I Do Understand Relativity

by

Douglas Ross FRS

Emeritus Professor of Physics,
University of Southampton

Contents

Acknowledgments

I am grateful to Hugh Lyon and Henry Philips for reading an earlier version of this manuscript and making a number of very helpful suggestions.

1 Introduction

This is a sequel to an essay I wrote a few years ago entitled "I don't Understand Quantum Physics". I used this title because *nobody* understands Quantum Physics. The microscopic world behaves in a way which is so orthogonal to our experiences in the macroscopic world that the human mind is simply incapable of conceptualising the microscopic world which is described by Quantum Physics. The best we can do is to assimilate a set of rules which enable us to determine the measurable behaviour of microscopic systems and compare our theoretical predictions with experimental data.

This is *not* the case for Special or General Relativity. I believe that I have a reasonable understanding of how systems behave when they travel at speeds comparable with the speed of light or are accelerating under the influence of gravity. It is my view that the reader will also be able to grasp the necessary concepts - at least for Special Relativity - provided (s)he is prepared to jettison certain pre-conceived ideas. i.e to accept that certain "reasonable" assumptions are in fact incorrect. We have only come to assume their validity because in the case of systems which are travelling (relative to us) at speeds which are much smaller than the speed of light, they are correct to very good (but *not* perfect) accuracy. I list these in the next section.

Meanwhile, let me explain what is meant by Special and General Relativity.

Special Relativity:
This is the study of systems which are moving relative to each other at constant speeds which are comparable to the speed of light, in the absence of a gravitational field.

General Relativity:
This is the study of systems which are moving in a gravitational field and therefore (in general) accelerating relative to each other. I will only touch this subject briefly - it is indeed conceptually difficult and to do it properly requires a University level of mathematics.

For Special Relativity, no sophisticated mathematics is required. What is required is an understanding of Pythagoras' theorem. There is a refresher in Appendix A for the reader who has only a vague recollection (or none at all) of this theorem, and discusses some of its applications. Some algebraic manipulations are required, such as the ability to use Pythagoras theorem to calculate the length of one of the sides of a right-angle triangle adjacent to the right-angle, given the length of the hypotenuse (the side opposite the right-angle) and the length of the other adjacent side. Don't worry - this is explained in Appendix A and, as always, the reader who is uncomfortable with this level of algebraic manipulation can still follow this essay provided (s)he is prepared to accept any algebraic results quoted.

I have also assumed familiarity with some aspect of classical mechanics, such as the concept of momentum and its conservation in collisions, centre-of-mass, and the projection of a vector into its components different directions¿ A brief summary of these is given in Appendix B.

The symbol c is used to denote the speed of light, in a vacuum. It takes the value

$$c \;=\; 299.792458 \text{ million metres per second } \;\approx\; 3 \times 10^8 \,\text{m.s.}^{-1}.$$

In comparison with this, the fastest jet aircraft today has a top speed of about 1000 metres per second. So we see why a subject which is only important at speeds comparable with the speed of light has little effect on our everyday experiences.

Relativity is about comparing the experiences (and for physicists this means the results of measurements) of two observers, A and B who are moving relative to each other. I will usually call these observers Albert (for obvious reasons) and Brenda and I deliberately use non-gender neutral pronouns so that the reader can easily identify which of these is being referred to.

I have tried to minimise the number of technical terms and to define them as they occur, but I cannot know with which of these terms the reader is likely to be familiar. Therefore I have appended an extensive glossary, which I hope is helpful. If there is a term with which the reader is unfamiliar, it can probably be found in the glossary.

2 Five Things We Assume but are Actually "Wrong"

1. If two events at different places are observed by one observer, Albert, to occur simultaneously, they will also be observed to occur simultaneously by a second observer, Brenda, moving relative to Albert.

2. If observer Brenda is moving relative to Albert, her clock will tick at the same time as Albert's. - i.e time is universal.

3. If observer Albert launches an object with a velocity u (in a given direction), then to an observer Brenda, moving relative to Albert (in the same direction) with velocity v, the object is seen to have a velocity $w = u - v$. For example, if you are following a car on the motorway travelling at 130km/hr and you are travelling at 110 km.hr then the car in front is pulling away from you at 130-110 $=$ 20 km/hr.

4. If an observer Albert measures the length of a 1 metre measuring rod, held by Brenda, who is moving relative to Albert in the direction of the length of the measuring rod, it would also be observed to be 1 metre in length.

5. If masses m_1 and m_2 of two chemical elements fuse to form a compound of the two elements (e.g. sodium combining with chlorine to produce common salt – sodium chloride), the mass of the compound will be exactly $m_1 + m_2$, i.e. the sum of the masses of the ingredients.

I have put "wrong" in inverted commas, for two reasons. Firstly, it really means "inconsistent with the postulates of Special Relativity". However, since there has been very substantial experimental verification of the results of Special Relativity, we can safely assume it to be correct - so this point is perhaps unnecessarily pedantic.

The other reason is because although the assumptions are actually wrong they work well to a fantastically high level of accuracy when applied to relative motion which is small compared with the speed of light. Let's look at all of these and see the discrepancies for usual relative velocity magnitudes, and see how fast things need to move for the discrepancies to become significant.

Assumption 1:
Consider a train travelling at 30 metres per second and a railway carriage 22 metres long. If there are flashes of light at the ends of the carriage, which are observed to be simultaneous to a passenger on the train, an observer on the platform will see them to be asynchronous by an amount 0.000000000000007 seconds (so (s)he probably wouldn't notice !). On the other hand if the railway carriage were 100000 kilometres long and the train was travelling at 200000 kilometres per second then the two flashes would be seen by the observer on the platform to be 0.22 seconds apart.

Assumption 2
Brenda is a pilot, flying an MiG Foxbat at 1000 metres per second. She synchronises her watch with ground control and flies for 10 hours. After the flight her watch will be behind that of ground control by 0.00000002 seconds. This is far too little to be observed at this time, but experiments carried out on a space station travelling at 7800 metres per second over a period of six months produced an observable time dilation of 7 milliseconds.

Assumption 3:
Albert is the pilot of an F-111 Aaadvark jet which only has a top speed of 750 metres per second, flying behind a MiG Foxbat which is flying at 1000 metres per second. The speed of the MiG relative to the F-111 is actually 250.000000000025 metres per second.

Assumption 4:
The ISS space station which measures 108 metres from end to end and travels at 7800 metres per second, is seen by an observer on Earth to contract by about 0.000004 metres.

Assumption 5:
If we take 23 gmd of sodium (1 mole) and 35 gms of chlorine (1 mole) and combine them together to make sodium chloride, the mass of the sodium chloride will be less than the sum of the masses of the sodium and the chlorine by 0.0000000045 gms. Thus mass difference (called a "mass defect") is converted onto the chemical binding energy of a mole (58 gmns) of salt, at the conversion rate given by the famous equation of Special Relativity $E = mc^2$. This is the energy released when the chemicla reaction takes place.

In a nuclear reaction the mass deficits are not so small. 2 gms of deuterium (hydrogen with a nucleus consisting of a proton and a neutron) and 3 gms of tritium (hydrogen with a nucleus consisting of a proton and two neutrons) can fuse to make helium and a neutron. In this case the mass deficit is 20 milligrams – releasing a very large quantity of energy (18 bIllion kiloJoules !).

3 A Few Definitions

Before we proceed, let us consider the definitions of some terms that will crop up a lot.

Event:
An event is something that occurs at a particular point at a particular time.

The use of this word is similar to its use in everyday parlance, but in physics we need to be more precise (i.e. pedantic). Whereas we might say that a concert in Hyde Park starting at 2 p.m. next Sunday is an event, in physics we need to state precisely where and when something happens. So if the concert starts with a drumbeat at precisely 2:04:15 then the event might be at the point of contact of the drumstick with the drum at precisely that time. We would specify the place that this occurred by precisely by giving a coordinates (x, y, z) of the point of contact of the drumstick and drum. This allows more precision than the statement "The concert is in Hyde Park, which is opposite the Royal Albert Hall." So the event is specified by four numbers x, y, z, t, where t is the number of seconds after our clock was set to zero, when the concert started. Probably the best thing to think about as an example of an event is an instantaneous flash of light.

Frame of Reference:
This is a coordinate system with measuring rods along three perpendicular axes x, y and z with an agreed orientation and agreed origin, together with a clock that has been set to zero at an agreed time. This way an event can be specified by the four numbers x, y, z, t.

"Frame of reference" is usually abbreviated to "frame."

At this point we need to be careful about what we mean by the time of an event that happened far away. An event is observed after it occurs because it takes time for the light from the event to reach the observer. For example, the nearest star to Earth after the sun is Alpha Centauri, which is 4.2 light years away, so when we observe it, we see it as it was 4.2 years ago. When we specify the time t that an event occurred we subtract the time taken for the light from the event to reach the observer from the time it was observed. In this way all observers who are at rest relative to us agree on the time of the event, even though they may be standing at different distance from the event. If the start of the concert described above is observed 85 km away in Winchester at 2:04:15.60741, then it actually started at

$$2:04:15.60741 - 0:0:\frac{85}{300000} = 2:04:15.60713.$$

At Exeter, which is 252 km away it is observed at 2:04:15.60797, but the observer in Exeter (who is not moving relative to the observer in Winchester) agrees with the observer in Winchester, that the concert started at

$$2:04:15.60797 - 0:0:\frac{252}{300000} = 2:04:15.60713.$$

Inertial frame:
This is a frame of reference which is moving, in the absence of a gravitational field, at constant velocity relative to any other inertial frame, i.e. it is not accelerating.

Special Relativity ignores gravity. Accounting for a gravitational field is the subject of General Relativity and for General Relativity this definition will need to be refined. Of course, we know that on Earth, the gravitational field is not absent but it is one hundred million times weaker than the gravitational field strengths at which General Relativity effects become significant.

4 Relative Motion - Galilean transformations

The main result of Special relativity is that mass m can be converted into energy E at the rate

$$E = mc^2. \tag{4.1}$$

Alternatively, it is possible to convert energy into matter using the same relationship. It is this that is exploited in high-energy particle physics experiments in which massive particles such as Higgs bosons are created through collisions of particles which have been accelerated up to sufficiently high energies and converted into particles whose mass exceeds the masses of the incident particles involved in the collisions.

The relation between energy and mass turns out to be a logical consequence of the observation that motion is relative and one has no way of knowing which inertial frame is "stationary" and which is moving. The connection between these two is certainly not *a priori* obvious, but if we follow strictly the logical steps that Einstein took in the development of the Special Theory of Relativity, we can see that the proposition that "all motion is relative" does indeed have the consequence that energy and mass can be converted into each other at a rate given by (4.1).

Most of us have experienced this by sitting on a train and observing the relative motion of the train on the adjacent railway line and being unable to determine if our train is moving and the other train is still stationary or vice versa (or both)[1]. This was first understood by Galileo and was refined by Einstein. Two inertial frames which are moving relative to each other at speeds much smaller than the speed of light, are (approximately) subject to "Galilean Transformations" with the following rules:

1. All (constant) motion is relative and cannot be detected without reference to an outside point.

2. The laws of motion are the same in any inertial frame.

3. The time at which an event occurs is the same in any inertial frame.

4. If an object C is travelling with velocity u as measured by an observer in inertial frame A, then an observer in inertial frame B which is moving with velocity v relative to A (in the same direction) measures its velocity to be $u - v$.

[1] Actually there is a difference when a train first moves off because the moving frame accelerates and is therefore not strictly an inertial frame. However in reality the acceleration is so small that we do not notice it.

5. If the clocks in inertial frames A and B are synchronised at the moment when their origins coincide and then an event occurs at time t and at a distance x_A as measured by an observer in frame A, then for an observer in frame B, which is moving with velocity v (in the same direction) the event is observed at distance

$$x_B \;=\; x_A - vt. \tag{4.2}$$

as can be seen from Figure 1.

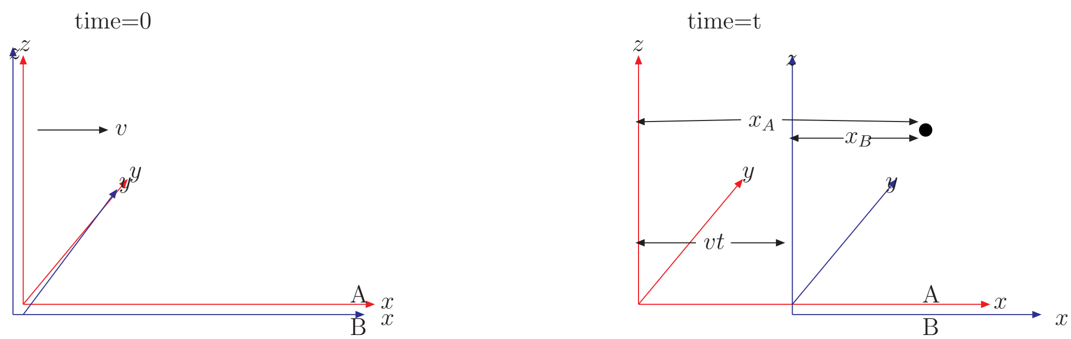


Figure 1: At time $t = 0$ origins of frames A and B coincide (left-hand diagram). At time t (right-hand diagram) an event (black dot) is observed at a distance x_A in frame A corresponding to a distance x_B in frame B. The origin of frame B has advanced a distance vt from the origin of frame A.

We have already stated that rules 3 and 4 are approximations which break down for speeds comparable with the speed of light, where relativistic effects have to be taken into account, and this also applies to rule 5.

At the age of sixteen, Einstein already noticed that there was something not quite right with Galileo's rules as we approach the speed of light. He considered a "thought experiment" in which an observer holds a shaving mirror at arms length and starts to accelerate. The light from his face is reflected in the mirror and he sees his reflection. This occurs despite the fact the the mirror is moving away and so the light approaches the mirror with a lower velocity than it does when the mirror is stationary. But when the person holding the mirror moves with the speed of light, the light never reaches the mirror and the reflection disappears. This contradicts rule 1 since it provides a method of distinguishing a frame moving with the speed of light without reference to any outside point.

At the end of the nineteenth century James Clerk Maxwell developed the theory of electromagnetism which describes the properties of electric and magnetic fields. In general, such fields are both time dependent and position dependent. Maxwell's theory provided equations which relate the dependence of electric fields to the space-dependence of magnetic fields and vice versa. Maxwell also showed that light consisted of wave of oscillating electric and magnetic fields and was able to relate the energy of a light beam to the amplitudes of these oscillating fields. Einstein used his theory of Special Relativity to determine how electric and magnetic fields change when measure in inertial reference frames which are moving relative to each other. This was an important step in the derivation of his equation relating energy, E, and mass, m: $E = mc^2$.

We can avoid using the mathematically complicated equations of Maxwell and instead use a result which Einstein himself had derived in his explanation of the photoelectric effect. Einstein' explanation was that light (or any other electromagnetic radiation) comes in discrete packers called "quanta" which have energy $h\,f$ where f is the frequency of the light and h is Planck's constant. These quanta behave like particles and are now called "photons", and as such they also carry momentum. The next section is a diversion which discusses the difficulties that existed in interpreting the results of experiments on the photoelectric effect and Einstein's explanation, which formed the foundation of Quantum Physics.

5 The Photoelectric Effect.

When light (visible or ultraviolet) is incident on certain metals, electrons can be emitted. This is called the "photoelectric effect". It is the mechanism by which a light meter for a camera works. It was discovered by Heinrich Hertz in 1887. We can understand this effect in terms of the wave nature of light. At any fixed point, an electromagnetic wave (e.g. visible light) is an oscillating electric and magnetic field. When incident upon a metal, which contains loosely bound electrons, the electric field applies a force on these electrons which can sometimes be large enough to liberate them from their binding in the metal. However, when we consider the quantitative effect of such a wave, we predict results which are at variance with experiment. The higher the frequency of the electromagnetic radiation the more often we would expect an electron to be emitted. The maximum kinetic energy of the emitted electrons is expected to increase with increasing intensity of the incident light. Furthermore for very low intensity incident light one would expect a delay before observing the emission of such electrons (known as "photoelectrons") in order to allow the metal to absorb enough energy to free the electrons from the metal.

The first quantitative experiment on the photoelectric effect was conducted in 1902 by Philip Lenard. The maximum energy of the emitted photoelectrons was determined by applying a retarding potential to the emitted electrons and adjusting the applied potential until it reached the "stopping potential", V_0, above which no electrons were collected at the collector electrode. A diagram of the experimental setup is shown in Figure 2. Lenard found that when he increased the frequency of the incident light, the stopping potential increased, i.e. the maximum kinetic energy of the photoelectrons increased. The rate at which the electrons were emitted was found to increase as one increases the intensity of the incident light, but keeping the frequency of the light constant. On the other hand, the electron emission rate was found to be (surprisingly) independent of the frequency of the illuminating light, whereas the classical wave interpretation would imply that at higher frequencies the rate of electron emission would be higher as the electrons were "shaken" with

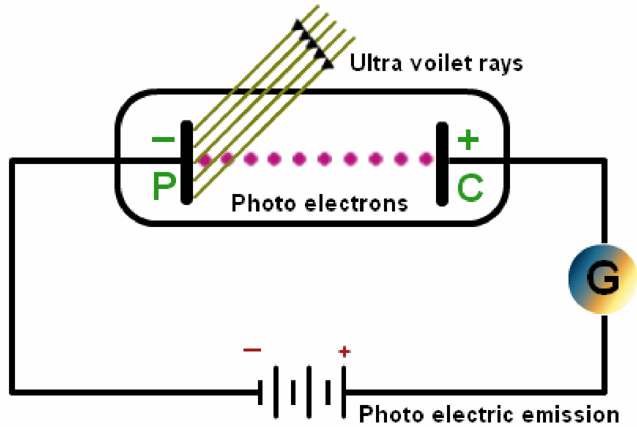


Figure 2: The experimental setup for the photoelectric effect. A voltage is applied so that the emitted electrons are accelerated towards the collecting anode, C. The photoelectric current is measured by the galvanometer, G. The stopping potential is determined by reversing the potential so that it retards the (negatively charged) electrons.

higher frequency. Furthermore, even with the lowest intensity light sources, there was no delay between the initial incidence of the radiation and the emission of the photoelectrons.

The explanation of this enigmatic behaviour was provided by Einstein in 1905[2]:

Light (or any other electromagnetic radiation) comes in packets or "quanta" of energy, E, given by

$$E \ = \ h\,f \tag{5.1}$$

where f is the frequency of the light and h is Planck's constant, which takes the value 6.63×10^{-34} Joule-seconds. These "quanta" behave like particles (they are now called "photons"). These particles scatter off the electrons in the metal. Some of the energy, ϕ , is used to free the electrons from the metal. This energy is called the "work function" and it is a property of the metal from which the electrons are emitted. Some or all of the remaining energy of the photon can be transferred to the emitted electrons, so the maximum energy which the photoelectrons can have is the difference $h\,f - \phi$. This maximum is the stopping potential multiplied by the charge, e, of an electron. We therefore have the simple formula relating the frequency, f, of the incident electromagnetic radiation to the stopping potential V_0

$$h\,f \ = \ \phi + eV_0 \tag{5.2}$$

This formula was tested in detail by Robert Millikan in 1916. His results for the stopping potential for various different frequencies is shown in Figure 3. The points lay on a very good straight line whose slope was proportional to Planck's constant. This was the most

[2]Together with Planck's theory of blackbody radiation, Einstein's theory of the photoelectric effect constitute the birth of Quantum Physics. It was *this* theory that won Einstein the Nobel prize in Physics in 1921 and *not* the Theory of Relativity.

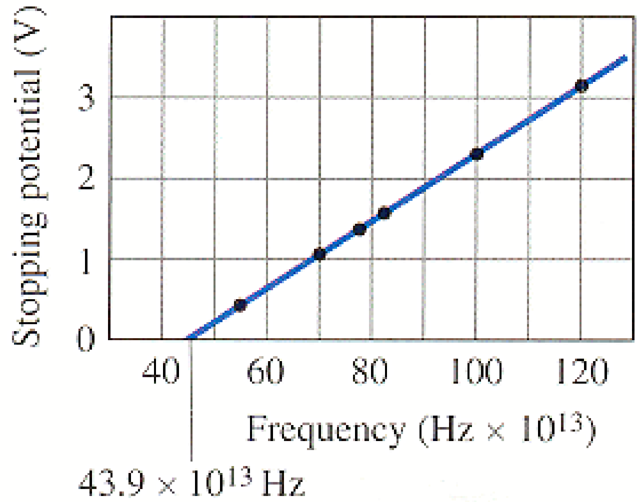


Figure 3: The result of Millikan's experiment on the photoelectric effect showing the stopping potential against the frequency of incident radiation. The slope, $\Delta V/\Delta f$ is Planck's constant divided by the electric charge of the electron. The intercept is equal the work function divided by Planck's constant.

accurate measurement of Planck's constant at that time yielding $h = (6.57 \pm 0.03) \times 10^{-34}$ Joule seconds. The intercept, i.e. the value of frequency, f_0, at which the line crosses the x-axis can be used to determine the work function, since $h\, f_0$ is the energy of a photon which causes the emission of an electron with zero kinetic energy.

From the wave viewpoint one might have expected the energy of the emitted electrons to increase as the intensity of the incident light increases. Increasing the intensity increases the rate of energy per unit area incident on the metal and this increases the total energy of the emitted electrons - but if the frequency of the incident light is held constant then this increase in total energy of emitted electrons manifests itself on an increase in the rage at which electrons ar emitted and *not* in the energy of the individual electrons. The increase in the rate of electrons emitted with increasing intensity of incident light (keeping the frequency is explained by the fact that increased intensity of light means a larger flux of incident photons – and therefore a larger number of interactions which can free an electron from the metal.[3]

6 Conversion of Energy to Matter and Vice-versa

Einstein's postulate that light is emitted in packets that behave like particles, already leads to the idea that energy and matter can be converted into each other.

[3]Einstein's explanation of the photoelectric effect can also be used to explain the observation made by Philipp Lenard in 1900 that light can ionise gases, for which he received the Nobel Prize in 1905. This is very ironic. Lenard was a staunch supporter of National Socialism and condemned the Theory of Relativity on the grounds that it was "Jewish physics."

A thought experiment is shown in Figure 4[4].

Figure 4: A photon is emitted from the left wall of the box. The recoil causes the box to move backwards until the photon impinges on the right wall. The walls of the box have moved backwards but the centre of mass of the box remains in the same place since there is no external force acting on the box. This means that some of the mass of the box has been transported by the photon from the left wall toi the right wall.

Suppose we have a cubic box of mass M and side L, and that a photon is emitted from the left-wall of the box. This photon (which behaves as a particle) must have momentum, p as well as energy, E. This means that by conservation of momentum the box must recoil with velocity v in the opposite direction (the recoil is clearly far too small to be observed which is why this has to be a thought experiment). The box moves backwards until the photon hits the opposite wall. At this point the box stops, but it has been displaced by a small amount Δx.

On the other hand, there has been *no* external force on the box and by Newton's first law of motion this means that the centre of mass (CM) of the box remains in the same place. This can only be possible if the photon transports a small mass Δm from the left wall to the right wall. So what happens is that the emission of the photon amounts to converting a small quantity of mass from the left wall into energy and when the photon is absorbed, the energy is converted back into mass (*The reader who is not familiar with mechanics is welcome to skip the rest of this section and just assimilate the fact that the interpretation of photons as particles leads to the concept of the convertibility of energy and mass.*)

Using conservation of momentum, and the time, L/c, taken for the photon, moving with velocity c, to cross the box of length L, it can be shown that the displacement of the box is

[4]Einstein was very keen on thought experiments but rather less interested in actual experiments. He failed the course on practical because he attended very few of the laboratory sessions.

given by

$$\Delta x \;=\; \frac{L}{Mc}p. \tag{6.1}$$

(6.1) follows from the fact that during the flight of the photon, the box has momentum $p = Mv$ in the opposite direction from the momentum, p, of the photon. The small quantity of mass transferred from one side of the box to the other can be shown to be related to the shift, Δx, of the box and its mass M, by

$$\Delta m L \;=\; M\Delta x \tag{6.2}$$

Combining (6.1) and (6.2), we find

$$\Delta m = \frac{p}{c} \tag{6.3}$$

The only remaining step in he derivation of $E = mc^2$ is to show that the momentum, p of a photon[5] is related to its energy, E by

$$E \;=\; pc.$$

This will emerge later, after we have derived the relativistic expressions for the energy and momentum of a moving particle.

7 Einstein's Postulates

Einstein's theory of Special Relativity is derived from two simple postulates.

1. The laws of physics are the same in any inertial frame.
2. The velocity of light in a vacuum is the same in any inertial frame

The first is a strengthening of Galileo's rules 2 and 3. It tells us that not only are the laws of motion the same in any inertial frame but that *all* laws of physics (and hence all laws of Nature) are the the same in all inertial frames. This means that you cannot perform any experiment which would distinguish between any two frames.

The second is not really an independent postulate. In 1865, James Clerk Maxwell developed the theory of electromagnetism, in which he showed that electric forces and magnetic forces were related and that light consists of a wave of an oscillating electric field and an oscillating magnetic at right-angles to each other and at right-angles to the direction of motion of

[5] In non-relativistic dynamics the momentum of a particle of mass m travelling with velocity v is given by $p = mv$, but this is only valid for particles with velocities much small tan the speed of light. A photon can have momentum, despite the fact that photons have no mass.

the light-wave. He showed that the speed of light[6] was related to the ratio of the electrostatic force, $F_{\mathrm{electric}}(Q_1, Q_2, r)$ between two charges, Q_1 and Q_2 separated by a distance r, and the magnetic force, $F_{\mathrm{magnetic}}(I_1, I_2, r)$, between two wires of length L, separated by a distance r carrying currents I_1 and I_2. We don't need to know the details of this relation - just the fact that the value of the speed of light follows from Maxwell's theory of electromagnetism, which is valid in any reference frame - implying that the speed of light must be the same in any reference frame.

For those who wish to see the relevant formulae - here they are - but the reader can skip to after (7.3) without losing the thread.

The electrostatic force $F_{\mathrm{electric}}(Q_1, Q_2, r)$ between two electric chrges Q_1 and Q_2 separated by a distance r obeys the inverse square law (Coulomb's law) – the force is proprtional to the product of the two charges and inversely proportional to the square of their separation:

$$F_{\mathrm{electric}}(Q_1, Q_2, r) \ = \kappa_E \frac{Q_1 Q_2}{r^2}, \tag{7.1}$$

The electromagnetic force per unit length between two wires carrying currents I_1 and I_2 separated by a distance r is proportional to the product of the currents and inversely proportional to the separation of the two wires: given by

$$\frac{F_{\mathrm{magnetic}}(I_1, I_2, r)}{L} \ = \ \kappa_M \frac{I_1 I_2}{2r} \tag{7.2}$$

Maxwell showed that light was a wave of an electric field and a magnetic field, perpendicular to each other, whose wave velocity c, was the square root of the the ratio between the two constants κ_E and κ_M of (7.1) and (7.2):

$$c \ = \ \sqrt{\frac{\kappa_E}{\kappa_M}} \tag{7.3}$$

This second postulate is then the assertion that Maxwell's equations are correct (they have been thoroughly tested by experiment) and are valid in any inertial frame of reference.

These two postulates appear benign. However, the statement that the speed of light is the same in all inertial frames is in conflict with rule 4 for Galilean transformations. If Albert throws an object with velocity u in a given direction, it "stands to reason" that the velocity of the object as measured by the observer Brenda moving at a velocity v relative to Albert is $u - v$. However much this "stands to reason", it is *wrong* if the object is a light beam, for which the velocity is c in *both* Albert's and Brenda's frames of reference (or any other frame).

[6] We have been using the terms "speed of light" and "velocity if light". Strictly, velocity is a vector which has a direction and the speed is the magnitude of that vector. Nevertheless, we will often use the term "velocity v" meaning velocity of *magnitude*, i.e. speed, v, in some specified direction.

Thus either the postulates of Special Relativity are incorrect, or distance and time differ in different frames so that the measured velocity of light (how far, Δx, does light travel in a given time interval Δt) can be still equal in two different inertial frames because Δx and Δt are *different* in the two inertial frames. The rules for transforming between inertial frames are only *approximately* valid for relative velocities which are small compared with the speed of light.

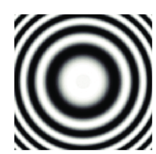
Strange as this may seem, it is not illogical. We will examine whether two events which are observed to be simultaneous in one inertial frame are simultaneous in another inertial frame. But first let's discuss a famous experiment which verified that the speed of light is the same in any inertial frame (despite the fact that the experiment actually set out to demonstrate the contrary).

8 Michelson-Morley Experiment

The universality of the velocity of light (in a vacuum) as postulated by the Special Theory of Relativity has to be verified by experiment.

In the nineteenth century, it was believed that there existed a medium called the the "aether" which carried light waves, in the same way that sound waves are carried by a medium (the air), and that the speed of light meant the speed relative to this aether.

In 1887 Albert Michelson and Edward Morley set out to measure the velocity of the Earth relative to the aether. It was a remarkable experiment and so it is worth spending a little time describing the experiment, which used an interferometer designed by Michelson and which is still used in optics today,



A schematic diagram of the interferometer is shown in Figure 5. Light is split by a half-silvered mirror such that one wave is reflected at an angle of 90° and is reflected by mirror 1, placed at a distance d from the splitting mirror. The other wave continues and is reflected from mirror 2, which is also at a distance d from the splitting point. If the distances are exactly the same then the two light-waves which are detected directly below the splitting point are in phase and will add, leading to a bright spot at the centre. But at a small angle away from directly below the splitting point the two waves of the (monochromatic) light will have travelled distances which differ by half a wavelength so that the two light-waves are out of phase and will interfere destructively, giving rise to an interference minimum, The resultant interference pattern consists of concentric circular fringes as shown above on the right.

Now suppose that the interferometer is travelling through the aether with velocity v in the direction of mirror 2. Since the velocity of light is c relative to the aether the light travels against the aether more slowly from the splitting point to mirror 2 than it does for the return journey from mirror 2 back to the splitting point. The other light-wave, which is travelling perpendicular to the direction of the aether will travel with yet a different speed. The result is that the times taken for the two waves to complete their journeys will be different – the

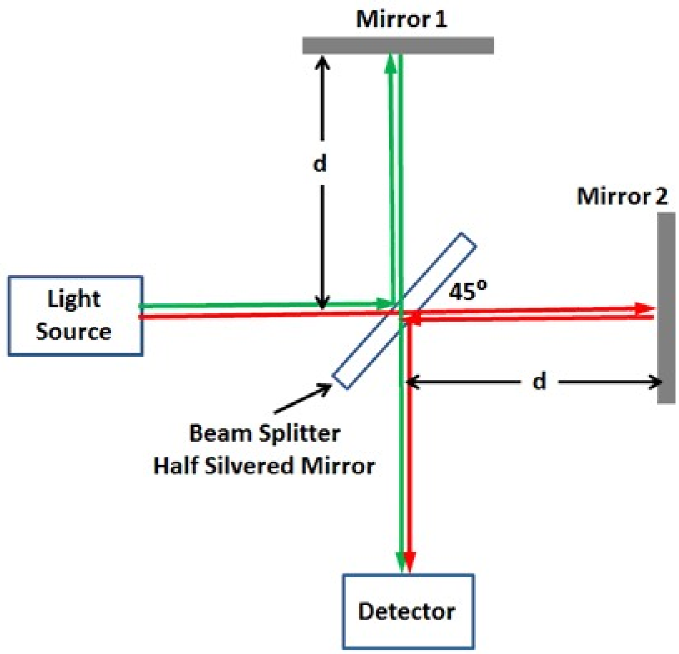


Figure 5: The Michelson interferometer.

two waves will have completed a different number of oscillations and so the fringe pattern will be different.

It is not possible to observe this different fringe pattern just by looking at the fringes. There is only a bright spot in the centre in the idealised situation where the mirror distances d are exactly the same - to within a small fraction of the wavelength of the light. Nevertheless we can see the fringe pattern and are able to measure their separation. If the interferometer is moving through the aether, then as we rotate the apparatus through $90°$ so that it is moving relative to the aether in the direction of mirror 1, one would observe a shift in the interference pattern fringes owing to the interchange of the different journey times of the two waves, as the apparatus is rotated. One expects to see the fringes migrating through the field of vision of the microscope used to observe the fringes.

The arms of the interferometer were 11 metres long and the entire interferometer was floated on mercury for ease of rotation. The microscope used to detect any shift in the interference fringes was able to detect a shift of one hundredth of a fringe-width. Nevertheless, *no shift in the fringes was observed when the apparatus was rotated through* $90°$*,*

The initial reaction to the null result of the Michelson-Morley experiment was to seek some explanation which maintains the hypothesis that light travels on the aether. The most famous of these was "aether drag". It was suggested that the Earth dragged the aether with it as it performed an orbit around the sun, so that the aether was always stationary relative to the Earth. However, if that were the case one would observe "stellar aberration", i.e. shifting of the observed positions of stars as the light from them is bent in different directions when it penetrates the moving aether.

A second explanation was provided by Hendrik Lorentz and George FitzGerald, and is called the "Lorentz-FitzGerald contraction". Their solution was that objects which are moving with velocity v in the direction of the aether contract by a factor of $\sqrt{1 - v^2/c^2}$ (for a derivation of this formula see the first worked problem in the Appendix). This would restore the equality of the times taken for light to complete their journey from the splitting mirror to the mirrors at the end of the arms and back, thereby explaining the null result. However, if this were true then this contraction would be observed in the changes of shapes of planets as they rotate in their own orbits around the sun. No such changes of shape are observed.

The only viable interpretation of the null result is that *there is no aether* and that light travels with the same velocity in a vacuum relative to *any observer*.

Einstein did not rely on the result of this experiment to formulate the Special Theory of Relativity, although he mentioned "unsuccessful attempts to discover any motion of the earth relatively to the 'light medium', He did not refer to this particular experiment in his paper of 1905, but he did say about it later: "If the Michelson–Morley experiment had not brought us into serious embarrassment, no one would have regarded the relativity theory as a (halfway) redemption."

9 Simultaneity

The first casualty of the universality of the speed of light is the notion that if two events occur simultaneously in one inertial frame, they will be observed to be simultaneous in another inertial frame.

That this is *not* the case can be seen from Figure 6, which depicts a railway carriage with the passenger on the train (Albert), seated exactly in the middle of the carriage passing an observer on the platform (Brenda) at time (a). At that moment, as seen by Brenda, lamps at either end of the carriage flash. The light from the two flashes arrives at the observer Brenda on the platform at time (c) and since the flashes occurred at equal distance from the observer, she concludes that the events themselves were simultaneous. However, at time (b) which was before time (c) the light from the front end of the carriage passes the observer, Albert, seated on the train whereas the light from the rear of the carriage has not yet reached him. Once again he is sitting in the middle of the carriage - exactly half-way between the events and therefore concludes (correctly in his frame) that the front flash occurred before the rear flash.

Conversely, if the light from the two flashes had reached the passenger, Albert, simultaneously, he would have concluded that the events were simultaneous, but the light from the flash at the rear of the carriage would already have passed Brenda, on the platform, whereas

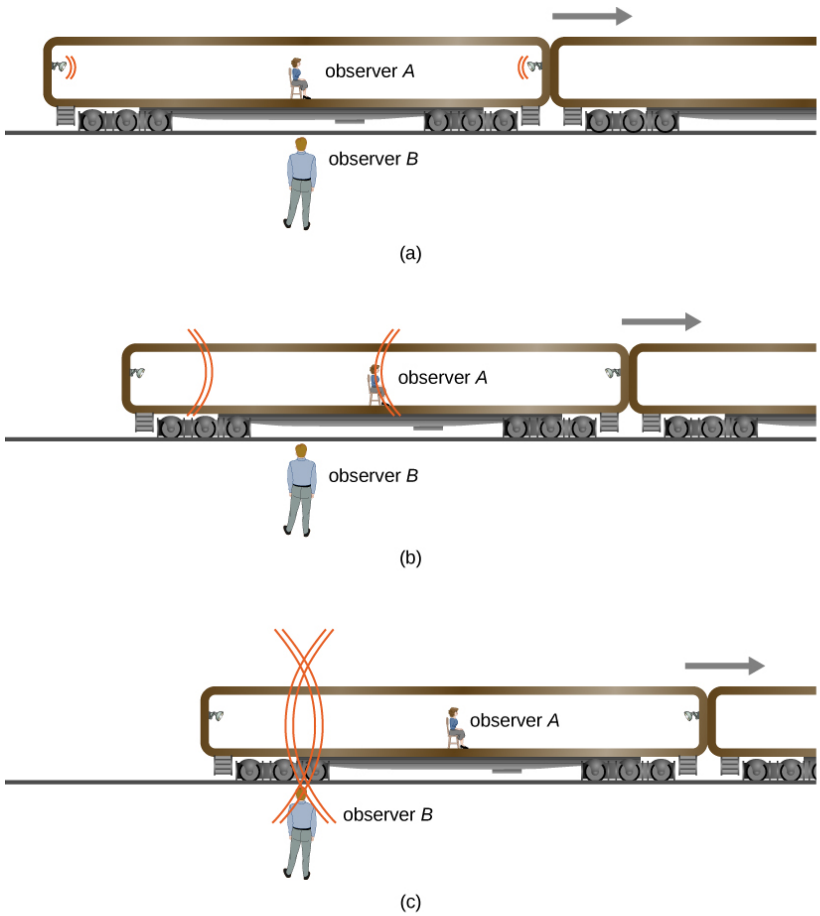


Figure 6: Two events seen as simultaneous by an observer B on the platform

the light from the flash at the front of the carriage would not yet have arrived – Brenda would conclude (again correctly in her frame) that the flash from the rear of the carriage occurred before the flash at the front.

Note that it makes *no* difference whether the source of the events (the lamps) are on the train or on the platform. The light from each of them travels at the same speed for *both* observers.

10 Time Dilation

Not only is the concept of simultaneity dependent on the inertial frame, a time interval itself depends on the inertial frame in which it is measured.

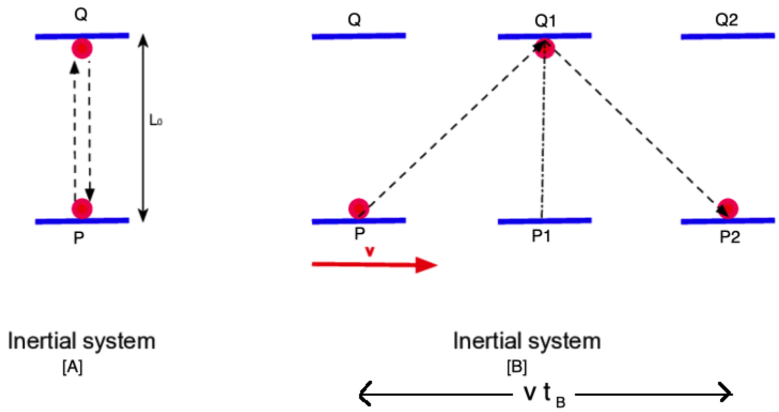


Figure 7: A light clock at rest in Albert's frame ticks when light emitted at P is reflected from a mirror at Q and detected again at P. But for the observer Brenda, who is moving with velocity v relative to A in a direction perpendicular to the direction of the light-clock PQ, the light has to travel further and so since the velocity of light is the same in both frames, Brenda measures a longer time-interval between ticks.

To see how this works, we consider a light-clock shown in Figure 7. A light-signal is sent from an emitter at P to a mirror at Q a distance L away. The light-signal is reflected and when it is detected at P, the clock ticks and another light-signal is emitted immediately. In Albert's frame in which the light-clock is stationary, the time interval between ticks is

$$t_A \;=\; \frac{2L}{c}. \tag{10.1}$$

However, for Brenda, who is moving with velocity v perpendicular to the arm of the light-clock (the line from the emitter P to the mirror Q), the interval t_B, between ticks is longer and during the time interval t_B, the light-clock has moved a distance vt_B relative to Brenda. The clock mirror moves half that distance, $vt_B/2$, between the time that the light is emitted from P and the time that the signal reaches the mirror at $Q1$.

The light travels further from P to $Q1$ and then from $Q1$ to $P2$, than it does in Albert's frame in which it just moved from P to Q and back. This means that the interval between ticks of this clock is longer than the interval between ticks measured by Albert. There is a time-dilation between the two frames.

To determine how this time dilation depends on the velocity of Brenda relative to Albert we will need Pythagoras' theorem. The reader who is daunted by this, but is prepared to accept the result can skip to (10.3).

From P to Q1, the light travels along the hypotenuse of a triangle whose adjacent sides are L and $vt_B/2$. The light travels with velocity c, so the total time taken for the light journey as measured by Brenda is given by

$$c\,t_B = 2\sqrt{L^2 + v^2 t_B^2/4}. \tag{10.2}$$

Substituting for L from (10.1) and re-arranging terms (see problem 2 in the Appendix), we see that in frame B the time interval t_A is dilated to

$$t_B \;=\; \frac{t_A}{\sqrt{1 - v^2/c^2}}. \tag{10.3}$$

This type of clock is rather unusual (and probably doesn't exist in reality). Nevertheless *all* possible ways of measuring time - an hour-glass, a sun-dial, the period of a pendulum, a quartz watch, the half-life of radioactive isotopes, the biological ageing of living organisms, etc. must display exactly the same time dilation, otherwise we would have a mechanism for identifying which frame is stationary and which is moving.

Note also, that t_A is the time interval between two events (the emission and detection of the light-signal) *at the same position.* This is referred to as the "proper-time" It is usually dnoted by the Greek letter τ ("tau"). . The time interval t_B is the dilated time interval between two events which, the the frame B, occur at different positions.

The most spectacular evidence for this time dilation is the observation on Earth of muons created by collisions with cosmic rays in the upper atmosphere, 15 km above the surface of the Earth, travelling at an average speed of 0.994c. Muons have an average lifetime of 2.2 microseconds before decaying, so that even though they travel at nearly the speed of light, a negligible fraction would reach the Earth if they decayed, on average, after 2.2 microseconds. However, the lifetime of 2.2 microseconds refers to the lifetime in the frame of the muons themselves (where the two events - the production of the muon and the decay of the muon are at the same place), whereas for an observer on Earth the lifetime is dilated to around 20 microseconds (measured by a clock on Earth), which is sufficient time for around 10% of the muons to reach the Earth before decaying.

The Twin "Paradox":

"Stayathome" (S) and "Traveller" (T) are two identical twins. On their tenth birthday, T sets off on a journey, moving away from S at four-fifths of the speed of light ($0.8c$) for nine years and then turns around and returns to Earth and to his twin brother, S. He agrees to send S a birthday greeting on each of his birthdays. These greetings are messages that travel from T to S at the speed of light. T sends a total of 18 such birthday greetings and when he returns he is 28 years old.

While T is moving away from S, S receives the birthday greetings at time intervals, $\Delta t_S^{\rightarrow}$ which are more than one year. This is for two reasons. Firstly, when T sends his next greeting he is further way by an amount $v\Delta t_S^{\rightarrow}$ and so takes $v/c\,\Delta t_S^{\rightarrow} \;=\; 4/5\,\Delta t_S^{\rightarrow}$ longer to reach S. In addition we need to account for the time dilation - T is sending messages at intervals of one year at the same place in his frame.

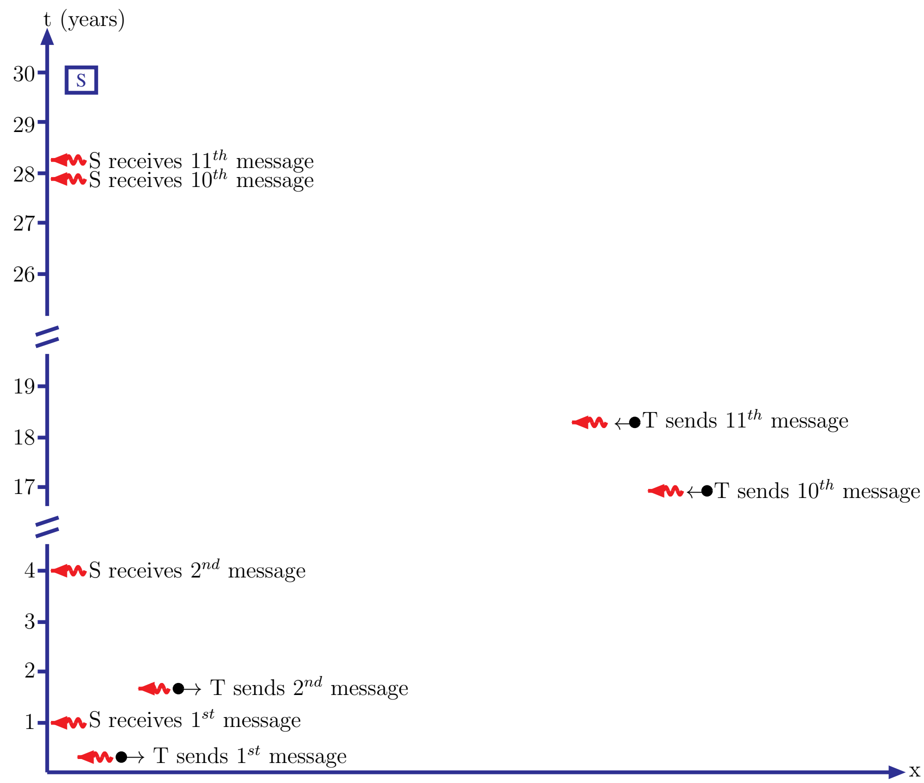


Figure 8: Stayathome's x-coordinate and clock. When Traveller sends the second message he is further away than he was when he sent the first message, so it takes longer for the signal to reach Stayathome - the interval between Stayathome receiving the first and the second massage is longer than the interval beteen the times that they were emitted. On the ther hand, by the time Traveller sends bois tenth message he is on his return journey and when he sends the eleventh message he is closer to Stayathome than when he sends the previous message. The time taken for the signal to reach Stayathome is shorter and is the tiime interval between Stayathome's receipt of messages 10 and 11 is shorter than the difference between times that Traveller sent the messages.

The upshot of this is that S receives messages from T at intervals of

$$\Delta t_S^{\rightarrow} \;=\; \frac{1}{\sqrt{1-(4/5)^2}}\left(1+\frac{4}{5}\right) \;=\; 3 \text{ years}$$

He receives nine such messages and this takes a total time of 27 years. This is illustrated in the lower section of Figure 8.

When T is travelling back to Earth, S receives birthday greetings at intervals of less than a year, because for the return journey T is nearer to S than he was when he sent the previous greeting. The interval $\Delta t_S^{\leftarrow}$ between the reception of T's greetings is now given by

$$\Delta t_S^{\leftarrow} \;=\; \frac{1}{\sqrt{1-(4/5)^2}}\left(1-\frac{4}{5}\right) \;=\; \frac{1}{3} \text{ year}$$

S receives 9 such greetings over a period of 3 years. This illustrated in the upper sections of Figure 8.

This means that as measured by S, the total time taken for the journey is 27+3 = 30 years. S is now 40 years old whereas his twin brother is only 28 !!

This phenomenon is sometimes (erroneously) called the “twin paradox” because at first sight it would appear that the equivalence of the inertial frames of the two twins is broken by the fact that one of the twins is older than the other. This is a trap – the two twins are *not* always both inertial frames, since in order for T to return to Earth, he needs to decelerate and then accelerate in the opposite direction, during which time he is *not* an inertial frame. Because of this, Stayathome and Traveller are *not* equivalent observers. Traveller has experienced a period of deceleration and acceleration which Stayathome has not.

As mentioned above, this dilation has been observed by comparing the time measured by an atomic clock on the ISS space station which travels at 7800 m/s ($v/c \approx 0.000026$) with and atomic clock on Earth. The clock on the space station gains 7 milliseconds every six months.

11 The Relativistic Doppler Effect

The Doppler effect occurs in any wave-motion and is caused by the motion of a wave source relative to the receiver.

We start by considering sound. The source is an oscillation in the pressure of the medium in which the sound-wave is carried, with frequency f_S, meaning that the time interval Δt_S between two successive maxima of the pressure, known as the “period of oscillation” is the reciprocal of the frequency, $1/f_S$. However, if the observer is moving away from the source at a velocity v, the sound from successive maxima have a further distance, $v\Delta t$, to travel and therefore the interval between the reception of successive maxima by the observer is

$$\Delta t_O = \Delta t_S + \frac{v}{w}\Delta t_S, \tag{11.1}$$

where w is the velocity of the sound-wave. Using the fact that the time interval between successive maxima is the reciprocal of the frequency, this gives the relation between the emitted frequency, f_S, and the observed frequency, f_O, as

$$f_O \;=\; \frac{f_S}{(1+v/w)}. \tag{11.2}$$

There is a further correction to this formula due to the time dilation. The period of oscillation of the source is the proper-time between successive maxima as these occur at the same place, but the period between maxima in the frame of the observer is dilated, so that

(11.1) should really be written

$$\Delta t_O = \frac{\Delta t_S}{\sqrt{1 - v^2/c^2}} \left(1 + \frac{v}{w}\right), \tag{11.3}$$

and the relation between frequencies becomes

$$f_O \;=\; \sqrt{1 - v^2/c^2} \frac{f_S}{(1 + v/w)} \tag{11.4}$$

If the wave-motion is light (travelling in a vacuum) rather than sound then we simply replace the wave-velocity, w by c to obtain

$$f_O \;=\; \sqrt{1 - v^2/c^2} \frac{f_S}{(1 + v/c)} \;=\; \sqrt{(1 - v/c)(1 + v/c)} \frac{f_S}{(1 + v/c)} \;=\; \sqrt{\frac{1 - v/c}{1 + v/c}}\, f_S \tag{11.5}$$

In the case of sound and other waves that travel on a medium, one can easily distinguish between the case in which the source is moving relative to the medium or the observer is moving relative to the medium (in the opposite direction). On the other hand, with light travelling through a vacuum ($w = c$) there is no way to distinguish between a source moving with velocity v relative to the observer and the observer relative to the source with velocity $-v$. The result (11.5) is consistent with this, as it can be inverted to give

$$f_S \;=\; \sqrt{\frac{1 + v/c}{1 - v/c}}\, f_O, \tag{11.6}$$

so that the frequency of the source in terms of the frequency measured by the observer is obtained simply by reversing the sign of v. This reciprocity, required by the postulates of Special Relativity, only holds because of the time dilation between the source's clock and the observer's clock.

12 Length Contraction

By now, we should be aware that in Special Relativity space and time are intertwined and if there is a time dilation between two inertial frames moving with a velocity v relative to each other then we expect lengths to be dilated also.

The length, L_A of a measuring rod which is stationary in Albert's frame can be measured by running a probe along it. The probe produces a light flash when it arrives at the near end of the rod at time $t = 0$ and again when it leaves the far end at time $t = \Delta t_A$. The length of the rod is therefore

$$L_A \;=\; v \Delta t_A.$$

But Δt_A is *not* the proper-time: the probe has moved so that the two flashes do not occur at the same place.

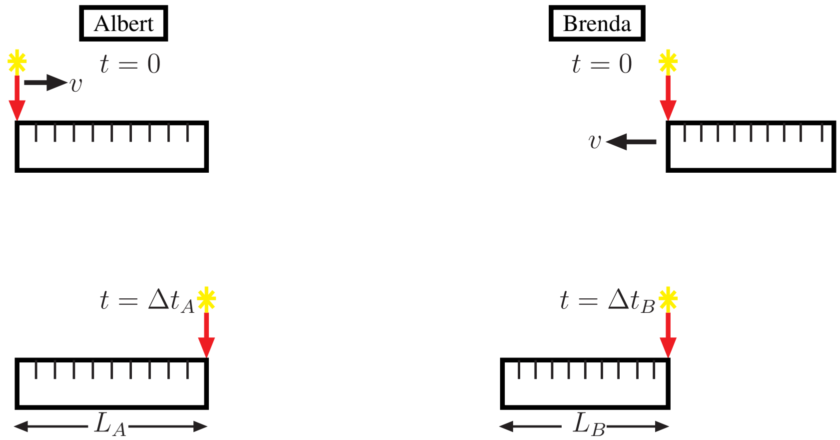


Figure 9: Measuring the length of a rod, in Albert's frame by running a probe along it at velocity v. Brenda is stationary relative to the probe but sees the rod moving with velocity v.

The proper-time is the interval Δt_B between the two flashes in Brenda's frame, as Brenda is moving along with the probe. Brenda sees the measuring rod as moving in the direction of its length at a velocity v (in the opposite direction), so the flashes occur at the same place in her frame (see the right-hand side of Figure 9). From the time dilation formula relating proper-time to the time in another frame,

$$\Delta t_A = \frac{\Delta t_B}{\sqrt{1 - v^2/c^2}},$$

Brenda measures the length of the rod to be

$$L_B = v\Delta t_B = v\sqrt{1 - v^2/c^2}\Delta t_A = \sqrt{1 - v^2/c^2}L_A. \tag{12.1}$$

A rod of length L_A in its rest frame, but which is moving with velocity v, is seen to contract by a factor of $\sqrt{1 - v^2/c^2}$.

This is the same factor as was derived by Lorentz and FitzGerald to explain the null result of the Michelson-Morley experiment. However, its interpretation is different. The Lorentz–FitzGerald contraction refers to contraction of an object which is travelling relative to the aether (which does not exist), whereas the contraction (12.1) refers to the contraction in *any* frame in which the object is moving.

A train whose length is L (at rest) is approaching a tunnel whose length is also precisely L. According to the train driver the tunnel has contracted and is shorter than the train so that it is impossible to fit the train inside the tunnel. On the other hand, according to the signal operator in the signal box beside the tunnel, the train has contracted and she momentarily encloses the train in the tunnel by shutting two gates at the end of the tunnel.

The key to this apparent dichotomy, is to do with simultaneity. The signal operator closes the front and rear gates of the tunnel simultaneously and because the moving train is shorter than the tunnel, the gates momentarily enclose the train in the tunnel, before it crashes through the front gate (if the train slows down and stops in the tunnel it is no longer an inertial frame). As seen by the train driver, the front gate is closed before the rear gate – by the time the rear gate is closed the front of the train has already crashed through the front gate, since in the train driver's frame the tunnel is shorter than the train.

13 Lorentz Transformations

Suppose that Albert releases a flash of light at his origin, $x_A = y_A = z_A = 0$, at the time when he sets his clock to zero $t_A = 0$.

The light travels outwards at speed c in all directions, so that at time t_A the light's wavefront is a sphere of radius $r = c\,t_A$.

By Pythagoras theorem, (see the Appendix), the coordinates, (x_A, y_A, z_A), of a point on the wave-front obeys the relation

$$x_A^2 + y_A^2 + z_A^2 \; = \; r^2 \; = c^2 t_A^2. \tag{13.1}$$

We can imagine a sensor situated at the point (x_A, y_A, z_A), which triggers an event when the light wave-front reaches it. This means that the time of position of the event are related by (13.1).

At time $t = 0$ Brenda passes Albert at a velocity v in the x-direction and sets her clock to zero at the moment that her origin coincides with Albert's. The wave-front of the light at time t_B on Brenda's clock is also (in Brenda's frame) a sphere whose radius is $c\,t_B$ and it is also the case that the same event is observed by Brenda at time t_B at the point (x_B, y_B, z_B) where

$$x_B^2 + y_B^2 + z_B^2 \; = c^2 t_B^2. \tag{13.2}$$

If Brenda moves in the x-direction relative to Albert, the coordinates of the event in the perpendicular, $y-$ and $z-$directions, must be equal, i.e. $y_B = y_A$ and $z_B = z_A$, otherwise we would have a means of distinguishing between the two inertial frames. This means that the combination $x^2 - c^2 t^2$, must take the same value in both frames

$$x_A^2 - c^2 t_A^2 \; = \; x_B^2 - c^2 t_B^2 \tag{13.3}$$

even though x_A and x_B are different and so are t_A and t_B.

Note that we have chosen the example of a light front for which $x^2 + y^2 + z^2 = c^2 t^2$ in order to derive the relation (13.3) we know that if the relative motion between Albert and Brenda is only in the x-direction then for *any* event the y- and z-components of the event will be the same – $y_A = y_B, z_A = z_B$ even if the two event is not triggered by a light wavefront. The relation (13.3) is therefore valid for *any* event if the relative motion between the two frames is in the x-direction.

The relations between the locations of an event x_A and x_B in Albert's and Brenda's frames and between the times t_A and t_B of the event are called "Lorentz transformations". They were first studied in 1889 by the Dutch physicist Hendrik Lorentz.

These transformations have to obey (13.3). Furthermore, since Brenda moving away from Albert with velocity v (in the positive x-direction) is equivalent to Albert moving away from Brenda with velocity $-v$ (in the same direction), one must be able to invert the transformation equations such that x_A and t_A can be found in terms x_B and t_B using the same relations as those giving x_B and t_B in terms of x_A and t_A but with the sign of the velocity reversed.

Before writing down the explicit expressions for the Lorentz transformations, we should note that the Galilean transformations, (4.2), which hold at low velocities,

$$x_B \;=\; x_A - vt_A$$

$$t_B = t_A,$$

cannot satisfy the condition (13.3). If we construct the combination $x^2 - c^2t^2$ and apply the Galilean transformations to the qunatity $x_A^2 - c^2t_A^2$ in Albert's frame, whereas we get in Brenda's frame

$$x_B^2 - c^2t_B^2 \;=\; x_A^2 - (c^2 - v^2)t_A^2 - 2vx_At_A,$$

so that even if v^2 is much less than c^2, there is a cross-term proportional to x_At_A. This can only be cancelled if, contrary to rule 3 of the Galilean transformations, which asserts (erroneously) that time is the same in all frames, the time transforms between frames, such that the expression for t_B, contains a term proportional to x_A, generating a cancelling cross-term, in t_B^2. This dependence of t_B on x_A reflects the fact that a time interval between two events depends on the separation of the two events (as discussed above when comparing a time interval in one frame with the proper-time in another). This has no analogue in non-relativistic dynamics, but it is a necessary consequence of the requirement that two inertial frames are equivalent (a wave-front which is spherical in Albert's frame with a radius that increases at the speed of light is also spherical in Brenda's frame with radius increasing at the same speed). As we shall see, it is this dependence of t_B on x_A that is responsible for the fact that the familiar rule for the addition of velocities, does not hold for velocities comparable with the speed of light.

The Lorentz transformations which satisfy the required conditions discussed above conditions are

$$x_B \;=\; \frac{x_A - vt_A}{\sqrt{1 - v^2/c^2}} \tag{13.4}$$

$$t_B \;=\; \frac{t_A - vx_A/c^2}{\sqrt{1 - v^2/c^2}} \tag{13.5}$$

The reader who is skilled at algebraic manipulations can check that these relations respect (13.3), and also that one can make x_A and t_A the subjects of these equations to obtain (as expected)

$$x_A = \frac{x_B + vt_B}{\sqrt{1 - v^2/c^2}} \tag{13.6}$$

$$t_A = \frac{t_B + vx_B/c^2}{\sqrt{1 - v^2/c^2}} \tag{13.7}$$

Suppose we have two events separated by small distances Δx_A, Δy_A, Δz_A and by a time interval Δt_A in Albert's frame, then we define the "proper-time" interval, $\Delta\tau$, between these events as

$$\Delta\tau^2 = \Delta t_A^2 - \frac{1}{c^2}\left(\Delta x_A^2 + \Delta y_A^2 + \Delta z_A^2\right) \tag{13.8}$$

This is consistent with our previous definition of a proper-time interval, namely the time interval between two events in the frame in which the events occur at the same place. If they occur at the same place in Albert's frame then this means that $\Delta x_A = \Delta y_A = \Delta z_A = 0$, and we see that we can reproduce the time dilation expression (10.3) by setting x_A to zero (Albert holds his clock at his origin) in (13.5),to get

$$t_B = \frac{1}{\sqrt{1 - v^2/c^2}} t_A.$$

The invariance under a Lorentz transformation (13.3) tells us that the proper-time is the same in Brenda's frame, or indeed in any inertial frame moving with any velocity in any direction relative to Albert's frame. It is known as a "Lorentz invariant quantity".

14 Addition of Velocities

Albert launches an object with velocity u_A in the x-direction. The object flashes regularly. He observes the flashes at time intervals, Δt_A, (*after* allowing for the time taken for for light from the flashes to travel the extra distance between Albert and the object). The distance between successive flashes is measured by Albert to be Δx_A, The velocity, u_A of this object as measured by Albert is therefore

$$u_A = \frac{\Delta x_A}{\Delta t_A} \tag{14.1}$$

(The object travels a distance $\Delta x_A = u_A \Delta t_A$ in the time interval Δt_A).

The velocity of the object as measured by Brenda who is moving with velocity v in the same direction is

$$u_B = \frac{\Delta x_B}{\Delta t_B}.$$

Δx_B and Δt_B can be expressed in terms of Δx_A and Δt_A using the Lorentz transformations discussed in the previous section. We find

$$\frac{\Delta x_B}{\Delta t_B} = \frac{(\Delta x_A - v\Delta t_A)}{(\Delta t_A - v\Delta x_A/c^2)}.$$

Dividing throughout by Δt_A and using the expression (14.1) for u_A, we arrive at the relativistic law for the addition of two velocities (in the same direction)

$$u_B = \frac{(u_A - v)}{(1 - u_A v/c^2)} \tag{14.2}$$

There are four noteworthy features of (14.2).

1. The deviation from the "expected" result $u_B = u_A - v$, which we get by applying a Galilean transformation, arises because what we mean by u_B is the distance that the object under investigation travels, as measured by a measuring rod held by Brenda, in a time interval of one second as measured by Brenda's clock. Since we have established that the length of Albert's and Brenda's measuring rods are different and the rate of ticks of Brenda's and Albert's clocks are also different, we have a qualitative understanding of why Brenda's measurement of the velocity is not simply equal to Albert's measurement minus Brenda's velocity relative to Albert.

2. If we invert it (i.e. make u_A the subject of the equation) we find

$$u_A = \frac{(u_B + v)}{(1 + u_B v/c^2)}. \tag{14.3}$$

   In other words the velocity u_A is related to u_B by the same rule but with the opposite sign for the velocity, v, of one frame relative to the other, as required by the equivalence of the two frames.

3. If either v or u_A is equal to the speed of light c, then $u_B = c$. The speed of light is the same in all inertial frames.

4. If u_A or v is much smaller than c then we approximate the familiar rule for adding velocities – namely if a missile passes Albert and he measures the velocity to be u_A, then Brenda, who is moving with velocity v relative to Albert in the same direction, will measure the velocity of the missile to be (approximately) $u_A - v$, provided v and v_A are both much smaller than the speed of light.

Nothing can travel faster than the speed of light. If Albert flies past us in the direction due West at a velocity of nine tenths of the speed of light and Brenda flies past us in the direction due East also at a velocity of nine tenths of the speed of light, then Albert's measurement of Brenda's velocity, v_A^B (or vice versa) is

$$v_A^B = \frac{(0.9c + 0.9c)}{(1 + 0.9 \times 0.9)} = \frac{180}{181}c,$$

– nearly the speed of light but very slightly less.

15 Relativistic Momentum

The relativistic expression for the addition of velocities has ramifications for the conservation of momentum in a collision between two particles of mass m. In Albert's frame the two particles initially head towards each other with equal and opposite velocity, v_A, whereas in Brenda's frame the target particle in initially at rest.

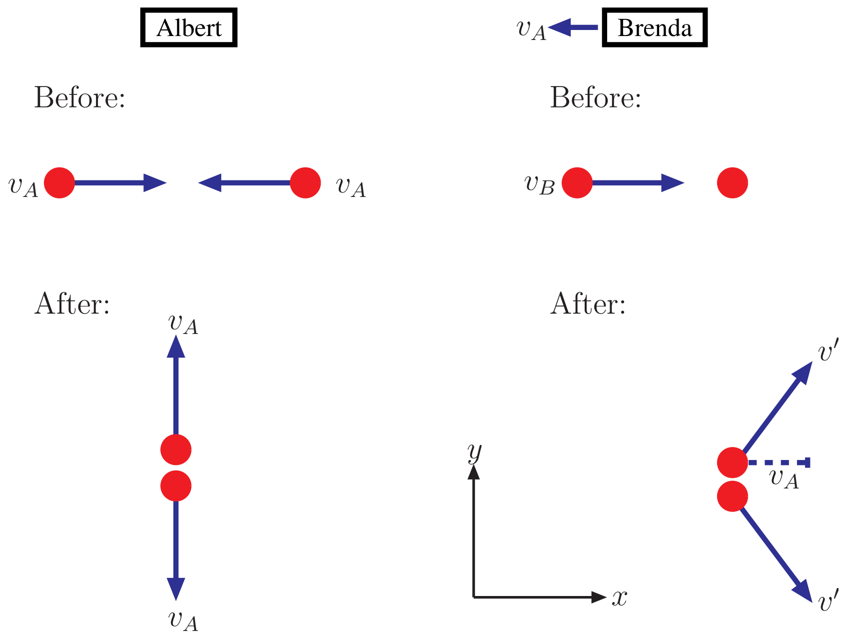


Figure 10: Collision of two equal mass particles in two different inertial frames

In Albert's frame the two particles are moving toward each other in the x-direction with equal and opposite velocity v_A and the collision is such that after the collision they move in opposite directions in the y-direction (i.e. perpendicular to their initial direction of motion). The total momentum of the two particles is zero before the collision and after the collision, so that momentum is conserved as required.

Brenda is moving with velocity v_A relative to Albert in the negative x-direction so she sees a particle of mass m moving with velocity

$$v_B = \frac{2v_A}{1 + v_A^2/c^2}, \tag{15.1}$$

(using (14.2)), colliding with a particle of equal mass which is initially at rest. After collision the particles move off on the positive x-direction with velocity of magnitude v', whose component in the x-direction must be v_A - since in Albert's frame, the particles have zero velocity in the x-direction.

But now the x-component of the velocity times by the mass before the collision is

$$mv_B \;=\; 2m\frac{v_A}{1 + v_A^2/c^2}.$$

whereas after the collision it is

$$2mv_A.$$

These are unequal and so something has gone wrong with momentum conservation.

The problem is repaired if we redefine the momentum p of a particle as

$$\mathbf{p} \;=\; m\frac{\mathbf{v}}{\sqrt{1-v^2/c^2}} \tag{15.2}$$

($\mathbf{p}$ and $\mathbf{v}$ are written in boldface to remind us that they are vectors and have direction as well as magnitude, p and v respectively).

Substituting the relation (15) in the expression for the momentum of the incident particle in Brenda's frame

$$p \;=\; m\frac{v_B}{\sqrt{1-v_B^2/c^2}}$$

and performing some tedious algebra, we can show that

$$m\frac{v_B}{\sqrt{1-v_B^2/c^2}} \;=\; 2m\frac{v_A}{\sqrt{1-v_A/c^2}}$$

so that momentum is conserved also in Brenda's frame.

16 Relativistic Kinetic Energy

In non-relativistic dynamics we learn that a particle of mass m travelling with velocity v has momentum mv and kinetic energy $\frac{1}{2}mv^2$. However, these two formulae are *not* independent. If an object is subjected to a constant force F, then by Newton's second law of motion the rate of change of momentum (in the direction of the force) is F, so that if the object starts from rest at time zero then at time t, its momentum is $F \times t$. During this time, the force has moved through a distance x and the work that the force does is $F \times x$. This work goes to increase the kinetic energy of the particle, so we see that momentum and kinetic energy are indeed related.

In non-relativistic dynamics a constant force generates a uniform acceleration, (as can be seen on the left-hand graph of Figure 11 in which the velocity is increasing linearly in time). However if we use the relativistic expression for momentum (15.2) then the acceleration decreases as the object gains speed and the velocity never exceeds c. We can explain this by interpreting the quantity

$$\frac{m}{\sqrt{1-v^2/c^2}}$$

as the "relativistic mass" which increases with increasing velocity v so that the "inertia" increases with velocity, making it accelerate more slowly.

The right-hand graph of Figure 11 shows the velocity against the work done by the force. For non-relativistic dynamics this leads to the formula $\frac{1}{2}m\,v^2$ for the kinetic energy. But in

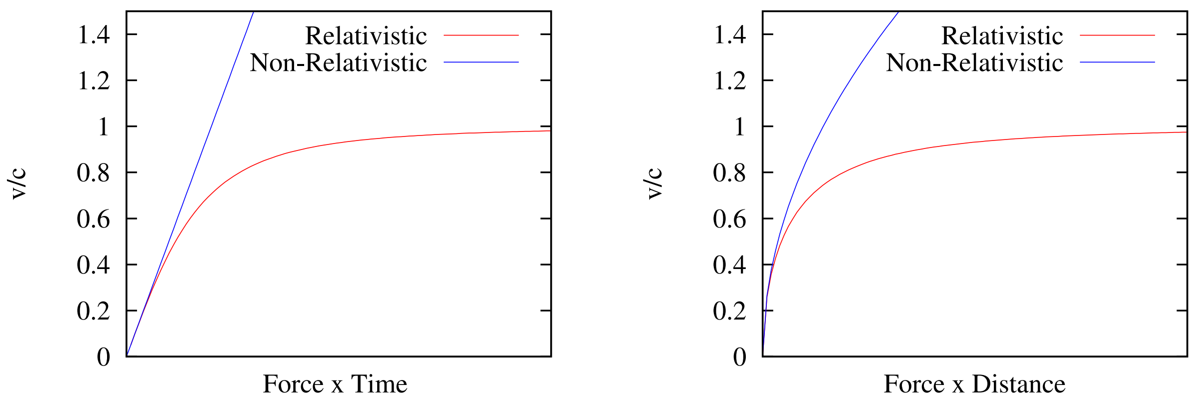


Figure 11: Velocity against time (multiplied by the constant force) in the non-relativistic and relativistic cases (left) and velocity against work done by applied constant force in the non-relativistic and relativistic cases (right).

the relativistic case the velocity increases more slowly as the particle picks up speed - again reflecting an increase in the inertia of the particle. The velocity gets close to c but does not actually reach c for any finite kinetic energy – it requires an infinite amount of energy to reach c.

The derivation of the expressions for kinetic energy from the expressions for momentum require the use of integral calculus, so we ask the reader to accept the relativistic formula for the kinetic energy, T,

$$T = \frac{mc^2}{\sqrt{1 - v^2/c^2}} - mc^2 \tag{16.1}$$

For velocities which are much smaller than c, the right-hand side of (16.1) is approximately $\frac{1}{2}mv^2$[7]. We can see in *both* the graphs of Figure 11 that the relativistic and non-relativistic expressions almost coincide for small velocities. We can check this with a numerical example. A body of mass 1 kg travelling at a speed of 10^8 metres per second (one third of the speed of light) has a kinetic energy (using the relativistic expression) of 5.46×10^8 Joules, whereas the non-relativistic formula $\frac{1}{2}mv^2$ yields the value 5×10^8 Joules – the error of the Newtonian formula is less than 10% even for speeds as large as one third of the speed of light!

17 Relativity and Maxwell's Equations

This section is included for historical reasons. The reader may skip this without losing the narrative.

[7]Showing this requires use of the binomial theorem so once again the reader is asked to accept this result.

The most significant development in Physics at the end of the nineteenth century was Maxwell's Theory of Electromagnetism, which describes the relation between electric and magnetic fields. More precisely it provided relations between the variation of an electric field in space and the variation of a magnetic field in time – and conversely between the variation of an magnetic field in space and the variation of a electric field in time. The mathematics of these relations is very complicated and requires advanced calculus, so henceforth we will only discuss the results qualitatively.

A static distribution of electric charge generates an electric field and a distribution of electric current generates a magnetic field. However, an electric current is nothing other than a moving electric charge. This means that if in Albert's inertial frame, we have a static electric charge and consequently an electric field but *no* magnetic field, then in Brenda's inertial frame, moving relative to Albert, we have an electric current and consequently a magnetic field.

From the Lorentz transformations which relate space an time in different inertial frames and the fact that, from the first postulate of Special Relativity, Maxwell's equations, which relate the space and time dependence of electric and magnetic fields, must be valid in all inertial frames, Einstein was able to determine how electric and magnetic fields transform between inertial frames, moving relative to each other.

One of the main results of Maxwell.s theory was that electromagnetic waves (e.g. light) is a wave of oscillating electric and magnetic fields and that the energy flux passing through a given area in a given time interval, was related to the magnitude of these fields. Furthermore, Maxwell showed that electromagnetic radiation carried momentum which was proportional to the energy flux.

From all this, Einstein showed (using a great deal of tortuous mathematics) that the energy of light signal varied from one inertial frame to another in precisely the same way as the frequency of the wave varied, which is given by the relativistic Doppler effect discussed above.

Ironically, this follows immediately from the interpretation of the photoelectric effect, which he had published earlier that year. If the "light-signal" consists of a single photon, th is energy is proportional to the frequency of the photon.

18 The Master Equation

We are now ready to piece all of this together to get to the relation between energy and mass.

Einstein considered an atom losing energy, ΔE, by emitting two light signals. If the atom remains at rest then the two light-signals must have equal and opposite momenta and hence the same energy, $\Delta E/2$ but in another inertial frame in which the atom is moving such that it possesses some kinetic energy the energy of the light beams differ by an amount given by the Doppler effect. Einstein equated the change in the energy of the light signals with the difference in the kinetic energy of the atom before and after the light emission, and conclude that for energy to be conserved the atom must lose mass when it emits energy in the form of light.

At that time, unstable elementary particles were unknown. Today we know examples of particles which decay entirely into two photons, thereby converting its entire mass into the energy of the two photons. One such example is the neutral pion, π^0, whose mass is $m_\pi = 2.4 \times 10^{-28}$ kg.

Figure 12: Particle decaying into photons from rest (left-hand) and moving at velocity v in the direction of the emitted photons (right-hand).

Consider a particle of mass m at rest in Albert's frame, which decays into two photons. The photons behave like particles and therefore have a momentum whose magnitude depends on the energy of the photons, which is proportional to their frequency (as shown by Einstein's explanation of the photoelectric effect). Therefore for a particle at rest (momentum zero), by conservation of momentum, the two photons must have the same frequency, f, and move in opposite directions, so that the combined momentum of the two photons is zero. The energy, E, liberated in the decay is the sum of the energies of the two photons

$$E = 2hf,$$

(h is Planck's constant).

Now consider the same process in Brenda's frame. Brenda is moving with velocity v in the direction of one of the emitted photons. In her frame, the photon moving towards her has a frequency f_1 which is smaller than f, owing to the (relativistic) Doppler shift. From (11.5)

$$f_1 = f\sqrt{\frac{1-v/c}{1+v/c}} = f\frac{(1-v/c)}{\sqrt{1-v^2/c^2}}$$

whereas the photon moving away from her has a frequency f_2, which is larger than f owing to the (relativistic) Doppler effect, and is also obtained from (11.5), reversing the sign of v:

$$f_2 = f\sqrt{\frac{1+v/c}{1-v/c}} = f\frac{(1+v/c)}{\sqrt{1-v^2/c^2}}$$

Therefore, the combined energy of the photons in Brenda's frame is

$$h(f_1 + f_2) \;=\; \frac{2hf}{\sqrt{1 - v^2/c^2}}$$

This exceeds the energy, E, of the two photons in Albert's frame by

$$\Delta E \;=\; E\left(\frac{1}{\sqrt{1 - v^2/c^2}} - 1\right). \tag{18.1}$$

Historically this was *not* the way Einstein derived this result. The decay of a particle into two photons was unknown in 1905. Instead he considered a particle emitting light with energy E in equal and opposite directions. Using the fact that Maxwell's equations of electromagnetism were valid in any inertial frame and the Lorentz transformations of space and time, he determined how electric an magnetic fields varied from one reference frame to another and showed that the energy given off by the light in a frame moving with velocity v (in the direction of the emitted light, increases by ΔE given by (18.1), so that actually Einstein's derivation did not make direct use of his explanation of the photoelectric effect and the existence of photons.

This excess energy has to be equal too the excess energy of the decaying particle, i.e. its kinetic energy, T, (in Brenda's frame) before it decayed. From (16.1) this is

$$T \;=\; \frac{mc^2}{\sqrt{1 - v^2/c^2}} - mc^2. \tag{18.2}$$

Equating T and ΔE and comparing (18.1) and (18.2), we finally arrive at

$$E \;=\; mc^2 \tag{18.3}$$

In the case of the decay of the π^0 from rest, the energy released is

$$E \;=\; 2.4 \times 10^{-28} \times \left(3 \times 10^8\right)^2 \;=\; 2.16 \times 10^{-11}\,\text{Joules}$$

The frequency, f, of each of the photons is

$$f \;=\frac{E}{2h} \;=\; \frac{1}{2} \times \frac{2.16 \times 10^{-11}}{6.63 \times 10^{-34}} \;=\; 1.63 \times 10^{22}\,\text{Hertz}$$

(corresponding to a wavelength $\approx\, 2 \times 10^{-14}$ metres - well into the γ-ray region of the spectrum).

If Brenda is moving with a velocity 0.6 c in the direction of one of these photons, then the photon towards her has frequency

$$f_1 \;=\; \sqrt{\frac{(1-0.6)}{(1+0.6)}}f \;=\; \sqrt{\frac{0.4}{1.6}} \times (1.63 \times 10^{22}) \;=\; 8.15 \times 10^{21}\,\text{Hertz}$$

and the photon moving towards her her has frequency

$$f_2 \;=\; \sqrt{\frac{(1+0.6)}{(1-0.6)}}f \;=\; \sqrt{\frac{1.6}{4}} \times (1.63 \times 10^{22}) \;=\; 3.26 \times 10^{22}\,\text{Hertz}$$

The total energy of the photons is

$$h(f_1 + f_2) \;=\; 6.63 \times 10^{-34} \times \left(8.15 \times 10^{21} + 3.26 \times 10^{22}\right) \;=\; 2.7 \times 10^{-11}\,\text{Joules}$$

This exceeds the total energy of the photons is Albert's frame by

$$\Delta E \;=\; 2.7 \times 10^{-11} - 2.16 \times 10^{-11} \;=\; 5.4 \times 10^{-12}\,\text{Joules}$$

The kinetic energy of the π^0 in Brenda's frame before it decays is

$$T \;=\; 2.4 \times 10^{-28} \times \left(3 \times 10^8\right)^2 \times \left(\frac{1}{\sqrt{1-0.6^2}} - 1\right) \;=\; 5.4 \times 10^{-12}\,\text{Joules}$$

We see that $T = \Delta E$, as expected.

A deuteron (hydrogen nucleus containing one proton and one neutron) has a mass, m_d, of 3.344×10^{-27} kg. Two such nuclei can fuse to make a nucleus of ^{4}He, which has a mass, m_{He}, of 6.646×10^{-27} kg – slightly less than the mass of two deuterons. The mass difference is converted into energy, so that the fusion reaction releases energy ΔE, where

$$\begin{aligned}\Delta E \;&=\; \left(2m_d - m_{He}\right)c^2 \\ &=\; \left(2 \times (3.344 \times 10^{-27}) - (6.646 \times 10^{-27})\right) \times \left(3 \times 10^8\right)^2 \;=\; 3.78 \times 10^{-12}\,\text{Joules}\end{aligned}$$

This may not seem much, but one gramme of deuterium generates over half a million kilowatt-hours of energy !

It has been quite a long haul to get to $E = mc^2$ – but all the steps were logical consequences of the postulates of Special Relativity. Let us recapitulate all six of the steps which led to this "master equation".

1. From the postulates of Special Relativity, we can derive the Lorentz transformation rules for an event which relate the position and time of the event in one inertial frame with the position and time of the event in another frame.

2. From the time dilation (included in the Lorentz transformations) we can derive the expression for the relativistic Doppler effect of light.

3. From the Lorentz transformations we can derive the relativistic expression for the addition of velocities.

4. The relativistic correction to the addition of velocities implies a relativistic modification of the expression for the momentum of a particle moving with velocity v, so that the law of conservation of momentum in elastic collisions is obeyed in all inertial frames.

5. Since by Newton's second law, the force on an object is equal to the rate of change of its momentum and the work done by that force is the change of its kinetic energy, there is a also a relativistic modification of the expression for the kinetic energy of a particle moving with velocity v.

6. We consider a particle of mass m which decays from rest by the emission of two photons with equal frequencies f. The total energy of the photons is $2h\,f$. When this decay is observed in a frame moving with velocity v, in the direction of the emitted photons, the frequencies of the photons undergo a Doppler shift and consequently the total energy of the photons increases. This increase must be equal to the kinetic energy of the particle in the moving frame, which is only possible if the total energy of the photons in the rest-frame of the decaying particle is $m\,c^2$.

As we have now reached the master equation $E = mc^2$, the exhausted reader may like to stop here. But I would like to outline some of the ideas of General Relativity. This was the real genius of Albert Einstein and was described by the Indian theoretical physicist Sivaramakrishna Chandrasekhar as "probably the most sublime creation of the human mind". Even if the reader does not follow the theoretical ideas, it is worth reading about some of the remarkable experiments that have been conducted to test the surprising predictions of this theory.

Before we can launch into General Relativity, we need to discuss a few more details of Special Relativity.

19 Transformations of Momentum and Relativistic Energy

If we add the "mass energy", mc^2, sometimes called the "rest energy" to the kinetic energy we get the (total) "relativistic energy", E, which, in a frame in which the particle is moving with velocity v is:

$$E \;=\; mc^2\left(\frac{1}{\sqrt{1-v^2/c^2}}-1\right)+mc^2 \;=\; \frac{mc^2}{\sqrt{1-v^2/c^2}} \tag{19.1}$$

Comparing this with the relativistic expression for the (magnitude of) momentum, p, (15.2), we see that

$$E^2 \;=\; p^2c^2+m^2c^4. \tag{19.2}$$

This immediately answers the question of the momentum of a photon. A photon of frequency f has energy hf and has zero mass (only massless particles travel at the speed of light - and they always move with the speed of light[8]). Therefore, a photon has momentum p and energy pc. This is equal to hf, where f is the frequency of the photon. The momentum of a photon of frequency f is therefore

$$p = \frac{hf}{c} \tag{19.3}$$

A particle of mass m with momentm p_A and relativistic energy E_A in Albert's frame will have a different momentum, p_B, and different relativistic energy, E_B, in Brenda's frame, moving relative to Albert. But from (19.2), we see that there is a combination of energy and momentum that is Lorentz invariant - it takes the same value in all inertial frames.

$$E_A^2 - p_A^2c^2 = E_B^2 - p_B^2c^2 = m^2c^4 \tag{19.4}$$

This is clearly the same since mass of the particle is a property of the particle itself and not the frame in which it is observed.

Furthermore since Albert's and Brenda's frames are equivalent the expressions for the energy E_A and momentum p_A is Albert's frame can be obtained in terms of the energy E_B and momentum p_B in Brenda's frame, by using the expressions for E_B, p_B in terms of E_A, p_A by reversing the sign of the relative velocity v. This is analogous to the Lorentz transformations between x_B, t_B and x_A, t_A for an event. Strictly speaking we need to divide by factors of c to get the dimensions correct - so it is actually the quantities, p and E/c^2 that transform in the same way as x and t. This means that observer Brenda, moving relative to observer Albert with velocity v in the x-direction, the x-component of momentum p_x and relativistic energy E in the two frames are related by

$$\begin{aligned} p_{x\,B} &= \frac{p_{x\,A} - vE_A/c^2}{\sqrt{1 - v^2/c^2}} \\ E_B &= \frac{E_A - vp_{x\,A}}{\sqrt{1 - v^2/c^2}} \end{aligned} \tag{19.5}$$

As with the Lorentz transformations for the position and time of an event, we can check

1. $E^2 - p_x^2c^2$ is the same in both frames.
2. (19.5) can be inverted to make E_A and $p_{x\,A}$ the subject of the equations, yielding a set of relations similar to (19.5) with v replaced by $-v$.

The components of momentum perpendicular to the direction of motion of Brenda relative to Albert are the same in both frames - as required by the equivalence of the two frames.

[8]The only known massless particles are photons and gravitons - the quanta of a gravitational field. Originally neutrinos were also thought to be massless, but in 1998 a very small mass for neutrinos was detected.

For a particle at rest in Albert's frame, $E_A = mc^2$ and $p_{x\,A} = p_{y\,A} = p_{z\,A} = 0$. In Brenda's frame the same particle is moving with velocity $-v$ in the x-direction and we reproduce the relativistic expressions for total energy and momentum (using (19.5)):

$$E_B = \frac{mc^2}{\sqrt{1 - v^2/c^2}}$$

$$p_{x\,B} = -\frac{mv}{\sqrt{1 - v^2/c^2}},$$

(the sign indicating that in Brenda's frame the particle is moving in the opposite direction to the direction of Brenda's motion relative to Albert). We also have $p_{y\,B} = p_{z\,B} = 0$.

20 Minkowski Diagrams

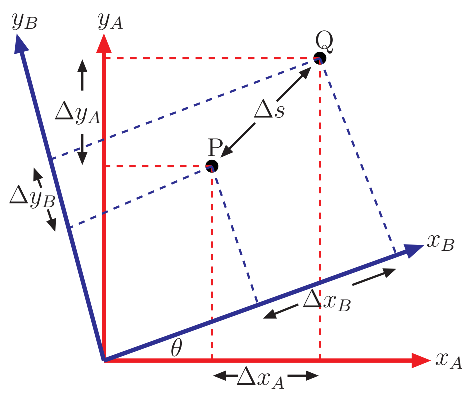


Figure 13: Two points P and Q whose x- and y-coordinates differ by Δx_A and Δy_A in coordinate system A (red), and by Δx_B and Δy_B in frame B (blue), which is rotated (about the z-axis relative to coordinate system A. The distance, Δs, between the two points is the same in both coordinate systems.

Let's consider two neighbouring points, P and Q, whose locations differ by Δx_A in the $x-$direction and Δy_A in the $y-$direction, as measured by Albert's coordinate system (we consider only two dimensions for simplicity, i.e. we assume that the two points lie in the $x-y$ plane with the same value of z). Brenda has her own coordinate system which is rotated through angle, θ, relative to Albert's. In Brenda's system the points differ by Δx_B in her x-direction and by Δy_B in her y-direction (see Figure 13). Whereas these are both different $\Delta\, x_B \neq \Delta\, x_A$ and $\Delta y_B \neq \Delta y_A$, the distance, Δs, between the two point is the same for both Albert and Brenda. By Pythagoras' theorem

$$\Delta s^2 = \Delta x_A^2 + \Delta y_A^2 = \Delta x_B^2 + \Delta y_B^2. \tag{20.1}$$

The distance between two points does does depend on the angle from which you look at them.

Therefore, under a rotation the coordinates Δx_A and Δy_A transform into Δx_B and Δy_B in a way that preserves the quantity Δs, the distance between the two points. The transformation that does this is[9]

$$\begin{aligned} x_B &= \frac{x_A + w y_A}{\sqrt{(1+w^2)}} \\ y_B &= \frac{y_A - w x_A}{\sqrt{(1+w^2)}} \end{aligned} \tag{20.2}$$

Trigonometry can be used to show that the parameter w is related to the angle of rotation, θ, of Brenda's coordinate system relative to Albert's by

$$w = \tan\theta.$$

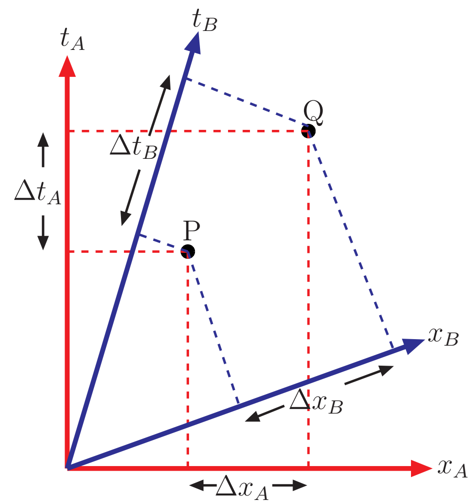


Figure 14: Two events P and Q which occur a distance Δx_A apart and at a time interval Δt_A in Albert's frame (red). The corresponding separation of events is Δx_B and time interval Δt_B in Brenda's frame (blue) which is moving relative to Albert (in the direction of the spatial separation of the two events). The proper-time interval is the same in both frames. In Brenda's frame, the x-axis and t-axis are rotated towards each other by an amount which increases with increasing relative velocity.

All this is reminiscent of the difference in the x-coordinate and time, t, of two events in Albert's and Brenda's frame of reference, where Brenda is moving relative to Albert with velocity v in the x-direction. We assume that the two events occur at the same point in the $y - z$ plane, i.e. $y_B = y_A$ and $z_B = z_A$. In this case, the analogue of distance between

[9]We have assumed that Albert and Brenda have a common origin, i,e, the point $x_A = y_A = 0$ corresponds to the point $x_B = y_B = 0$. As we are discussing the *differences* in coordinates of two neighbouring points, this assumption is not necessary as a shift in origin cancels between the two points.

two points is the difference in proper-time between the two events. This is the same in both frames.

$$\Delta\tau^2 \ = \ \Delta t_A^2 - \left(\frac{\Delta x_A}{c}\right)^2 \ = \ \Delta t_B^2 - \left(\frac{\Delta x_B}{c}\right)^2 \tag{20.3}$$

and the transformations which achieve this invariance of the proper-time are the Lorentz transformations (13.4) and (13.5). These differ from the rotation (20.2) in the assignment of the signs, reflecting the difference in sign of the expression for proper-time difference in terms of Δx and Δt compared with the corresponding expression for the distance between two points in terms of Δx and Δy when we consider a rotation in the $x-y$ plane.

Nevertheless, we can represent a Lorentz transformation diagrammatically, in the same way that we represent a rotation diagrammatically in Figure 13 for rotations, but in this case the axes are x/c and t (we have divided x by c so that the axes have the same dimension, but this only affects the scaling of the diagram). Such a diagram is called a "Minkowski diagram" and an example is shown in Figure 14. A point in the diagram represents an event. In Albert's frame, we draw the x- and time axes at right-angles to each other with the time-axis vertical. The axes representing x-coordinate and time in Brenda's frame are not simply rotations of the axes in Albert's frame, but rather the x-axis is rotated anticlockwise, but the t-axis is rotated clockwise, through an angle related to the relative velocity v. In Brenda's frame the axes are not perpendicular to each other. This reflects the sign difference in the transformations rules for rotation and for Lorentz transformations (rotations preserve $\Delta x^2 + \Delta y^2$ whereas Lorentz transformations preserve $\Delta t^2 - (\Delta x/c)^2$.)

We can see that such a representation immediately encodes the relativistic nature of simultaneity. Two events represented on a horizontal in the diagram are simultaneous in Albert's frame, but occur at different times in Brenda's.

As the velocity of Brenda relative to Albert increases Brenda's x-axis and t-axis rotate towards each other. The maximum possible relative velocity is the one for which Brenda's x- and t- axes coincide. The magnitude of this velocity is the speed of light.

The fact that we can represent transformations in space-time (i.e. Lorentz transformations of an event between two inertial frames) with diagrams that are similar to rotations, leads us to consider an enlarged space in which time is the fourth coordinate. A graph which represented all four coordinates would have to be four-dimensional. Most of us (perhaps all of us) cannot envisage a four-dimensional space, but it exists as a mathematical entity. We can, however, imagine two or three dimensional subspaces, such as the subspace shown in Figure 14. We therefore treat time as a fourth dimension in a four dimensional space called "Minkowski space" in which the invariant quantity is proper-time interval, $\Delta\tau$, given (in the full 4-dimensions) by

$$\Delta\tau^2 \ = \ \Delta t^2 - \frac{1}{c^2}\left(\Delta x^2 + \Delta y^2 + \Delta z^2\right). \tag{20.4}$$

This quantity is invariant under a rotation of any of the space-axis (a rotation does not affect the time axis) as well as under Lorentz transformations in which the time axis and one of the space axes rotate in opposite senses.

We now go on to discuss some of the features of General Relativity and we will see that the definition of this invariant quantity is modified by the presence of a gravitational field.

21 Fundamentals of General Relativity

So far, we have been dealing with inertial frames in the absence of any gravitational field, which are moving relative to each other with constant velocity.

Actually, we are *not* in an inertial frame but rather we live in a gravitational field which leads to an acceleration of about 10 metres per second per second. We are used to this gravitational field. It makes us feel comfortable and keeps our breakfast in the pit of our stomach where it belongs. If we were very far away from any massive body which produces a gravitational field we would experience the unfamiliar feeling of weightlessness – we can experience this for short periods on a roller-coaster when the vehicle is in free-fall. Astronauts in space stations experience this because the force of gravity produces their centrifugal acceleration – so that they are effectively in free fall. They need to adjust their lifestyle to cope with this.

If we found ourselves very far away from any gravitating body, so that we were weightless, the familiar feel of the Earth's gravitational field can be simulated by travelling in a rocket which is accelerating upwards at 10 metres per second per second. This would totally restore the effect of the Earth's gravity and there is no way to distinguish between the two scenarios.

This is essentially Einstein's strong equivalence principle, which states that in a small region around an observer one cannot perform an experiment which distinguishes between a gravitational field and an acceleration.

Since we now wish to discuss reference frames in a gravitational field, we need to amend the definition of an inertial frame. The amended definition of an inertial frame is a reference frame which is in free-fall (hence we are only an approximation to an inertial frame). It is in two such free-falling inertial frames that the result of any experiment must be identical. In the absence of a gravitational field two inertial frames are moving with constant velocity relative to each other (as we have been considering hitherto), but if the frames are at positions with different gravitational potential they will be accelerating (or decelerating) relative to each other.

An immediate consequence of this is that light is bent by gravity as shown in Figure 15 in which the right-hand diagram shows an upwardly accelerated frame. Between the time that the light is emitted from the source and the time it hits the far wall, the upward velocity of the module, in which the experiment is being conducted, has increased and so the light beam lands at a place which is lower down in the module. Since upward acceleration is equivalent to a downward gravitational field, this means that light is bent by gravity.

Unfortunately, this is another example of a seemingly benign phenomenon, which leads to a possible inconsistency, whose resolution is the essence of the General Theory of Relativity.

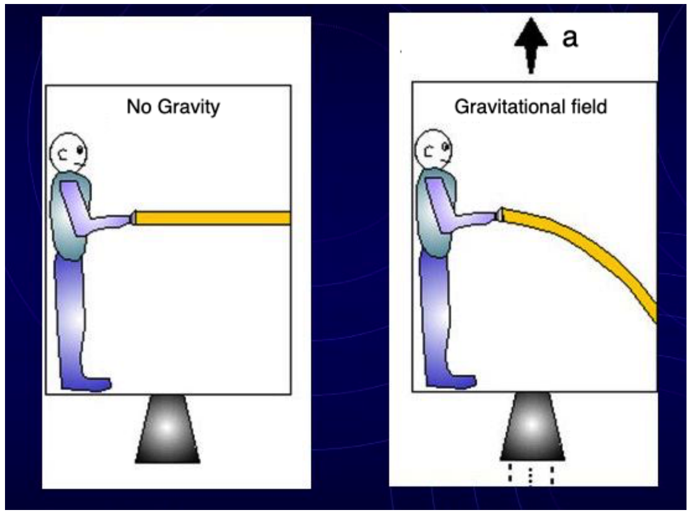


Figure 15: Path of a light beam in no gravitational field (left) and in an upward accelerating frame equivalent to a downward gravitational field (right). Note that the observer on the right is *not* in an inertial frame. She is either stationary relative the surface of the surface of a planet producing a gravitational field or is in an accelerating frame in outer space. In either case this is different from an observer in free fall in a gravitational field.

Suppose Brenda and Albert are orbiting the Earth in space stations and Brenda's space station is further away from the centre of the Earth, so that the gravitational field is weaker than in Albert's space station. She jumps out of her space station in a landing module and hurtles towards the Earth in free fall. Just before she passes Albert's space station, Albert also jumps out in a landing module *(don't worry – both the landing modules are equipped with retro-rockets which will fire before the landing modules crash to the Earth)*. At the time that Brenda passes Albert she is moving with some velocity relative to him, but at that moment they are accelerating at the same rate as they are both the same distance from the centre of the Earth. However, some time later Brenda is closer to the Earth than Albert, where the gravitational field is stronger and therefore she is accelerating relative to Albert. She now fires a light-beam across her landing module. Because of her downward acceleration Albert sees this beam as being bent upwards whereas Brenda, who is also an inertial observer in free fall sees it as a straight line. One of the laws of Physics is that light travels in straight lines and as both Albert and Brenda are in free fall they *both* should observe straight lines.

The "fix" requires a careful definition of a "straight line" as the shortest distance between two points On a flat surface this is what we usually mean by a straight line but of we look at a curved surface then the shortest distance between two points is *not* a straight line but a curve. Mathematicians (who love inventing words that other people do not understand) call such curves "geodesics".

Therefore, we modify the statement "light travels in straight lines" to "light takes the shortest path between to two points". Now we interpret a gravitational field as the agent that causes space to be curved (actually it's both space and time that is curved in the four-dimensional Minkowski space described in the previous section) and a gravitational field can be described by specifying the properties of this curved space-time. Since a gravitational field is caused by the presence of matter, Einstein's equations describing gravity are actually equations which relate certain properties of curved space-time to the density and energy of matter. We cannot go into the details of this but we can discuss some of the features of curved space.

21.1 Properties of Curved Space

The geometry that we learn at school is geometry in flat space and is a set of theorems concerning the properties of shapes drawn on a flat sheet of paper, such as the sum of the angles of a triangle, the circumference of a circle of a given radius, Pythagoras' theorem etc.

The study of geometry in curved space is called "Riemannian geometry" and it is very advanced mathematics.

We cannot imagine curved space in three dimensions, let alone a curved four-dimensional space-time, but we can consider geometry in a two-dimensional sub-space which is curved, such as the surface of a balloon or a globe. On such a curved surface, many of the theorems which we learned about the properties of shapes in flat space (called "Euclidean geometry" as the foundation of the subject was due to the ancient Greek mathematician Euclid of Alexandria), do not hold.

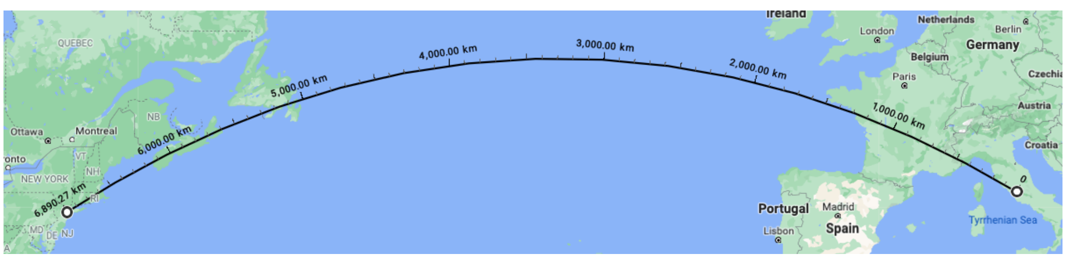


Figure 16: Shortest path between Rome and New York. It is *not* due West, but has a starting direction 55° North

If you were a pilot plotting a route from Rome to New York, which are approximately on the same latitude, you might be tempted to head due West. But that route is 7130 km long. The shortest route (geodesic on the Earth's curved surface) starts in a direction which is approximately North-West and has length 6891 km.

The sum of angles in a triangle drawn on the Earth's surface depends on the size of the triangle. If the triangle is much smaller than the radius of the Earth then it is almost indistinguishable from a triangle drawn on a flat surface, so the sum of their angles is very close to 180°, but for larger triangles we get a larger angle sum. For example, start at the North pole and draw a line down the Greenwhich meridian (longitude 0°) as far as

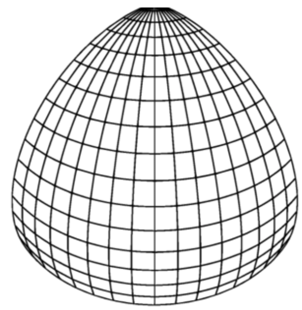


Figure 17: Octant of a sphere whose boundary is a triangle with three right-angles

the equator, turn through a right-angle and go due East until you get to the meridian at longitude 90°, turn again through a right-angle and return along the 90° meridian to the North pole. You approach the North pole at 90° to the Greenwich meridian. This means that you have drawn a triangle with three right-angles - a sum of 270°.

The circumference of a circle of radius R is less than $2\pi R$ and depends on the size of the circle. Again, for small circles whose radius is much smaller than the Earth's surface the curvature, has a negligible effect. On the other hand, suppose we draw a circle whose radius, R, is $\frac{1}{4}$ of the circumference of the Earth, with its centre at the North pole. Such a circle is, in fact, the equator and the circumference is therefore $4R$ - somewhat less than $2\pi R$ for a circle of the same radius drawn in flat space.

Albert is on vacation in a village called Avang-Biloug, which is on the equator in Gabon. He sets the origin of his coordinate system at the centre of Avang-Biloug, with the x axis pointing East and the y-axis pointing North. He has bought a map of Gabon, which is flat - but this is okay since the entire size of Gabon is much smaller than the radius of the Earth and so the curvature of the Earth's surface can be neglected. He travels 12 km due North and 16 km due East to reach the neighbouring village of Andokh Ngang. Using Pythagoras' theorem he deduces that the shortest distance between the villages is $\sqrt{12^2+16^2}=20$ km. Furthermore his GPS tells him that it makes no difference if he first goes North and then East or the other way around.

He now wishes to visit Brenda who is on vacation in the remote Russian village of Bobrovka, which is 8000 km due East and then 6000 km due North. Not understanding the effect of the curvature of the Earth, he uses Pythagoras' theorem to deduce (incorrectly) that Bobrovka is $\sqrt{8000^2+6000^2}=10000$ km away. In fact, the distance of Bobrovka from Avang-Biloug is only 9600 km. Even more bizarre is the fact that if Albert first travels 6000 km due North and then 8000 km due East he ends up in Novaya – which is 1300 km away from Bobrovka to the East. Over large distances it matters in which order you travel given distances. This is a consequence of the curvature of the surface of the Earth.

Brenda sets her coordinate system up in the same way as Albert - namely she places her origin at the centre of Bobrovka with the x-axis pointing East and the y-axis pointing North.

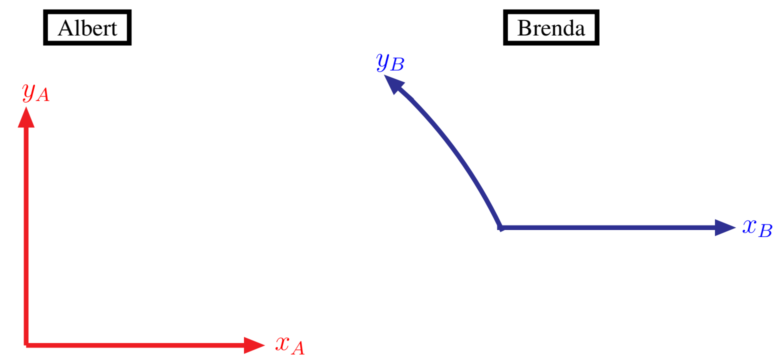


Figure 18: Albert's view (from Avang-Biloug) of Brenda's axes (in Bobrovka). Note that Brenda's y-axis is seen to be inclining to the West and curved. We can picture the y-axes in Albert's and Brenda's frames as two meridian lines at different longitudes.

But if Albert looks at her coordinate system (while he is still in Gabon) he sees her x-axis pointing in the same direction as his, whereas he sees her y-axis as pointing at an angle of 125° to the x-direction rather than at right-angles. Moreover, the inclination of Brenda's (North-pointing) y-axis increases as one goes North and this means that Albert's view of Brenda's y-axis is actually curved (see Figure 18).

All this unfamiliar behaviour can be encoded by amending the expression for the distance, Δs, between two neighbouring points (20.1) to

$$\Delta s^2 \;=\; g_{xx}\Delta x^2 + g_{yy}\Delta y^2 \tag{21.1}$$

where the coefficients g_{xx} and g_{yy} depend on position. For small x and y - where we are near the origin, they are both very close to one, so that flat-space (Euclidean) geometry is a very good approximation.

In the case of a spherical surface of radius R, g_{xx} and g_{yy} are

$$g_{xx} \;=\; \sqrt{1-\frac{y^2}{R^2}}$$

$$g_{yy} = 1,$$

so we see that if y is much smaller than the radius R the effect of the curvature is negligible. But for points with larger values of y, the x-axis "shrinks". On a globe this is equivalent to the fact that the distance between places with a given difference in longitude, decreases as one moves away from the equator. The distance between two places on the equator whose longitudes differ by 1° is 111 km, whereas the distance beween Pisa and Florence which are almost at the same latitude and approximately 1° apart in longitude, is only 80 km.

We can, of course, have other types of curved surfaces for which g_{xx} has a different dependence of position (x, y), and g_{yy} differs from one.

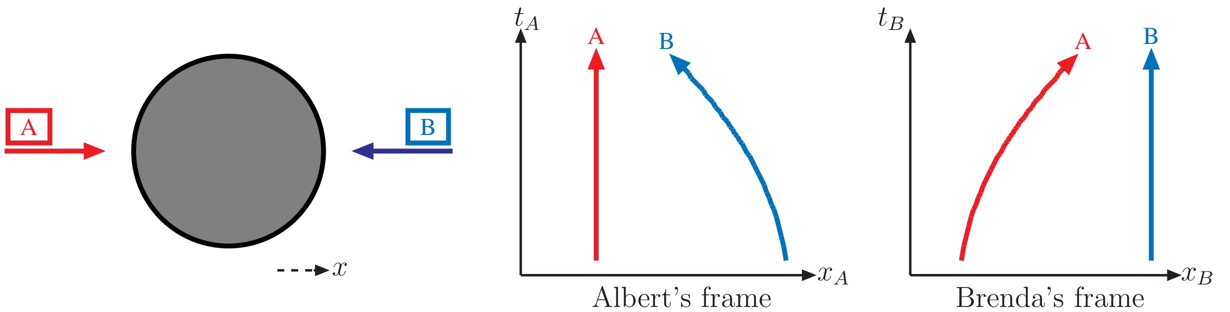


Figure 19: Albert and Brenda are *both* inertial observers in free fall towards the Earth from opposite sides.

In General Relativity, the effect of the curvature is to amend the definition of proper-time (20.4). For a static gravitational field in the x-direction the proper-time is amended to

$$\Delta\tau^2 \ = \ \left(1 + 2\Phi/c^2\right)\Delta t^2 - \frac{1}{(c^2 + 2\Phi)}\Delta x^2 - \frac{1}{c^2}\left(\Delta y^2 + \Delta z^2\right) \tag{21.2}$$

where Φ is the (position dependent) gravitational potential. We see that in the absence of gravity ($\Phi = 0$) the proper-time coincides with (20.4).

An example of how a gravitational field leads to curvature of space-time can be seen in Figure 19, which depicts Albert and Brenda in free fall towards the Earth but from opposite sides (n direction x), so that as viewed by a distant observer they are accelerating towards each other. Both are inertial observers. However, in Albert's frame his world-line plotted on his space-time Minkowski diagram, is a straight line parallel to th time-axis, since his x-position remains unchanged as his clock ticks, but he observes Brenda to be moving towards him with increasing speed world so that he sees her world-line as a curve in a similar way to the view from Avang-Bilong of a north pointing ruler located further East at Bobrovka, which is seen to be curved as a result f the curvature f the surface of the Earth. On the other hand, Brenda views her world line t be a straight line parallel to her time axis whereas she sees Albert's world-line too be a curve representing th fact that she sees Albert as accelerating towards her. This is analogous to the curvature of a north pointing rules in Avang-Bilong as seen by an observer in Bobrovka.

Very large gravitational fields are required for this curvature to make any noticeable difference. For example, the Earth's gravitational potential, Φ at the Earth's surface is such that Φ/c^2 is about one millionth of a millionth.

For gravitational potentials that are not too large, the determination of the path in space-time taken by an object moving in a space whose proper-time is given by (21.2) is almost identical to the path taken by an object moving in the same potential (e.g. an apple falling from a tree), calculated using Newton's law of gravity. But for a larger gravitational potential, which is no longer negligible compared with c^2, there are corrections to Newton's law of gravity due to General Relativity.

21.2 Gravitational Time Dilation

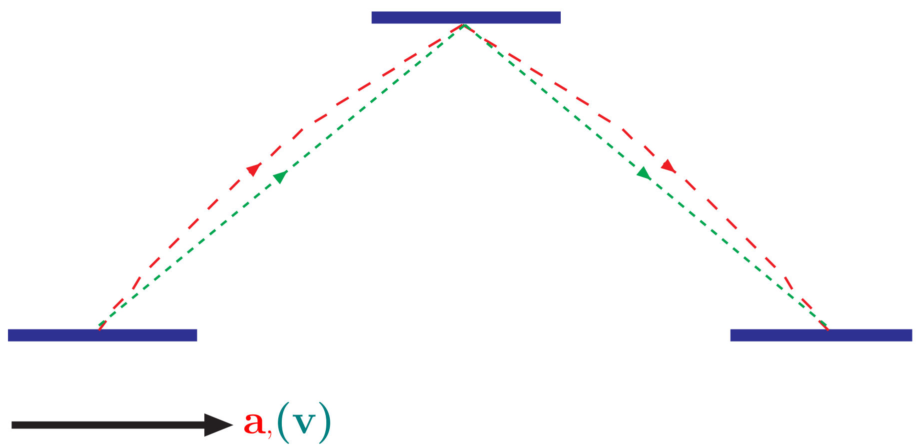


Figure 20: Path of light between two mirrors of a light clock in a frame which is accelerating relative to the observer (red) compared with moving with constant velocity relative to the observer (green).

The curvature of space-time means that there is a further contribution to time dilation arising from the presence of a gravitational field. If Brenda is accelerating towards the Earth relative to Albert (she is closer to the Earth and has a smaller gravitational potential) and is holding a light-clock then Albert's observation of the light path between the mirrors will be a curve (even though it is seen by Brenda to be a straight line - see Figure 20). Since light travels with the same speed in all frames it will take longer, according to Albert, to travel the curved distance between the mirrors of Brenda's light clock. This is in addition to the time dilation owing to the motion of Brenda's light-clock in Albert's frame.

Although for the Earth's gravitational field this is a tiny effect, it has actually been observed in atomic clocks held at different altitudes and the effect has to be accounted for in modern GPS devices.

21.3 The Gravitational Red-shift

The wavelength of light moving away from a massive body is shifted towards the red end of the spectrum as a result of the interaction of light with a gravitational field. This gravitational red-shift can be understood qualitatively by the realisation that gravity actually couples to the total relativistic energy of a particle and not its mass. For slow-moving particles the relativistic energy is approximately the mass energy, mc^2, so the Newtonian theory of gravity in which the force of gravity on an object is proportional to its mass, is a very good approximation. But photons have energy even though they have no mass and therefore photons couple to gravity – meaning that light is bent by gravity – and the red-shift is a result of the loss of energy of a photon as it moves away from a massive body which generates a gravitational potential. This is the exact analogy to the loss of kinetic energy of a massive particle which is thrown up in the air and is moving away from Earth. There is also a gravitational blue-shift experienced by light which is falling towards a massive body.

The bending of light by gravity was demonstrated by Arthur Eddington in 1919 in Brazil, where there was a total eclipse of the sun. During this eclipse Eddington was able to observe a group of stars known as Hyades, which was directly behind the sun, but could nevertheless be observed owing to the bending of the light from Hyades caused by the gravitational field of the sun. There is some controversy today over whether Eddington's quantitative measurements were consistent with the predictions of General Relativity.

The gravitational red-shift was first observed in 1925 by Walter Adams from observations of the white dwarf star Sirius B. He measured one of the spectral lines of hydrogen (the H_α line). It was found to have a wavelength which is larger than that observed from a terrestrial measurement of the same spectral line by 0.03%. From this red-shift and the radius of Sirius B we can deduce its mass (which is approximately the same as the mass of the sun.)

21.4 The Rebka-Pound experiment

In 1959, Robert Pound and Glen Rebka detected this gravitational shift using a technique developed by Mössbauer. A radioactive isotope which emits γ-rays can also absorb the γ-rays with exactly the same frequency. This is called "resonance absorption". But it only works if the absorber is stationary relative to the source. If source and absorber move relative to each other then the frequency of the emitted γ-rays, as measured by the absorber is shifted due to the Doppler shift and the absorption does not take place. This technique can detect shifts in γ-ray frequencies of one part in a hundred million million. On the other hand, if the γ-ray frequency is shifted by a tiny amount due to the gravitational red-shift, then by adjusting the velocity of the absorber until the red-shift is cancelled by the Doppler effect due to the moving absorber, resonance absorbtion is recovered, The velocity needed to recover this resonance absorption can then be used to determine the gravitational shift in the frequency of the γ-rays.

They placed a source sample of the iron isotope ^{56}Fe on the top of the tower of the Jefferson Laboratory at Harvard University, 22.6 metres above ground-level, and an absorber sample at the bottom of the tower, with a scintillation counter below it. The velocity at which the source must move relative to the absorber to recover resonance absorption, was 0.00074 mm per second. This very small velocity, at which absorption was observed, was measured by placing the sample at the top of the tower in the cone of a loudspeaker to which they applied a pure signal with frequency that ranged between 10 and 50 Hertz. Absorption occurs once every cycle when the velocity of the loudspeaker membrane is exactly equal to the resonant velocity. The determination of the precise phase of the oscillation at which this absorption is observed allowed them to calculate the velocity of the membrane of the loudspeaker at which absorption occurs. In order to improve the accuracy, the experiment was conducted both with the upper sample as the source and the lower sample as the absorber and the other way around.

They obtained a result which was within 10% of the theoretical result. This accuracy was later improved to 1%.

21.5 The Orbit of Mercury

Kepler's laws of planetary motion were originally derived from astronomical observations but were later derived by Newton using Newtonian mechanics and Newton's law of gravitation.

One of these laws states that planets revolve around the sun in fixed elliptic orbits. The orbit of the Earth around the sun is almost circular – the Earth is 3% close to the sun in January than in July. But the orbit of Mercury is a much more eccentric ellipse. At its point of closest approach (called its "perihelion") it is two-thirds of the distance when it is furthest away (called the "aphelion').

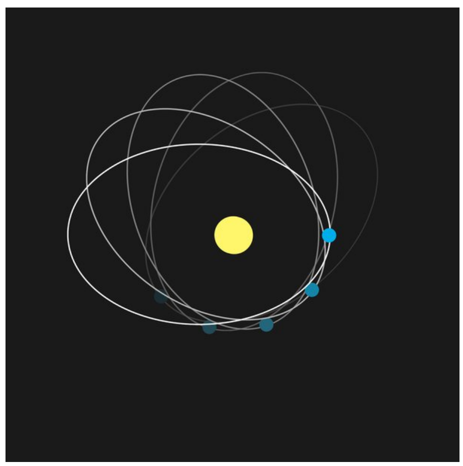


Figure 21: Precession of the orbit of Mercury

Although Kepler's laws predict a fixed elliptic orbit, the orbit of Mercury has been observed to precess very slowly at a rate of 575″ of arc every century. This was known in the nineteenth century and it was assumed that this was caused by the gravitational pull of other nearby planets in the solar system (such perturbations are not accounted for in the derivation of Kepler's laws). Much work was done in calculating the effect of such gravitational interaction with neighbouring planets, but the result yielded a precession rate of only 532″ of arc every century.

If Einstein's equations are applied to determine the path of the planet Mercury in a space, which is curved due to the gravitational potential of the sun, we almost find the ellipse expected from Kepler's laws, but with some very small correction. Einstein calculated that the effect of this General Relativity correction contributes 43″ per century of the precession of Mercury's orbit. This accounts for the shortfall of the calculation of the precession from perturbations caused by other planets and the observed rate of precession.

21.6 Black Holes

The gravitational potential at a distance x away from the centre of a sphere of mass M is

$$\Phi(x) = -\frac{GM}{x}, \tag{21.3}$$

where G is the gravitational constant (6.67×10^{-11} Joules metres per kilogram). This expression holds for values of x larger than the radius of the sphere, i.e. it only applies outside the surface of the sphere.

This is one of the assumptions of Newtonian gravity, but it can be derived from Einstein's equation of General Relativity, which relates the curvature properties of a curved space to the energy and momentum distribution of matter.

This means that the expression for the proper-time difference, $\Delta\tau$, between two events at a distance x from the centre of the massive object is (see (21.2))

$$\Delta\tau^2 = \left(1 - \frac{2GM}{x\,c^2}\right)\Delta t^2 - \frac{1}{c^2}\left(1 - \frac{2GM}{x\,c^2}\right)^{-1}\Delta x^2 - \frac{1}{c^2}\left(\Delta y^2 + \Delta z^2\right) \tag{21.4}$$

This is only valid outside the radius of the gravitating sphere. In most cases the radius of the sphere is sufficiently large that the gravitational potential only has a small effect on the proper-time. For example at the surface of the sun, mass $M_\odot$ and radius $R_\odot$ we have

$$\frac{G\,M_\odot}{R_\odot c^2} \approx 0.000004.$$

However, there are objects which are so massive that all the matter in the object is confined to within a radial distance, R_S, from the centre, for which the coefficients of Δt^2 and Δx^2, in the expression for the proper-time, change sign. Such objects are called "black holes". The radial distance from the centre, R_S, which is *outside* the black hole, is called the "horizon" given by

$$R_S = \frac{2GM}{c^2}, \tag{21.5}$$

(the subscript 'S' stands for Schwarzschild after Karl Schwarzschild who first derived this solution to Einstein's equations). Any matter approaching the black hole from outside, that crosses the horizon falls into the black hole and can never get out again.

A way to understand this is to recall that a rocket which can escape the Earth's gravitational field has to exceed a certain velocity known as the "escape velocity". For the Earth, this escape velocity is about 11 km per second. The escape velocity increases with increasing ratio of the mass to the radius of the body. Since no rocket can travel faster than the spped of light, c, any body which is sufficiently dense that the escape velocity exceeds c is a black hole - the rocket (or any other object) simply can't get out.

Clearly, nobody has been inside a black hole and then come back to tell us about it, but it seems that the world inside a black hole is very strange. From the expression for the proper-time, corrected by General Relativity, we see that for values of distance x from the centre which are less than the horizon radius, R_S, the coefficient of Δt^2 in the expression for proper-time is negative, whereas the coefficient of Δx^2 is positive - contrary to the situation outside the black hole. It appears then that distance outside the horizon becomes time inside the horizon and time outside the horizon becomes distance inside the horizon. Viewed from outside the horizon, a black hole appears to occupy a finite volume in space, but is time independent – it lasts for ever – whereas inside the horizon the black hole appears to last for a finite time interval, but exists everywhere. *Don't even try to picture this!*

There are now many examples of black holes which have been detected. They cannot be seen directly as light cannot get out of them, but they are usually accompanied by a visible star which rotates rapidly around it, due to the large gravitational field surrounding the black hole. From the period of the rotation and the distance of the visible star from the centre of rotation one can deduce the mass black hole and hence calculate the radius of its horizon (using (21.5)).

Actually, black holes do not last for ever. Stephen Hawking showed that they eventually evaporate through "Hawking radiation". The mechanism for this makes use of the fact that close to the horizon of a black hole, the gravitational field is so intense that the gravitational energy can exceed $2mc^2$. This is sufficient energy so that a particle of mass m can be created out of the empty space surrounding the black hole together with its anti-particle. The particle falls into the black hole but the anti-particle is radiated outwards (or vice versa). The net effect of this is that the radiated particle carries off some of the energy of the black hole – so the mass of the black hole decreases, until it becomes below the critical mass and the black hole disappears.

21.7 Gravitational Waves

Einstein's equation of General Relativity generalises Newton's law of gravity in the same way that Maxwell's equations of electromagnetism generalise Coulomb's inverse square law for the electric field at a distance x from a charged particle.

It is an equation for the curvature of space-time, and hence for the gravitational field, in the presence of moving massive objects, in terms of their energy and momentum. This is the gravitational analogue of Maxwell equations for electric and magnetic fields in terms of moving electric charges.

Maxwell's equations predict the the existence of electromagnetic waves – such as light – in which electric and magnetic fields are oscillating and advancing in the direction of the wave.

Likewise the equations of General Relativity predict the existence of waves of a gravitational field, which can be thought of as oscillations in the curvature of space advancing in the direction of the wave. Because gravity is such a weak force, these waves have very low intensity and are impossible to detect unless the source of the gravitational wave is a very violent event in space.

In 2015 the LIGO collaboration detected gravitational waves from the collision of two black holes 1.3 billion light-years away. They used an interferometer which was similar in design the Michelson interferometer used in the Michelson-Morley experiment, but in this case the arms were 4 km long. The gravitational wave can be thought of as a wave disturbance in the curvature of space. This resulted in a change in the length of one of the arms of the interferometer by one million-million-millionth (10^{-18}) of a metre. Nevertheless this remarkable experiment was able to detect the minute shift in the interference pattern of the laser beams in the two arms, confirming the passage of the gravitational wave caused by the collision of the two black holes

22 Appendix A: A Pythagoras Refresher

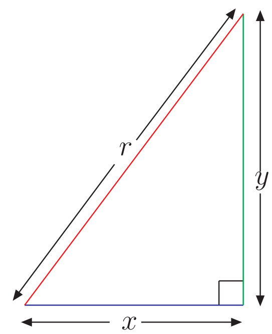


Figure 22: A right-angle triangle with hypotenuse (side opposite the right-angle) length r and sides adjacent to the right-angle with sides of length x and y.

Figure 22 shows a right-angle triangle. The longest side (called the "hypotenuse") is the one opposite the right-angle and has length r. The other two sides have lengths x and y. Pythagoras' theorem states simply that

$$r^2 \;=\; x^2 + y^2. \tag{22.1}$$

So for example of x is 3 metres and y is 4 metres, then the length, r, of the hypotenuse is given by

$$r \;=\; \sqrt{3^2 + 4^2} \;=\; \sqrt{9 + 16} \;=\sqrt{25} \;= 5 \text{ metres}.$$

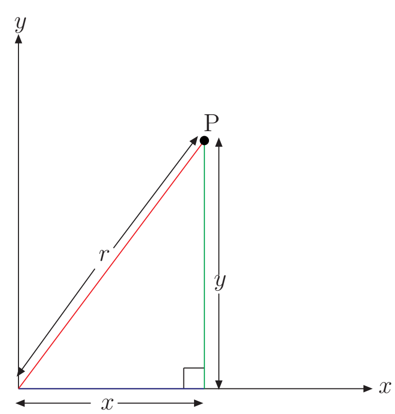


Figure 23: The point P in the $x - y$ plane (i.e. $z = 0$) with coordinates (x, y)

If a point P has coordinates (x, y), as shown in Figure 23, then the distance, r, of P from the origin can be obtained by Pythagoras' theorem.

$$r \;=\; \sqrt{x^2 + y^2} \tag{22.2}$$

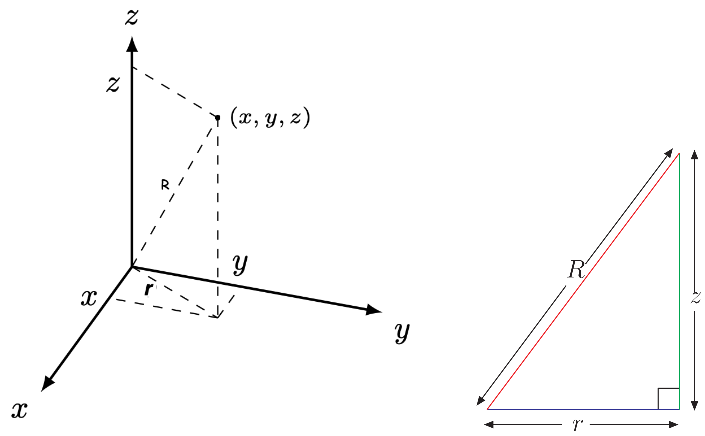


Figure 24: A point in three dimensions with coordinates x, y, z.

We can extend this to calculate the distance from the origin of a point in three dimensions with coordinates (x, y, z). The projection of the point in the $x - y$ plane has coordinates $(x, y, 0)$. Its distance from the origin is r, where

$$r^2 \ = x^2 + y^2.$$

But the hypotenuse in the $x-y$ plane is at right-angles to the z-axis, and so we have a further right-angle triangle shown on the right of Figure 24. We can therefore use Pythagoras again to show that the total distance R of the point (x, y, z) from the origin is given by

$$R^2 \ = \ r^2 + z^2 \ = \ x^2 + y^2 + z^2 \tag{22.3}$$

Pythagoras' theorem is also used in the determination of the magnitude of a vector in terms of its components. If $\mathbf{p}$ is a vector (a quantity which has a direction as well as a magnitude, such a velocity or momentum), which has component p_x in the x-direction, p_y in the y-direction and p_z in the z-direction. The magnitude $|\mathbf{p}|$ of the vector is given by

$$|\mathbf{p}| \ = \ \sqrt{p_x^2 + p_y^2 + p_z^2}.$$

We often need to use Pythagoras's theorem to calculate the length, y, of one of the adjacent sides in terms of the lengths of the hypotenuse, r, and the other adjacent side x:

$$y^2 \ = \ r^2 - x^2 \ = \ (r + x)\,(r - x)\,,$$

where is the last step we have used the expression for the difference of two squares to show that the square of the length of one of the adjacent sides is the sum of the lengths of the hypotenuse and the remaining adjacent side, times by their difference.

Now let's do two problems which use Pythagoras' theorem for their solution.

Problem 1: Adam and Eve go down to the river to bathe. The river has a width w_A and is flowing with speed v. Both swimmers can swim at a speed u relative to the flow of the river. They start their swim at the same time and Adam swims across the river and back, whereas Eve swims downstream a distance w_E and then turns back and swims upstream back to the point of departure. They find that they both arrive at the starting point at the same time (their elation at this is responsible for the birth of Cain !). Show that the distance downstream, w_E, that Eve swam is given by

$$w_E \;=\; w_A\sqrt{1 - v^2/u^2}$$

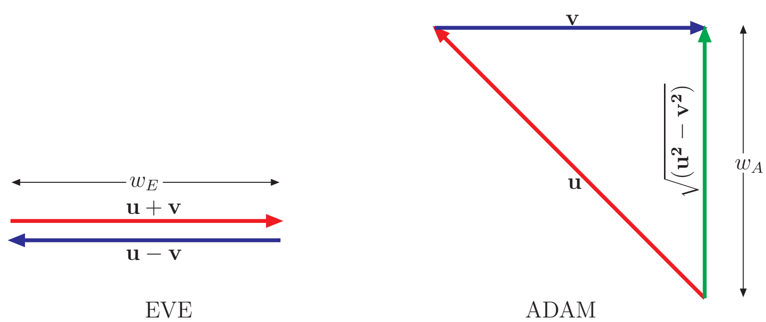


Figure 25: Eve swims downstream a distance w_E and then upstream again, whereas Adam points himself in such a direction that his resultant velocity (the velocity relative to the river plus the velocity of the flow of the river) is exactly across the river.

Solution:
Clearly we can neglect any relativistic effects.

Eve swims downstream a distance w_E with a velocity relative to the river bank $(v+u)$ (see left-hand part of Figure 25) so that the time taken for her to reach her destination is $w_E/(u+v)$. When she swims back upstream her velocity relative to the river bank is $(u-v)$. Therefore time taken for her to swim back top her starting point is $w_E/(u-v)$. The total time, t_E, taken for Eve's swim is

$$t_E \;=\; \frac{w_E}{(u+v)} + \frac{w_E}{(u-v)} \;=\; \frac{2u\,w_E}{(u+v)(u-v)} \;=\; \frac{2u\,w_E}{(u^2-v^2)}$$

Adam swims with velocity $\mathbf{u}$ relative to the steam in such a direction that his resultant velocity is exactly across the river, perpendicular to the stream. As can be seen from the right-hand part of Figure 25, his velocity is the hypotenuse of a right-handed triangle of length u and the length of one of the adjacent sides is v. By Pythagoras' theorem, Adam's velocity across the river is therefore $\sqrt{u^2-v^2})$. The time taken for Adam to swim across the river and back is therefore

$$t_A \;=\; \frac{2w_A}{\sqrt{u^2-v^2}}$$

If the two swims take the same total time we can equate t_A and t_B to obtain

$$\frac{2w_A}{\sqrt{u^2 - v^2}} = \frac{2u\,w_E}{(u^2 - v^2)}$$

A little rearranging yields the desired result

$$w_E = w_A\sqrt{1 - v^2/u^2} \tag{22.4}$$

Problem 2:

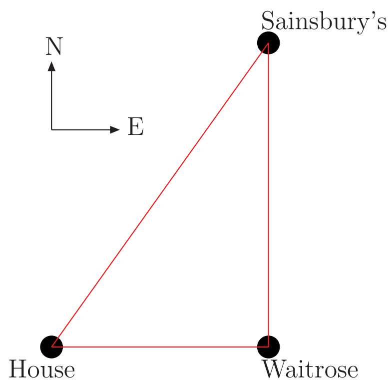


Figure 26: Waitrose is due East of the house where Albert and Brenda live. Sainsbury's is due North of Waitrose.

Albert and his wife Brenda go shopping. They leave the house together, but whereas Brenda goes to Waitrose which is due East, Albert goes to Sainsburys which is due North of Waitrose, but there is a path leading directly there. Because Albert walks at a speed v_A which is faster than Brenda's speed v_B, they both arrive at their destination shops at the same time.

One day Albert tells Brenda that after he has finished his shopping at Sainsburys's, he wants to come to Waitrose to buy something else (not available at Sainsbury's) and asks Brenda to wait for him after she has finished her shopping at Waitrose (which takes exactly the same amount of time as Albert's shopping at Sainsbury's). Brenda agrees and finds she has to wait a time T for Albert to join her.

Show that the time taken for either of them to reach their usual shop is

$$\frac{T}{\sqrt{1 - v_B^2/v_A^2}}. \tag{22.5}$$

Solution:
There are three possible approaches for the reader:

1. This is far too complicated for me. It makes my head spin. I will simply accept that this is the correct answer.

2. I am not going to derive the result but I am prepared to use Pythagoras' theorem to check the answer:

   The house, Waitrose, and Sainsbury's form a right-angle triangle with the right-angle at Waitrose. The path from the house to Sainsbury's is the hypotenuse If the time taken for Albert to walk to Sainsbury's at speed v_A is given by (22.5) then the distance to Sainsbury's along the hypotenuse has length

   $$\text{Distance to Sainsburys'} \ = \ \frac{v_A T}{\sqrt{1 - v_B^2/v_A^2}}.$$

   Likewise, the distance that Brenda walks to Waitrose, at speed v_B, is

   $$\text{Distance to Waitrose} \ = \ \frac{v_B T}{\sqrt{1 - v_B^2/v_A^2}},$$

   The distance from Sainsburys to Waitrose is $v_A T$, since it takes time T for Albert to walk it, at speed v_A.

   So by Pythagoras, theorem, the length of the hypotenuse (the distance from the house to Sainsbury's is

   $$v_A^2 T^2 + \frac{v_B^2 T^2}{(1 - v_B^2/v_A^2)} \ = \ \frac{T^2}{(1 - v_B^2/v_A^2)} \left(v_A^2 \left(1 - v_B^2/v_A\right)^2 + v_B^2 \right) \ = \ \frac{v_A^2 T^2}{(1 - v_B^2/v_A^2)},$$

   which is indeed the square of the distance to Sainsburys, so the answer is correct..

3. Derive the result from first principles:

   Let the time taken to reach their shops be X, so that the distance between the house and Sainsbury's is $v_A X$, the distance to Waitrose is $v_B X$, and the distance from Sainsbury's to Waitrose is $v_A T$ (it takes Albert time T to walk it as speed v_A). Now by Pythagoras' theorem

   $$v_A^2 T^2 + v_B^2 X^2 \ = \ v_A^2 X^2$$

   Rearranging

   $$v_A^2 T^2 \ = \ (v_A^2 - v_B^2) X^2$$

   Dividing both sides by $(v_A^2 - v_B^2)$

   $$X^2 \ = \ \frac{v_A^2 T^2}{(v_A^2 - v_B^2)}$$

   Divide numerator and denominator by v_A^2 and take the square root to get

   $$X \ = \ \frac{T}{\sqrt{1 - v_B^2/v_A^2}}$$

23 Appendix B: Some Results from Classical Mechanics

In this Appendix, we always assume that particles are moving with velocities that are much smaller tna the speed of light, so that all relativistic effects are neglected.

- **Momentum**
  A particle with mass m moving with velocity v has momentum $p = m \times v$
- **Vectors**
  Velocity $\mathbf{v}$ and momentum, $\mathbf{p}$ are examples of vectors which have direction as well as magnitude. They have components in each of the x-, y-, and z-directions. These components transform into each other if the axes are rotated, but the magnitude of the vector remains invariant.
- **Newton's second law of motion**
  The momentum of a particle which is subjected to as force, F, increases at a rate equal to the force. For a body whose mass, m, is constant, this means that the force is equal to the mass times the acceleration of the particle.
- **Conservation of Momentum** In a collision between several particles in which there is no external force, the total momentum of the particles (in each direction) is the same before and after the collision.
- **Kinetic Energy** A particle of mass m with velocty v has kinetic energy $T = \frac{1}{2}mv^2$.
- **Conservation of Energy** In an elastic collision between several particles the total kinetic energy is the same before and after the collision.
- **Centre-of-Mass** A solid body has a centre of mass. For an ensemble of particles of mass m_i at a distances x_i in the x-direction from a given origin, the distance from the origin in the x_i-direction of the cetnre of mass is the sume of $m_i \times x_i$ divided by the total mass.

  For exmaple if we have a dumbbell with a weight of mass 1 kg at one end (the origin) and a weight of mass 2 kg at a distance of 025 m., then the centre of mass or the dumbbell is at a distance, x from the smaller weight, where

$$x = \frac{0 \times 1 + 2 \times 0.25}{1+2} = 0.167\,\mathrm{m}.$$

24 Glossary

- **Acceleration:** The rate at which a body gains speed.
- **Aether:** A postulated medium on which light travels. This has been shown not to exist.
- **Anti-particle:** of a given particle, is a particle with precisely the same mass but opposite electric charge (and other properties). The positron is a positively charged particle with the same mass as the electron – it is the anti-particle of the electron.
- **Aphelion** The point of furthest distance of an ellipse from a focus.
- **Blackbody Radiation:** Radiation from an object which absorbs all of the radiation incident upon it.
- **Black hole:** An object which is so dense that the associated gravitational field is so strong that any object which enters cannot escape.
- **Component:** of a vector. The projection of a vector in a given direction.
- **Coordinate:** The distance of a point along a given axis.
- **Electromagnetic Wave:** A wave of an oscillating electric field and a perpendicular osicllating magnetic field, both perpendicular to the direction of motion of the wave. Light is an electromagnetic wave with a wavelength in the visible range.
- **Escape velocity:** Velocity required by a body in a gravitational field to escape that gravitational field. In other words the velocity which is such that the kinetic energy is equal to the gravitational potential energy.
- **Euclidean Geometry:** A set of theorems about properties of shapes in flat space.
- **Event:** Something that takes place at a point in space and a given instant in time. It is specified by the three coordinates, indicating its position, and a time. These values are different in different inertial frames.
- **Frame of Reference:** A set of measuring rods, which constitute a coordinate system with a given origin and given orientation, together with a clock which has been set to zero at some given time.
- **Frequency:** Number of oscillations performed in unit time (per second).
- **Galilean Transformations:** Non-relativistic transformation of the location of an event between two different frames of reference, in which the time of the event is the same in both frames.
- **General Relativity:** The mechanics of bodies moving in the presence of a gravitational field and are accelerating relative to each other.

- **Geodesic:** The generalisation of a straight line in curved space, defined as the shortest distance between two points in the curved space.
- **Gravitational Potential:** The potential energy of a body at a given location in a gravitational field, divided by the mass of the body.
- **Gravitational Constant:** Gravitational potential due to a particle of mass 1 kg at a distance of 1 metre ftom the particle.
- **Gravitational Red-shift:** The increase in wavelength of light (or other electromagnetic radiation) radiated from a body, due to the interaction of the radiation with the gravitational field produced by the body.
- **Greenwhich Meridian:** The meridian with longitude 0°. It passes through Greenwich.
- **Hawking Radiation:** The mechanism by which black holes evaporate. Particles and their anti-particles are produced out of the vacuum surrounding the black hole as a result of the intense gravitational field.
- **Hertz:** Unit of frequency equal to one oscillation per second.
- **Inertial Frame:** A frame of reference which is in free fall in a gravitational field. In the absence of a gravitational field an inertial frame moves with constant velocity relative to any other inertial frame. Inertial frames at different locations in a gravitational field are accelerating relative to each other.
- **In Phase:** Two monochromatic waves of the same frequency are said to be in phase if their maxima and minima coincide.
- **Interference pattern:** A configuration of bright and dark rings or fringes at positions where the motion of two light-waves are in phase so that the waves add leading to a maximum, and positions where the two waves are out of phase so that they cancel.
- **Interferometer:** A device for detecting interference pattern. These are often used to measure the wavelength of a light wave by measuring the separation of the bright and dark fringes or rings.
- **Kepler's Laws:** The description of the paths of planets orbiting the sun.
- **Kinetic Energy:** The energy that a moving body possesses as a result of its motion.
- **LIGO:** Laser Interferometer Gravitational-Wave Observatory. An experiment designed to detect the existence of gravitational waves.
- **Lorentz Invariant:** Takes the same value in all inertial frames.
- **Lorentz Transformations:** Relativistic transformations of the position and time of an event in different inertial frames.

- **Lorentz-FitzGerald Contraction:** The contraction of the length of an object due to its motion. Originally intended to refer to motion relative to the aether in an attempt to explain the null result of the Michelson-Morley experiment, but now used to describe the contraction of any object in the direction of its motion relative to the observer.

- **Mass Energy:** The energy of a body at rest, due to its mass.

- **Maxwell's Equations:** A set of equations relating electric and magnetic fields to each other and to distributions of electric charge and electric current.

- **Meridian:** A line from the North pole to the South pole of fixed longitude.

- **Minkowski Diagram:** A graph of one space dimension and time. The axes can be perpendicular in one inertial frame, but will not be perpendicular in another inertial frame.

- **Minkowski Space:** a four-dimensional space, in which the fourth dimension is time. An event at a given time and location is a point in this space. The invariant quantity under rotations of both the space and time axes is the proper-time difference of two neighbouring events.

- **Momentum:** A property of a moving body. It is a vector in the direction of the velocity of the body. The momentum of a body remains unchanged if there is no external force acting on it. Total momentum is conserved in elastic collisions.

- **Monochromatic Light:** Light of a single wavelength.

- **Neutrino:** A particle with mass much smaller than that of an electron, which is emitted together with an electron in radioactive β-decay.

- **Out of Phase:** Two monochromatic waves of the same frequency are said to be out of phase if the maxima of one wave coincides with the minima of the other.

- **Perihelion:** The closest distance of a point on an ellipse from a focus.

- **Period of Oscillation:** Time interval between two subsequent maxima, (or two subsequent minima) of an oscillating system, which could be a wave at a fixed point in space.

- **Photoelectric Effect:** The emission of electrons from certain metals when illuminated with visible or ultraviolet light.

- **Photo-electron:** An electron emitted from an irradiated target, due to the photoelectric effect.

- **Photon:** A particle of light (or other electromagnetic radiation).

- **Planck's constant:** The ratio between the energy of a photon and its frequency.

- **Proper-time:** A property of the difference between two neighbouring events. In the absence of gravity the square of the proper-time difference is the square of the time difference between the two events minus the square of the spatial distance between the two events, divided by the speed of light. In the presence of a gravitational field there are corrections to account for the curvature of space-time in the presence of a gravitational field. For a given pair of events, it takes the same value on all inertial frames.
- **Pythagoras' Theorem:** The relation between the length of the hypotenuse of a right-angled triangle and the lengths of the sides adjacent to the right-angle.
- **Refraction:** The change in the direction of light when it passes from one medium to another.
- **Relativistic Energy:** The kinetic energy of a body plus its mass energy
- **Resonance:** Enhanced response of a system at a particular frequency.
- **Rest Energy:** Alternative term for mass energy.
- **Resultant:** The sum of two vectors, usually pointing in different directions.
- **Riemannian Geometry:** The study of the properties of curved spaces.
- **Special Relativity:** Mechanics of objects moving with constant velocity relative to each other, in the absence of a gravitational field.
- **Spectral Lines** of an atom or molecule are the discrete frequencies of the radiation absorbed or emiterd by the atom or molecule when it makes a transition from one energy level to another.
- **Stellar Aberration;** The shift in the observed positions of stars due to the refraction of the light from the star by the Earth's atmosphere.
- **Vector:** A quantity which has direction as well as magnitude.
- **Velocity:** A vector property of a body, whose magnitude is the speed of the body and whose direction is the direction of motion of the body.
- **Wavelength:** The distance between adjacent corresponding points on a wave, at a given instant in time.
- **Work Function:** The energy required to free an electron from the surface of a metal.