gamma(v)v \cos(\theta). \]

Also,

\[ g(X, V) = g(X' - \tau V, V) = -\tau c^2. \]

Solving for \( \tau \) gives

\[ \tau = R\gamma(v)v \cos(\theta). \]

Then the length of the rod \( R \) in the new frame is:

\[
R^2 = -g(X, X) = -g(X' - \tau V, X' - \tau V) = -(g(X', X') + g(\tau V, \tau V)) = R_0^2 - \tau^2 c^2
\]

\[
= R_0^2 - \frac{R^2 v^2 \gamma(v)^2 \cos^2(\theta)}{c^2}
\]

Then combining like terms gives

\[
R^2 \left( 1 + \frac{v^2 \gamma(v)^2 \cos^2(\theta)}{c^2} \right) = R_0^2
\]

\[
R^2 \left( 1 + \frac{v^2}{1 - (v/c)^2} \cos^2(\theta) \right) = R_0^2
\]

\[
R^2 \left( 1 + \frac{v^2 \cos^2(\theta)}{c^2 - v^2} \right) = R_0^2
\]

\[
R^2 \left( \frac{c^2 - v^2 + v^2(1 - \sin^2(\theta))}{c^2 - v^2} \right) = R_0^2
\]

Then

\[ R = \frac{R_0 \sqrt{c^2 - v^2}}{\sqrt{c^2 - v^2 \sin^2(\theta)}} \]
If \( \theta = 0 \), then \( \sin^2(0) = 0 \), and
\[
R = \frac{R_0 \sqrt{c^2 - v^2}}{\sqrt{c^2 - 0}} = R_0 c \sqrt{1 - \frac{v^2}{c^2}} = R_0 \sqrt{1 - \frac{v^2}{c^2}}.
\]

It is interesting to note that time dilation, the Lorentz transformation, simultaneity, and the Lorentz contraction do not have to be introduced in this order. We could have very well introduced the concept of simultaneity first, and then shown time dilation (many sources actually do this). In a sense, these four things are nearly equivalent statements; they are all implications of the fact that \( c \) is constant!

### 4.2 Mass, Momentum, and Energy

In the *Principia*, Newton introduced his three laws of motion. These laws are flawed for two main reasons: first, they assume the velocity of light changes depending upon the motion of the observer and second, they only hold in inertial coordinate systems (we could have thus defined inertial coordinate systems as reference frames which adhere to Newton’s Laws of Motion). The second problem is fine for our purposes, since we are only considering inertial reference frames. Yet we need to examine each law and do some fine-tuning if necessary, to comply with our results up to this point.

**Definition 21.** Newton’s Laws of Motion

1. In the absence of forces, every body remains in its state of rest or moves in a straight line at constant speed [?].

2. The *change of motion* is proportional to the *motive force* impressed, and is made in the direction of the straight line in which that force is impressed [?].

3. For every action there is an equal and opposite reaction [?].

To understand the second law, we need to know what is meant by “change of motion.” To tackle the change of motion, we must be familiar with mass, momentum, and force. Newton defines momentum \( \mathbf{p} \) as \( \mathbf{p} = m \mathbf{v} \) where \( \mathbf{v} \) is the vector velocity of a moving particle, \( m \) is its mass, and change of motion is simply \( d\mathbf{p}/dt \). From this, we see that the force \( \mathbf{f} \) is equal to the change of motion or
\[
\mathbf{f} = \frac{d\mathbf{p}}{dt} = \frac{d(m\mathbf{v})}{dt}.
\]

If \( m \) is constant, we have
\[
\mathbf{f} = m \frac{d\mathbf{v}}{dt} = ma.
\]
Here we come across our first problem. He assumes this mass \( m \), which he calls the inertial mass of the body, is constant.

**Theorem 7.** A body’s inertial mass \( m \) is not constant. Furthermore, \( m \) is a function of \( v \) the velocity of the body in an inertial coordinate system, and can be rewritten as \( m = m(v) \).

**Proof.** Suppose otherwise. Then there exists \( M \) such that \( m \leq M \). By Newton’s Second Law

\[
v(t) = \frac{k}{m} t + b
\]

where \( k \) is a constant force and \( b \) is a constant. If \( k = \frac{d(mv)}{dt} \), then

\[
kt + b = mv \leq Mv \Rightarrow \frac{kt + b}{M} = \frac{k}{M} t + \frac{b}{M} \leq v(t).
\]

Then we can find \( T \) such that

\[
\frac{k}{M} T + \frac{b}{M} = c \leq v(t)
\]

which contradicts Lemma ??.

**Definition 22.** The rest mass \( m_0 = m(0) \) of a body is the mass of a body measured in an inertial coordinate system in which the body is at rest.

**Definition 23.** The four-momentum of a particle is \( P = mV \), where \( m \) is the inertial mass of the particle and \( V \) is the four-velocity of the particle.

**Theorem 8.** The inertial mass of a body moving with velocity \( v \) is \( m = m(v) = \gamma(v)m_0 \).

**Proof.** Let \( O \) be an observer to which the particle is moving, and \( O' \) one to whom the body is at rest. Then according to \( O' \), the mass of the body is \( m_0 \), and its four-momentum is \( P = m_0 V \), where \( V = (c, 0) \). Then

\[
P' = L_v P = L_v m_0 V = \gamma(v) \begin{pmatrix} c & v/c \\ v/c & c \end{pmatrix} \begin{pmatrix} m_0 c \\ 0 \end{pmatrix} = m_0 \gamma(v) \begin{pmatrix} c^2 \\ v \end{pmatrix} = m \begin{pmatrix} c^2 \\ v \end{pmatrix}.
\]

(using the \( L_v \) rather than \( L \) changes nothing, since it is still a Lorentz transformation). So \( m = m_0 \gamma(v) \)

This is another reason why it is impossible to travel at a velocity greater than \( c \), because a body’s inertial mass (or just mass for short) becomes infinitely large as \( v \to c \). And again, as we have seen with time and length, mass is not an absolute quantity, but a relative one. Because of this, \( m \) can be referred to as the relativistic mass of a body. Now that we have corrected Newton’s second law, we can move onto the third. Newton’s third law is referring to collisions between two
bodies (this is what is meant by “action”). An action can be as simple as pushing on a boat (which according to Newton will push back on you) to the collision of moving vehicles, to the splitting of an atom. Collisions come in two types: inelastic, where two bodies come together after colliding (they coalesce); or elastic, where two bodies bounce off one another after colliding.

The four-momentum is invariant under a Lorentz transformation. We assume by Newton’s third law that total momentum is conserved in a collision, or that the total momentum \( P \) before the collision is equal to the total momentum \( \mathcal{P} \) after the collision.

**Theorem 9.** Take two bodies \( G_1 \) and \( G_2 \) in \( O \)'s frame. Let the bodies collide with one another, and let \( O \) say the total four-momentum is conserved. Then a second observer \( O' \) says the total four-momentum is also conserved in his frame.

**Proof.** Let \( P_1, P_2 \) be the four-momentum vector of \( G_1 \) and \( G_2 \), respectively. Then in \( O' \)'s system, \( P_1' = LP_1 \) and \( P_2' = LP_2 \). So

\[
P_1' + P_2' = LP_1 + LP_2 = L(P_1 + P_2) = L\mathcal{P} = L\overrightarrow{P_1} + L\overrightarrow{P_2} = \overrightarrow{Q_1} + \overrightarrow{Q_2} \quad [?] .
\]

This theorem is looking at an elastic collision. Yet in an inelastic collision we arrive at a similar property. Consider two bodies travelling towards each other with equal velocities, and whose rest masses \( m_1 \) and \( m_2 \) are equal (for simplicity we will assume they are moving in only one dimension, so their velocities are \( v_1 \) and \( v_2 = -v_1 \)). Now, according to Taylor’s Theorem,

\[
\gamma(v_1) = \gamma(v_2) = 1 + \frac{1}{2} \frac{v^2}{c^2} + O\left(\frac{v^4}{c^4}\right) \quad [?].
\]

Now the momentum of each body before the collision can be given by

\[
P_1 = (cm_1 \gamma(v_1), m_1 v_1 \gamma(v_1))
\]

and

\[
P_2 = (cm_2 \gamma(v_2), -m_1 v_1 \gamma(v_2)) \quad [?].
\]

Yet after the collision we have

\[
P = P_1 + P_2 = (c(m_1 + m_2) \gamma(v), 0) . \quad (4)
\]

Since the particle is now at rest, it means that the first term is the rest mass of the new object. But this mass

\[
(m_1 + m_2) \gamma(v) > m_1 + m_2
\]
which was the total rest mass of the two bodies before the collision. Extra rest mass was created from the collision!

This extra mass came from the kinetic energy of the objects; it was converted to mass once the bodies came to rest. If we let \( m_0 = m_1 + m_2 \),

\[
m_0 \gamma(v) = m_0 \left( 1 + \frac{1}{2} \frac{v^2}{c^2} + O \left( \frac{v^4}{c^4} \right) \right) = m_0 + m_0 \frac{1}{2} \frac{v^2}{c^2} = \frac{1}{c^2} \left( m_0 c^2 + \frac{1}{2} m_0 v^2 \right).
\]

On the right-hand side, the term \( m_0 c^2 \) is oftentimes called the rest energy, while every term after that is known as the kinetic energy. The entire expression is known as the total energy. From (??) only \( P_0 \) is conserved in a collision, so this is our total energy. Energy must always be given in terms of \( \text{mass} \times \text{velocity}^2 \), so the \( c^2 \) in the denominator can be pulled over and we have

\[
cP_0 = E = \gamma(v) m_0 c^2 \quad [?].
\]

But from Theorem ?? we know \( m = \gamma(v) m_0 \), and we end up with perhaps the most famous equation of all time:

\[
E = mc^2.
\]

The fact that the total energy in a given inertial coordinate is equal to \( mc^2 \) is another direct consequence of the speed of light being constant! Perhaps the idea of the ether being the one absolute standard to which all things be measured was not so crazy after all. The predictions which the mathematics of special relativity gives are all based upon one simple postulate about the velocity of light.

5 Future of Relativity

As was said, special relativity is special because it is limited to inertial coordinate systems. In 1916, Einstein released the General Theory of Relativity, in which he considers all observers in relative motion, be they travelling with constant velocity or not. The mathematics of general relativity is concerned with differential geometry and manifolds, representing how acceleration causes spacetime to curve. General relativity is considered by many to be the greatest achievement of the last century.

However, the world of physics is at another monumental crisis. General relativity appears to explain the motion of bodies of a certain size, yet fails when looking at the quantum level. Thus we have a problem; the theory of relativity does not comply with the laws of quantum mechanics. This is strikingly similar to the problem of Maxwell. The current state of physics rests on two pillars, and one (if not both) must fall [?].

So why study relativity? What if the future of relativity is doomed to become like classical mechanics; reshaped and twisted to the point of becoming an entirely new subject? Just because
there are problems with relativity, does not mean it is unimportant. Global positioning systems use relativity to pinpoint our exact location on earth [?]. The dawn of the atomic age was brought about by dropping two nuclear bombs on Hiroshima and Nagasaki, bombs which rested entirely upon the theory of relativity. As was mentioned, the predictions made by relativity have all been confirmed (except on the quantum level) by experiment. In fact, only those theories which are consistent with relativity need be considered in physics [?]. Relativity has changed, and continues to change our world.

But more importantly, the theory of relativity is one of the greatest examples of the wonderful connections between mathematics and physics. Without the physics, the fascinating mathematics used by relativity (the Lorentz group $\mathcal{L}$, the inner product $g$, the fine topology $T^F$) would never have been created. Yet without the mathematics, finding solutions to explain phenomena like time dilation or the equivalence of mass and energy is hopeless. Much like how Einstein showed the ideas of the ether needed to be better examined, the mathematics of special relativity is beautiful enough that it rightfully deserves a closer look.