Introduction to Relativity

MP352

Charles Nash,
Department of Mathematical Physics,
National University of Ireland,
Maynooth.

Preface

Le savant doit ordonner; on fait la science avec des faits come une maison avec des pierres; mais une accumulation de faits n'est pas plus une science qu'un tas de pierres n'est une maison.

Henri Poincaré [1]

This is a set of notes which supplement my lectures on relativity. They are reasonably self contained but should be read *as well as* other library material. There are so many books on relativity in the library that I shall recommend just three all of which are in the library. These are:

1. Rindler W., *Introduction to Special Relativity*, Oxford University Press, (1991).

2. Rindler W., *Relativity, Special, General and Cosmological*, Oxford University Press, (2001).

3. Einstein A., *The meaning of relativity*, Methuen, (1956).

The first of these three would be an adequate book for this course but do read as widely as you can.

Relativity is a beautiful subject which, for mainly pedagogical purposes, is divided into the following two parts:

(i) Special relativity or *the special theory*

(ii) General relativity or *the general theory*

We shall deal almost exclusively with the special theory but I have included some material on the general theory in chapter 7 where the experimental tests of relativity are discussed.

The special theory completely revolutionises the physics of mass, space, time, energy and momentum; it is also the origin of the celebrated equation

$$E = mc^2$$

relating the energy E of a body to its mass m and c, where c is the velocity of light in a vacuum. It is this equation which predicts the gigantic release of energy in an atomic fission or fusion reaction.

The general theory is a very far reaching new theory of gravity which replaces Newton and Hooke's inverse square law theory of gravity. As well as being a new theory of gravity—or gravitation if you prefer a longer word—it includes the special theory as a special case and points the way forward to a new approach in studying physical problems: this is to emphasise the coordinate independence of the mathematics used to treat physics problems and to formulate theoretical physics in a way that makes this manifest.

Charles Nash

[1] The scientist must organise; one constructs a science with facts like a house with stones; but an accumulation of facts is no more a science than a pile of stones is a house.

Henri Poincaré, *La Science et l'Hypothèse*, p. 168, Flammarion, (1908).

CHAPTER I

How it all began

§ 1. Two special problems of late nineteenth century physics

The world of physics faced two very special problems in the last quarter of the nineteenth century: one was theoretical and concerned the theory of a radiating black body, and the other was experimental and concerned certain velocity of light measurements made by Michelson and Morley in 1887. [1]

The subsequent solving of these two problems changed the face of physics for the indefinite future.

Planck, his constant h, and quantum theory

The first problem—*the black body problem*—was solved in 1900 by Max Planck in the famous paper [2] which introduces the constant h which ever afterwards has been known as *Planck's constant.* This then was the birth of *Quantum Theory* which did not take its final form until the 1920's.

The second problem—*the Michelson–Morley velocity of light problem*— was solved by Albert Einstein in 1905 in his famous paper [3] entitled

On the electrodynamics of moving bodies

This is the paper which founded [4] the theory of *Special Relativity.*

Since quantum theory does not concern us in this set of lectures we now turn our attention to special relativity.

Einstein's 1905 papers

Chapter 8 consists of *English translations of Einstein's two papers of 1905.* The reader should peruse these from time to time: for example after completing each chapter. [5] Note

[1] Michelson A. A. and Morley E. W., *On the Relative Motion of the Earth and the Luminiferous Ether*, Am. J. Sci., **134**, 333–345 (1887)

[2] *Uber des Gesetz der Energieverteilung im Normalspectrum (On the Law of Energy Distribution in Normal Spectra)*, Annalen der Physik, **4**, 553, 1901. Note that the publication date of this paper is 1901 but Planck completed the work in 1900.

[3] *Zur Elektrodynamik bewegter Körper (On the electrodynamics of moving bodies)*, Annalen der Physik. **322**, 891–921, (1905).

[4] Actually Einstein published two papers on relativity in 1905: one long one in June founding what we now call *Special Relativity*, and a short three page paper in September pointing out that his first paper implies the famous relationship written nowadays as $E = mc^2$, although Einstein did not write the equation in quite that way.

[5] Be reassured, though, that the reading of Einstein's papers is *not obligatory*, the lecture material is self contained and does not require one to read Einstein's papers.

that this incredibly influential paper does not use any sophisticated mathematics to obtain its revolutionary results; indeed a modest appearance made by calculus using the interpretation of a velocity as a derivative is as sophisticated as it gets. [6]

These lectures can be regarded as simply an explanation of this celebrated 1905 paper. The difficulties that a newcomer faces are *not* really mathematical but *physical* since the results overturn all previous intuition about length and time, and physical quantities constructed from them such as velocity, momentum and energy.

§ 2. The ether

The reader may have noticed the phrase *luminiferous ether* in the title of Michelson and Morley's paper when we quoted the reference, cf. the footnote on p. 1. We now explain what this phrase means.

Up to the advent of relativity in 1905 most physicists believed that there was a substance called *the ether, aether or luminiferous ether* [7] which filled up all space and was the medium that "vibrated" as light waves passed through it. The speed of travel of these vibrations is then the speed of light c.

Other waves, such as sound waves and water waves, need a medium for them to be transmitted—they cannot travel through a vacuum—it was believed that this was also necessary for the transmission of light. In particular, scientists were reluctant to believe that light could travel all the way from the Sun to the Earth through a complete vacuum.

To accomplish all that was required of it the ether needed to be a medium of great elasticity and was supposed to pervade all space and the interiors of solid and liquid bodies.

Let us emphasise then that the ether was meant to carry all light signals whether they travel in empty space, or in some medium, and the ether was believed to be at absolute rest.

Accepting the existence of the ether, then one could conceive of measuring the orbital velocity v of the Earth round the Sun by using light rays entirely within a laboratory. The main idea is as follows.

Simply apply the usual laws of relative velocities to light rays: If one moves with velocity v *towards* a light ray whose velocity is

$$c \tag{1.1}$$

then the light ray should appear to be approaching one with velocity

$$c+v \tag{1.2}$$

Similarly if one moved *away* from the light ray with velocity v, the ray should appear to approach with velocity

$$c-v \tag{1.3}$$

[6] Calculus is used a little more in the paper's discussion of Maxwell's equations mainly because these latter are differential equations; however by then the remarkable results about mass, length and time have all been derived.

[7] The word *luminiferous* is used because it means *light carrying* from the Latin *lumen*: light or lamp, and ferre: to carry; the point being that the ether was supposed to be a medium for light transmission.

Hence appropriate measurements of two such rays, as we explain in the next section, should allow one to deduce v the velocity of the Earth. This velocity v of the Earth was also referred to as the *velocity of the Earth through the ether*.

A famous experiment—the one we mentioned at the beginning of the section—to make precisely such a measurement was done in 1887 by Michelson and Morley. This experiment, and its *failure to measure* v, led to the founding of relativity.

The failure of this experiment, and particularly his disquiet about asymmetries of Maxwell's equations found when comparing electric and magnetic phenomena, led Einstein to realise that the ether could be dispensed with altogether; though not all the scientific community accepted this when his 1905 paper was published.

For example a common way to try and avoid discarding the ether was to consider that the Earth *dragged* the ether along with it as it travelled through space. Though this would have explained Michelson and Morley's result, there were other difficulties with the notion of ether drag and so the ether was eventually abandoned after 1905. Nowadays the ether is not believed in and has become a thing of the past.

The demise of the ether

It is now time to describe the work of Michelson and Morley.

§ 3. The problem unearthed by Michelson and Morley

With the hindsight offered by history one can pinpoint a single experimental result as being crucial in drawing physics along the path to relativity. This is the famous light interference experiment of Michelson and Morley carried out in 1887 which we now describe.

The experiment of Michelson and Morley

Michelson and Morley, in their experiment, used an experimental apparatus located on the Earth's surface and so it was the velocity of the Earth round the Sun that became the velocity v referred to in the previous paragraph.

The experiment carried out by Michelson and Morley in 1887 is illustrated below in figure 1. Here is a brief description of the experimental equipment. What we see is a light ray from a source falling on a half silvered mirror set at 45^0; this ray is thereby split into two and falls on the two mirrors labelled I and II, these mirrors in their turn reflect their rays back to the half silvered mirror allowing them to enter the measuring apparatus which is known as a Michelson interferometer.

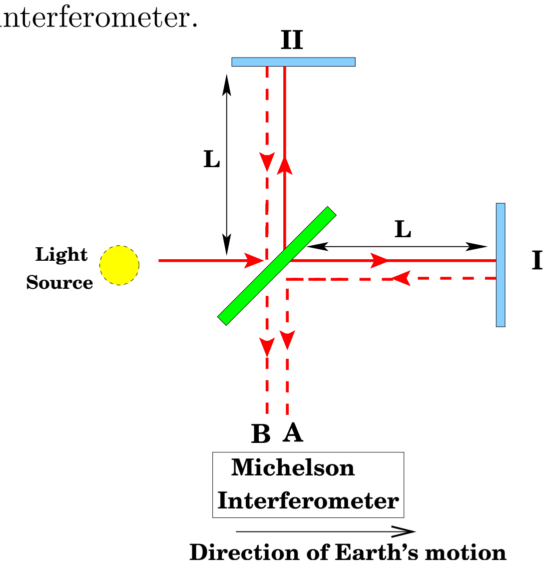

Fig. 1: The Michelson–Morley apparatus

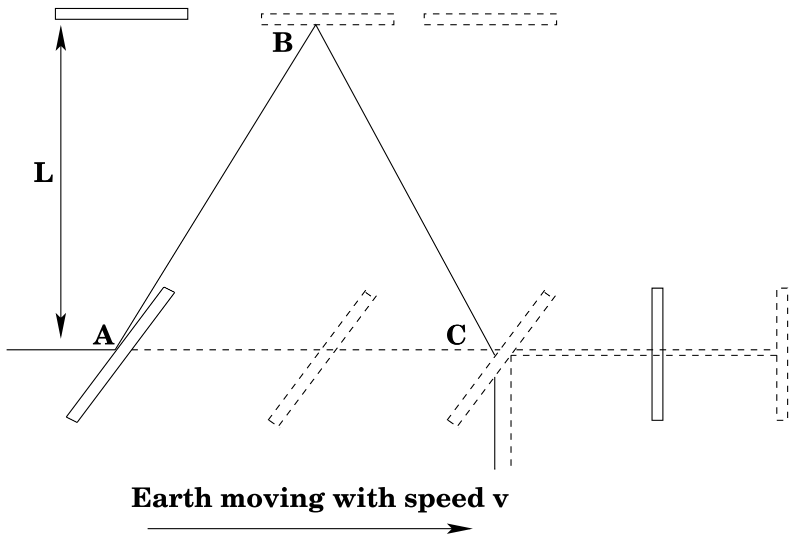

Fig. 2: The Michelson–Morley apparatus in motion

The next figure (figure 2) shows how the mirrors move as the Earth orbits the Sun; the path ABC of one of the light rays is also shown.

Now we can analyse the experiment as follows: Figure 2 shows the half silvered mirror moving from A to C; let this take a time

$$t \tag{1.4}$$

Since its velocity is v then this mirror travels a distance

$$vt \tag{1.5}$$

so the distance AC is precisely vt as is shown in figure 3. It is now clear that triangle ABC has the lengths as shown in figure 3.

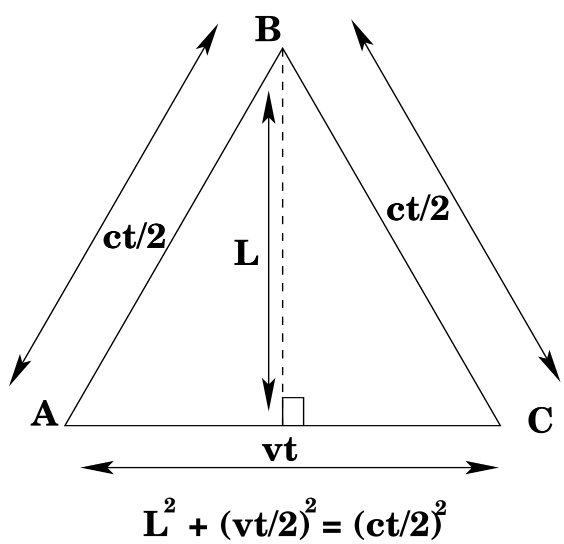

Fig. 3: The light path ABC

Let us now *compare* the times taken by the two light rays to reach the measuring apparatus i.e. the Michelson interferometer. One ray follows the path ABC, and takes a time

$$t \tag{1.6}$$

However, from figure 3, we see that

$$\begin{aligned} L^2 + \left(\frac{vt}{2}\right)^2 &= \left(\frac{ct}{2}\right)^2 \\ \Rightarrow \frac{1}{4}\left(c^2t^2 - v^2t^2\right) &= L^2 \\ \Rightarrow t^2 &= \frac{4L^2}{c^2 - v^2} \end{aligned} \tag{1.7}$$

So that

$$t = \frac{2L}{\sqrt{c^2 - v^2}} \tag{1.8}$$

The second light ray is the one that travels horizontally a distance L with velocity relative to its target $c - v$ and then a distance L back again with velocity relative to its target $c + v$ so it should take a total time t' where

$$\begin{aligned} t' &= \frac{L}{c - v} + \frac{L}{c + v} \\ &= \frac{2Lc}{(c - v)(c + v)} \end{aligned} \tag{1.9}$$

So that

$$t' = \frac{2Lc}{c^2 - v^2} \tag{1.10}$$

Summarising we find that the times t and t' for the two light rays to get back to the Michelson interferometer are

$$\begin{aligned} t &= \frac{2L}{\sqrt{c^2 - v^2}} \\ t' &= \frac{2Lc}{c^2 - v^2} \end{aligned} \tag{1.11}$$

and we observe that these are unequal i.e.

$$t \neq t' \tag{1.12}$$

The *problem* is that the experiment showed that the *times were exactly the same*: had there been any time difference the two rays would have been out of phase when they reached the Michelson interferometer. This phase difference was to be detected by rotating the apparatus through 90^0 and repeating the experiment; such a rotation interchanges the two arms of the interferometer and should cause a shift in the interference pattern; however *no such shift in the interference pattern was detected.* [8]

§ 4. The significance of the negative result of Michelson and Morley

Michelson and Morley's findings were inexplicable by physics as understood at that time—i.e. 1887. However some partial explanations of the result were offered: the most important one being due to Fitzgerald, in 1889; the same explanation was offered by Lorentz independently, in 1892. [9]

The explanation of Fitzgerald and Lorentz was that somehow or other the interferometer *contracted* along the direction of its motion—the horizontal direction—and that this contraction changed its length, in the horizontal direction, by a factor

$$\sqrt{1-\frac{v^2}{c^2}} \tag{1.13}$$

This meant that the length of the horizontal arm changed from

$$L \tag{1.14}$$

to

The Fitzgerald-Lorentz contraction

$$L\sqrt{1-\frac{v^2}{c^2}} \tag{1.15}$$

This contraction is often referred to as the *Fitzgerald-Lorentz contraction*—it is a *contraction* because $v < c$ and so $\sqrt{1 - v^2/c^2} < 1$. We shall see, in chapter 4, that length contraction

[8] One might reasonably wonder if the experimental apparatus was sensitive enough to detect the expected result; it certainly was and here are the details: The velocity v in 1.11 is the orbital velocity of the Earth round the Sun; this orbit, though elliptical, can be approximated by a circle of radius 93 million miles which is traversed in 365.25 days. This gives

$$v = \frac{2\pi \cdot 93,000,000}{365.25 \cdot 24 \cdot 60 \cdot 60} = 18.51, \quad \text{miles per second} \equiv 29.78\, km/sec$$

This should have given a shift in the interference pattern of 0.4 fringes which would have been easily measurable. In fact the 1887 experiment (cf. Michelson A. A. and Morley E. W., *On the Relative Motion of the Earth and the Luminiferous Ether*, Am. J. Sci., **134**, 333–345 (1887)) found the shift to be *at least* 80 *times smaller*—i.e. less than 0.005 fringes—consistent with zero shift. Subsequent experiments have refined this result enormously and always found the shift to be consistent with zero. Michelson received the 1907 Nobel prize for physics for his work.

[9] Fitzgerald was Irish and Lorentz was Dutch. Fitzgerald limited himself to the observation that had the *horizontal* arm of the interferometer *somehow contracted* by the amount $\sqrt{1-v^2/c^2}$, then Michelson and Morley's result would have found no fringe shift, as was the case. Lorentz, as well as independently making the same point, went on, in later years, to pursue some of the mathematical consequences of such a contraction, particularly those for electromagnetism. In doing this he derived for the first time, in 1905, the transformations now known as the Lorentz transformations; these latter emerge naturally in Einstein's theory.

is predicted by special relativity, but, by itself, this is not quite the full explanation of the Michelson-Morley result; one also needs what is called *time dilation*, cf. chapter 4.

To see that this contraction would explain Michelson and Morley's result is easy: this is simply because the time t' has now *changed* whereas t has *not*; t and t' now have the values

$$\begin{aligned} t &= \frac{2L}{\sqrt{c^2 - v^2}} \\ t' &= \frac{2\sqrt{1 - \frac{v^2}{c^2}}Lc}{c^2 - v^2} \end{aligned} \tag{1.16}$$

But we now easily calculate that

$$\begin{aligned} \frac{2\sqrt{1 - \frac{v^2}{c^2}}Lc}{c^2 - v^2} &= \frac{2\sqrt{1 - \frac{v^2}{c^2}}Lc}{c^2(1 - \frac{v^2}{c^2})} \\ &= \frac{2L}{c\sqrt{1 - \frac{v^2}{c^2}}} \\ &= \frac{2L}{\sqrt{c^2 - v^2}} \end{aligned} \tag{1.17}$$

Hence we do have

$$t = t' \tag{1.18}$$

as claimed.

The contraction explanation is ad hoc and left the laws of physics in an awkward state: there had to be a deeper reason for the behaviour of the light rays in the Michelson Morley experiment. This was supplied by Einstein's work to which we turn in the next section.

§ 5. Einstein enters: his Relativity Principle

Einstein simply decided to interpret the negative result of the Michelson Morley experiment as meaning that if one travels *towards* or *away from* a light ray that it always has the *same velocity c*.

This bold hypothesis is counterintuitive but experiment shows that—counterintuitive or not—it is indeed correct.

Einstein combined this hypothesis with what he called *The Relativity Principle* and thereby derived many striking consequences[10] which revolutionised[11] physics for ever afterwards.

[10] We shall see in chapter 4 that these consequences include the contracting of lengths and the slowing down of times as an object moves, as well as the emergence of the famous equation $E = mc^2$ relating mass and energy. This equation of course providing the prediction that enormous amounts of energy can be released in nuclear fission and fusion.

[11] Einstein did not receive a Nobel prize for his work on relativity due, perhaps, to caution on the part of the Nobel committee. However he did receive the 1921 Nobel prize for physics for his use of quantum theory to explain the photoelectric effect.

Before giving Einstein's *relativity principle* we first introduce and explain the term *frame of reference* which is used frequently in relativity. A frame of reference (not a precise term) can be thought of as three rigid rods representing the x, y and z axes, together with a clock which provides a time t. Then an *inertial frame* is a frame of reference that is moving at constant velocity.

An inertial frame

More precisely what Einstein did what was to introduce two postulates; and these are:

Einstein's two postulates (preliminary form)

(i) The Relativity Principle *The laws of electrodynamics and optics are the same in all frames of reference for which the laws of mechanics hold good*

(ii) Constancy of the velocity of light *The velocity c of light is a constant independent of the state of motion of the emitting body.*

Now the *Relativity Principle* requires some elucidation: we need to define what is meant by the phrase *"frames of reference for which the laws of mechanics hold good"*.

What Einstein means by this is that if one takes two laboratories and allows one to move at a constant velocity relative to the other, then the laws of mechanics, established by doing whatever experiments one wishes, will be found to be the same in both laboratories. One need not restrict oneself to the *laws of electrodynamics and optics* either. In other words *the laws of physics are the same in all inertial frames.*

It is very convenient to use what we have just said to shorten Einstein's *relativity principle*. So here are the two postulates again where the *relativity principle* has been shortened but the second postulate is unchanged.

Einstein's two postulates (final form)

(i) The Relativity Principle *The laws of physics are the same in all inertial frames.*

Einstein's two postulates

(ii) Constancy of the velocity of light *The velocity c of light is a constant independent of the state of motion of the emitting body.*

The second of Einstein's postulates was very hard to accept at first but it did explain the result of the Michelson Morley experiment, though not, one should note, in precisely the same way as envisaged by Fitzgerald and Lorentz. We shall see in chapter 3 that the constancy of the velocity of light is absolutely vital to make special relativity work.

Let us turn next to the matter of the coordinates that we use to discuss physics.

§ 6. Physics and coordinates

Einstein's relativity principle should be seen as an important first step in trying to separate physical phenomena from the coordinates which are used to discuss them.

This separation is taken a step further in *General Relativity* which is the theory Einstein formulated in 1915 to replace Newtonian gravity.

Geometry is also placed in the foreground in relativity, both special and general, and this is something that fits in well with the desire to separate coordinates from the phenomena that they are describing: one knows that geometrical objects exist and are definable quite separately from any coordinates used to analyse them.

This point about the separation of coordinates from the phenomena that they describe may seem simple and unlikely to lead to much extra knowledge; however this is quite wrong: Physics since 1905, and mathematics since that time, both emphasise the importance of formulations which are as coordinate free as possible.

The geometry that we need for special relativity is the four dimensional geometry of Minkowski which we shall study in chapter 3; however before that we turn to some results in two dimensional geometry; they form the material of the next chapter.

CHAPTER II

Some two dimensional geometry

§ 1. Rotations: some two dimensional geometry revisited

We shall need some simple results from two dimensional geometry later so we go through them now; the first part will be just revision of familiar material. Consider two sets of orthogonal axes inclined at an angle θ to each other as shown in 4 (a). Each set of axes possesses a pair of orthonormal vectors which we have denoted by

$$\{\mathbf{i}, \mathbf{j}\} \qquad \text{and} \qquad \{\mathbf{e_1}, \mathbf{e_2}\} \tag{2.1}$$

Fig 4 (b) shows a vector $\mathbf{v}$ whose components we display relative to each set of axes in Figs. 5 (a) and (b).

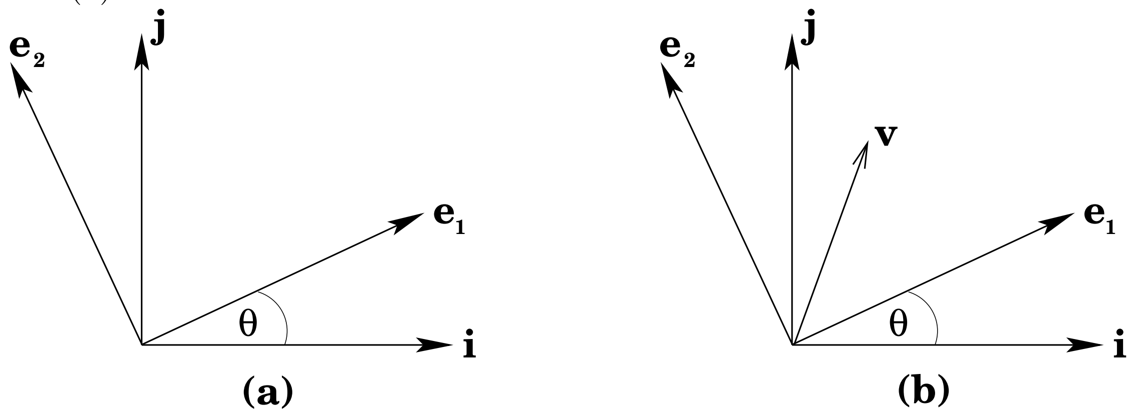

Fig. 4: (a) Two sets of axes **(b) The axes with a vector v added**

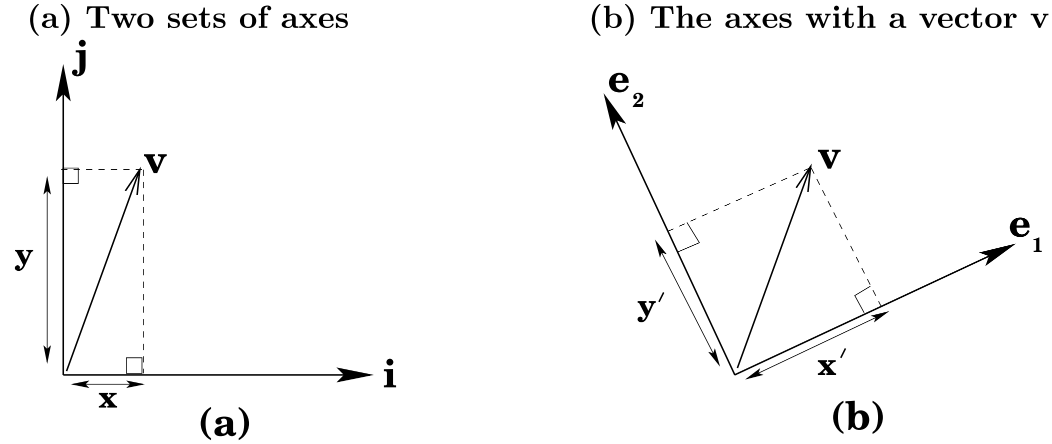

Fig. 5: The same vector v and its components relative to two sets of axes

Consulting figure 5 yields two expressions for $\mathbf{v}$ namely

$$\begin{aligned} \mathbf{v} &= x\,\mathbf{i} + y\,\mathbf{j} \\ \mathbf{v} &= x'\mathbf{e_1} + y'\mathbf{e_2} \end{aligned} \tag{2.2}$$

We want to express x' and y' in terms of x and y: taking the dot product with $\mathbf{e_1}$ gives

$$\begin{aligned} x'\mathbf{e_1} + y'\mathbf{e_2} &= x\,\mathbf{i} + y\,\mathbf{j} \\ \Rightarrow (x'\mathbf{e_1} + y'\mathbf{e_2}) \cdot \mathbf{e_1} &= (x\,\mathbf{i} + y\,\mathbf{j}) \cdot \mathbf{e_1} \\ \Rightarrow x' &= x\,\mathbf{i} \cdot \mathbf{e_1} + y\,\mathbf{j} \cdot \mathbf{e_1} \\ \Rightarrow x' &= \cos(\theta)x + \sin(\theta)y \end{aligned} \tag{2.3}$$

Similarly, taking the dot product with $\mathbf{e_2}$ gives

$$\begin{aligned} (x'\mathbf{e_1} + y'\mathbf{e_2}) \cdot \mathbf{e_2} &= (x\,\mathbf{i} + y\,\mathbf{j}) \cdot \mathbf{e_2} \\ \Rightarrow y' &= x\,\mathbf{i} \cdot \mathbf{e_2} + y\,\mathbf{j} \cdot \mathbf{e_2} \\ \Rightarrow y' &= \cos(\theta + \pi/2)x + \cos(\theta)y \\ &= -\sin(\theta)x + \cos(\theta)y \end{aligned} \tag{2.4}$$

Summarising, we have found that (x, y) and (x', y') are related by the equations

$$\begin{aligned} x' &= \cos(\theta)x + \sin(\theta)y \\ y' &= -\sin(\theta)x + \cos(\theta)y \end{aligned} \tag{2.5}$$

which we often write in matrix form as

$$\begin{bmatrix} x' \\ y' \end{bmatrix} = \begin{bmatrix} \cos(\theta) & \sin(\theta) \\ -\sin(\theta) & \cos(\theta) \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} \tag{2.6}$$

Lengths are also preserved by rotations; we just remind the reader that the length of the vector $\mathbf{v}$ which we denote by

$$|\mathbf{v}| \tag{2.7}$$

can be computed in terms of (x, y) or (x', y'). So we have two expressions for $|\mathbf{v}|$, and these are

$$\begin{aligned} |\mathbf{v}| &= \sqrt{x^2 + y^2} \\ |\mathbf{v}| &= \sqrt{(x')^2 + (y')^2} \end{aligned} \tag{2.8}$$

This is trivial to check as we might at well do it: Using 2.5 we can calculate that

$$\begin{aligned} (x')^2 + (y')^2 &= \{\cos(\theta)x + \sin(\theta)y\}^2 + \{-\sin(\theta)x + \cos(\theta)y\}^2 \\ &= \left\{\cos^2(\theta) + \sin^2(\theta)\right\} x^2 + \left\{\cos^2(\theta) + \sin^2(\theta)\right\} y^2 \\ &= x^2 + y^2 \end{aligned} \tag{2.9}$$

So we do indeed have

$$(x')^2 + (y')^2 = x^2 + y^2 \tag{2.10}$$

Rotations preserve the length $\sqrt{x^2+y^2}$

§ 2. Trigonometric and hyperbolic functions

We shall also need to make use of some of the standard properties of trigonometric and hyperbolic functions and so we shall go through these now.

Recall that sin and cos can be expressed in terms of exponentials as

$$\sin(\theta) = \frac{e^{i\theta} - e^{-i\theta}}{2i}, \qquad \cos(\theta) = \frac{e^{i\theta} + e^{-i\theta}}{2} \tag{2.11}$$

In contrast, the definitions of the *hyperbolic functions* cosh and sinh contain *no factors of* i, and are given by the equations

$$\sinh(\theta) = \frac{e^{\theta} - e^{-\theta}}{2}, \qquad \cosh(\theta) = \frac{e^{\theta} + e^{-\theta}}{2} \tag{2.12}$$

One also has the usual definition

$$\tanh(\theta) = \frac{\sinh(\theta)}{\cosh(\theta)} \tag{2.13}$$

From these facts follow some useful relations between the trigonometric and hyperbolic functions which the reader can easily verify, namely

$$\cos(i\theta) = \cosh(\theta), \qquad \sin(i\theta) = i\sinh(\theta), \qquad \tan(i\theta) = i\tanh(\theta) \tag{2.14}$$

Also certain well known identities for trigonometric functions have similar, *but not identical*, counterparts expressed in terms of hyperbolic functions. We list what we have in mind below: For trigonometric functions we have

$$\cos^2(\theta) + \sin^2(\theta) = 1, \qquad \tan(\theta + \phi) = \frac{\tan(\theta) + \tan(\phi)}{1 - \tan(\theta)\tan(\phi)} \tag{2.15}$$

While the counterparts for hyperbolic functions are

$$\cosh^2(\theta) - \sinh^2(\theta) = 1, \qquad \tanh(\theta + \phi) = \frac{\tanh(\theta) + \tanh(\phi)}{1 + \tanh(\theta)\tanh(\phi)} \tag{2.16}$$

where one should note the vital changes of sign.

§ 3. Transformations preserving the quantity $x^2 - y^2$

Of great importance for special relativity is the quantity

$$x^2 - y^2 \tag{2.17}$$

rather than

$$x^2 + y^2 \tag{2.18}$$

Also any transformations which preserve $x^2 - y^2$ play a central rôle in special relativity. It turns out that if we rotate through an imaginary angle, that is through the angle

$$i\theta, \quad (i^2 = -1 \text{ as usual}) \tag{2.19}$$

then we get such a transformation. Let us spell out the details since the result is so important:

First note that if we substitute $i\theta$ for θ in 2.6 we get

$$\begin{bmatrix} x' \\ y' \end{bmatrix} = \begin{bmatrix} \cos(i\theta) & \sin(i\theta) \\ -\sin(i\theta) & \cos(i\theta) \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} \tag{2.20}$$

But if we use the relations between $\sin, \cos, \sinh, \cosh$ of 2.14 this becomes

$$\begin{bmatrix} x' \\ y' \end{bmatrix} = \begin{bmatrix} \cosh(\theta) & i\sinh(\theta) \\ -i\sinh(\theta) & \cosh(\theta) \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} \tag{2.21}$$

which is the pair of equations

$$\begin{aligned} x' &= \cosh(\theta)x + i\sinh(\theta)y \\ y' &= -i\sinh(\theta)x + \cosh(\theta)y \end{aligned} \tag{2.22}$$

Now if we adjust the notation slightly by defining (u, v) and (u', v') according to the equations

$$\begin{aligned} u = x, &\quad v = iy \\ u' = x', &\quad v' = iy' \end{aligned}$$

then we obtain

$$\begin{aligned} u' &= \cosh(\theta)u + \sinh(\theta)v \\ v' &= \sinh(\theta)u + \cosh(\theta)v \end{aligned} \tag{2.23}$$

which we can also write in matrix form as

$$\begin{bmatrix} u' \\ v' \end{bmatrix} = \begin{bmatrix} \cosh(\theta) & \sinh(\theta) \\ \sinh(\theta) & \cosh(\theta) \end{bmatrix} \begin{bmatrix} u \\ v \end{bmatrix} \tag{2.24}$$

Finally, as we now show by a direct check, the quadratic form $u^2 - v^2$ is preserved by such a change of variable; i.e.

$$(u')^2 - (v')^2 = u^2 - v^2 \tag{2.25}$$

Carrying out the check (remember from 2.16 that $\cosh^2(\theta) - \sinh^2(\theta) = 1$) we compute that

$$\begin{aligned} (u')^2 - (v')^2 &= \{\cosh(\theta)u + \sinh(\theta)v\}^2 - \{\sinh(\theta)u + \cosh(\theta)v\}^2 \\ &= \left\{\cosh^2(\theta) - \sinh^2(\theta)\right\} u^2 - \left\{\cosh^2(\theta) - \sinh^2(\theta)\right\} v^2 \\ &= u^2 - v^2 \end{aligned} \tag{2.26}$$

Hence we have indeed successfully checked that

$$(u')^2 - (v')^2 = u^2 - v^2 \tag{2.27}$$

$u^2 - v^2$ is preserved by "hyperbolic rotations".

CHAPTER III

Minkowski space and Lorentz transformations

Die Anschauungen über Raum und Zeit, die ich Ihnen entwicklen möchte, sind auf experimentell-physikalischem Boden erwachsen. Darin liegt ihre Stärke. Ihre Tendenz ist eine radikale. Von Stund an sollen Raum für sich und Zeit für sich völlig zu Schatten herabsinken, und nur noch eine Art Union der beiden soll Selbständigkeit bewahren.

Hermann Minkowski [1]

§ 1. Minkowski space

Minkowski was responsible, in 1908, for a fundamental step in physics: he introduced the spacetime approach to relativity: he suggested using time

$$t \tag{3.1}$$

as a fourth dimension together with the three spatial variables

$$x, y, z \tag{3.2}$$

The point being that relativity became much simpler to understand if one used this four dimensional setting. So one no longer considers time and space separately but combines them into a single four dimensional geometry.

This four dimensional space is then called *spacetime* or *Minkowski space* and we shall denote it by

Minkowski space

$$\mathbf{Mk} \tag{3.3}$$

Any point in Minkowski space **Mk** has four coordinates, namely

$$x, y, z, t \tag{3.4}$$

[1] The views of space and time which I wish to develop have sprung from the soil of experimental physics. Therein lies their strength. They have a radical tendency. Henceforth space by itself, and time by itself, will fade away into mere shadows, and only a kind of union of the two will preserve an independent existence.

Hermann Minkowski, Conference talk in Cologne, 21 September 1908, published in *Raum und Zeit*, *Physikalische Zeitschrift*, 104, (1909).

and a point in Minkowski space represents an *event*—an alternative description of Minkowski space is that it is the set of all possible events.

Minkowski space as the set of all events

Special relativity changes radically our understanding of space and time, and the justification of this four dimensional approach is that it makes this understanding natural, elegant and much easier to acquire.

We now turn from mathematics to physics and consider Einstein's two postulates applied to light rays.

§ **2. Light rays and frames of reference**

Let us use coordinates

$$(x, y, z, t) \tag{3.5}$$

which belong to a frame

$$F \tag{3.6}$$

with origin O.

Consider a light ray moving in a straight line along, or parallel to, the x-axis; the ray obeys one of the two equations

$$\begin{array}{ll} x = ct & \text{(the ray is moving from left to right)} \\ x = -ct & \text{(the ray is moving from right to left)} \end{array} \tag{3.7}$$

Let us now consider that both the left-moving and the right-moving rays are simultaneously present then we have

$$\begin{aligned} x &= \mp ct \\ \Rightarrow x^2 &= c^2t^2 \end{aligned} \tag{3.8}$$

which we find convenient to write as

$$x^2 - c^2t^2 = 0 \tag{3.9}$$

Having understood this now we consider a spherical light ray starting out from the origin

$$(x, y, z) = (0, 0, 0) \tag{3.10}$$

and spreading out in a spherical manner as t increases. Since light travels a distance ct, in time t, then the radius of this sphere, at time t, will be

$$ct \tag{3.11}$$

Hence the spherical wave front has the equation

$$x^2 + y^2 + z^2 = c^2t^2 \tag{3.12}$$

that is

$$x^2 + y^2 + z^2 - c^2t^2 = 0 \tag{3.13}$$

and we recognise 3.13 as a three dimensional version of 3.9.

A pair of inertial frames

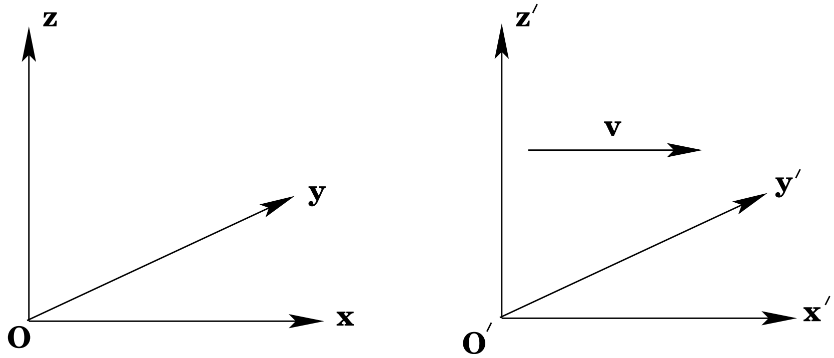

Fig. 6: The two frames F and F'

Now consider what this equation

$$x^2 + y^2 + z^2 - c^2t^2 = 0 \tag{3.14}$$

looks like in another frame F' which moves with constant velocity v relative to the frame F. The two frames are depicted in figure 6 with coordinates in F' being denoted by

$$(x', y', z', t') \tag{3.15}$$

For later convenience we also arrange that the two spatial origins O and O' coincide when $t = t' = 0$.

Now Einstein's second postulate asserts that the velocity of light, seen in F', will be *identical* to that of F—so the velocity of light in F' will still be

$$c \tag{3.16}$$

But this means that the equation of the spherical wave in the frame F' must *precisely replicate* that of the frame F. Therefore, in F', the light wave simply has the equation

Einstein's second postulate applied to F'

$$(x')^2 + (y')^2 + (z')^2 - c^2(t')^2 = 0 \tag{3.17}$$

This being so, we can immediately observe that

$$x^2 + y^2 + z^2 - c^2t^2 = (x')^2 + (y')^2 + (z')^2 - c^2(t')^2 \tag{3.18}$$

This tells us that the mathematical task we have to accomplish is to find all transformations Λ of Minkowski space

$$\begin{aligned} \Lambda : \mathbf{Mk} &\longrightarrow \mathbf{Mk} \\ (x, y, z, t) &\longmapsto (x', y', z', t') \end{aligned} \tag{3.19}$$

which have the property that [2]

$$x^2 + y^2 + z^2 - c^2t^2 = (x')^2 + (y')^2 + (z')^2 - c^2(t')^2 \tag{3.21}$$

[2] Actually the careful reader may have noticed that we could replace the property 3.21 of Λ by the more general condition

$$x^2 + y^2 + z^2 - c^2t^2 = \lambda\left[(x')^2 + (y')^2 + (z')^2 - c^2(t')^2\right] \tag{3.20}$$

where λ is some constant. But, if we did so, we would find that it would emerge in the end that $\lambda = 1$. The argument to establish this first uses the fact that 3.20 also has to hold for the *inverse* Lorentz transformation Λ^{-1} so that $\lambda = \lambda^{-1}$, thus $\lambda = \pm 1$; finally the the choice $\lambda = -1$ is impossible because it would force x', y', z', t' to become imaginary. Hence we suffer no loss of generality by not including λ.

These transformations Λ are called *Lorentz transformations* and we derive their form in the next section.

§ 3. The Lorentz transformation Λ derived

We shall now *assume* that Lorentz transformations are linear. This is only done to make the argument less cluttered with detail—it is not difficult to prove—the reader who wishes to see a proof will find one in § 7.

Now let us temporarily pretend that spacetime is two dimensional with coordinates (x, t) instead of (x, y, z, t) so that Λ sends

Temporarily work with a two dimensional spacetime

$$(x, t) \mapsto (x', t') \tag{3.22}$$

and just has to satisfy

$$x^2 - c^2t^2 = (x')^2 - c^2(t')^2 \tag{3.23}$$

This is familiar territory and was discussed in section 1 of this chapter. More precisely such transformations Λ are discussed in 2.23–2.25. Hence this means that x' and t' are given by the equations

$$\begin{aligned} x' &= \cosh(\theta)x + \sinh(\theta)ct \\ ct' &= \sinh(\theta)x + \cosh(\theta)ct \end{aligned} \tag{3.24}$$

or, in matrix form, we can write

$$\begin{bmatrix} x' \\ ct' \end{bmatrix} = \begin{bmatrix} \cosh(\theta) & \sinh(\theta) \\ \sinh(\theta) & \cosh(\theta) \end{bmatrix} \begin{bmatrix} x \\ ct \end{bmatrix} \tag{3.25}$$

All we have left to do is find out what the angle θ can be.

To do this we just look at figure 6 and consider the point from the point of view of both frames F and F': Viewed from F' the point O' is just the origin and is specified by the equation

$$x' = 0 \tag{3.26}$$

But viewed from F the point O' is moving with constant speed v and so is specified by the equation

$$x = vt \tag{3.27}$$

Now if we substitute 3.26 and 3.27 into the first equation of 3.24 we obtain

$$\begin{aligned} 0 &= \cosh(\theta)vt + \sinh(\theta)ct \\ \Rightarrow \tanh(\theta) &= -\frac{v}{c} \end{aligned} \tag{3.28}$$

so θ is now known.[3] We proceed at once to fill in the remaining details of the Lorentz

[3] Some authors write $\tanh(\beta) = +v/c$ (so that $\beta = -\theta$); the variable β is then called the *rapidity*; we shall return to this matter in Chapter 5, cf. p. 42

transformation Λ. Since

$$\begin{aligned}
\tanh(\theta) &= -\frac{v}{c} \\
\Rightarrow \frac{\sinh(\theta)}{\cosh(\theta)} &= -\frac{v}{c} \\
\Rightarrow \frac{\sqrt{\cosh^2(\theta)-1}}{\cosh(\theta)} &= -\frac{v}{c}, \text{ (remember } \cosh^2(\theta)-\sinh^2(\theta)=1) \\
\Rightarrow \frac{\cosh^2(\theta)-1}{\cosh^2(\theta)} &= \frac{v^2}{c^2}
\end{aligned} \tag{3.29}$$

So we immediately deduce that [4]

$$\cosh(\theta) = \frac{1}{\sqrt{1-\frac{v^2}{c^2}}}, \quad \sinh(\theta) = -\frac{v}{c}\frac{1}{\sqrt{1-\frac{v^2}{c^2}}}, \tag{3.30}$$

Using these values in 3.24 above we find that

$$\begin{aligned}
x' &= \frac{1}{\sqrt{1-\frac{v^2}{c^2}}}\,(x-vt) \\
t' &= \frac{1}{\sqrt{1-\frac{v^2}{c^2}}}\left(-\frac{v}{c^2}x+t\right)
\end{aligned} \tag{3.31}$$

Note that it is standard notational practice to abbreviate things by defining the quantity $\gamma(v)$ where

$$\gamma(v) = \frac{1}{\sqrt{1-\frac{v^2}{c^2}}} \tag{3.32}$$

and [5], when the context allows, $\gamma(v)$ is often abbreviated to just γ.

In any case, with this notation, our Lorentz [6] transformation Λ is given by the more compact pair of equations [7]

[4] Notice that, in 3.30, the positive sign on the RHS of the equation for $\cosh(\theta)$ is needed because the function $\cosh(\theta)$ is always positive; on the other hand a minus sign is needed on the RHS of the equation for $\sinh(\theta)$ because of the minus sign in the equation 3.29 for $\tanh(\theta)$.

[5] It is amusing to note that what we call γ here Einstein called β; cf. p. 96 of these notes where you will find the relevant equation of Einstein's first 1905 paper.

[6] Lorentz and Zeeman received the 1902 Nobel prize for physics for their work on the effect of magnetic fields on electromagnetic radiation. Zeeman was Lorentz's student, obtaining his doctorate in 1893, and and doing his fundamental experiment on the magnetic splitting of spectral lines in 1897. Zeeman was the experimentalist who complemented the theorist Lorentz; somewhat unfairly, perhaps, this phenomenon is now usually called the *Zeeman effect*.

[7] The Lorentz transformation Λ of 3.33 is sometimes called a *boost* because its application to coordinates (x, y, z, t) produces coordinates (x', y', z', t') which move with velocity v relative to (x, y, z, t).

The Lorentz transformation Λ

$$\begin{aligned} x' &= \gamma(v)\,(x - vt) \\ t' &= \gamma(v)\left(-\frac{v}{c^2}x + t\right) \end{aligned} \tag{3.33}$$

Observe that

$$\frac{1}{\sqrt{1-\frac{v^2}{c^2}}} = \infty \tag{3.34}$$

when $v = c$; this means that something goes wrong if something, other than light itself, travels at the speed of light. We shall see in chapter 5 that no particle of positive mass m can attain the speed of light, it would require infinite energy to do so. Thus the speed of light is an upper limit on the velocity for ordinary objects of mass m.

There is also an *inverse Lorentz transformation* Λ^{-1} of the form

$$\begin{aligned} \Lambda^{-1} &: \mathbf{Mk} \longrightarrow \mathbf{Mk} \\ (x', y', z', t') &\longmapsto (x, y, z, t) \end{aligned} \tag{3.35}$$

Specialising to 2 dimensions, so that y and z drop out, we can easily calculate Λ^{-1}: Since, from the point of view of F', F is moving with velocity

$$-v \tag{3.36}$$

along the x-axis; Λ^{-1} can be got by simply changing the sign of v in the formula 3.33 for Λ.

If we carry out this sign change we get

The inverse Lorentz transformation Λ^{-1}

$$\begin{aligned} x &= \gamma\,(x' + vt') \\ t &= \gamma\left(\frac{v}{c^2}x' + t'\right) \end{aligned} \tag{3.37}$$

§ 4. Minkowski space again

Back to four dimensional spacetime

Now let us drop our restriction to two dimensional spacetime and reinstate the missing variables y and z.

In the next section we shall introduce some important notation due to Einstein. This notation both shortens calculations and makes them clearer. In order to unify space and time into a four dimensional setting we shall denote

$$x, y, z, t \tag{3.38}$$

by

$$x^0, x^1, x^2, x^3 \tag{3.39}$$

where

$$x^0 = ct, x^1 = x, x^2 = y, x^3 = z \tag{3.40}$$

Note carefully that $x^0 \neq t$ but rather we have

$$x^0 = ct \tag{3.41}$$

this inclusion of the factor c is natural given the occurrence of ct in the equation for a light ray and the importance of light propagation in relativity.

When we want to refer to *all four* of these coordinates we shall always use a *Greek letter* and write

$$x^{\mu} \tag{3.42}$$

So μ runs through the *four* values $0, \ldots, 3$ so the possible values[8] of the Greek index μ are

$$\mu = 0, 1, 2, 3 \tag{3.43}$$

When we only want to refer to the *three spatial coordinates* x, y, z we shall use a Latin letter and write

$$x^{i} \tag{3.44}$$

So i runs through the *three* values $1, 2, 3$ and the possible values of the Latin index i are

$$i = 1, 2, 3 \tag{3.45}$$

The Lorentz transformation Λ should also be expressible as a 4×4 matrix when we use all 4 variables x^0, x^1, x^2, x^3. One should realise that we want to have

$$\begin{bmatrix} ct' \\ x' \\ y' \\ z' \end{bmatrix} = \Lambda \begin{bmatrix} ct \\ x \\ y \\ z \end{bmatrix}, \qquad \text{with } \Lambda = \left[\Lambda^{\alpha}_{\beta}\right]_{4\times 4} \text{ a } 4 \times 4 \text{ matrix} \tag{3.46}$$

All we have to do is to take the 2×2 matrix form of Λ given in 3.25, *taking note that we have now written ct as the first variable instead of x*, and add a 2×2 identity matrix at the bottom so as to leave y and z unchanged; then we fill in the remainder of the matrix with zeroes.

Hence we have (remembering that ct is now the first variable instead of x)

The Lorentz transformation as a 4×4 matrix

$$\Lambda = \begin{bmatrix} \sinh(\theta) & \cosh(\theta) & 0 & 0 \\ \cosh(\theta) & \sinh(\theta) & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \tag{3.47}$$

and, if we use the values of $\cosh(\theta)$ and $\sinh(\theta)$ provided in 3.30, we can further write

$$\Lambda = \begin{bmatrix} -\gamma\frac{v}{c} & \gamma & 0 & 0 \\ \gamma & -\gamma\frac{v}{c} & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}, \quad \gamma = \frac{1}{\sqrt{1 - \frac{v^2}{c^2}}} \tag{3.48}$$

[8] The reader will find that some textbooks use the slightly different notation $x^1 = x, x^2 = y, x^3 = z, x^4 = ct$ so that Greek indices still take four values but these values are $1, 2, 3, 4$ instead of $0, 1, 2, 3$. Yet another convention is to interchange the rôles of the Greek and Latin indices, that is to cause Greek indices to take three values and Latin indices to take four values. Clearly all notations will do the same job.

and if we want to read off the various values of the matrix entries Λ^α_β for the Lorentz transformation we can use the equation

$$\begin{bmatrix} \Lambda^0_0 & \Lambda^0_1 & \Lambda^0_2 & \Lambda^0_3 \\ \Lambda^1_0 & \Lambda^1_1 & \Lambda^1_2 & \Lambda^1_3 \\ \Lambda^2_0 & \Lambda^2_1 & \Lambda^2_2 & \Lambda^2_3 \\ \Lambda^3_0 & \Lambda^3_1 & \Lambda^3_2 & \Lambda^3_3 \end{bmatrix} = \begin{bmatrix} -\gamma\frac{v}{c} & \gamma & 0 & 0 \\ \gamma & -\gamma\frac{v}{c} & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \tag{3.49}$$

Finally if we want the corresponding formulae for the inverse Lorentz transformation Λ^{-1} we just have to change the sign of v in the above expressions; by the way the reader can check that this is the same as changing the sign of θ. In any case—for the record—the result of doing this is that

$$\Lambda^{-1} = \begin{bmatrix} -\sinh(\theta) & \cosh(\theta) & 0 & 0 \\ \cosh(\theta) & -\sinh(\theta) & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \tag{3.50}$$

and

Matrix form of the inverse Lorentz transformation

$$\Lambda^{-1} = \begin{bmatrix} \gamma\frac{v}{c} & \gamma & 0 & 0 \\ \gamma & \gamma\frac{v}{c} & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}, \quad \gamma = \frac{1}{\sqrt{1-\frac{v^2}{c^2}}} \tag{3.51}$$

Incidentally this matrix form makes it very quick to check that Λ^{-1} *is* an inverse, i.e. that

$$\Lambda\Lambda^{-1} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} = \Lambda^{-1}\Lambda \tag{3.52}$$

§ 5. Length in Minkowski space

The quantity $x^2 + y^2 + z^2 - c^2t^2$ which arises naturally when discussing light rays has a geometric interpretation in terms of a "length" in Minkowski space.

Recall first that in ordinary Euclidean 3 space a length L is given by the equation

$$L^2 = x^2 + y^2 + z^2 \tag{3.53}$$

or if were to have an n-dimensional Euclidean space, with coordinates $x^1, x^2, \ldots x^n$, then L would be determined by the similar equation

$$L^2 = (x^1)^2 + (x^2)^2 + \cdots + (x^n)^2 \tag{3.54}$$

Well, in Minkowski space **Mk**, we want to interpret the quantity $L_{\mathbf{Mk}}$ defined by

$$L_{\mathbf{Mk}} = \sqrt{c^2t^2 - x^2 - y^2 - z^2} \tag{3.55}$$

or more conveniently in squared form

$$L_{\mathbf{Mk}}^2 = c^2t^2 - x^2 - y^2 - z^2 \tag{3.56}$$

as a “length” [9].

We can write the formula for the Minkowski length $L_{\mathbf{Mk}}$ by using a matrix known as the *metric tensor* $g_{\mu\nu}$ or simply the metric. The metric tensor $g_{\mu\nu}$ is a rather simple 4×4 matrix whose definition is

$$g_{\mu\nu} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{bmatrix} \tag{3.57}$$

The metric $g_{\mu\nu}$ is introduced

Now $g_{\mu\nu}$ is an important, but simple, object in special relativity. The word tensor should not intimidate the reader, we shall not need to explain it here: the tensorial properties of $g_{\mu\nu}$ are only needed in general relativity where $g_{\mu\nu}$ is a more complicated object. Hence the reader need not worry that any mathematics, beyond that of matrices, is required to understand $g_{\mu\nu}$: in special relativity we can just regard it as a matrix.

The metric $g_{\mu\nu}$ has an inverse which is denoted by using *upper indices* and is written as

$$g^{\mu\nu} \tag{3.58}$$

Actually $g_{\mu\nu}$ is such a simple matrix that it is its own inverse: so for $g^{\mu\nu}$ we also have

$$g^{\mu\nu} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{bmatrix} \tag{3.59}$$

$g^{\mu\nu}$ is the inverse of $g_{\mu\nu}$

Despite this fact it is still important to use $g^{\mu\nu}$ in some formulae rather than $g_{\mu\nu}$. This is because $g^{\mu\nu}$ and $g_{\mu\nu}$ are used for the task of *raising and lowering indices*, a task which we explain below, cf. 3.72 and what follows. Also all spacetimes possess metrics $g_{\mu\nu}$ and, for spacetimes other than Minkowski's, such as those of *General Relativity*, $g_{\mu\nu}$ is *not* usually its own inverse.

Now we use the metric $g_{\mu\nu}$ to obtain our promised new formula for the Minkowski length $L_{\mathbf{Mk}}$. Consider the expression

$$\begin{bmatrix} ct & x & y & z \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{bmatrix} \begin{bmatrix} ct \\ x \\ y \\ z \end{bmatrix} \tag{3.60}$$

[9] Care is needed here since the minus signs in the definition 3.55 of $L_{\mathbf{Mk}}$ means that $L_{\mathbf{Mk}}^2$ can be *negative* as well as positive—a negative value of $L_{\mathbf{Mk}}^2$ means that $L_{\mathbf{Mk}}$ is pure imaginary and this is impossible for a normal Euclidean length. Indeed, as we shall see below, negative values of $L_{\mathbf{Mk}}^2$ do occur all the time and have an important physical meaning. It turns out that one is safe enough in thinking of $L_{\mathbf{Mk}}$ as a “length” as long as one does not try and apply results like Pythagoras' theorem to it; these latter are only true for the normal Euclidean length L of 3.54

We readily compute that

$$\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{bmatrix} \begin{bmatrix} ct \\ y \\ y \\ z \end{bmatrix} = \begin{bmatrix} ct \\ -x \\ -y \\ -z \end{bmatrix} \tag{3.61}$$

So

$$\begin{bmatrix} ct & x & y & z \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{bmatrix} \begin{bmatrix} ct \\ x \\ y \\ z \end{bmatrix} = \begin{bmatrix} ct & x & y & z \end{bmatrix} \begin{bmatrix} ct \\ -x \\ -y \\ -z \end{bmatrix} \tag{3.62}$$

and finally we have

$$\begin{bmatrix} ct & x & y & z \end{bmatrix} \begin{bmatrix} ct \\ -x \\ -y \\ -z \end{bmatrix} = c^2t^2 - x^2 - y^2 - z^2 \tag{3.63}$$

So we have found that

$$\begin{bmatrix} ct & x & y & z \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{bmatrix} \begin{bmatrix} ct \\ x \\ y \\ z \end{bmatrix} = L^2_{\mathbf{Mk}} \tag{3.64}$$

This expression for $L^2_{\mathbf{Mk}}$ is crying out for some more compact notation and this is easy to supply. Since $[x^\mu] \equiv [\, ct \quad x \quad y \quad z \,]$ we see that

$$\sum_{\mu=0}^{3} g_{\mu\nu} x^\mu = [\, ct \quad -x \quad -y \quad -z \,] \tag{3.65}$$

and multiplying by a further factor x^ν and summing over ν we find that

$$\sum_{\nu=0}^{3} \sum_{\mu=0}^{3} g_{\mu\nu} x^\mu x^\nu = c^2t^2 - x^2 - y^2 - z^2 \tag{3.66}$$

i.e. we have the more compact formula

$$L^2_{\mathbf{Mk}} = \sum_{\nu=0}^{3} \sum_{\mu=0}^{3} g_{\mu\nu} x^\mu x^\nu \tag{3.67}$$

It turns out that the summation signs in this formula can actually be dispensed with making it even more compact. This is made possible by using what is called the *Einstein summation convention* and we explain this in the next section.

§ 6. The Einstein summation convention

Einstein introduced a very useful notational rule for avoiding the need to write the summation sign

$$\Sigma \tag{3.68}$$

This is called the *Einstein summation convention* and says that *any index which is repeated is summed over.* We now illustrate it by some examples.

If we take our expression 3.67 for the Minkowski length which was

$$L^2_{\mathbf{Mk}} = \sum_{\nu=0}^{3}\sum_{\mu=0}^{3} g_{\mu\nu}x^\mu x^\nu \tag{3.69}$$

we notice that both μ and ν are repeated so, using the summation convention we have the really compact formula

$$L^2_{\mathbf{Mk}} = g_{\mu\nu}x^\mu x^\nu \tag{3.70}$$

The metric $g_{\mu\nu}$, and its inverse $g^{\mu\nu}$, have a second important use which is to *raise* and *lower* indices on linear algebraic objects.

$g_{\mu\nu}$ and $g^{\mu\nu}$ are used when raising and lowering indices

Example *Index Raising*

For example an object with an *lower* index such as

$$A_\alpha \tag{3.71}$$

can have it *raised* by multiplication with $g^{\alpha\beta}$ and summation over β: more precisely we *define* A^α—that is A with the upper index—by writing

An index raised

$$A^\alpha = g^{\alpha\beta}A_\beta \tag{3.72}$$

Suppose then that we take a specific example

$$(A_0, A_1, A_2, A_3) = (a, b, c, d) \tag{3.73}$$

then, on raising the index, we straightaway compute that

Index raising changes three signs

$$(A^0, A^1, A^2, A^3) = (a, -b, -c - d) \tag{3.74}$$

and one sees that the index raising operation leaves the A_0 *component unchanged* but *changes the sign of A_1, A_2 and A_3.*

Example *Index lowering*

Conversely the lowering of an index is done with the metric $g_{\mu\nu}$, if we begin with

$$B^\nu \tag{3.75}$$

then B_μ is *defined* by the equation

An index lowered

$$B_\mu = g_{\mu\nu} B^\nu \tag{3.76}$$

Again it is clear that index lowering changes some signs; in fact it changes exactly the same signs as the index is done by the index raising operation. If we start with

$$(x^0, x^1, x^2, x^3) = (e, f, g, h) \tag{3.77}$$

then, on applying the definition 3.76, we find that

Index lowering also changes three signs

$$(x_0, x_1, x_2, x_3) = (e, -f, -g, -h) \tag{3.78}$$

as claimed. Next we illustrate what happens when an index is repeated.

Example *Some examples of repeated indices*

If we take the expression

$$A_\alpha B^\beta \tag{3.79}$$

and set α equal to β giving [10]

$$A_\alpha B^\alpha \tag{3.80}$$

then the repetition of the *Greek index* α means that $A_\alpha B^\alpha$ is actually a sum of four terms so one has

A repeated Greek index stands for four terms

$$A_\alpha B^\alpha = \sum_{\alpha=0}^{3} A_\alpha B^\alpha = A_0 B^0 + A_1 B^1 + A_2 B^2 + A_3 B^3 \tag{3.81}$$

On the other hand if we repeat a *Latin index* then we sum over only *three values.* For example one has

A repeated Latin index stands for only three terms

$$x_k x^k = \sum_{k=1}^{3} x_k x^k = x_1 x^1 + x_2 x^2 + x_3 x^3 \tag{3.82}$$

In calculus if

$$f = f(y^0, y^1, y^2, y^3) \tag{3.83}$$

and one changes variables from

$$(y^0, y^1, y^2, y^3) \quad \text{to} \quad (x^0, x^1, x^2, x^3) \tag{3.84}$$

then the chain rule for partial derivatives often throws up expressions such as

$$\frac{\partial f}{\partial x^\alpha} = \frac{\partial f}{\partial y^0}\frac{\partial y^0}{\partial x^\alpha} + \frac{\partial f}{\partial y^1}\frac{\partial y^1}{\partial x^\alpha} + \frac{\partial f}{\partial y^2}\frac{\partial y^2}{\partial x^\alpha} + \frac{\partial f}{\partial y^3}\frac{\partial y^3}{\partial x^\alpha} \tag{3.85}$$

which we can now abbreviate this to (note β is repeated but α is not)

[10] The setting of one index equal to another, thereby producing a repeated index, is called *contraction.*

$$\frac{\partial f}{\partial x^\alpha} = \frac{\partial f}{\partial y^\beta}\frac{\partial y^\beta}{\partial x^\alpha} \tag{3.86}$$

The chain rule written using the repeated index rule

and so on; we shall meet more examples in subsequent calculations.

Finally, when calculating with repeated indices, it is important to notice that one only ever contracts an upper index with a lower index: One *never contracts* two lower indices or two upper indices. In fact the reader should not worry much about this matter because, in routine calculations of *invariant* quantities—that is basis independent quantities—one will only encounter instances of repeated indices where one index is upper and the other lower; indeed an instance of a failure of this assertion is almost certainly an algebraic error.

Contractions are always of upper indices with lower indices

Terminology (Four vectors or 4 vectors) *The various mathematical objects*

Four vectors

$$x_\mu, A^\alpha, B^\alpha \tag{3.87}$$

discussed above are all vectors in Minkowski space with 4 *components; such objects are often referred to as four vectors or* 4 *vectors.*

Example *The length of a four vector A*

It is also useful to write out the length (squared) A^2 of an arbitrary four vector A in terms of its components. We readily compute that

$$\begin{aligned} A &= (A^0, A^1, A^2, A^3) \\ \Rightarrow A^2 \equiv g_{\mu\nu}A^\mu A^\nu &= (A^0)^2 - (A^1)^2 - (A^2)^2 - (A^3)^2 \\ &= A_\mu A^\mu \end{aligned} \tag{3.88}$$

Summarising we write

$$A^2 = (A^0)^2 - (A^1)^2 - (A^2)^2 - (A^3)^2 \tag{3.89}$$

which the reader will find a useful formula to remember.

Example *The Lorentz transformation abbreviated*

The summation convention allows to write the Lorentz transformation in a satisfyingly abbreviated form: if we use four vector notation and write

$$(x^0, x^1, x^2, x^3) = (ct, x, y, z) \quad \text{and} \quad (y^0, y^1, y^2, y^3) = (ct', x', y', z') \tag{3.90}$$

then statement 3.46 of the Lorentz transformation becomes simply

$$y^\alpha = \Lambda^\alpha_\beta x^\beta \tag{3.91}$$

§ 7. Lorentz transformations are linear

We want to show that the Lorentz transformations Λ are *linear*: that is the x', y', z', t' are linear functions of the x, y, z, t. This is accomplished by using the fact that a particle moving with constant velocity in the frame F must also have constant velocity in the other inertial frame F'. In other words if the particle is moving with constant velocity in F it cannot appear to be accelerating in F'.

For brevity we shall again use four vector notation; so, as before, we have $(x^0, x^1, x^2, x^3) = (ct, x, y, z)$ and $(y^0, y^1, y^2, y^3) = (ct', x', y', z')$.

To this end, then, let the particle have coordinates

$$\begin{aligned} &x^\alpha(\tau) \quad \text{in} \quad F \\ &y^\alpha(\tau) \quad \text{in} \quad F' \end{aligned} \tag{3.92}$$

where τ is a parameter which determines the position of the particle on its path; it is clear that, for constant velocity, we can assume that

$$\begin{aligned} \frac{d\tau}{dt} &= A, \quad A \text{ a constant} \\ \Rightarrow \frac{dt}{d\tau} &= \frac{1}{A} \end{aligned} \tag{3.93}$$

Now its velocity is given by $dx^i(\tau)/dt$, $i = 1, \ldots 3$, and so we compute that

$$\begin{aligned} \frac{dx^i(\tau)}{dt} &= \frac{dx^i(\tau)}{d\tau}\frac{d\tau}{dt} \\ &= A\frac{dx^i(\tau)}{d\tau} = \text{a constant} \end{aligned} \tag{3.94}$$

But $x^0 = ct$ and so

$$\frac{dx^0}{d\tau} = \frac{c}{A} = \text{a constant} \tag{3.95}$$

and so $dx^\alpha/d\tau$ is constant for all four values $\alpha = 0, \ldots 3$. Hence constant velocity in F means that

$$\begin{aligned} \frac{dx^\alpha}{d\tau} &= \text{a constant} \\ \Rightarrow \frac{d^2x^\alpha}{d\tau^2} &= 0 \end{aligned} \tag{3.96}$$

But, using the same reasoning for F', we *also* have

$$\begin{aligned} \frac{dy^\alpha}{d\tau} &= \text{another constant} \\ \Rightarrow \frac{d^2y^\alpha}{d\tau^2} &= 0 \end{aligned} \tag{3.97}$$

However the chain rule tells us that

$$\begin{aligned} \frac{dy^\alpha}{d\tau} &= \frac{\partial y^\alpha}{\partial x^\beta}\frac{dx^\beta}{d\tau} \\ \Rightarrow \frac{d^2y^\alpha}{d\tau^2} &= \frac{\partial^2 y^\alpha}{\partial x^\beta \partial x^\gamma}\frac{dx^\beta}{d\tau}\frac{dx^\gamma}{d\tau} + \frac{\partial y^\alpha}{\partial x^\beta}\frac{d^2x^\beta}{d\tau^2} \end{aligned} \tag{3.98}$$

But we have just seen in 3.96 and 3.97 that

$$\frac{d^2y^\alpha}{d\tau^2} = 0, \quad \frac{d^2x^\beta}{d\tau^2} = 0 \tag{3.99}$$

So we deduce that

$$\frac{\partial^2 y^\alpha}{\partial x^\beta \partial x^\gamma} \frac{dx^\beta}{d\tau} \frac{dx^\gamma}{d\tau} = 0 \tag{3.100}$$

and this must be true for *arbitrary* (constant) particle 4 vectors $dx^\alpha/d\tau$ so we conclude that

$$\frac{\partial^2 y^\alpha}{\partial x^\beta \partial x^\gamma} = 0 \tag{3.101}$$

and integrating twice we get

$$\begin{aligned} \frac{\partial y^\alpha}{\partial x^\beta} &= C^\alpha_\beta, \quad (C^\alpha_\beta \text{ constant}) \\ y^\alpha &= C^\alpha_\beta x^\beta + D^\alpha, \quad (D^\alpha \text{ constant}) \end{aligned} \tag{3.102}$$

i.e. y^α is a linear function of x^β as we claimed.

We can set D^α to zero since, as we said on p. 16, the spatial origins of the two frames F and F' coincide when $t = t' = 0$; this means that that $x^\alpha = 0 \Rightarrow y^\alpha = 0$ so that $D^\alpha = 0$. Hence and so we are just left with

$$y^\alpha = C^\alpha_\beta x^\beta \tag{3.103}$$

and this shows that the constants C^α_β are precisely the matrix entries Λ^α_β of 3.49.

CHAPTER IV

New properties for length and time

§ 1. Simultaneity is relative

One of the first very striking and counter intuitive results of relativity is that the simultaneity of two events is *not an absolute notion* but is *only relative*. This is easy to prove.

We shall work with the frames F and F' of chapter 3 as this will make our formulae depend only on x and t rather than x, y, z, t; there is no loss of generality in doing this.

Consider the Lorentz transformation as given in 3.33, that is

$$\begin{aligned} x' &= \gamma\,(x - vt) \\ t' &= \gamma\left(-\frac{v}{c^2}x + t\right) \end{aligned} \tag{4.1}$$

Take two events 1 and 2, with locations and times (x_1, t_1) and (x_2, t_2) respectively in F. Now let 1 and 2 be *simultaneous* in F, so we have

Simultaneous in F

$$t_1 = t_2 \tag{4.2}$$

The snag is that, as we shall now see,

$$t_1 = t_2 \tag{4.3}$$

does *not imply*

$$t_1' = t_2' \tag{4.4}$$

so 1 and 2 will not also be simultaneous in F'. Let us verify this: using 4.1 gives us

$$\begin{aligned} t_1' &= \gamma\left(-\frac{v}{c^2}x_1 + t_1\right) \\ t_2' &= \gamma\left(-\frac{v}{c^2}x_2 + t_2\right) \end{aligned} \tag{4.5}$$

Subtracting the two equations, and remembering that $t_1 = t_2$, we get

$$\begin{aligned} t_1' - t_2' &= \gamma\left(-\frac{v}{c^2}\right)(x_1 - x_2) \\ &\neq 0, \quad \text{because } x_1 \neq x_2 \end{aligned} \tag{4.6}$$

Hence 1 and 2 are *not simultaneous* in F'.

Not simultaneous in F'

So we have found a very important result: Simultaneity is only a relative concept, relative, that is, to an inertial frame. Events which are simultaneous in one frame are not, in general, simultaneous in another frame.

This relativity of simultaneity is also the key to understanding properly the famous length contraction of a moving object. It is the subject of the next section.

§ 2. The Fitzgerald-Lorentz length contraction

We have already seen in chapter 1 that the negative result of the Michelson-Morley experiment can be given a kind of ad hoc explanation by a contraction of the horizontal arm of the interferometer by a factor

$$\sqrt{1-\frac{v^2}{c^2}}=\frac{1}{\gamma}, \quad \gamma=\frac{1}{\sqrt{1-\frac{v^2}{c^2}}} \tag{4.7}$$

where v is the orbital velocity of the Earth. It is now time to show that length contraction does happen in relativity and examine carefully its nature.

Take our customary pair of frames F and F' of figure 6. Next lay a rod of length L' at rest along the x'-axis of frame F'.

Now let us calculate: Let the coordinates of the ends of the rod be

$$\begin{array}{lllll} (x_1,t_1) & \text{and} & (x_2,t_2) & \text{in} & F \\ (x'_1,t'_1) & \text{and} & (x'_2,t'_2) & \text{in} & F' \end{array} \tag{4.8}$$

So we have

$$L'=x'_2-x'_1 \tag{4.9}$$

What will an observer at rest in F measure the length to be? Well, first we *define* the *length of a moving rod* to be the distance obtained by a *simultaneous* measurement of the positions of its ends.

Note that we don't have to worry about the values of the times t'_1 and t'_2 at which F' measures the ends of the rod: since the rod is at rest in F' he will get the same answers for x'_1 and x'_2 whatever the values of t'_1 and t'_2.

So the observer at rest in F will say that the length of the moving rod is L where

$$L=x_2-x_1 \tag{4.10}$$

and he will also say that

$$t_1=t_2 \tag{4.11}$$

since he must measure both ends at the same time. But we can calculate this at once using the Lorentz transformation 4.1. which gives us

$$x'_1=\gamma(x_1-vt_1), \quad x'_2=\gamma(x_2-vt_2) \tag{4.12}$$

Hence we calculate that

$$x_2' - x_1' = \gamma\{(x_2 - x_1) - v(t_2 - t_1)\} \tag{4.13}$$

But

$$\begin{aligned}(x_2' - x_1') &= L' \\ (x_2 - x_1) &= L \\ t_1 &= t_2\end{aligned} \tag{4.14}$$

so we have

$$L' = \gamma L \tag{4.15}$$

and we see that the *moving length* L and the *static length* L' obey the relation

The Lorentz-Fitzgerald contraction derived

$$L = \frac{L'}{\gamma} \tag{4.16}$$

and

$$L < L', \quad \text{since } \frac{1}{\gamma} < 1 \tag{4.17}$$

which is precisely the Lorentz-Fitzgerald result of chapter 1: the moving rod appears to have a length shortened by the factor $\sqrt{1 - v^2/c^2} = 1/\gamma$.

Symmetry of length result between in F and F'

It is important to realise that a rod at rest in F would also appear contracted to an observer at rest in F' and this does not contradict our first conclusion 4.16; thus the length contraction result is symmetric under interchange of the two frames F and F'.

The point is that neither observer will ever agree that both have measured the rods properly since the measurements require a simultaneous measurements of both ends and, as we have seen, if the measurements are *simultaneous in one of the frames* they are *not simultaneous in the other.*

Moving rods appear shorter

Hence there is no contradiction. We can summarise the result by saying *moving rods appear contracted.*

§ 3. Time dilation

Now we investigate the consequences of the Lorentz transformations for time. Consider a clock at rest in the frame F'; let an observer in F' read its time twice, comparing simultaneously with a standard clock of his own, thus obtaining the values

$$t_1', t_2' \tag{4.18}$$

Now an observer in F, reading the same clock, will obtain the values

$$t_1, t_2 \tag{4.19}$$

So the spacetime coordinates of the clock are

$$\begin{array}{lllll}(x_1, t_1) & \text{and} & (x_2, t_2) & \text{in} & F \\ (x_1', t_1') & \text{and} & (x_2', t_2') & \text{in} & F'\end{array} \tag{4.20}$$

So, in F', since the clock is at rest, we have

$$x'_1 = x'_2 \tag{4.21}$$

and—because the algebra works better this way—we now use the *inverse* Lorentz transformation Λ^{-1}, cf. 3.37 rather than Λ and compute that

$$\begin{aligned} t_1 &= \gamma\left(\frac{v}{c^2}x'_1 + t'_1\right) \\ t_2 &= \gamma\left(\frac{v}{c^2}x'_2 + t'_2\right) \end{aligned} \tag{4.22}$$

Subtracting these equations now gives

$$\begin{aligned} t_2 - t_1 &= \gamma\left\{\left(\frac{v}{c^2}\right)(x'_2 - x'_1) + (t'_2 - t'_1)\right\} \\ &= \gamma(t'_2 - t'_1), \quad \text{because } x'_1 = x'_2 \end{aligned} \tag{4.23}$$

So setting

$$T = t_2 - t_1, \quad T' = t'_2 - t'_1 \tag{4.24}$$

we have found that

$$T = \gamma T' \tag{4.25}$$

so

Moving clocks run slow

$$T > T', \quad \text{since } \gamma > 1 \tag{4.26}$$

Hence the observer in F records *longer* times than those recorded in F'; therefore he will say that the *moving clock is slow.*

Symmetry of clock result between in F and F'

Again, as in the length contraction discussion, the result is symmetric under interchange of the two frames F and F'. That is, a clock at rest in F would also appear slow, by the same factor, to an observer in F'. There are no contradictions in the physics just as there were none in the length contraction case; the relativity of simultaneity is again the essential ingredient which prevents there being a contradiction: in this case the observer in F' would not see the comparisons with a standard clock made by the observer in F as being simultaneous.

§ 4. Worldlines

As a particle both moves and exists in space and time it traces out a path in spacetime; this path is called a *worldline*. Figure 7 shows two worldlines Γ_1 and Γ_2. We also show just two spacetime axes x and t but of course there are four in reality; our two dimensional representation is only for simplicity. In addition to x and t we show the axes x' and t' of a another reference frame; these are drawn oblique to x and t so as to obey the Lorentz transformation 4.1.

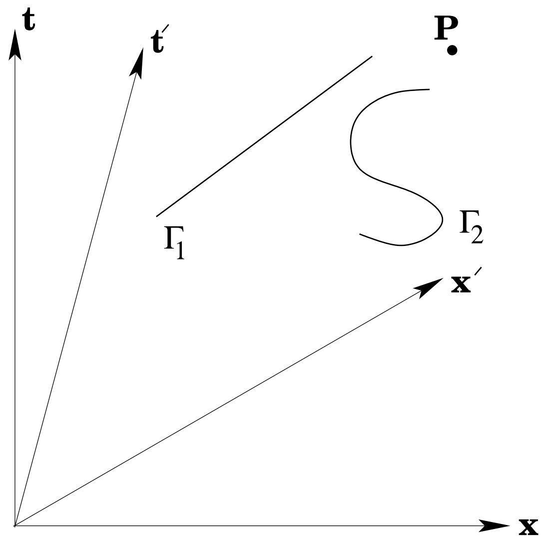

Fig. 7: Two worldlines Γ_1 and Γ_2 and a point P

A spacetime diagram

The worldline Γ_1 is straight and is that of a particle moving with constant velocity, the other one Γ_2 is curved and is traced out by a particle of varying velocity. Figure 7 is an example of what is called a *spacetime diagram*.

A *point* in spacetime, such as P shown in figure 7, represents an *event* and it can be given coordinates, say, (x, t); but these coordinates are *not unique*. For example if we have our usual two reference frames F and F' then any event, like P, will have two sets of coordinates

$$(x, t), \quad \text{and} \quad (x', t') \tag{4.27}$$

and these will be connected to one another by a Lorentz transformation.

However, remember that the *worldline itself* is a representation of the *history of a particle* which is *independent of any frame of reference.*

Worldlines independent of frame of reference

This is the same as saying that a circle drawn on a plane has an equation which depends on a choice of axes; but the circle itself is something absolute and independent of any set of axes. So, in fact, spacetime diagrams give a representation of events which is independent of any frame of reference.

A useful piece of terminology when discussing frames of reference is that of a *rest frame* so we now introduce it.

A rest frame

Terminology (Rest frame) *If a particle, or object, is at rest in an inertial frame F, then F is called the rest frame of the object.*

§ 5. Minkowski space: four vectors and light cones

Consider any four vector v^μ. From the definition 3.55, its length squared in Minkowski space is

$$g_{\mu\nu} v^\mu v^\nu \tag{4.28}$$

Incidentally by our index lowering rule 3.76 we have

$$v_\mu = g_{\mu\nu} v^\nu \tag{4.29}$$

so it is true that

$$g_{\mu\nu}v^\mu v^\nu = v_\mu v^\mu \tag{4.30}$$

This is a useful fact to remember.

We want to point out that that there are three possibilities for the magnitude of v^μ; these are described by three terms *spacelike, timelike and lightlike* which we now unveil.

Spacelike,timelike and lightlike

Terminology (Spacelike, timelike and lightlike four vectors)

$$\begin{aligned} g_{\mu\nu}v^\mu v^\nu > 0, &\quad v^\mu \text{ is then called } \textit{timelike} \\ g_{\mu\nu}v^\mu v^\nu = 0, &\quad v^\mu \text{ is then called } \textit{lightlike} \text{ or } \textit{null} \\ g_{\mu\nu}v^\mu v^\nu < 0, &\quad v^\mu \text{ is then called } \textit{spacelike} \end{aligned} \tag{4.31}$$

There is an important geometrical meaning to each of these three possibilities which we now explain. It has to do with light cones. The equation

$$x^2 + y^2 + z^2 - c^2t^2 = 0 \tag{4.32}$$

for a light wave is actually the equation for a cone in Minkowski space. If we suppress the z coordinate, so that we can make a drawing, this cone is shown if figure 8

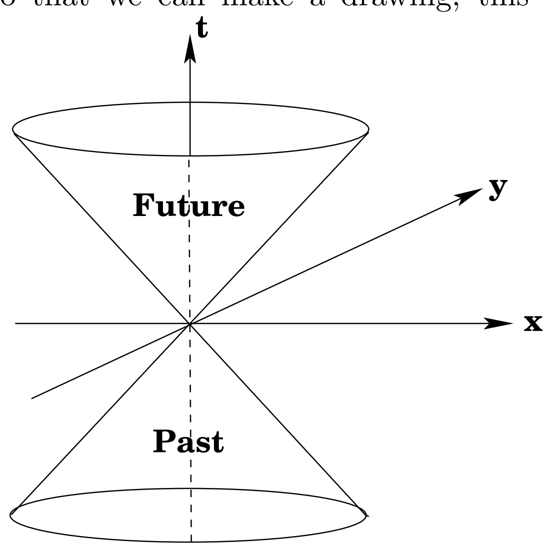

Fig. 8: A lightcone

We also show the future and the past.

It is now easy to check—for example just draw the picture with only two coordinates (x, t)—that the *sign* of the magnitude $v_\mu v^\mu$ of a four vector determines whether it is inside, outside or on the light cone. We display the situation graphically in figure 9

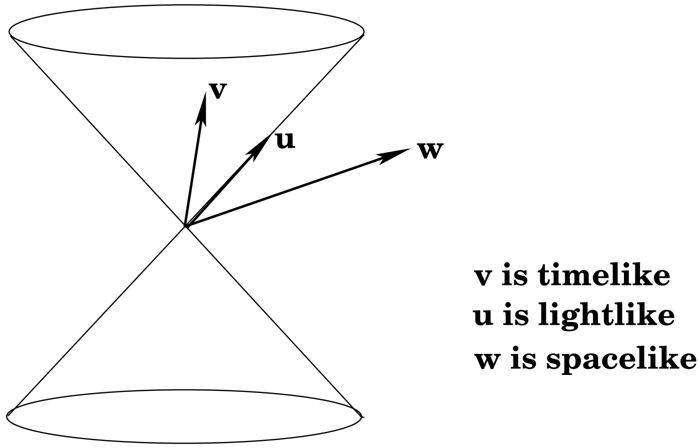

Fig. 9: Spacelike, timelike and lightlike four vectors

We can now relate these facts to worldlines. Let ds be an element of length on the worldline of a moving particle so that we have

$$ds^2 = g_{\mu\nu} dx^\mu dx^\nu = c^2 dt^2 - dx^2 - dy^2 - dz^2 \tag{4.33}$$

and the velocity $\mathbf{v}$ of the particle is given by

$$\mathbf{v} = \left(\frac{dx(t)}{dt}, \frac{dy(t)}{dt}, \frac{dz(t)}{dt} \right) \tag{4.34}$$

and suppose that ds^2 is *timelike*. Then we have

$$\begin{aligned} ds^2 = c^2 dt^2 - dx^2 - dy^2 - dz^2 > 0 \\ \Rightarrow c^2 - \left(\frac{dx}{dt}\right)^2 - \left(\frac{dy}{dt}\right)^2 - \left(\frac{dz}{dt}\right)^2 > 0 \\ \Rightarrow c^2 > v^2 \end{aligned} \tag{4.35}$$

In other words

$$v < c \tag{4.36}$$

so the particle is moving *slower than the velocity of light.*

Timelike worldline $v < c$

A worldline which permanently has this property—i.e. that ds^2 is always *positive*—is called a *timelike worldline.* So the velocity of a particle with a timelike worldline, though it can vary, always has magnitude less than c. It is also true that the worldline is always *inside* any light cone that has its vertex on it—cf. figure 10.

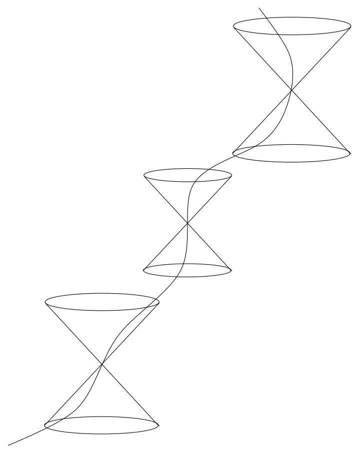

Fig. 10: A timelike worldline

Conversely if we consider a worldline with $ds^2 < 0$ always—a spacelike worldline— then the same argument with appropriate changes of sign would show that

Spacelike worldline $v > c$

$$v > c \tag{4.37}$$

so that a particle with this worldline would have exceeded the velocity of light which we know is impossible. Hence physical objects *cannot* have *spacelike worldlines.*

Finally a worldline with $ds^2 = 0$ always—a lightlike worldline—must be that of something travelling with velocity

Lightlike worldline $v = c$

$$v = c \tag{4.38}$$

Furthermore these properties of worldlines are independent of any reference frames: this is clear since the properties all depend on the Lorentz invariant object

$$ds^2 \tag{4.39}$$

which has the same value in all reference frames.

§ 6. Proper time

Time is no longer absolute in relativity; we have seen that it is local to a particular reference frame. The absence of an absolute time in relativity creates problems; these can be partially remedied by the introduction of what is called *proper time* which is denoted by

$$\tau \tag{4.40}$$

Proper time is defined so that it has the *same value* in all *inertial frames* and is therefore a Lorentz invariant quantity.

It is very convenient to define proper time infinitesimally; that is to say one defines

$$d\tau \tag{4.41}$$

first rather than defining τ.

Here is how it goes first we define the spacetime separation ds between the two points

$$(x, y, z, t), \qquad \text{and} \qquad (x + dx, y + dy, z + dz, t + dt) \tag{4.42}$$

by writing

$$ds^2 = c^2dt^2 - dx^2 - dy^2 - dz^2 \tag{4.43}$$

Then $d\tau$ is defined by

$$\begin{aligned} d\tau = \frac{ds}{c}, \quad &\text{provided } ds \text{ is } \textit{real} \\ &\text{i.e. provided } ds^2 > 0 \end{aligned} \tag{4.44}$$

Note carefully the restriction to timelike separation ds—i.e. $ds^2 > 0$—this means that ds is a portion of a *timelike worldline* and hence represents the history of something travelling with less than the velocity of light.

Now consider a particle moving—for example consider the worldlines Γ_1 or Γ_2 of figure 7—then x, y, z all depend on the time t and we have

$$\begin{aligned} \mathbf{v} &= \left(\frac{dx(t)}{dt}, \frac{dy(t)}{dt}, \frac{dz(t)}{dt}\right) \\ |\mathbf{v}| &= \sqrt{\left(\frac{dx(t)}{dt}\right)^2 + \left(\frac{dy(t)}{dt}\right)^2 + \left(\frac{dz(t)}{dt}\right)^2} \\ &= v \end{aligned} \tag{4.45}$$

where v need not be constant.

But if ds is an element of length one of these worldlines then we can extract a factor of dt^2 and work as follows

$$\begin{aligned} ds^2 &= c^2dt^2 - dx^2 - dy^2 - dz^2 \\ &= \left\{c^2 - \left(\frac{dx}{dt}\right)^2 - \left(\frac{dy}{dt}\right)^2 - \left(\frac{dz}{dt}\right)^2\right\} dt^2 \\ &= (c^2 - v^2)dt^2 \\ \Rightarrow ds^2 &= c^2\left\{1 - \frac{v^2}{c^2}\right\} dt^2 \end{aligned} \tag{4.46}$$

Now we divide both sides by c^2 and take the square root obtaining

$$\frac{ds}{c} = \sqrt{1 - \frac{v^2}{c^2}}dt \tag{4.47}$$

that is

$$\begin{aligned} & d\tau = \sqrt{1 - \frac{v^2}{c^2}}dt, \qquad \text{by definition 4.44} \\ \text{i.e.} \qquad & d\tau = \frac{dt}{\gamma}, \quad \gamma = \frac{1}{\sqrt{1 - \frac{v^2}{c^2}}} \end{aligned} \tag{4.48}$$

Note that, when $v = 0$, we have

$$d\tau = dt, \quad (\text{for } v = 0) \tag{4.49}$$

and t and τ coincide i.e. τ is the normal time measured in the rest frame of the particle; this is a good way of thinking about τ.

In the next section we discuss the very interesting subject known as the *twin paradox.*

§ 7. The twin paradox

Consider the following story: A pair of twins live on the Earth; one goes on a very high speed journey into space and eventually returns to the Earth. At their reunion the twins compare clocks and find, that, because of time dilation, much less time has elapsed for the twin who made the journey than for the one who stayed behind. Hence the twin who travelled is *younger*.

This appears to be a paradox because we saw, in § 3, that time dilation was symmetrical between a pair of inertial frames F and F'; so does the twin story contradict that assertion?

The answer is *no*: indeed there is no symmetry because the twin story does *not* involve two *inertial* frames but only *one*. The twin who stays at home is in an inertial frame but the one who travels undergoes acceleration to reach a high velocity and deceleration to come back. Hence the second twin, because of the varying velocity, was not in an inertial frame. This absence of symmetry means that there is no contradiction and the travelled twin *is younger.* We now prove this.

The ingredients we need for the proof are worldlines and proper time. Let the twin on the Earth have worldline Γ_1 and the other have word line Γ_2. Figure 11 shows the worldlines; note that they must both begin and end at the same pair of spacetime points.

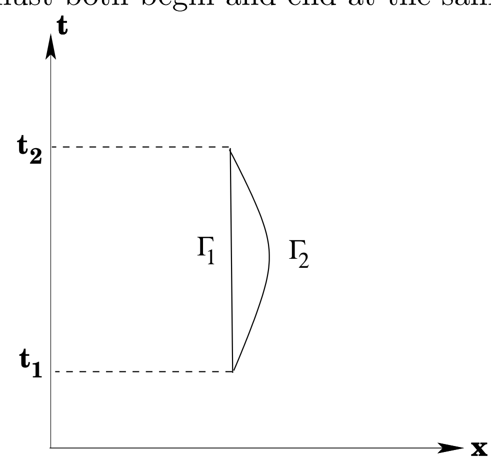

Fig. 11: The worldlines Γ_1 and Γ_2 of the twins

Let the journey begin at time t_1 and end at time t_2; this time will be the same for both twins. Now if T_1 is the time elapsed for the Earthbound twin and T_2 the time elapsed for the travelling twin; then T_1 and T_2 are given by the total proper time for each worldline. Thus we have

$$T_1 = \int_{\Gamma_1} d\tau, \quad T_2 = \int_{\Gamma_2} d\tau \tag{4.50}$$

But along Γ_1, the twin has zero velocity v so

$$\begin{aligned} d\tau &= dt, \qquad \text{using 4.49} \\ \Rightarrow \int_{\Gamma_1} d\tau &= \int_{t_1}^{t_2} dt \\ &= t_2 - t_1 \end{aligned} \tag{4.51}$$

So we have

$$T_1 = t_2 - t_1 \tag{4.52}$$

But for the travelling twin we have

$$\int_{\Gamma_2} d\tau = \int_{t_1}^{t_2} \sqrt{1 - \frac{v^2}{c^2}} dt, \qquad \text{using 4.48} \tag{4.53}$$

giving

$$T_2 = \int_{t_1}^{t_2} \sqrt{1 - \frac{v^2}{c^2}} dt \tag{4.54}$$

However, remembering that

$$\sqrt{1 - \frac{v^2}{c^2}} < 1, \qquad \text{since} \qquad 0 < v < c \tag{4.55}$$

it is immediate that,

$$\int_{t_1}^{t_2} \sqrt{1 - \frac{v^2}{c^2}} dt < \int_{t_1}^{t_2} dt \tag{4.56}$$

since the integrand $\sqrt{1 - \frac{v^2}{c^2}}$ of the first integral is less than that of the second integral for that part where $v \neq 0$.

So we have proved that

$$T_2 < T_1 \tag{4.57}$$

meaning that the travelled twin *is indeed younger*.

CHAPTER V
Relativistic kinematics

§ 1. Relativistic addition of velocities

In this chapter we want to study the kinematics of particles and see what changes relativity brings. The first thing we examine is the matter of *relative velocities*.

Consider three inertial frames F, F_1 and F_2. Let

$$\begin{aligned} &F_1 \text{ have a velocity } v_1 \text{ relative to } F \\ &F_2 \text{ have a velocity } v_2 \text{ relative to } F_1 \end{aligned} \tag{5.1}$$

Then, in Newtonian physics, we would say that the velocity of the frame F_2 relative to F is

Newtonian result

$$v_1 + v_2 \tag{5.2}$$

However, in special relativity, this is *false* and the *correct result* is

Relativistic result

$$\frac{v_1 + v_2}{1 + \frac{v_1 v_2}{c^2}} \tag{5.3}$$

We now prove this.

We shall simplify matters by working in a two dimensional spacetime so that, using 3.25 and 3.28, a Lorentz transformation Λ has the form

$$\Lambda = \begin{bmatrix} \cosh(\theta) & \sinh(\theta) \\ \sinh(\theta) & \cosh(\theta) \end{bmatrix}, \quad \tanh(\theta) = -\frac{v}{c} \tag{5.4}$$

Now consider two Lorentz transformations Λ_{FF_1} and $\Lambda_{F_1F_2}$. Let

$$\begin{aligned} &\Lambda_{FF_1} \text{ transform coordinates from } F \text{ to } F_1 \\ &\Lambda_{F_1F_2} \text{ transform coordinates from } F_1 \text{ to } F_2 \end{aligned} \tag{5.5}$$

So this means that

$$\begin{aligned} \Lambda_{FF_1} &= \begin{bmatrix} \cosh(\theta_1) & \sinh(\theta_1) \\ \sinh(\theta_1) & \cosh(\theta_1) \end{bmatrix}, \quad \tanh(\theta_1) = -\frac{v_1}{c} \\ \Lambda_{F_1F_2} &= \begin{bmatrix} \cosh(\theta_2) & \sinh(\theta_2) \\ \sinh(\theta_2) & \cosh(\theta_2) \end{bmatrix}, \quad \tanh(\theta_2) = -\frac{v_2}{c} \end{aligned} \tag{5.6}$$

Now we want to Lorentz transform from F to F_2; this we can do by applying Λ_{FF_1} followed by $\Lambda_{F_1F_2}$; calling this Lorentz transformation Λ_{FF_2}, we have

$$\begin{aligned} &\mathbf{Mk} \xrightarrow{\Lambda_{FF_1}} \mathbf{Mk} \xrightarrow{\Lambda_{F_1F_2}} \mathbf{Mk} \\ &\Lambda_{FF_2} = \Lambda_{F_1F_2}\Lambda_{FF_1} \end{aligned} \tag{5.7}$$

So, using our expressions for $\Lambda_{F_1F_2}$ and Λ_{FF_1}, we readily compute that

$$\begin{aligned} \Lambda_{FF_2} &= \begin{bmatrix} \cosh(\theta_2) & \sinh(\theta_2) \\ \sinh(\theta_2) & \cosh(\theta_2) \end{bmatrix} \begin{bmatrix} \cosh(\theta_1) & \sinh(\theta_1) \\ \sinh(\theta_1) & \cosh(\theta_1) \end{bmatrix} \\ &= \begin{bmatrix} \cosh(\theta_2)\cosh(\theta_1) + \sinh(\theta_2)\sinh(\theta_1) & \sinh(\theta_2)\cosh(\theta_1) + \cosh(\theta_2)\sinh(\theta_1) \\ \sinh(\theta_2)\cosh(\theta_1) + \cosh(\theta_2)\sinh(\theta_1) & \sinh(\theta_2)\sinh(\theta_1) + \cosh(\theta_2)\cosh(\theta_1) \end{bmatrix} \end{aligned} \tag{5.8}$$

But cosh and sinh obey the addition formulae

$$\begin{aligned} \cosh(A+B) &= \cosh(A)\cosh(B) + \sinh(A)\sinh(B) \\ \sinh(A+B) &= \sinh(A)\cosh(B) + \cosh(A)\sinh(B) \end{aligned} \tag{5.9}$$

So, setting $\theta_2 = A$ and $\theta_1 = B$ we immediately conclude that

$$\Lambda_{FF_2} = \begin{bmatrix} \cosh(\theta_1+\theta_2) & \sinh(\theta_1+\theta_2) \\ \sinh(\theta_1+\theta_2) & \cosh(\theta_1+\theta_2) \end{bmatrix} \tag{5.10}$$

We see that Λ_{FF_2} is a Lorentz transformation whose velocity v_{rel} relative to F is given by

$$\tanh(\theta_1+\theta_2) = -\frac{v_{rel}}{c}, \qquad \text{using 5.4} \tag{5.11}$$

However we know, from 2.16, that

$$\tanh(\theta_1+\theta_2) = \frac{\tanh(\theta_1) + \tanh(\theta_2)}{1 + \tanh(\theta_1)\tanh(\theta_2)} \tag{5.12}$$

so, taking $\tanh(\theta_1)$ and $\tanh(\theta_2)$ from 5.6, we deduce that

$$-\frac{v_{rel}}{c} = \frac{-v_1/c - v_2/c}{1 + \frac{v_1v_2}{c^2}} \tag{5.13}$$

which tidies up to the desired equation, namely

Relativistic law for relative velocities

$$v_{rel} = \frac{v_1 + v_2}{1 + \frac{v_1v_2}{c^2}} \tag{5.14}$$

in agreement with 5.3.

Note that, as one might expect, if v_1 and v_2 are both much less that c, we have, to a good approximation, the old Newtonian result. More precisely

$$\left.\begin{aligned} v_1 << c \\ v_2 << c \end{aligned}\right\} \Rightarrow \frac{v_1 + v_2}{1 + \frac{v_1v_2}{c^2}} \approx v_1 + v_2 \tag{5.15}$$

Finally let us point out an alternative rather neat way of stating the velocity addition formula 5.14. Define the hyperbolic angles $\theta(v_1)$, $\theta(v_2)$ and $\theta(v_{rel})$ by writing

$$\tanh(\theta(v_1)) = -\frac{v_1}{c}, \quad \tanh(\theta(v_2)) = -\frac{v_2}{c}, \quad \tanh(\theta(v_{rel})) = -\frac{v_{rel}}{c}$$

then we know that velocity addition formula 5.14 just says that

$$\tanh(\theta(v_{rel})) = \tanh(\theta(v_1) + \theta(v_2)) \tag{5.16}$$

which implies the simple additive formula

$$\theta(v_{rel}) = \theta(v_1) + \theta(v_2) \tag{5.17}$$

which is equivalent to 5.14.

The variable $-\theta(v)$, or simply $-\theta$ if the context allows, is called the *rapidity*; cf. also the footnote on p. 17. So we have

Rapidity defined

$$-\theta(v) \quad \text{is called the } \textit{rapidity}$$
$$\text{where } \tanh(\theta(v)) = -\frac{v}{c}$$

and equation 5.17 is referred to as the *additivity of rapidities* between inertial frames.

§ 2. Relativity and equations of motion

In this section we study the equations of motion of a particle of mass m; while doing so we must remember that Einstein's relativity principle says that such equations must be the same in all inertial frames. This means that we shall need to reformulate the old Newtonian equations of motion in order to make them relativistically invariant.

What we require therefore is that the equations of motion of a particle should be *invariant under Lorentz transformations.*

Figure 12 shows a particle of mass m tracing out a path $\mathbf{r}(t)$ where

$$\mathbf{r}(t) = x(t)\mathbf{i} + y(t)\mathbf{j} + z(t)\mathbf{k} \equiv (x(t), y(t), z(t)) \equiv x^i(t) \tag{5.18}$$

This is a three dimensional picture: it does not show a worldline

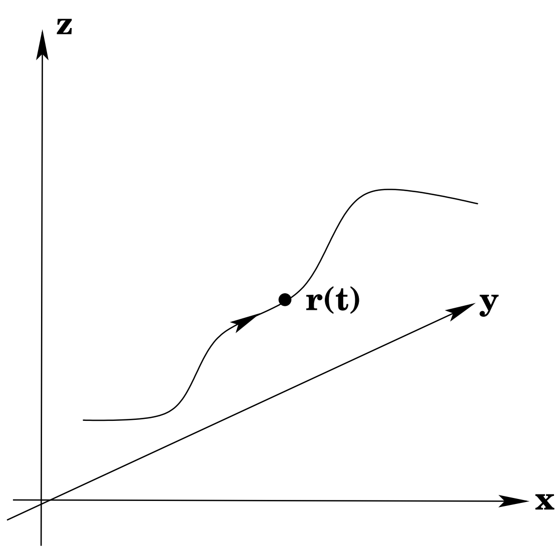

Fig. 12: The path r(t) of a particle subject to a force F

Thus, in Newtonian physics, if a force $\mathbf{F}$ is applied the particle, and the particle has momentum [1] $\mathbf{p}_N$ with

$$\mathbf{p}_N = m\mathbf{v} \tag{5.19}$$

Newton's second law gives the equation of motion for a particle as being

Newtonian equation of motion

$$\frac{d\mathbf{p}_N}{dt} = \mathbf{F} \tag{5.20}$$

This equation can be regarded as a definition of what we mean by a force $\mathbf{F}$. Now this equation is *not* invariant under a Lorentz transformation—it is an equation between three vectors; to obtain equations of motion which *are* invariant under Lorentz transformations the key idea is to use four vectors—*we must equate four vectors to four vectors.* A further key ingredient is that we should use the *proper time* τ; recall that τ is a Lorentz invariant quantity.

Figure 13 shows the particle's worldline Γ—depicted for simplicity in just two spacetime dimensions—on which the particle's position is denoted by $x^\mu(\tau)$, τ being the particle's proper time.

The particle's worldline

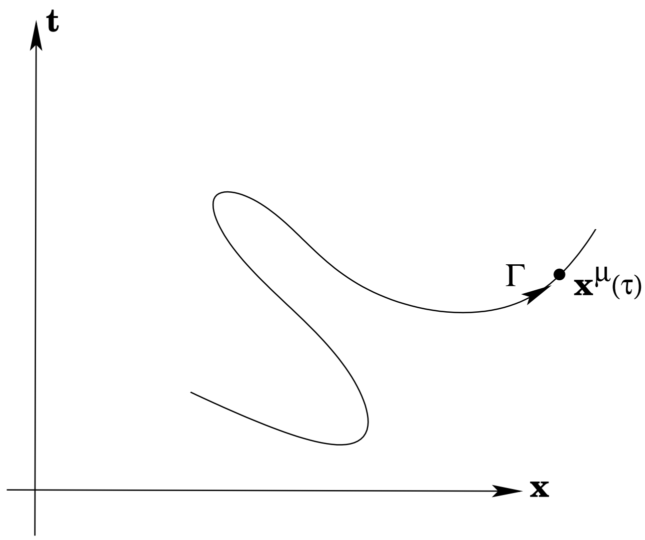

Fig. 13: The worldline Γ of the particle

The solution to our difficulties is to use Γ and to replace

(i) The time t by

$$\tau \tag{5.21}$$

(ii) The position three vector [2] $\mathbf{r}(t) = x^i(t)$ by the four vector

$$x^\mu(\tau) \tag{5.22}$$

[1] Note carefully that we denote the Newtonian momentum by $\mathbf{p}_N$ to remind the reader that it is the Newtonian momentum. Later we shall use the symbol $\mathbf{p}$ to denote three of the components of what is called the four momentum.

[2] Remember, from § 4 of chapter 3 , that Latin indices run through three values while Greek run through four.

(iii) The three velocity $\mathbf{v} = dx^i/dt$ by the four velocity

$$\frac{dx^\mu}{d\tau} \tag{5.23}$$

(iv) The three momentum $\mathbf{p}_N = m(dx^i/dt)$ by the four vector $p = (p^0, p^1, p^2, p^3)$, where

$$p^\mu = m\frac{dx^\mu}{d\tau} \tag{5.24}$$

Finally our new *relativistic* equation of motion—which replaces Newton's equation of motion 5.20— is

New relativistic equation of motion

$$\frac{dp^\mu}{d\tau} = F^\mu \tag{5.25}$$

which can be regarded as defining the new four vector F^μ known as the four force.

dx^i/dt is tangent to $\mathbf{r}(t)$ while $dx^\mu/d\tau$ is tangent to Γ

Notice that the velocity $\mathbf{v} = dx^i/dt$, which is *a three vector*, and is tangent to the *path* $\mathbf{r}(t) = x^i(t)$ *in three dimensions*, is replaced by the *four vector* $dx^\mu/d\tau$ which is tangent to the *path* Γ *in four dimensions*, that is the the worldline.

Actually, the most important new vector here is the four momentum p; so let us spell out some details. We do this in the next section.

§ 3. Energy, mass and four momentum

For the ordinary three momentum $\mathbf{p}_N$ we have

$$\mathbf{p}_N = m\mathbf{v} = m\left(\frac{dx}{dt}, \frac{dy}{dt}, \frac{dz}{dt}\right) \tag{5.26}$$

while for the new object, the four momentum $p = (p^0, p^1, p^2, p^3)$ we have

$$\begin{aligned} p^\mu &= m\frac{dx^\mu}{d\tau} \\ &= m\frac{dx^\mu}{dt}\frac{dt}{d\tau} \\ &= \gamma m\frac{dx^\mu}{dt}, \qquad \text{since } \frac{dt}{d\tau} = \gamma \text{ from 4.48} \end{aligned} \tag{5.27}$$

Continuing our elucidation of p^μ, we calculate that

$$\begin{aligned} \gamma m\frac{dx^\mu}{dt} &= \gamma m\left(\frac{d(ct)}{dt}, \frac{dx}{dt}, \frac{dy}{dt}, \frac{dz}{dt}\right) \\ &= (\gamma mc, \gamma m\mathbf{v}) \\ &= (\gamma mc, \gamma\mathbf{p}_N), \qquad \text{using 5.26} \end{aligned} \tag{5.28}$$

So we have found that the four momentum $p = (p^0, p^1, p^2, p^3)$ of a particle of mass m is given by

The four momentum p

$$p = (\gamma mc, \gamma \mathbf{p}_N) \tag{5.29}$$

where $\mathbf{p}_N = m\mathbf{v}$ is the ordinary three momentum.

Now we want to understand the physical meaning of the four components of the four momentum p: well the last three components are given by

$$\gamma \mathbf{p}_N = \gamma m \mathbf{v} \tag{5.30}$$

so they are like the Newtonian three momentum $\mathbf{p}_N$ except that the mass m is multiplied by the factor γ—thus the "effective" mass is now *not* m, but γm, where

$$\gamma m = \frac{m}{\sqrt{1 - \frac{v^2}{c^2}}} \tag{5.31}$$

using the value of γ. So the mass m for a moving particle gets replaced by

$$\frac{m}{\sqrt{1 - \frac{v^2}{c^2}}} \tag{5.32}$$

Moving particles get heavier

This is really what does happen experimentally: in other words we have the very striking fact that *moving particles get heavier* and have a mass given by 5.32.

The ordinary mass m is called the *rest mass* of the particle and is sometimes denoted [3] by

$$m_0 \tag{5.33}$$

where the zero reminds us that the particle is at rest.

Notice that

$$\frac{m}{\sqrt{1 - \frac{v^2}{c^2}}} \longrightarrow \infty, \qquad \text{for } v \longrightarrow c \tag{5.34}$$

Speed of light unattainable if $m > 0$

Thus a particle moving at the speed of light would have *infinite mass*; it would also have required an infinite amount of energy to accelerate it from rest up to velocity c, so this is a way of seeing that the speed of light is not attainable for a particle of positive rest mass, i.e. $m > 0$.

As well as having positive rest mass one can have *zero rest mass* as we now explain: A particle of light is known as a *photon* and, by definition, it travels at velocity c: a photon has the important property of having *rest mass zero*, i.e.

Photons have zero rest mass

$$\begin{gathered} m = 0, \qquad \text{for a photon} \\ (\text{or } m_0 = 0 \text{ in the alternative notation}) \end{gathered} \tag{5.35}$$

[3] Look out if you are reading a textbook which uses m_0 for rest mass; it is very likely to then use m to stand for $\frac{m_0}{\sqrt{1-v^2/c^2}}$; a quantity which we would denote by γm.

Looseness in the language used in relativity can sometimes result in the word mass being used to refer to the quantity γm; be careful that this does not confuse you—confusion is fairly unlikely as the context usually makes it quite clear whether m or γm is meant.

This fact that $m = 0$, for a photon, avoids any infinite energy argument such as the one above.

We have seen the physical significance of three of the components p^μ; what about p^0 which is the remaining component? This is the subject of the next section

§ 4. The equation $E = mc^2$

We have found that

$$p^0 = \gamma mc = \frac{mc}{\sqrt{1 - \frac{v^2}{c^2}}} \tag{5.36}$$

Note, first, that if we introduce the quantity E by writing

$$p^0 = \frac{E}{c} \tag{5.37}$$

then E has the dimensions of energy; we would like to find an expression for E. To this end expand γ in powers of v^2/c^2 yielding

$$\begin{aligned} \gamma &= \frac{1}{\sqrt{1 - \frac{v^2}{c^2}}} \\ &= \left(1 + \frac{1}{2}\frac{v^2}{c^2} + \cdots\right) \end{aligned} \tag{5.38}$$

Then combining 5.36 and 5.37 gives us

$$\begin{aligned} \frac{E}{c} &= \frac{mc}{\sqrt{1 - \frac{v^2}{c^2}}} \\ \Rightarrow E &= \frac{mc^2}{\sqrt{1 - \frac{v^2}{c^2}}} \\ \Rightarrow E &= mc^2\left(1 + \frac{1}{2}\frac{v^2}{c^2} + \cdots\right), \qquad \text{using 5.38} \\ \Rightarrow E &= mc^2 + \frac{mv^2}{2} + \cdots \end{aligned} \tag{5.39}$$

Now if we peruse the last line of 5.39 we see that the second term is just the usual *kinetic energy*

Kinetic energy contained in the relativistic energy

$$\frac{mv^2}{2} \tag{5.40}$$

but the first term is simply

$$mc^2 \tag{5.41}$$

However this term mc^2, *unlike the kinetic energy*, is independent of the velocity of the particle and is *non-zero* even when the particle is *at rest*.

It was this that led Einstein to boldly suggest[4] that

$$E = \frac{mc^2}{\sqrt{1-\frac{v^2}{c^2}}} \tag{5.42}$$

is the energy of a particle of mass m whether it is *at rest or in motion.*

$E = mc^2$, at last we have it

Thus a particle of mass m at rest has energy E given by

$$E = mc^2 \tag{5.43}$$

and we now have arrived at this famous equation.

Equivalence between mass and energy

The content of the statement $E = mc^2$ is that there is an *equivalence* between mass and energy, i.e. that they are just different forms of the same thing; further, these two forms can be converted into one another: One can convert mass into energy—and also energy into mass.

However, like all theoretical statements, experiments must provide the ultimate tests of their validity. The equivalence between mass and energy asserted by $E = mc^2$ has been tested many times—cf. § 8 of chapter 7—and has been found to be perfectly correct.

The amount of energy locked up in a small amount of matter is *gigantic*. For example if we take

$$m = 1 \text{ gram} \tag{5.44}$$

then, since $c = 3 \times 10^{10}\ cm/sec$, we compute that

Enormity of the energy locked up in matter

$$\begin{aligned} E &= mc^2 \\ &= 1 \cdot \left(3 \times 10^{10}\right)^2 \ \text{ergs} \\ &= 9 \times 10^{20} \ \text{ergs} \end{aligned} \tag{5.45}$$

Now a 100 watt light bulb consumes

$$100 \text{ joules/sec}, \ (1 \text{ joule} = 10^7 \text{ ergs.}) \tag{5.46}$$

and a reasonable electricity power station has a power measured in hundreds of megawatts (MW). So 1 gram of matter could provide enough energy for a $500MW$ power station to run for

$$\begin{aligned} \frac{9 \times 10^{20}}{500 \times 10^6 \times 10^7} &= 1.8 \times 10^5 \text{ secs} \\ &= 2.08 \ \text{ days} \end{aligned} \tag{5.47}$$

a pretty impressive statistic. Unfortunately for mankind the sheer size of these energies has also been a focus of military interest.[5]

[4] Turn to chapter 8, § 2 if you want to peruse the original three page paper

[5] This energy is about the same amount of energy released in an atomic explosion. The energy released by an atomic bomb is usually specified by quoting the number of tons of TNT which would explode with the same energy: exploding one ton of TNT releases 4.184×10^9 joules and so a 21.5 kiloton explosion is equivalent to 9×10^{20} ergs. The atomic bombs tragically dropped on Hiroshima and Nagasaki in 1945 had energies of 15 and 21 kilotons respectively. Bombs with sizes measured in tens of megatons were once exploded in various nuclear tests.

§ 5. Summary of four momentum properties

It is useful to have a small section which summarises the properties and notation of the four momentum of a particle of (rest) mass m.

First of all we denote the four momentum by

$$p \tag{5.48}$$

where

$$p = (p^0, p^1, p^2, p^3) \tag{5.49}$$

We shall sometimes abbreviate the last three components (p^1, p^2, p^3) by $\mathbf{p}$—in other words we may write

$$\mathbf{p} = (p^1, p^2, p^3) \tag{5.50}$$

This means that we can abbreviate the four momentum by writing

$$p = (p^0, \mathbf{p}) \tag{5.51}$$

Also p^0 and $\mathbf{p}$ are related to energy E and the Newtonian momentum $\mathbf{p}_N$ by

$$p^0 = \frac{E}{c}, \quad (E = \gamma mc^2) \qquad \mathbf{p} = \gamma \mathbf{p}_N, \quad (\mathbf{p}_N = m\mathbf{v}) \tag{5.52}$$

If $v = 0$—i.e. the particle is at rest—then then

In its rest frame the four momentum p takes the simple form $p = (p^0, 0, 0, 0)$.

$$\gamma = 1, \qquad \text{and } p = (p^0, 0, 0, 0) \tag{5.53}$$

so that the statement $E = \gamma mc^2$ becomes

$$E = mc^2 \tag{5.54}$$

as we saw in 5.42 and 5.43.

Finally the Minkowski squared length of the four momentum p—which, of course, the same value in all frames of reference is given, using 3.89, by

$$p^2 = (p^0)^2 - (p^1)^2 - (p^2)^2 - (p^3)^2 \tag{5.55}$$

But 5.52 tells us that

$$p^0 = \frac{E}{c} = \frac{\gamma mc^2}{c} \qquad \mathbf{p} = \gamma m\mathbf{v} \tag{5.56}$$

So

$$\begin{aligned} p^2 &= m^2\gamma^2c^2 - m^2\gamma^2v^2 \\ &= m^2\gamma^2(c^2 - v^2) \\ &= m^2\frac{(c^2 - v^2)}{\left(1 - \dfrac{v^2}{c^2}\right)}, \qquad \text{using } \gamma = \frac{1}{\sqrt{1 - \frac{v^2}{c^2}}} \\ &= m^2c^2 \end{aligned}$$

Hence we have learnt that

An important result: $p^2 = m^2c^2$

$$p^2 = m^2c^2 \tag{5.57}$$

and this is a very elegant and important result: the length of the four momentum vector p of a particle is determined solely by its mass m.

In the next section we want to look at the four momentum of a photon. Since the photon has zero rest mass this is something of a special case.

§ 6. Four momentum of photons

We suppose that we have a photon of four momentum

$$p = (p^0, p^1, p^2, p^3) \tag{5.58}$$

Then because a photon has *zero rest mass*—i.e. $m = 0$—we have

$p^2 = 0$ *for photons*

$$\begin{aligned} p^2 &= 0 \\ \text{i.e. } (p^0)^2 - (p^1)^2 - (p^2)^2 - (p^3)^2 &= 0 \end{aligned} \tag{5.59}$$

But $p^0 = E/c$ where E is the energy so we have

$$p = (E/c, p^1, p^2, p^3) \tag{5.60}$$

Suppose we specialise to the case where $p^1 = p^2 = 0$—which we can do without loss of generality by rotating our spatial axes so that the photon travels along the z axis—then we have

$$p = (E/c, 0, , 0, p^3) \tag{5.61}$$

But now the $p^2 = 0$ condition says that

$$\frac{E^2}{c^2} - (p^3)^2 = 0 \tag{5.62}$$

so that

$$p^3 = \pm\frac{E}{c} \tag{5.63}$$

We shall choose

$$p^3 = +\frac{E}{c} \tag{5.64}$$

which means that the photon is traveling along the z axis in the positive z direction. Hence its four momentum p is now given by

$$p = (\frac{E}{c}, 0, 0, \frac{E}{c}) \tag{5.65}$$

Finally if we know that the frequency of the photon is ν, then quantum theory tells us that

$$E = h\nu \tag{5.66}$$

where h is Planck's constant. This our final expression for the four momentum of a photon of frequency ν is

$$p = (\frac{h\nu}{c}, 0, 0, \frac{h\nu}{c}) \tag{5.67}$$

The four momentum of a photon of frequency ν.

This latter expression can often be useful.

No rest frame for a photon

Note, by the way, that a photon has no rest frame: it moves permanently at the velocity c of light now matter what frame it is viewed from; indeed this is simply Einstein's second postulate.

§ 7. A few useful four vector properties

Here is a piece of four vector terminology.

Terminology (Future pointing) *A four vector $v = (v^0, v^1, v^2, v^3)$ is called future pointing if*

$$v^0 > 0 \tag{5.68}$$

If one refers to figure 9 one sees that the vector $\mathbf{v}$ is future pointing; this reveals the origin of the phrase *future pointing.*

It is very useful to know that the sum of two, future pointing, timelike vectors is timelike and we know prove this.

Example *The sum of timelike, future pointing, vectors is timelike*

Let p and q be two timelike, future pointing, four vectors so we have, in some frame F,

$$p = (p^0, p^1, p^2, p^3), \qquad q = (q^0, q^1, q^2, q^3)$$

and

$$\begin{array}{ll} p^2 = (p^0)^2 - (p^1)^2 - (p^2)^2 - (p^3)^2 > 0 & \quad q^2 = (q^0)^2 - (q^1)^2 - (q^2)^2 - (q^3)^2 > 0 \\ p^0 > 0 & \quad q^0 > 0 \end{array} \tag{5.69}$$

Now what about $p + q$? Well

$$(p+q)^2 = p^2 + q^2 + 2p \cdot q \tag{5.70}$$

where $p \cdot q$ is the *four dimensional dot product in Minkowski space*—that is

Minkowski dot product

$$p \cdot q = p^0 q^0 - p^1 q^1 - p^2 q^2 - q^3 q^3 \tag{5.71}$$

Remember, too, that $p \cdot q$ is a Lorentz invariant object, hence it will have the same value in all rest frames. Now let the frame F be the *rest frame of the four vector p*; then [6] $p^1 = p^2 = p^3 = 0$ and p becomes

$$p = (p^0, 0, 0, 0) \tag{5.72}$$

[6] To this remember that because p is *timelike* it could be the 4 momentum p of some particle moving *less* than the velocity of light, but such a particle has a rest frame, and, in its rest frame, $p = (p^0, 0, 0, 0)$, cf. 5.53.

so we compute that

$$p \cdot q = p^0 q^0, \quad \text{since } F \text{ is the rest frame of } p \tag{5.73}$$

Hence

$$(p+q)^2 = p^2 + q^2 + 2p^0 q^0 \tag{5.74}$$

But, by assumption, p^2, q^2 and p^0 and q^0 are all positive therefore

$$(p+q)^2 > 0 \tag{5.75}$$

and $p + q$ is timelike as claimed.

§ 8. Conservation of energy and momentum

In Newtonian physics we are used to the conservation of both energy E and momentum $\mathbf{p}_N$; this gives us two conservation laws.

Four momentum is conserved

In relativistic physics E and $\mathbf{p}$ are still conserved but their conservation is described by a *single conservation law*. This single conservation law simply states that *four momentum* p is conserved.

In this way we see that the unification of space and time achieved by using Minkowski space has a parallel. This is that the unification of E and $\mathbf{p}$ into the single four vector $p = (p^0, p^1, p^2, p^3)$ allows both energy and momentum to be studied in one four dimensional space[7]: the space of all possible four momenta—geometrically speaking this space is just a copy of Minkowski space.

We shall illustrate four momentum conservation by considering the collision or scattering of two particles cf. figure 14.

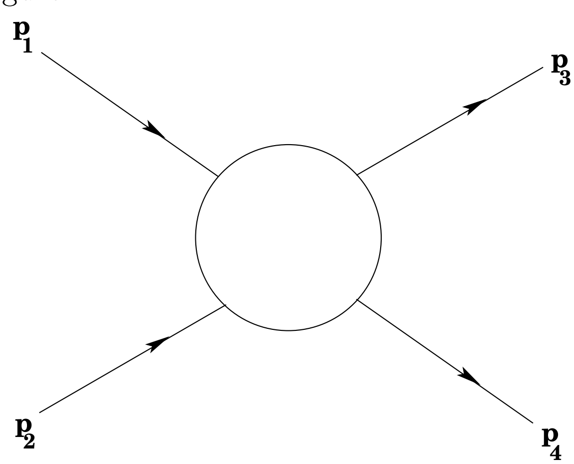

Fig. 14: Two to two particle scattering

[7] Here is a remark purely for readers who know quantum mechanics and how it makes use of the Fourier transform. The Fourier transform $\hat{f}$ of a function $f(t, \mathbf{x})$, in quantum mechanics, is defined by

$$\hat{f}(E, \mathbf{p}) = \frac{1}{(2\pi)^2} \int_{-\infty}^{\infty} e^{(iEt/\hbar)} e^{(i\mathbf{p}\cdot\mathbf{x}/\hbar)} f(t, \mathbf{x})\, dt\, d^3\mathbf{x}$$

Hence one should realise that, E is dual to, or paired with, t, as is $\mathbf{p}$ with $\mathbf{x}$. Hence the unification of t and $\mathbf{x}$ achieved by Minkowski space automatically induces a unification of E and $\mathbf{p}$.

The scattering consists of two initial particles, of four momenta p_1 and p_2; and two final particles, of four momenta p_3 and p_4. The two initial particles collide and scatter to give the two final particles. Conservation of four momentum means that

$$p_1 + p_2 = p_3 + p_4 \tag{5.76}$$

This guarantees that the

$$\begin{aligned} \textit{total initial energy} &= \textit{total final energy} \\ \textit{total initial three momentum} &= \textit{total final three momentum} \end{aligned} \tag{5.77}$$

To see this in more detail we display the components of the four momenta by writing

$$\begin{aligned} p_1 &= (p_1^0,\, \mathbf{p}_1) \qquad & p_3 &= (p_3^0,\, \mathbf{p}_3) \\ p_2 &= (p_2^0,\, \mathbf{p}_2) \qquad & p_4 &= (p_4^0,\, \mathbf{p}_4) \end{aligned} \tag{5.78}$$

So

$$p_1 + p_2 = (p_1^0 + p_2^0,\, \mathbf{p}_1 + \mathbf{p}_2) \quad \text{and} \quad p_3 + p_4 = (p_3^0 + p_4^0,\, \mathbf{p}_3 + \mathbf{p}_4) \tag{5.79}$$

Now, remembering that any four momentum p is given by $p = (E/c, \mathbf{p})$, and applying

$$p_1 + p_2 = p_3 + p_4 \tag{5.80}$$

gives

$$\begin{aligned} p_1^0 + p_2^0 &= p_3^0 + p_4^0 \\ \Rightarrow E_1 + E_2 &= E_3 + E_4, \qquad \textit{energy conservation} \\ \mathbf{p}_1 + \mathbf{p}_2 &= \mathbf{p}_3 + \mathbf{p}_4, \qquad \textit{three momentum conservation} \end{aligned} \tag{5.81}$$

So we do indeed get both energy and three momentum conservation from four momentum conservation.

One can also scatter more than 2 particles together—say n particles—which, after scattering become m particles, and m need not equal n; this is shown in figure 15.

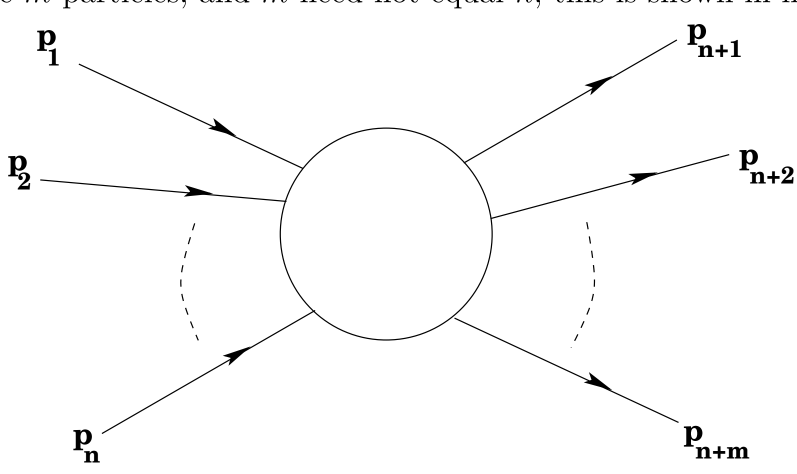

Fig. 15: n **to** m **particle scattering**

To see why m need not equal n remember that mass is not conserved in relativity and energy and mass are inter-convertible. For instance there are many unstable particles, one example being the π^+ meson. The π^+ meson decays into three particles—cf. figure 16—and this process has $n = 1$ and $m = 3$.

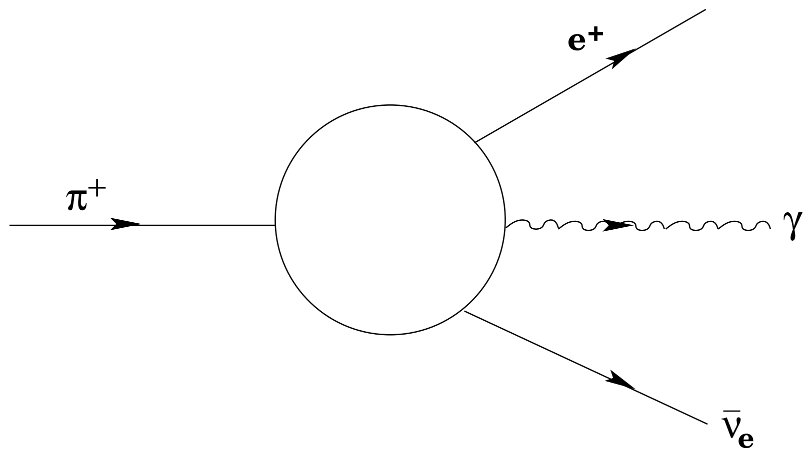

Fig. 16: The decay of a π^+ meson

The three decays products are: a positron, an anti-electron neutrino and a photon and these are denoted by the symbols e^+, $\bar{\nu}_e$ and γ respectively—cf. equation 5.82

$$\pi^+ \longrightarrow \quad e^+ + \bar{\nu}_e + \gamma \tag{5.82}$$

Another process where $n \neq m$ is that of particle production: for example the collision of two protons to produce a π^0 meson, cf. equation 5.83

$$p + p \quad \longrightarrow \quad p + p + \pi^0 \tag{5.83}$$

so that we have $n = 2$ and $m = 3$.

But, whatever the values n and m of the numbers of initial and final particles, the *total four momentum must be conserved.* In other words we must have

$$p_1 + p_2 + \cdots + p_n = p_{n+1} + p_{n+2} + \cdots + p_{n+m} \tag{5.84}$$

§ 9. The centre of mass frame

The Newtonian *centre of mass* of a set of (possibly moving) particles is not very useful in relativity. However we can introduce, not a *distinguished position vector*—which is what the centre of mass is—but a *distinguished frame* known as the *centre of mass frame.* In relativistic kinematics the centre of mass frame assumes the importance previously held by the Newtonian centre of mass. We now explain these remarks.

First recall that, in Newtonian physics, if we have n particles with masses and positions given by m_i $(m_i > 0)$ and $\mathbf{r}_i$ respectively then the centre of mass of this system is given by the vector $\bar{\mathbf{r}}$ where

$$\bar{\mathbf{r}} = \frac{m_1\mathbf{r}_1 + m_2\mathbf{r}_2 + \cdots + m_n\mathbf{r}_n}{m_1 + m_2 + \cdots + m_n} \tag{5.85}$$

However, if we allow each particle to have a velocity v_i, and set $\gamma_i = 1/\sqrt{1 - v_i^2/c^2}$, then each m_i will be replaced in the *relativistic equations of motion* (cf. 5.25) by

$$\gamma_i m_i = \frac{m_i}{\sqrt{1 - \frac{v_i^2}{c^2}}}, \ i = 1, \ldots, n \tag{5.86}$$

Hence, in a relativistic world, a more sensible quantity than $\bar{\mathbf{r}}$ would be obtained by replacing m_i by $\gamma_i m_i$ giving us the vector $\bar{\mathbf{r}}_{rel}$—the relativistic centre of mass—which is defined by

$$\bar{\mathbf{r}}_{rel} = \frac{\gamma_1 m_1 \mathbf{r}_1 + \gamma_2 m_2 \mathbf{r}_2 + \cdots + \gamma_n m_n \mathbf{r}_n}{\gamma_1 m_1 + \gamma_2 m_2 + \cdots + \gamma_n m_n} \tag{5.87}$$

We see, though, that $\bar{\mathbf{r}}_{rel}$ depends on the frame being used and so is not unique.

Now we come to the centre of mass frame: The *centre of mass frame* for a system of n particles is defined to be that frame for which the *total relativistic three momentum* $\mathbf{p}$ *is zero;*[8] i.e. one has

Centre of mass frame

$$\left.\begin{array}{ll} p_1 = (E_1/c, \mathbf{p_1}) \\ p_2 = (E_2/c, \mathbf{p_2}) \\ \vdots & \vdots \\ p_n = (E_n/c, \mathbf{p_n}) \end{array}\right\} \text{ but } \mathbf{p_1} + \mathbf{p}_2 + \cdots + \mathbf{p}_n = 0 \tag{5.88}$$

Notice that, in the centre of mass frame, the *total four momentum* is not zero, rather we have

$$\begin{aligned} p_1 + p_2 + \cdots + p_n &= (\frac{E_1}{c} + \frac{E_2}{c} + \cdots + \frac{E_n}{c}, \mathbf{p_1} + \mathbf{p_2} + \cdots + \mathbf{p}_n) \\ \Rightarrow p_1 + p_2 + \cdots + p_n &= (\sum_{i=1}^{n} \frac{E_i}{c}, 0, 0, 0) \end{aligned} \tag{5.89}$$

So the centre of mass frame is just the rest frame of the four vector p.

Returning to the contrast between the centre of mass and the centre of mass frame, we can now say the following.

If the particles behave like relativistic billiard balls—i.e. they just collide with one another and are subject to no other forces—then, in the centre of mass frame, the relativistic centre of mass, $\bar{\mathbf{r}}_{rel}$, is at rest. However, since relativistic physics allows the very number of particles to change—as we saw in figure 5.82 and 5.83—this billiard behaviour is very often not realised. This results in a consequent decline in importance for the relativistic centre of mass $\bar{\mathbf{r}}_{rel}$ in favour of the centre of mass frame.

Finally some material included just for the curious reader, who may wonder: is there still a centre of mass frame if some, or all, of the p_i in the expression

$$p = p_1 + p_2 + \cdots + p_n \tag{5.90}$$

[8] For such a frame to exist the four momentum p must be timelike. But all the p_i have $p_i^0 > 0$, since p_i^0 is the energy of a particle and must be positive; so all the p_i are future pointing; they are also all timelike since $p_i^2 = m_i^2 > 0$. Thus p, being a sum of timelike, future pointing, four vectors, is indeed timelike—cf. the proof in the example on p. 50.

are those of zero rest mass particles: that is to say some of the p_i are lightlike? The answer—which is easy to verify by redoing through the calculations of the example on p. 50—is that if at *least one* of the p_i is *timelike*, then p is timelike and the centre of mass frame still exists.

However if *all* of the p_i are lightlike then the centre of mass frame *may or may not exist.* For example if

$$n = 2 \tag{5.91}$$

and

$$p = p_1 + p_2, \quad \text{with } p_1^2 = p_2^2 = 0 \tag{5.92}$$

then it is a simple matter to check that, if p_1 is parallel to p_2, i.e.

$$p_1 = \lambda p_2, \ \ \lambda \text{ a constant} \tag{5.93}$$

then p is automatically lightlike and the centre of mass frame does *not exist.* But, if p_1 is *not parallel* to p_2, the centre of mass frame *does exist* because it is then easy to check that $(p_1 + p_2)^2$ is positive.

So a system of just two photons must be non-parallel for there to exist a centre of mass frame; if there are *more than two photons* it is enough that at least two of them are non-parallel for the centre of mass frame to exist.

This last point is easily proved as follows: without loss of generality we can take the two non-parallel photons to have four momenta p_1 and p_2. Then the four vector $p_1 + p_2$ is timelike and we can reuse the argument above that asserts that the centre of mass frame exists if at least one four momentum in the sum giving p is timelike.

§ 10. Threshold energies

In this section we study the creation of new particles by scattering such as in the process 5.83 which, we recall, was

$$p + p \longrightarrow p + p + \pi^0 \tag{5.94}$$

This process is christened π^0 production after the new particle π^0 which is being created.

Suppose that the experimental setup is that one proton is at rest and is bombarded by the other.

We now ask a simple kinematical question which is this: what is the minimum energy needed by the moving proton to be able to create a π^0 meson?

The threshold energy for π^0 production

This minimum energy is then called a *threshold energy* and we shall now calculate it. Denote the (rest) masses of the proton and the π^0 by M and m respectively and let the initial state protons have four momenta p_1 and p_2 where

$$\begin{array}{ll} p_1 = (E/c, p^1, p^2, p^3) & \text{The moving proton} \\ p_2 = (Mc, 0, 0, 0) & \text{The proton at rest} \end{array} \tag{5.95}$$

In the final state let the four momenta be

$$\begin{array}{ll} p_3 & \text{One of the protons} \\ p_4 & \text{The other proton} \\ p_5 & \text{The } \pi^0 \end{array} \tag{5.96}$$

Four momentum conservation says that

$$p_1 + p_2 = p_3 + p_4 + p_5 \tag{5.97}$$

so that it is also true that,

$$(p_1 + p_2)^2 = (p_3 + p_4 + p_5)^2 \tag{5.98}$$

But

$$\begin{aligned}(p_1 + p_2)^2 &= p_1^2 + 2p_1 \cdot p_2 + p_2^2 \\ &= 2M^2c^2 + 2EM\end{aligned} \tag{5.99}$$

using $p_1^2 = p_2^2 = M^2c^2$ and 5.95 Also, for the final state, we have

$$\begin{aligned}(p_3 + p_4 + p_5)^2 &= (p_3 + p_4)^2 + 2(p_3 + p_4) \cdot p_5 + p_5^2 \\ &= p_3^2 + 2p_3 \cdot p_4 + p_4^2 + 2p_3 \cdot p_5 + 2p_4 \cdot p_5 + p_5^2\end{aligned} \tag{5.100}$$

But, again, we know that

$$\begin{aligned}p_3^2 = p_4^2 = M^2c^2, &\quad M \text{ being the proton rest mass} \\ p_5^2 = m^2c^2, &\quad m \text{ being the } \pi^0 \text{ rest mass}\end{aligned} \tag{5.101}$$

so we have

$$(p_3 + p_4 + p_5)^2 = 2M^2c^2 + m^2c^2 + 2(p_3 \cdot p_4 + p_3 \cdot p_5 + p_4 \cdot p_5) \tag{5.102}$$

Now to compute the Minkowski dot products, which are all of the form

$$p_i \cdot p_j \tag{5.103}$$

we go to the rest frame of either p_i or p_j—it doesn't matter which—and for definiteness we shall choose the rest frame of p_i. To this end let

$$\mathbf{v}_{ij} \tag{5.104}$$

denote the velocity of p_j in the rest frame of p_i and let the mass of p_i be denoted by m_i then, in this frame, we have

$$p_i = (m_i c, 0, 0, 0), \quad p_j = (\gamma(v_{ij}) m_j c, \gamma(v_{ij}) m_j \mathbf{v}_{ij}) \tag{5.105}$$

Hence, for the dot product, we immediately calculate that

$$p_i \cdot p_j = \gamma(v_{ij}) m_i m_j c^2 \tag{5.106}$$

Now, using 5.106, we find that

$$(p_3 + p_4 + p_5)^2 = 2M^2c^2 + m^2c^2 + 2(\gamma(v_{34})M^2c^2 + \gamma(v_{35})Mmc^2 + \gamma(v_{45})Mmc^2) \tag{5.107}$$

Thus the statement

$$(p_1 + p_2)^2 = (p_3 + p_4 + p_5)^2 \tag{5.108}$$

of 5.98 becomes

$$2M^2c^2 + 2EM = 2M^2c^2 + m^2c^2 + 2(\gamma(v_{34})M^2c^2 + \gamma(v_{35})Mmc^2 + \gamma(v_{45})Mmc^2) \tag{5.109}$$

and we now want the *minimum* value of E—the incident proton's energy—for which this is true. But the only variables are the $\gamma(v_{ij})$ factors on the RHS; and since all terms on the RHS are positive the minimum is obtained by *minimising* these $\gamma(v_{ij})$, this is trivial to do since the minimum value of any $\gamma(v_{ij})$ is unity which is achieved when $v_{ij} = 0$. Hence the minimum energy is given by setting all the $\gamma(v_{ij})$ to unity in 5.109. This is what we have called the *threshold energy* above and so it is given by E where

$$2M^2c^2 + 2EM = 4M^2c^2 + m^2c^2 + 4Mmc^2 \tag{5.110}$$

from which we compute that

Threshold energy formula

$$E = Mc^2 + 2mc^2 + \frac{m^2}{2M}c^2 \tag{5.111}$$

and this is finally the threshold energy of the proton for π^0 production.

Note that the final state particles, when produced at this threshold energy, have a common rest frame; but they are not at rest relative to the target proton in the laboratory since, by three momentum conservation, they must have some three momentum because the incident proton does.

CHAPTER VI

Relativity, optics and electromagnetism

§ 1. Stellar aberration

When determining the position of stars *stellar aberration* is important. Stellar aberration originates in the the fact that any fixed star observed from the Earth is being observed from a moving object, and this must be taken into account when using light from the star to deduce its position.

The underlying mechanism for stellar aberration is the same as that which makes rain run diagonally down the window on the side of a moving train.

The path of rain on a train window

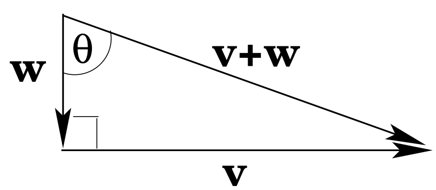

Fig. 17: The path of rain on a train window

Figure 17 shows rain running diagonally down the window of a train at an angle θ to the vertical: the train is moving horizontally with velocity $\mathbf{v}$ and the rain is falling vertically with velocity $\mathbf{w}$; one sees that the angle θ is determined by the equation

$$\tan(\theta) = \frac{v}{w}, \quad \begin{cases} v = |\mathbf{v}| \\ w = |\mathbf{w}| \end{cases} \tag{6.1}$$

As can be seen from figure 17 this mechanism is simply the addition of velocities; however, since the relativistic velocity addition formula differs from the classical one, relativity modifies this phenomenon.

We shall now explain that stellar aberration also results from the addition of velocities—the two velocities being that of light and the orbital velocity of the Earth—so relativity will modify the classical version of this result too.

When one views a fixed star from the Earth one is using a moving telescope: this means that one has to tilt the telescope slightly to centre the star in the telescope. Figure 18 shows two positions of a star according to whether the telescope moves relative to the star or not.

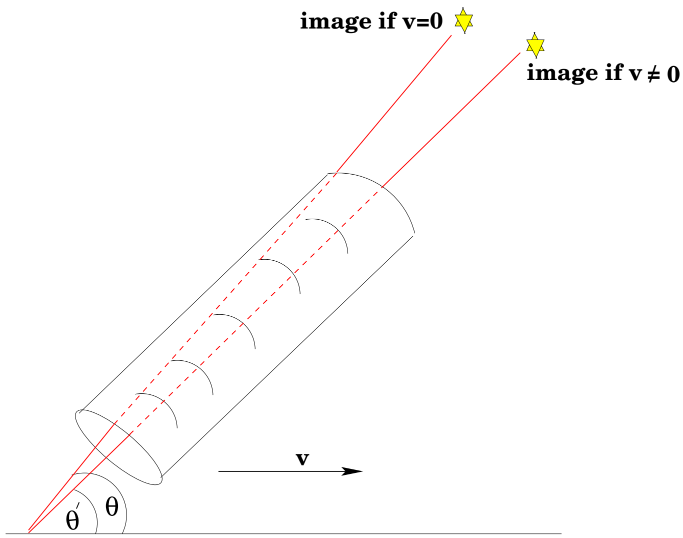

Fig. 18: A moving telescope viewing a star

Stellar aberration

The angular position of the star if the telescope is *stationary* is

$$\theta \tag{6.2}$$

and if the telescope *moves* it is

$$\theta' \tag{6.3}$$

The stellar aberration, or simply the *aberration*, is the difference

$$\theta - \theta' \tag{6.4}$$

We now compute the aberration. As we said above it is a relative velocity phenomenon and so we can calculate what we need from the velocity triangle shown in figure 19.

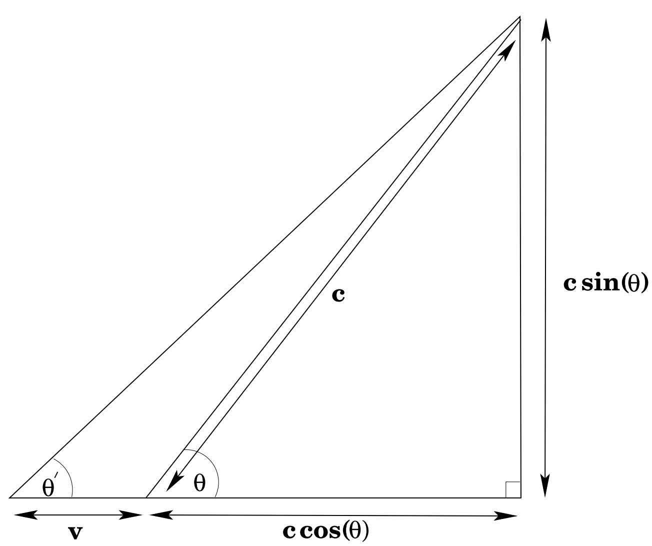

Fig. 19: The Newtonian velocity diagram for stellar aberration

The most convenient object to calculate from figure 19 is $\cot(\theta')$ and we find that

$$\begin{aligned}\cot(\theta') &= \frac{c\cos(\theta)+v}{c\sin(\theta)} \\ &= \frac{\cos(\theta)+\frac{v}{c}}{\sin(\theta)}\end{aligned} \tag{6.5}$$

So the classical, or non-relativistic, aberration formula is

Classical stellar aberration formula

$$\cot(\theta') = \frac{\cos(\theta)+\frac{v}{c}}{\sin(\theta)} \tag{6.6}$$

However we can take advantage of the fact that, since v/c is small [1], $\cot(\theta')$ is very close to $\cot(\theta)$ and there is a useful approximate form of 6.6 which is very accurate. We now derive this approximate form.

All that is needed is to apply Taylor's theorem to $\cot(x)$; doing this we have

$$\cot(x+h) = \cot(x) + h\cot'(x)h + O(h^2) \tag{6.7}$$

Recall that

$$\cot'(x) = -\csc^2(x) = -\frac{1}{\sin^2(x)} \tag{6.8}$$

and set

$$x = \theta, \quad h = \theta' - \theta \tag{6.9}$$

Using this information 6.7 gives

$$\cot(\theta') \approx \cot(\theta) - \frac{\theta'-\theta}{\sin^2(\theta)} \tag{6.10}$$

and, substituting this into our aberration formula 6.6 we find that

$$\begin{aligned}\cot(\theta) - \frac{\theta'-\theta}{\sin^2(\theta)} &= \frac{\cos(\theta)+\frac{v}{c}}{\sin(\theta)} \\ \Rightarrow \cot(\theta) - \frac{\theta'-\theta}{\sin^2(\theta)} &= \cot(\theta) + \frac{v}{c}\frac{1}{\sin(\theta)}\end{aligned} \tag{6.11}$$

which we write as

Approximate classical aberration formula

$$\theta - \theta' = \frac{v}{c}\sin(\theta) \tag{6.12}$$

and this—i.e. 6.12—is a well known formula in non-relativistic astronomy. Using the value $v/c = 9.92 \times 10^{-5}$ which we just quoted in our last footnote, and the fact that $\sin(\theta)$ varies at most between $+1$ and -1, we see that the aberration $\theta - \theta'$ does not exceed [2]

$$\frac{2v}{c} = 19.84 \times 10^{-5}\ \textit{radians} \equiv 40.92\ \textit{seconds of arc} \tag{6.13}$$

[1] Since $v = 29.78\,km/sec$—cf. the footnote on p. 6—and $c = 3 \times 10^8\,km/sec$, the value of v/c is 9.92×10^{-5}.

[2] Recall that one second of arc is 1/3600 of a degree.

a number which is indeed small, as claimed above.

We finish this section by giving the relativistic account of stellar aberration. Fortunately this requires little extra work.

All we need to do is to use the fact that the classical formula followed from velocity addition applied to the telescope and the light ray. More precisely, figure 19 shows that the horizontal velocity components v (which was already horizontal) and $c\cos(\theta)$ add together to give

$$c\cos(\theta) + v \tag{6.14}$$

and then one computes $\cos(\theta')$. Relativity has its own velocity addition formula for two parallel velocities, namely

$$v_{rel} = \frac{v_1 + v_2}{1 + \frac{v_1 v_2}{c^2}} \tag{6.15}$$

so applying this we write

$$v_{rel} = \frac{c\cos(\theta) + v}{1 + \frac{vc\cos(\theta)}{c^2}} \tag{6.16}$$

for the horizontal velocity component of the light ray in the moving frame; but this latter is just

$$c\cos(\theta') \tag{6.17}$$

so we deduce that

$$c\cos(\theta') = \frac{c\cos(\theta) + v}{1 + \frac{vc\cos(\theta)}{c^2}} \tag{6.18}$$

which we rewrite as

Relativistic stellar aberration formula

$$\cos(\theta') = \frac{\cos(\theta) + \frac{v}{c}}{1 + \frac{v}{c}\cos(\theta)} \tag{6.19}$$

and this is the desired *relativistic stellar aberration formula*. We note that it does differ from the classical one of 6.6.

As might be expected the classical and the relativistic formula agree for small v/c: in fact they both reduce to the approximate form 6.12 for small v/c. One way of making this really obvious is too rewrite the relativistic formula 6.19 as a formula for $\cot(\theta')$ rather than for $\cos(\theta')$. We now do this. Evidently we just need $\sin(\theta')$ which we get by using $\sin(x) = \sqrt{1 - \cos^2(x)}$. Hence we find that

$$\begin{aligned}
\sin(\theta') &= \sqrt{1 - \left\{\frac{\cos(\theta) + \frac{v}{c}}{1 + \frac{v}{c}\cos(\theta)}\right\}^2} \\
&= \sqrt{\frac{1 + \frac{2v}{c}\cos(\theta) + \frac{v^2}{c^2}\cos^2(\theta) - \cos^2(\theta) - \frac{2v}{c}\cos(\theta) - \frac{v^2}{c^2}}{\left\{1 + \frac{v}{c}\cos(\theta)\right\}^2}} \\
&= \sqrt{\frac{\{1 - \cos^2(\theta)\} - \frac{v^2}{c^2}\{1 - \cos^2(\theta)\}}{\left\{1 + \frac{v}{c}\cos(\theta)\right\}^2}} = \frac{\sin(\theta)\sqrt{1 - \frac{v^2}{c^2}}}{1 + \frac{v}{c}\cos(\theta)}
\end{aligned} \tag{6.20}$$

Now, combining 6.19 and 6.20, we get

$$\cot(\theta') = \frac{1}{\sqrt{1-\frac{v^2}{c^2}}}\left\{\frac{\cos(\theta)+\frac{v}{c}}{\sin(\theta)}\right\} \tag{6.21}$$

or

Alternative version of relativistic stellar aberration formula

$$\cot(\theta') = \gamma\frac{\cos(\theta)+\frac{v}{c}}{\sin(\theta)}, \quad \gamma = \frac{1}{\sqrt{1-\frac{v^2}{c^2}}}, \quad \text{(relativistic formula)} \tag{6.22}$$

and the non-relativistic formula 6.6 was

$$\cot(\theta') = \frac{\cos(\theta)+\frac{v}{c}}{\sin(\theta)}, \quad \text{(non-relativistic formula)} \tag{6.23}$$

We see that the relativistic one 6.22 only differs by the presence of the multiplicative factor γ; thus to *first order* in v/c they will both agree and reduce to the same formula 6.12 as we said above. This ends our discussion of stellar aberration.

§ 2. The Doppler effect

Suppose a wave be emitted by a source S and received by an observer O; we can then distinguish two cases

(i) The source S and the observer O are both at rest

(ii) There is relative motion between the source S and the observer O.

The *Doppler effect* is the fact that the frequency measured by O is *different* in the two cases.

If the wave is one of sound then this fact is familiar to most of us who have heard an ambulance siren: its frequency rises, as it approaches us, and falls, as it moves away from us; while if the ambulance is stationary the frequency is unchanged.

The Doppler effect also applies to light, and it is light waves which we wish to consider. We shall find below that relativity modifies the classical calculation of the Doppler effect; but it also predicts a *new Doppler effect* where there is none classically: this latter is called the *transverse Doppler effect.*

It will turn out that *time dilation* is responsible for both the modification of the classical result and the existence of the new transverse Doppler effect.

Consider now figure 20 which shows a light source S moving with constant velocity $\mathbf{u}$ along the x axis of a frame F towards and observer O at the origin.

Doppler effect for S moving horizontally towards O

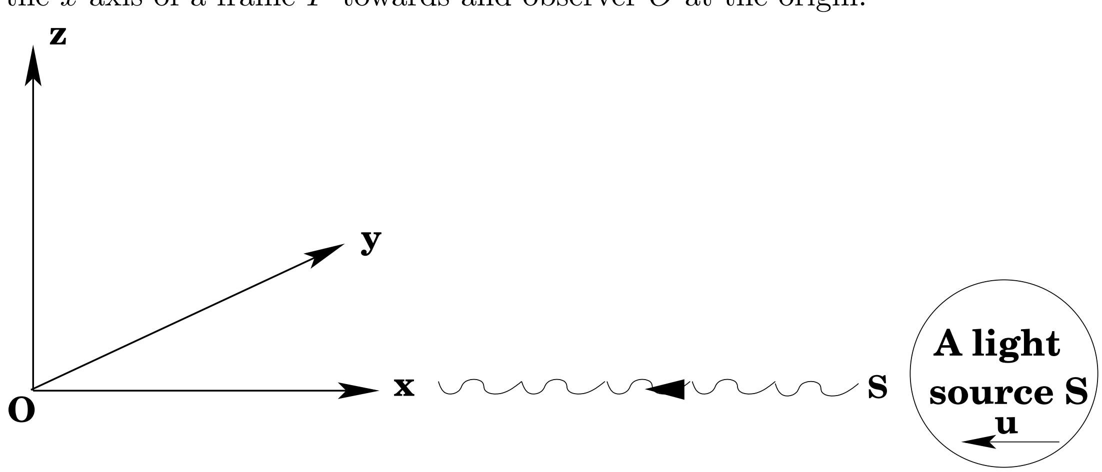

Fig. 20: S is moving with velocity u towards 0

First we reason classically—that is ignoring special relativity. Let the frequency of the light, as measured by S, be

$$\nu \tag{6.24}$$

then S measures each pulse to be t seconds apart, where $t = 1/\nu$. But, during this time interval t, S moves a distance ut nearer to O; and so the pulses they arrive earlier at O by the amount ut/c seconds. Hence O measures the time between pulses to be t_O where

$$t_O = t - \frac{ut}{c} = \left(1 - \frac{u}{c}\right) t \tag{6.25}$$

Let ν_O be the frequency measured by O then $\nu_O = 1/t_O$ so that

$$\nu_O = \frac{1}{\left(1 - \frac{u}{c}\right) t} \tag{6.26}$$

Thus, remembering that $t = 1/\nu$, we find that the frequencies ν measured by S and ν_O measured by O are related by

The non-relativistic Doppler effect

$$\nu_O = \frac{1}{\left(1 - \frac{u}{c}\right)}\nu, \quad u > 0 \tag{6.27}$$

This—i.e. equation 6.27—is the *non-relativistic* or *classical Doppler effect.*

We have assumed that $u > 0$: that is the source S is moving *towards* the observer O and so we deduce that

$$\nu_O > \nu, \quad u > 0 \tag{6.28}$$

If S moves *away* from the observer O, then we change the sign of u in 6.27 giving

$$\nu_O = \frac{1}{\left(1 + \frac{u}{c}\right)}\nu, \quad u < 0 \tag{6.29}$$

and so we will have

$$\nu_O < \nu, \quad u < 0 \tag{6.30}$$

This, of course, agrees with experiment and our experience with ambulance sirens.

Now we must take account of relativity; this is accomplished by simply including time dilation in the above calculation.

When time dilation is included, the observer O measures these pulses as being t'_O seconds apart where t'_O has acquired a factor of $\gamma(u) = 1/\sqrt{1 - u^2/c^2}$. Hence we have

$$\begin{aligned} t'_O &= \gamma(u) t_O \\ &= \frac{t_O}{\sqrt{1 - \frac{u^2}{c^2}}} \end{aligned} \tag{6.31}$$

Hence, with relativity included, the frequency measured by O is ν'_O where

$$\begin{aligned} \nu'_O &= \frac{1}{t'_O} \\ &= \frac{\sqrt{1-\frac{u^2}{c^2}}}{t_O} \\ \Rightarrow \nu'_O &= \sqrt{1-\frac{u^2}{c^2}}\,\nu_O \end{aligned} \tag{6.32}$$

Combining this with 6.27 we obtain the *relativistic Doppler effect* which is

The relativistic Doppler effect

$$\nu'_O = \frac{\sqrt{1-\frac{u^2}{c^2}}}{\left(1-\frac{u}{c}\right)}\nu \tag{6.33}$$

If we expand in powers of u/c we find that

$$\begin{aligned} \nu'_O &= \left(1-\frac{1}{2}\frac{u^2}{c^2}+\cdots\right)\left(1+\frac{u}{c}+\frac{u^2}{c^2}+\cdots\right)\nu \\ &= \left(1+\frac{u}{c}+\frac{1}{2}\frac{u^2}{c^2}+O\left(\frac{u^3}{c^3}\right)\right)\nu \end{aligned} \tag{6.34}$$

and we see that the classical Doppler effect provides the correction $(u/c)\nu$—the *pure Doppler term*—while relativity provides the further correction $(1/2)(u^2/c^2)\nu$, and corrections involving higher powers of u/c if numerically desired.

Note, too, that the relativity correction factor $\sqrt{1-\frac{u^2}{c^2}}$ is *independent* of whether the source S is moving *towards* or *away* from O as it is a function of u^2 rather than u.

§ 3. The transverse Doppler effect

Thus far we have restricted the discussion to the case where the the light source S is moving *horizontally* towards, or away from, the observer O. W now relax this restriction.

To this end let S emit a ray of light towards O but let S have velocity $\mathbf{u}$, where the angle between $\mathbf{u}$ and a line joining S to O is θ—cf. figure 21.

Doppler effect for S moving at an angle θ relative to O

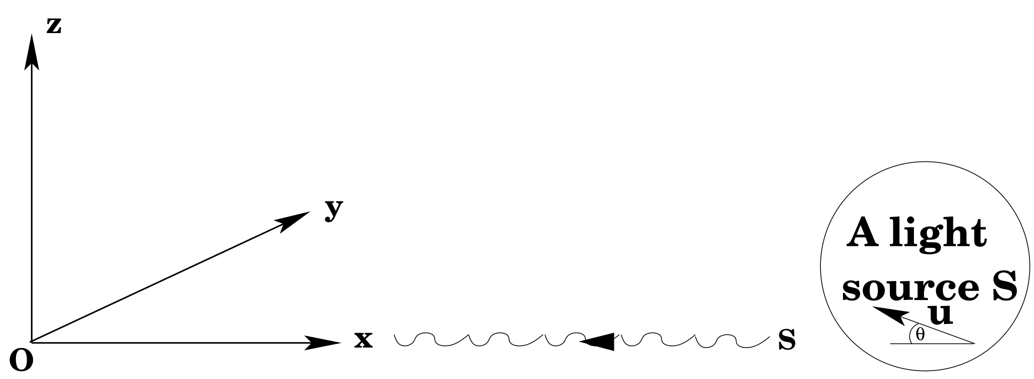

Fig. 21: S is moving with velocity u at an angle θ relative to 0

Provided we replace u by $u\cos(\theta)$ in the *pure Doppler factor* $(1 - u/c)$, and leave the relativistic factor $\sqrt{1 - u^2/c^2}$ unchanged, the calculation of the Doppler effect goes through exactly as before.[3]

Hence our final, fully relativistic, formula is[4]

Doppler effect when S moves at an angle θ relative to O

$$\nu'_O = \frac{\sqrt{1-\frac{u^2}{c^2}}}{\left(1-\frac{u\cos(\theta)}{c}\right)}\nu \tag{6.37}$$

However when

$$\theta = \frac{\pi}{2} \tag{6.38}$$

the classical Doppler term switches off since we then have

$$\left(1 - \frac{u\cos(\theta)}{c}\right) = 1, \quad (\theta = \tfrac{\pi}{2}) \tag{6.39}$$

But we still see that

$$\nu'_O \neq \nu \tag{6.40}$$

for when $\theta = \pi/2$ we have

The transverse Doppler effect

$$\nu'_O = \sqrt{1 - \frac{u^2}{c^2}}\nu \tag{6.41}$$

This is called the *transverse Doppler effect* and is *a purely relativistic phenomenon.*

[3] The point is that what matters for the pure Doppler term is just the component of $\mathbf{u}$ in the direction of O, and this is just $u\cos(\theta)$; while for the relativistic term the factor $\sqrt{1-u^2/c^2}$ just depends on the magnitude u of $\mathbf{u}$ and this is independent of θ

[4] Purely for the really interested we add the following. This angle θ is measured in the rest frame of O. If one does not realise this then confusion can result. The formula 6.37 will change if one uses the angle—call it θ'—measured in the rest frame of S. Using θ', 6.37 is replaced by

$$\nu'_O = \frac{\left(1+\frac{u\cos(\theta')}{c}\right)}{\sqrt{1-\frac{u^2}{c^2}}}\nu \tag{6.35}$$

The two formulae are consistent and one can transform one into the other by using the fact that θ and θ' are related by

$$\cos(\theta') = \frac{\cos(\theta) - \frac{u}{c}}{1 - \frac{u}{c}\cos(\theta)} \tag{6.36}$$

a fact which follows from our stellar aberration formula 6.19 if you set $v = -u$.

In fact, if you look at p. 104 of Einstein's 1905 paper, you will see that Einstein derives the relativistic Doppler formula but obtains the other version of this formula, i.e. equation 6.35 of this footnote. In other words Einstein measures the angle in the rest frame of the source, if you are looking at Einstein's paper you will see that he calls the angle ϕ, you need to know that this angle is $\pi - \theta'$ in our notation.

§ 4. Maxwell's equations in relativistic notation

Maxwell's equations for electric and magnetic fields are invariant under Lorentz transformations although we do not have the time to show that here. However we do intend to quote the equations in a manifestly relativistic notation since this may be useful to the reader elsewhere.

Let $\mathbf{E}$ and $\mathbf{B}$ denote electric and magnetic and fields respectively then recall that Maxwell's four equations are

Maxwell's equations

$$\begin{aligned}\nabla\cdot\mathbf{E} &= \frac{\rho}{\epsilon_0}\\ \nabla\cdot\mathbf{B} &= 0\\ \nabla\times\mathbf{E} &= -\frac{\partial\mathbf{B}}{\partial t}\\ \nabla\times\mathbf{B} &= \frac{1}{\epsilon_0 c^2}\mathbf{J}+\frac{1}{c^2}\frac{\partial\mathbf{E}}{\partial t}\end{aligned} \tag{6.42}$$

where $\mathbf{J}$ and ρ are the current density and charge density respectively.

Now introduce the *scalar potential* ϕ for $\mathbf{E}$ and the *vector potential* $\tilde{\mathbf{A}}$ for $\mathbf{B}$. These potentials obey the equations

The potentials ϕ and $\tilde{\mathbf{A}}$

$$\begin{aligned}&\mathbf{E} = -\nabla\phi-\frac{\partial\tilde{\mathbf{A}}}{\partial t}, \qquad \text{or } E_i = \partial_i\phi-\partial_t\tilde{A}_i,\ i=1,2,3\\ &\mathbf{B} = \nabla\times\tilde{\mathbf{A}}, \qquad \text{or } \begin{cases} B_1 = \partial_2\tilde{A}_3-\partial_3\tilde{A}_2\\ B_2 = \partial_3\tilde{A}_1-\partial_1\tilde{A}_3\\ B_3 = \partial_1\tilde{A}_2-\partial_2\tilde{A}_1\end{cases}\end{aligned} \tag{6.43}$$

It is important to realise that the introduction of these two potentials ϕ and $\tilde{\mathbf{A}}$ automatically generates an $\mathbf{E}$ and a $\mathbf{B}$ which satisfy equations the second and third of Maxwell's four equations. To see this one just substitutes 6.43 into the second and third Maxwell equations: doing this one computes that

$$\begin{aligned}\nabla\cdot\mathbf{B} &= \nabla\cdot\nabla\times\tilde{\mathbf{A}}\\ &= 0, \qquad \text{as required}\end{aligned} \tag{6.44}$$

thus dealing with $\nabla\cdot\mathbf{B}$. While for $\nabla\times\mathbf{E}$ we have

$$\begin{aligned}\nabla\times\mathbf{E} &= \nabla\times(-\nabla\phi-\frac{\partial\tilde{\mathbf{A}}}{\partial t})\\ &= 0-\nabla\times\frac{\partial\tilde{\mathbf{A}}}{\partial t}\\ &= -\frac{\partial(\nabla\times\tilde{\mathbf{A}})}{\partial t}\\ &= -\frac{\partial\mathbf{B}}{\partial t}, \qquad \text{as required}\end{aligned} \tag{6.45}$$

So we have now polished off $\nabla \times \mathbf{E}$.

The key relativistic step is to combine the potentials ϕ and $\tilde{\mathbf{A}}$ into a *single four potential* A^μ whose definition is

$$A^\mu = (\frac{\phi}{c}, \tilde{\mathbf{A}}) \tag{6.46}$$

and so if we lower the index using

The four potential A_μ

$$A_\mu = g_{\mu\nu} A^\nu \tag{6.47}$$

we get

$$A_\mu = (\frac{\phi}{c}, -\tilde{\mathbf{A}}) \tag{6.48}$$

or, spelling it out in full,

$$\begin{aligned} A_0 &= \frac{\phi}{c} \\ A_1 &= -\tilde{A}_1 \\ A_2 &= -\tilde{A}_2 \\ A_3 &= -\tilde{A}_3 \end{aligned} \tag{6.49}$$

Finally define the tensor $F_{\mu\nu}$ by writing

The tensor $F_{\mu\nu}$

$$F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu \tag{6.50}$$

where

$$\partial_\mu = \frac{\partial}{\partial x^\mu} \tag{6.51}$$

Notice that $F_{\mu\nu}$ is *antisymmetric* in μ and ν so that we have

$$F_{\mu\nu} = -F_{\nu\mu} \tag{6.52}$$

It turns out that $F_{\mu\nu}$ contains all the components of the electric and magnetic fields; for, if we calculate $F_{\mu\nu}$, we find that (remember $x^0 = ct$)

$$\begin{aligned} F_{0i} &= \partial_0 A_i - \partial_i A_0 \\ \Rightarrow F_{0i} &= -\frac{1}{c}\partial_t \tilde{A}_i - \frac{1}{c}\partial_i \phi \,, \qquad \text{valid for i=1,2,3} \\ &= \frac{1}{c} E_i \end{aligned} \tag{6.53}$$

and

$$\begin{aligned} F_{ij} &= \partial_i A_j - \partial_j A_i \\ \Rightarrow F_{ij} &= -\partial_i \tilde{A}_j + \partial_j \tilde{A}_i \\ &\Rightarrow \begin{cases} F_{12} = -B_3 \\ F_{31} = -B_2 \\ F_{23} = -B_1 \end{cases} \end{aligned} \tag{6.54}$$

This is enough information to write out $F_{\mu\nu}$ as an *antisymmetric* 4×4 matrix which gives us

The way in which $\mathbf{E}$ *and* $\mathbf{B}$ *are contained in* $F_{\mu\nu}$

$$\begin{aligned} F_{\mu\nu} &= \begin{bmatrix} F_{00} & F_{01} & F_{02} & F_{03} \\ F_{10} & F_{11} & F_{12} & F_{13} \\ F_{20} & F_{21} & F_{22} & F_{23} \\ F_{30} & F_{31} & F_{32} & F_{33} \end{bmatrix} \\ &= \begin{bmatrix} 0 & \frac{E_1}{c} & \frac{E_2}{c} & \frac{E_3}{c} \\ -\frac{E_1}{c} & 0 & -B_3 & B_2 \\ -\frac{E_2}{c} & B_3 & 0 & -B_1 \\ -\frac{E_3}{c} & -B_2 & B_1 & 0 \end{bmatrix} \end{aligned} \tag{6.55}$$

With all this in place, one combines ρ and $\mathbf{J}$ into the single four vector J_μ where

$$J_\mu = (\frac{\rho}{\epsilon_0 c}, \frac{\mathbf{J}}{\epsilon_0 c^2}) \tag{6.56}$$

and Maxwell's remaining two equations take the compact form

$$\partial_\mu F_{\mu\nu} = J_\nu \tag{6.57}$$

Let us check these last two Maxwell equations: Setting $\nu = 0$ in 6.57 we get for the LHS

$$\begin{aligned} \partial_\mu F_{\mu 0} &= \partial_0 F_{00} - \partial_i F_{i0}, \ \text{(the Minkowsi metric generated the minus sign in front of } \partial_i F_{i0}) \\ &= 0 + \partial_i \left(\frac{E_i}{c}\right), \qquad \text{using 6.53} \\ &= \frac{1}{c} \nabla \cdot \mathbf{E} \end{aligned} \tag{6.58}$$

But the RHS of 6.57 with $\mu = 0$ is

$$\frac{\rho}{\epsilon_0 c} \tag{6.59}$$

so the $\nu = 0$ part of 6.57 asserts that

$$\nabla \cdot \mathbf{E} = \frac{\rho}{\epsilon_0} \tag{6.60}$$

which is indeed Maxwell's first equation.

Now setting $\nu = i$, (remember the range of i is $i = 1, 2, 3$) on the LHS of 6.57 we obtain

$$\begin{aligned} \partial_\mu F_{\mu i} &= \partial_0 F_{0i} - \partial_j F_{ji} \\ &= \frac{1}{c} \partial_t \left(\frac{E_i}{c}\right) - \partial_j F_{ji}, \qquad \text{using 6.53} \\ &= \frac{1}{c^2} \left(\frac{\partial \mathbf{E}}{\partial t}\right)_i - (\nabla \times \mathbf{B})_i, \qquad \text{using 6.54 and writing out } \partial_j F_{ji} \text{ carefully} \end{aligned} \tag{6.61}$$

where the suffix i in the last line of 6.61 above denotes the i^{th} component of the vector to which it is applied. But now we see that the RHS of 6.57 with $\nu = i$ is just

$$\frac{J_i}{\epsilon_0 c^2} \tag{6.62}$$

so the $\nu = i$ part of 6.57 asserts that

$$\nabla \times \mathbf{B} = \frac{1}{\epsilon_0 c^2}\mathbf{J} + \frac{1}{c^2}\frac{\partial \mathbf{E}}{\partial t} \tag{6.63}$$

which is correct.

In summary Maxwell's four equations are all satisfied by the choice

Relativistic formulation of Maxwell's equations

$$\begin{aligned} A_\mu &= (\frac{\phi}{c}, -\tilde{\mathbf{A}}) \\ F_{\mu\nu} &= \partial_\mu A_\nu - \partial_\nu A_\mu \\ \partial_\mu F_{\mu\nu} &= J_\nu \end{aligned} \tag{6.64}$$

CHAPTER VII

Experimental tests of relativity

§ 1. Time dilation and length contraction

The existence of time dilation is easily proved by measuring the lifetimes of unstable particles. What one does is to select an unstable particle and measure its lifetime in more than one frame and note that the results differ. A celebrated example of this is provided by the decay of muons as we now explain.

§ 2. Muon decay and time dilation

Introducing the muon

Lifetime and half life of a muon

A muon is an unstable particle of (rest) mass about 210 times that of an electron. When it decays at rest its lifetime t_L and half life $t_{1/2}$ are measured to be [1]

$$\begin{aligned} t_L &= 2.19 \times 10^{-6} \text{ seconds} \\ t_{1/2} &= 1.52 \times 10^{-6} \text{ seconds} \end{aligned} \tag{7.2}$$

But very fast moving muons are available as in the upper atmosphere of the Earth where they form part of the *cosmic rays*: i.e. the collection of particles and radiation of various kinds which arrive continuously from outside the planet. Such muons have been observed with speed v given by

$$\begin{aligned} v &= 2.94 \times 10^8 \; m/sec \\ \Rightarrow v &= 0.98c, \text{ taking } c = 3 \times 10^8 \; m/sec \end{aligned} \tag{7.3}$$

We can immediately calculate that

$$\begin{aligned} \gamma(v) &= \frac{1}{\sqrt{1 - \frac{v^2}{c^2}}} \\ &= 5.02 \end{aligned} \tag{7.4}$$

[1] We remind the reader that the lifetime and half life of unstable nuclei or unstable particles are always related by the simple equation

$$t_{1/2} = \ln(2) t_L \tag{7.1}$$

Hence we only need to know one of the two quantities t_L and $t_{1/2}$ and can use equation 7.1 to obtain the other.

Hence, by time dilation, the lifetime of the fast muons measured by a ground based observer should be

Time dilation

$$\gamma(v)t_L \tag{7.5}$$

which we find has the value

Lifetime of a fast muon extended by time dilation

$$5.02 \cdot 2.19 \times 10^{-6} = 10.99 \times 10^{-6} \text{ seconds} \tag{7.6}$$

So the muon should live about five times longer than it does when measured at rest: This is what is found by experiments—cf. the papers quoted in the footnote on p. 74—and is a striking confirmation of time dilation.

§ 3. Muon decay and length contraction

One can also regard this experiment as an indirect confirmation of *length contraction* as we now explain below.

Let us go to the rest frame of the muon as it travels down towards the Earth's surface. Let an Earth based observer O_{Earth} measure the muons at a height L above the Earth's surface where

$$L = 10\,km \tag{7.7}$$

and measure them again at the Earth's surface.

Well, an observer O_{muon}, in the rest frame of the muon, will claim that the Earth has moved, with velocity 2.94×10^8 m/sec towards the muon a total distance *not of* $10\,km$, but a distance of

Length contraction seen in the muon rest frame

$$\begin{aligned}\frac{L}{\gamma(v)} &= \frac{10}{5.02}\,km \\ &= 1.99\,km\end{aligned} \tag{7.8}$$

Also O_{muon} will see no time dilation and will measure the muon lifetime t_L to have the normal value given by

$$t_L = 2.19 \times 10^{-6} \text{ seconds} \tag{7.9}$$

We summarise the situation by pointing out that

$$\begin{aligned} O_{Earth} \text{ sees } &\begin{cases} \text{Time dilation sending } t_L \text{ to } \gamma(v)t_L \\ \textit{No length contraction} \end{cases} \\ O_{muon} \text{ sees } &\begin{cases} \textit{No time dilation} \\ \text{Length contraction sending } L \text{ to } \dfrac{L}{\gamma(v)} \end{cases} \end{aligned} \tag{7.10}$$

Both observers are happy that their measurements are entirely consistent; and we can make this even more convincing by calculating the number of muon decays each observer sees. Let us now do this.

§ 4. O_{Earth} and O_{muon} measure the same number of muon decays

Suppose N muons begin travelling downwards at speed v, a distance L so as to reach the Earth's surface. Remember also that their population halves every half life number of seconds.

Now observer O_{Earth} says that the journey takes

$$\frac{L}{v} \text{ seconds} \tag{7.11}$$

and O_{Earth}, who also sees *time dilation*, says that the half life is $\gamma(v)t_{1/2}$. So, dividing by this half life, he says that the muon journey takes

$$\frac{L}{v\gamma(v)t_{1/2}} \text{ half lives} = p \text{ say} \tag{7.12}$$

So he says that the number of muons has halved p times and is therefore now

$$\frac{N}{2^p}, \qquad \text{where } p = \frac{L}{v\gamma(v)t_{1/2}} \tag{7.13}$$

This means that the number of muons that have decayed should be

$$N - \frac{N}{2^p}, \qquad (p = \frac{L}{v\gamma(v)t_{1/2}}) \tag{7.14}$$

Number of decays measured by O_{Earth}

In contrast observer O_{muon} sees length contraction and claims that the distance travelled by the Earth, at speed v, towards the muons is only

$$\frac{L}{\gamma(v)} \tag{7.15}$$

and so the time for the muons to touch the Earth surface is, he claims,

$$\frac{L}{v\gamma(v)} \tag{7.16}$$

O_{muon} sees *no time dilation* and asserts that the half life is just

$$t_{1/2} \tag{7.17}$$

and so, on division by $t_{1/2}$, he says that the Earth journey takes

$$\frac{L}{v\gamma(v)t_{1/2}} \text{ half lives} = q \text{ say} \tag{7.18}$$

Thus O_{muon} says that the muon population is

$$\frac{N}{2^q}, \qquad \text{where } q = \frac{L}{v\gamma(v)t_{1/2}} \tag{7.19}$$

and so the number of decayed muons is

$$N - \frac{N}{2^q}, \qquad (q = \frac{L}{v\gamma(v)t_{1/2}}) \tag{7.20}$$

Number of decays measured by O_{muon}

But we observe that

$$p = q \tag{7.21}$$

O_{Earth} and O_{muon} agree on the number of decays

and so both observers will measure the same number of decays as indeed they should.

Do note, however, that this number of decays which is

$$N - \frac{N}{2^p}, \quad \text{with } p = \frac{L}{v\gamma(v)t_{1/2}} \tag{7.22}$$

is *very different* from that predicted without special relativity because relativity inserts the quantity $\gamma(v)$ into the formula 7.22. Hence these measurements do constitute a striking verification of a relativistic law—a verification which can be viewed as one of *time dilation* if measured by O_{Earth}; or *length contraction* if measured by O_{muon}.

§ 5. A worked example

We finish this muon discussion by putting in some numbers to show what happens. All we have to do is to specify N since the values of the other quantities have already been quoted above. Let us take

$$N = 1 \text{ million particles} \tag{7.23}$$

then the number of decays according to relativity is

$$\begin{aligned} 10^6 - \frac{10^6}{2^p} \quad &\text{with } p = \frac{10^4}{2.94 \times 10^8 \cdot 5.02 \cdot 1.52 \times 10^{-6}} = 4.45 \\ &= 10^6\left(1 - \frac{1}{2^{4.45}}\right) \\ &= 954,247.32 \quad \text{decays} \end{aligned} \tag{7.24}$$

leaving 45,752 surviving muons.

Relativity says 45,752 muons reach the Earths' surface

However, *without relativity*, the value of p would be

$$\begin{aligned} \frac{L}{vt_{1/2}} &= \frac{10^4}{2.94 \times 10^8 \cdot 1.52 \times 10^{-6}} \\ &= 22.33 \end{aligned} \tag{7.25}$$

so that the number of decays $N - N/2^p$ becomes

$$10^6\left(1 - \frac{1}{2^{22.33}}\right) = 999,999.98 \quad \text{decays} \tag{7.26}$$

Non-relativistic physics says no muons reach the Earths' surface

leaving 0.02 of a surviving muon—hence all the particles have decayed.

Thus the predictions of relativity are vastly different and experiment[2] has decided in favour of relativity.

Experiment agrees with relativity

§ 6. Gravitational time dilation

General relativity[3] predicts that a clock is slowed by being in a gravitational field: the stronger the field the more the clock is slowed down. More precisely we have the following formula.

If t_0 the clock time with zero gravitational field and t the time in a gravitational field then, for the Earth, the two clock times are related by the equation

$$t = \frac{t_0}{\sqrt{1 - \frac{2GM}{Rc^2}}} \qquad \left(\begin{array}{l} G = \text{Newton's constant} = 6.672 \times 10^{-11}\ m^2 kg^{-1} s^{-2} \\ R = \text{ distance of clock from centre of the Earth} \\ M = \text{ mass of the Earth} = 5.973 \times 10^{24}\ kg \\ c = \text{ speed of light} = 3 \times 10^8\ ms^{-1} \end{array} \right) \tag{7.27}$$

The gravitational red shift

This phenomenon is also called the *gravitational red shift* because it means that the frequencies of emitted radiation from atoms is shifted towards the red end of the spectrum by the gravitational field. Knowledge of the gravitational red shift is vital for astronomers when observing radiation from stars.

This gravitational effect is *quite distinct* from the time dilation of special relativity which requires a moving clock; gravitational time dilation occurs for a clock *at rest* in a gravitational field.

Gravitational effects due to general relativity are very small for the Earth when compared with those of a star like the Sun—because of the Sun's much greater mass—nevertheless they have been experimentally confirmed. Here are some details.

Experiment *The scout-D rocket experiment*

In 1976 the Smithsonian Astrophysical laboratory sent a rocket to an altitude of $10,000\,km$ and then allowed it to fall back to the Earth for two hours while simultaneously transmitting pulses from a *hydrogen maser* to the surface. This maser oscillates at a frequency ν which is accurate to 1 part in 10^{16}; but ν should vary during the fall.

Also, bearing in mind that $\nu = t^{-1}$, then formula 7.27 shows us that the maser frequencies ν and ν_0, corresponding to t and t_0, will obey the formula

$$\frac{\nu}{\nu_0} = \sqrt{1 - \frac{2GM}{Rc^2}} \tag{7.28}$$

[2] If the reader is interested, experimental details can be found in the following two papers: Rossi B. and Hall D. B., *On muon time dilation*, Phys. Rev. **59**, 223–, (1941) and Bailey, J., Borer, K., Combley, F., Drumm H., Krienen F., Lange F., Picasso E., von Ruden W., Farley F. J. M., Field J. H., Flegel W., and Hattersley P. M., *Measurements of relativistic time dilation for positive and negative muons in a circular orbit*, Nature, **268**, 301–305, (1977).

[3] We do not require any previous knowledge of general relativity here; we just quote the formulae that we need and the rest is surprisingly simple.

If we denote the radius of the Earth by R_E, and the altitude of the rocket by H, then $R = R_E + H$ so that, *at its highest point*, the frequency is ν_H with

$$\frac{\nu_H}{\nu_0} = \sqrt{1 - \frac{2GM}{Rc^2}}, \quad \text{with } R = R_E + H \tag{7.29}$$

while, *at the Earth's surface*, we set $R = R_E$ and the frequency is ν_S with

$$\frac{\nu_S}{\nu_0} = \sqrt{1 - \frac{2GM}{R_E c^2}} \tag{7.30}$$

Hence the frequency on the Earth's surface and at an altitude H are related by the equation

$$\frac{\nu_H}{\nu_S} = \frac{\sqrt{1 - \dfrac{2GM}{(R_E + H)c^2}}}{\sqrt{1 - \dfrac{2GM}{R_E c^2}}}, \quad \text{with } H = 10,000\, km \tag{7.31}$$

But since

$$R_E = 6.378 \times 10^3 \; km = 6.378 \times 10^6 \, m, \quad H = 10,000 \; km = 10^7 \, m \tag{7.32}$$

so one easily computes that

A very successful gravitational time dilation experiment

$$\frac{\nu_H}{\nu_S} = 1.000000000423 \tag{7.33}$$

that is a shift of *just 4.23 parts in* 10^{10} in the maser frequency—in other words the maser at height $10,000$ km has a slightly higher frequency than the one on the Earth's surface. This was very satisfactorily confirmed to within 0.01% by the experiment [4]

Here is another famous gravitational red shift experiment.

Experiment *The Harvard tower experiment*

In the Harvard tower [5] experiment a frequency shift is observed and measured when the source and receiver atoms are *only* 22.6 *metres apart* rather than $10,000$ kilometres. This means that one uses the formula of 7.31 with $H = 22.6\, m$ giving us

$$\frac{\nu_H}{\nu_S} = \frac{\sqrt{1 - \dfrac{2GM}{(R_E + H)c^2}}}{\sqrt{1 - \dfrac{2GM}{R_E c^2}}}, \quad \text{with } H = 22.6\, m \tag{7.34}$$

[4] The results of the experiment are described in: Vessot R. F. C. and Levine M. W., *A test of the equivalence principle using a space-borne clock*, J. Gen. Rel. and Grav., **10**, 181–204, (1979).

[5] The phrase *Harvard tower* and the height 22.6 metres are due to the fact that 22.6 metres was the distance within an elevator shaft in the Jefferson Tower physics building in Harvard used for the experiment.

and we find that
$$\frac{\nu_H}{\nu_S} = 1.0000000000000002460000 \tag{7.35}$$

This is an *incredibly small* frequency shift of *2.46 parts in* 10^{15} and it is truly amazing that it is measurable. The measurement was made possible by using what is called *the Mössbauer effect.*[6] The successful experiment, confirming general relativity to within 1%, was done in 1960 by Pound and Rebka.[7]

Pound and Rebka's successful experiment

In the actual experiment a photon in the γ-ray region is emitted from a source crystal at a height of 22.6 m and allowed to *fall* in the Earth's gravitational field to the target where it should be absorbed via the Mössbauer mechanism if its frequency has not shifted. Pound and Rebka made the source oscillate up and down—thus adding a large Doppler shift and thereby periodically creating and destroying the conditions for absorption—they also had to take account of temperature effects and differences between the source and target crystals.

§ 7. Combined special and general relativistic time dilation

If one takes a very accurate atomic clock and places it on an airplane flying above the Earth's surface then it should tick at a different rate to an identical Earth bound clock. Moreover there are *two* sources for this rate difference: one special relativistic and one general relativistic. These are simply

(i) The special relativistic time dilation given by the factor

$$\frac{1}{\sqrt{1-\frac{v^2}{c^2}}}, \qquad \text{where } v \text{ is the horizontal speed of the airplane} \tag{7.36}$$

(ii)

Special and general relativity both at work

The general relativistic red shift time dilation due to the factor

$$\sqrt{1-\frac{2GM}{(R_E+H)c^2}}, \qquad \text{where } H \text{ is the altitude of the airplane} \tag{7.37}$$

Thus when such an airplane lands its clock should differ from the one of the ground by an amount which is calculable from both general and special relativity. Such an experiment was done by Haefele and Keating in 1971 and we now provide a brief description.

[6] Some information about the the Mössbauer effect may be of interest to the curious reader. When an atom of substance A, say, emits a γ-ray it recoils, thereby reducing somewhat the energy of the γ-ray; this photon has therefore too low a frequency to be reabsorbed by another atom of substance A.

However, Mössbauer discovered, that if substance A is a crystal at a *low enough temperature* then any recoil energy must be taken up by the whole crystal and, if the "would be recoil energy" is *lower than the lowest vibrational energy state (i.e. what is called the lowest phonon energy)* of the crystal then it *cannot be taken up at all* and *no energy is lost to a recoil on photon emission.* This means that reabsorption *is now possible* and, when it occurs, it is called resonant absorption. This phenomenon is referred to as the *the Mössbauer effect.*

[7] The results are published in the following short, but very important, paper: Pound R. V. and Rebka Jr. G. A.., *Apparent weight of photons*, Phys. Rev. Lett., **4**, 337–341 (1960)

Experiment *The Haefele and Keating experiment*

This experiment was done in 1971 and involved placing four Caesium atomic clocks on a commercial jet airliner and flying it *twice round the world:* once *eastwards* and once *westwards*. The clocks *did differ* from a ground based clock—the results[8] were roughly as follows (all quoted times are in nanoseconds):

	Eastward Journey	**Westward Journey**	
Kinematic (Special theory)	-184 ± 18	$+96 \pm 18$	
Gravitational (General theory)	$+144 \pm 14$	$+179 \pm 18$	(7.38)
Total predicted effect	-40 ± 23	$+275 \pm 21$	
Total measured effect	-59 ± 10	$+273 \pm 21$	

The twin paradox explanation confirmed

Happily we see agreement between theory[9] and experiment within the experimental errors. Note that this experiment constitutes a confirmation of our account of the twin paradox given in chapter 4.

§ 8. Verification of $E = mc^2$: nuclear power and stellar energy

The first experimental verification of the equation $E = mc^2$ was by Cockroft and Walton in 1932 when they split the lithium nucleus.[10]

We now want to remind the reader of some basic matters relevant to atomic nuclei and the understanding of the equation $E = mc^2$.

If we take the nucleus of any stable atom then the mass of the nucleus is always *less* than the sum of the masses of its individual protons and neutrons. This mass difference is convertible into an energy using $E = mc^2$, and this energy is called the *binding energy* of the nucleus.

Here is an example.

Example *The binding energy of Helium*

Take the ordinary stable Helium nucleus ^4He which consists of two protons and two neutrons—this is also known as an α-particle. The mass of the ^4He, nucleus is M where

$$M = 4.001506179\, u \tag{7.39}$$

where u denotes what are called *unified atomic mass units* which are defined by

$$1\, u = 1.66053873 \times 10^{-27}\, kg \tag{7.40}$$

[8] The reference for this work is: Haefele J. C. and Keating R. E., *Around-the-world atomic clocks: Predicted relativistic time gains*, Science, **177**, 166–170, (1972).

[9] We do not show how to calculate the relativistic effects though this is quite easy; rather we conserve our efforts till we come to the much more interesting subject of *global positioning by satellites, or GPS* (cf. § 9) where we do give the details.

[10] Their work appeared in the paper: Cockroft J. D. and Walton E. T. S., *Experiments with high velocity positive ions II: the disintegration of elements by high velocity protons*, Proc. Roy. Soc. London, **A137**, 229–, (1932). Walton was Irish and Cockroft English. Cockroft and Walton received the 1951 Nobel prize for physics for this work.

Now if m_p and m_n denote the masses of a single proton and neutron respectively then we know that

$$m_p = 1.007276466\,u, \quad m_n = 1.008664915\,u \tag{7.41}$$

So the sum of the masses of the four nucleons of ^4He is

$$2m_p + 2m_n = (2 \cdot 1.007276466 + 2 \cdot 1.008664915)\,u = 4.031882762\,u$$

But this is *more* than M the mass of ^4He, the difference being ΔM—and known as the *mass defect*—where

Binding energy, $E = mc^2$ and and the mass defect

$$\Delta M = (4.031882762 - 4.001506179)\,u = 0.030376583\,u \tag{7.42}$$

So the binding energy is now ΔE where

$$\Delta E = \Delta M c^2 \tag{7.43}$$

giving us, for ^4He, a binding energy of

$$\Delta E = 0.030376583 \times 1.66053873 \times 10^{-27} \times (3 \times 10^8)^2\,J = 4.53597 \times 10^{-12}\,J \tag{7.44}$$

We have given ΔE above in Joules, but it is also common to use electron volts, which are denoted by eV; the relation between the two sets of units is that

$$1\,eV = 1.60217 \times 10^{-19}\,J \tag{7.45}$$

It is also very useful to know that

$$1\,eV/c^2 = 1.7826 \times 10^{-36}\,kg \tag{7.46}$$

In any case, in electron volts, we find that the binding energy of ^4He is given by

$$\begin{aligned} \Delta E &= \frac{4.53597 \times 10^{-12}}{1.60217 \times 10^{-19}}\,eV = 2.8334 \times 10^7\,eV \\ \text{or} \qquad \Delta E &= 28.33\,Mev \end{aligned} \tag{7.47}$$

where 1 Mev denotes 10^6 eV.

This binding energy ΔE is the energy that must be supplied to overcome the strong nuclear forces and break it up into its four constituent nucleons.

Binding energy is the energy that gets released in nuclear fission and fusion via the following mechanism. Let a nucleus have N nucleons and corresponding mass

$$M_N \tag{7.48}$$

If its binding energy is ΔE, then the binding energy *per nucleon* is

$$\frac{\Delta E}{N} \tag{7.49}$$

The crucial point is that the binding energy per nucleon $\Delta E/N$ is *not a constant* but is a function of N—cf. the sketch comprising figure 22.

Binding energy per nucleon not constant

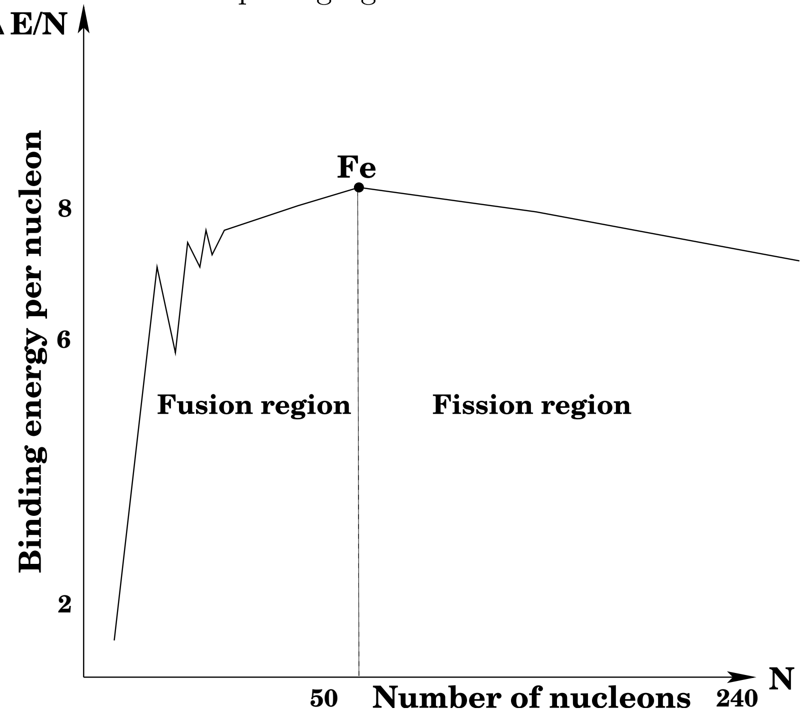

Fig. 22: The binding energy per nucleon in the periodic table

Now if we take the nucleus of an element A and divide it into two pieces, then, depending on where A is in the periodic table, the sum of masses of the two pieces may be *bigger or smaller* than the mass of the nucleus of A.

In the *smaller* case an A nucleus can release energy by splitting up into the two pieces: this is called *nuclear fission*, alternatively, if the *bigger* case pertains, then energy will be released by joining two such pieces together to form an A nucleus: this is called *nuclear fusion*.

In each case—fission or fusion—the final nucleus or nuclei have a *higher* binding energy than the initial nucleus or nuclei. The net increase in energy ΔE is related to the mass difference Δm by

$$\Delta E = \Delta m\, c^2 \tag{7.50}$$

ΔE is then the energy released in the fission or fusion process.

We see a maximum in the graph at the element iron: ^{56}Fe. What this means is that, for the lighter elements, to the left of the dotted line in figure 22, one tends to have *fusion*; while, for the heavier elements, to the right of the dotted line one tends to have *fission*.

Example *A fission reaction*

Uranium has an unstable radioactive isotope $^{235}_{92}U$ (mass 235.04392 u) which can undergo fission when hit by a slow, or thermal, neutron. One has a reaction of the form

$$n + {}^{235}_{92}U \longrightarrow A + B + n + n \tag{7.51}$$

where n denotes a neutron and A and B are the two main fission fragments. There are many possibilities for A and B but one example is where $A = {}^{90}_{37}Rb$ (Rubidium, mass 89.914813 u) and $B = {}^{144}_{55}Cs$ (Caesium, mass 143.93202 u) giving

A typical Uranium fission reaction

$$n + {}^{235}_{92}U \longrightarrow {}^{90}_{37}Rb + {}^{144}_{55}Cs + n + n \tag{7.52}$$

Remembering that the neutron mass is $1.008664915\,u$ the values for the total initial and final state masses $M_{initial}$ and M_{final} are

$$\begin{aligned} M_{initial} &= 1.008664915 + 235.04392 = 236.0525849 \\ M_{final} &= 2 \cdot 1.008664915 + 89.914813 + 143.93202 = 235.8641628 \end{aligned} \tag{7.53}$$

The mass difference $\Delta M = M_{initial} - M_{final}$ is thus given by [11]

$$\Delta M = 236.0525849 - 235.8641628 = 0.190956 \tag{7.54}$$

and so the fission energy release $\Delta M\, c^2$ is

$$0.190956\, c^2 \tag{7.55}$$

and this is equal to

Fission energy yield

$$178.121\, MeV \tag{7.56}$$

if we use the conversion information of 7.40 and 7.46; and we note that this is an energy yield of $178.121/236 = 0.754\, MeV$ per nucleon.

Example *A fusion reaction*

If we take the two heavy hydrogen isotopes deuterium ${}^2_1H$ (mass 2.0141017 u) and tritium ${}^3_1H$ (mass 3.0160293 u) then they can combine in a fusion reaction giving helium ${}^4_2He$ (mass 4.0026032 u) plus a neutron. Thus we have

A fusion reaction favoured by fusion reactor designers

$${}^2_1H + {}^3_1H \longrightarrow {}^4_2He + n \tag{7.57}$$

and for $M_{initial}$ and M_{final} we find [12]

$$\begin{aligned} M_{initial} &= 2.0141017 + 3.0160293 = 5.030131 \\ M_{final} &= 4.0026032 + 1.008664915 = 5.0112681 \end{aligned} \tag{7.58}$$

Thus for $\Delta M = M_{initial} - M_{final}$ we find

$$\Delta M = 0.0188629 \tag{7.59}$$

[11] Note that the masses I am using here for Uranium, Caesium and Rubidium are not the masses of the respective *nuclei* but the *atomic masses*—i.e. the electron masses are included. This would matter were it not for the fact that the electron masses exactly cancel in the difference ΔM—thus the calculation of the energy yield is indeed correct despite the fact that $M_{initial}$ and M_{final} should really have been the *nuclear masses.* The reason for using atomic masses in the first place is because they are much more readily accessible than nuclear masses: recall that they are the ones that are quoted in periodic tables of the elements.

[12] Again we are using *atomic masses* instead of *nuclear masses* here but this does not matter, cf. the footnote to the previous example.

giving a fusion energy release of

$$0.0188629\,c^2 \tag{7.60}$$

which, in MeV, is

Fusion energy yield

$$17.595\,MeV \tag{7.61}$$

and this is an energy yield of $17.95/5 = 3.5\,MeV$ per nucleon—i.e. considerably bigger than the fission yield quoted above. The above reaction is one favoured by the designers of *fusion reactors* which may be realised in the future.

We provide one more example for those interested in astrophysics.

Example *Stellar fusion*

There are three main *stellar fusion* processes that fuel the stars depending on the age A, core temperature T and mass M: the proton-proton cycle ($T < 10^6\,K$), the carbon cycle (larger M and $T > 10^6\,K$) and the helium cycle (larger A, centre collapsing and $T > 10^8\,K$).

In brief the proton-proton cycle looks like this (e^+ denotes a positron and ν_e an electron-neutrino):

The stellar proton-proton fusion reaction

$$\begin{aligned} {}^1_1H + {}^1_1H &\longrightarrow {}^2_1H + e^+ + \nu_e \\ {}^2_1H + {}^1_1H &\longrightarrow {}^3_2He \\ {}^3_2He + {}^3_2He &\longrightarrow {}^4_2He + {}^1_1H + {}^1_1H, \qquad \text{(energy yield about } 25\,MeV) \end{aligned} \tag{7.62}$$

§ 9. Global positioning by satellites: GPS

The GPS system consists of the use of 24 satellites to obtain the position of objects on, or above, the Earth. The position of the object—a person, a ship, a building etc—is a *three dimensional position* i.e. height as well as latitude and longitude are given; it is also *very accurate indeed* yielding positions to within an error of only *several metres*.

The 24 satellites are divided into six groups of four and each satellite contains a Caesium atomic clock whose accuracy must be at least as good as 1 nanosecond per second. Each group of four lie in a common orbital plane and the six orbital planes are all at 55^0 to the Earth's equator. Finally the orbital period of each satellite is 12 hours which determines[13] its distance from the centre of the Earth to be $26608\,km$—we assume a circular orbit for simplicity which is a fair assumption, but see our remarks on p. 83 for details on what happens without this assumption.

These 24 satellites give global coverage so that a receiver on or near the surface can always communicate with about four satellites. These four satellites continually transmit

[13] This distance is computed from Kepler's third law which says that the orbital period T and the radius of the orbit R are related by the equation

$$R = \left(\frac{GMT^2}{4\pi^2}\right)^{1/3} \tag{7.63}$$

where G is Newton's constant and M is the mass of the Earth; their values are given in 7.27.

their position and their time and allow the receiver to deduce its own position to a great accuracy.

Both special and general relativistic effects are large, and their inclusion is crucial to obtaining the correct receiver positions. We now calculate both of these relativistic effects and show just how important[14] they are.

The special relativistic effect comes from the factor

$$\frac{1}{\sqrt{1-\frac{v^2}{c^2}}}, \qquad \text{where } v \text{ is speed of the satellite} \tag{7.64}$$

Now if ω is the satellite's angular velocity then

$$\omega = \frac{2\pi}{12} \text{ rads per hour} \equiv 1.45 \times 10^{-4} \text{rads per sec} \tag{7.65}$$

and since $v = R\omega$ we compute that

$$\begin{aligned} v &= 26.608 \times 10^6 \cdot 1.45 \times 10^{-4}\, m/sec \\ &= 3858.16\, m/sec \end{aligned} \tag{7.66}$$

Hence we find that

$$\begin{aligned} \frac{1}{\sqrt{1-\frac{v^2}{c^2}}} &= \frac{1}{\sqrt{1-\frac{(3858.16)^2}{(3\times 10^8)^2}}} \\ &= 1.00000000008269 \end{aligned} \tag{7.67}$$

This represents a change of 0.82 parts in 10^{10} and, a moving clock ticks slow, so the satellite clock must be corrected *upwards* by

Special relativistic correction

$$0.082 \text{ nanoseconds per second} \tag{7.68}$$

Thus, *over a whole day*, the clock would be wrong by

$$24 \cdot 3600 \cdot 0.082 \times 10^{-9} = 7.08 \times 10^{-6} \text{ secs} \tag{7.69}$$

So much for the correction due to special relativity.

Now, coming to general relativity, we know that the clock will tick *faster* in the satellite due to the weaker gravitational field at 26608 km from the Earth's centre. Thus the general relativistic correction is *opposite* to the special relativistic one. The correction is given, as we saw in 7.31 by the quantity

General relativistic correction

[14] One should also take into account the fact that the Earth is rotating and so a surface clock is in a rotating frame; this does affect matters and is known as the *Sagnac effect* but it is numerically small and, though we shall not include it here, it is included in a full treatment of the GPS corrections.

$$\frac{\sqrt{1-\dfrac{2GM}{Rc^2}}}{\sqrt{1-\dfrac{2GM}{R_E c^2}}}, \quad \text{with } R = 2.6608\times 10^6 \tag{7.70}$$
$$= 1.0000000005278$$

and this is a change of 5.27 parts in 10^{10} which we note is about 6.4 times *bigger* than the special relativistic correction. Also, *over a whole day*, this effect would make the clock be fast by

$$24\cdot 3600\cdot 5.27\times 10^{-10} = 45.6\times 10^{-6}\text{ secs} \tag{7.71}$$

Finally the *total effect* on the satellite clock in a day, *if uncorrected*, would be

$$(-7.08+45.6)\times 10^{-6} = 38.52\times 10^{-6}\text{ secs} \tag{7.72}$$

Hence the satellite clock must be corrected to run *slower* by

Combined special and general relativistic correction

$$\begin{aligned}(-0.82+5.27)\times 10^{-10} &= 4.45\times 10^{-10}\text{ seconds per second}\\ &= 0.445\text{ nanoseconds per second}\end{aligned} \tag{7.73}$$

which is the same as 38.52×10^{-6} secs per day.

Note that the consequence of *not* applying this correction would be *disastrous*: for example, if left uncorrected for a whole day, the clock would be out by

$$38.52\times 10^{-6}\,secs \tag{7.74}$$

but the speed of light is $3\times 10^8\,m/sec$ and so, in $38.52\times 10^{-6}\,secs$, light travels

$$3\times 10^8\cdot 38.52\times 10^{-6} = 11556\,m \tag{7.75}$$

and so a position determination would be out by

Possible position error without relativity

$$11556\,metres \equiv 11.55\,km \tag{7.76}$$

a quite unacceptable amount, rendering the whole GPS scheme useless.

Non circularity of satellite orbits a small matter

We now comment, as promised, on what happens if one takes account of the fact that the GPS satellite orbits are not circular. The GPS satellite orbits, though not all the same shape, are all close to being circular, their eccentricity is about 0.01 or less. However the fact that they are not exactly circular means that the distance of the satellites from the centre the Earth varies slightly over a single orbit. This causes a periodic change in clock rate whose size, for an eccentricity of 0.01, is about 23 nanoseconds over a 12 hour orbit. This can be taken account of fairly easily. We note that 23 nanoseconds is nothing like the size of the corrections that have to applied because of relativity over a 12 hour period which are three orders of magnitude bigger.[15]

GPS receivers: inexpensive verifiers of special and general relativity

Nowadays handheld GPS receivers cost as little as €200 and they are used for navigation on airliners and ships, as well as by climbers, hill walkers and for many other purposes; they can even estimate speed of movement as well as three dimensional position.

When the first satellites were launched in 1977 some people doubted the necessity for relativistic corrections but these doubts were very quickly dispelled. Thus we can say, without any exaggeration, that GPS receivers are a very useful, commonplace and cheap device in which incorporation of both *special and general relativity* are essential to their operation.

§ 10. Bending of light by a gravitational field

Light bends in gravitational field

Relativity predicts that light bends in a gravitational field. This famous prediction of relativity was confirmed in 1919 during a famous expedition to photograph a total eclipse of the sun. The amount of the bending is small, but it was measured for light, originating in a certain star, and passing close to the Sun. The amount of the bending was found to be precisely as predicted by Einstein and caused quite a sensation in the public and scientific world of the time. The figure below shows the bending; the size θ of the deflection is about 1.75 seconds of arc.

[15] If one sticks to the circularity approximation then one can obtain quite a neat formula for the combined special and general relativistic corrections as follows: Using 7.64 and 7.70 we see that the correction we have calculated in 7.73 above is given by the combined formula

$$\sqrt{1-\frac{v^2}{c^2}}\,\frac{\sqrt{1-\dfrac{2GM}{Rc^2}}}{\sqrt{1-\dfrac{2GM}{R_Ec^2}}} \tag{7.77}$$

But we know that, for a circular orbit, $T = 2\pi/\omega$ and so Kepler's formula 7.63 tells us that

$$R^3 = \frac{GM}{\omega^2} \tag{7.78}$$

and since $v = R\omega$ we have

$$v^2 = \frac{GM}{R} \tag{7.79}$$

giving $\sqrt{1-v^2/c^2} = \sqrt{1-GM/Rc^2}$. Hence our combined formula 7.77 becomes the rather neat expression

$$\frac{\sqrt{1-\dfrac{GM}{Rc^2}}\sqrt{1-\dfrac{2GM}{Rc^2}}}{\sqrt{1-\dfrac{2GM}{R_Ec^2}}} \tag{7.80}$$

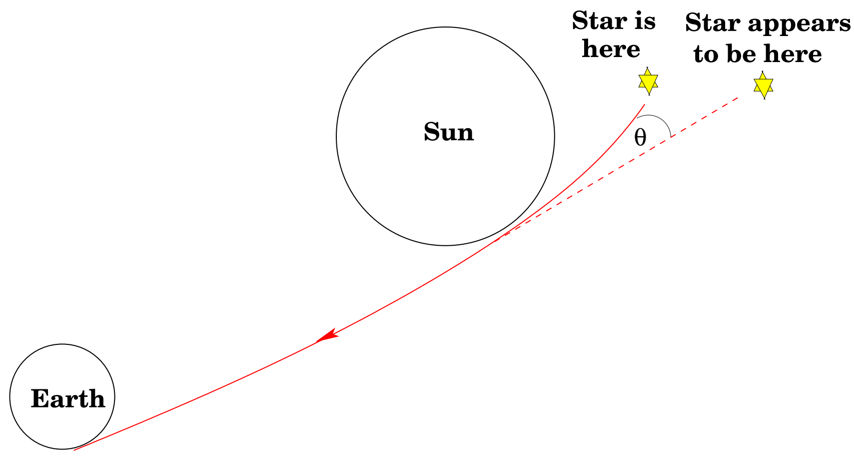

Fig. 23: The bending of starlight by the Sun

This was an early but very important experimental success for general relativity as it was a *prediction* and no one had ever thought that such a thing could happen before.

§ 11. Advance of the perihelion of Mercury

Before the advent of general relativity the orbit of the planet mercury—the nearest planet to the Sun—had been a puzzle to scientists for some considerable time. Newtonian gravity could not explain it properly. The elliptical orbit of mercury rotates very, very slowly as it goes round the Sun; this is shown in figure 24 below.

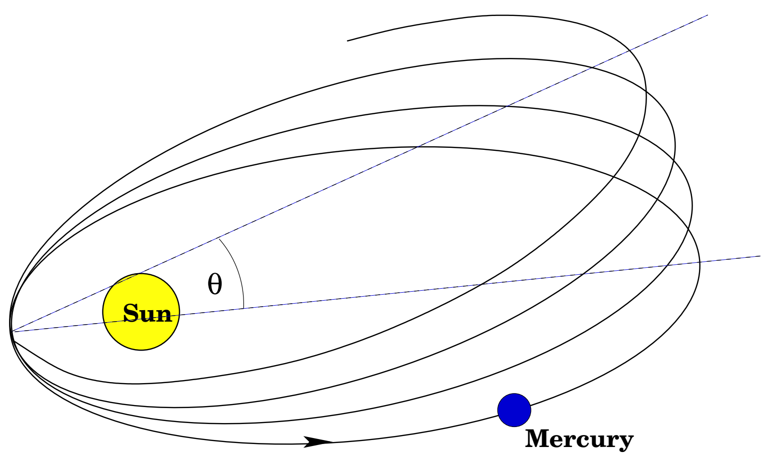

Fig. 24: The orbit of Mercury around the Sun

Relativity successfully predicts the size of Mercury's perihelion advance

This phenomenon is referred to as the *advance of the perihelion* of mercury and it is measured by the angle θ shown in figure 24. The measured value of θ, which is accurately known, is about 5750 seconds of arc per century.

Newtonian gravity predicts such an advance of the perihelion—the advance being caused by the gravitational fields of neighbouring planets—but the value of θ obtained by the Newtonian calculation is only about 5707 seconds of arc; and this is therefore 43 seconds of arc too small.

Relativity successfully predicts the extra 43 seconds of arc for the perihelion advance that Newtonian gravity failed to do; this was one of the earliest experimental successes of general relativity.

§ 12. Gravitational radiation

Relativity predicts gravitational radiation

The photon is a zero rest mass object which is the quantum of electromagnetic radiation. Physicists also believe that there is a gravitational analogue of the photon called the *graviton*; the graviton also has zero rest mass and is the quantum of *gravitational radiation*. General relativity predicts the existence of gravitational radiation; indeed gravitational radiation should be emitted from an accelerating mass in the same way as electromagnetic radiation is emitted from an accelerating charged particle. So a successful experimental measurement of such radiation would be another experimental confirmation of relativity.

Unfortunately, while the photon and electromagnetic radiation do exist and are easily measured, the existence of gravitons and gravitational radiation has not yet been confirmed experimentally.

Gravitons, quantum gravity, string theory and M theory

A proper discussion of *gravitons* requires a *quantum theory of gravity*; such a theory is still under construction but may finally emerge from what is called *string theory* or a related theory called *M theory*; however we must now leave the subject because it is so far outside the scope of our current discussion.

However, returning to *gravitational radiation* we want to show that its existence is strongly supported by very striking indirect astronomical evidence. This evidence is provided by measurements of the orbits of a *binary pulsar* by Hulse and Taylor.

Pulsars and neutron stars

We now give a brief account of their work. A *pulsar* is a rotating *neutron star*. The neutron star [16], which is also very strongly magnetised, has a magnetic axis that does not coincide with its rotational axis. Nearby electrons get magnetically trapped and spiral in the magnetic field thereby emitting radiation along the direction of magnetic axis cf. figure 25

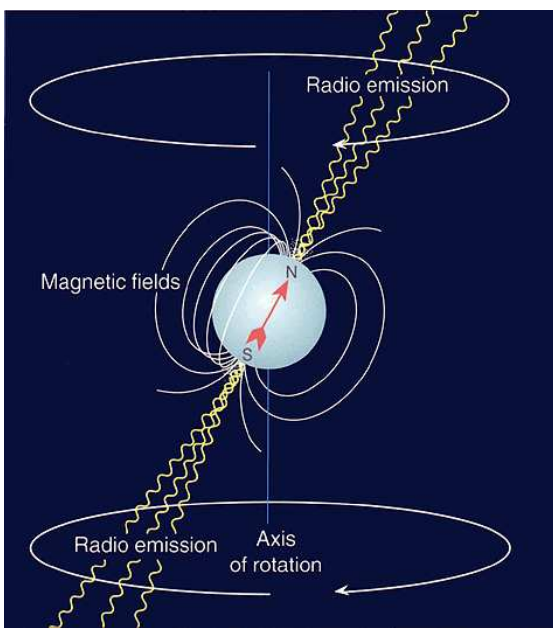

Fig. 25: A pulsar and its radio beacon

[16] A neutron star is a collapsed star which is too light to have collapsed to form a black hole (for black holes cf. § 13) but which has all its electrons and protons crushed together by its strong gravitational field so that is consists entirely of neutrons.

This fact, together with its rotation, causes the neutron star to emit a beam of electromagnetic radiation which sweeps around in a circular manner similar to a lighthouse beacon. This radiation can be detected on Earth as a pulse whose frequency determined by the rate of rotation of the neutron star—hence the name *pulsar.*

The frequency of these pulses can be constant to an enormously high accuracy and some pulsars have a companion star so that the pair orbit round their common centre of mass. Such a pulsar is called a *binary pulsar* and this is the type we are concerned with here.

Hulse and Taylor detected a pulsar in 1974—usually referred to as $PSR1913+16$—which they deduced was a binary pulsar by a variation in its pulse frequency caused by its common orbiting of the companion star. The orbital period of about 7.75 hours could also be deduced.

Orbital slowing caused by emission of gravitational radiation

The pulsar $PSR1913+16$ was kept (and is still kept) under continuous observation and its orbital period was found to be *slowing* at the rate of about 76×10^{-6} *sec* per year. General relativity predicts that such a slowing—which also means that the orbit is shrinking in size—would occur due to the loss of energy [17] from the binary system due to gravitational radiation. Furthermore relativity predicts *very precisely* the amount of slowing and this is in very good agreement indeed with the experimentally measured amount. Figure 26 shows a picture of the binary system.

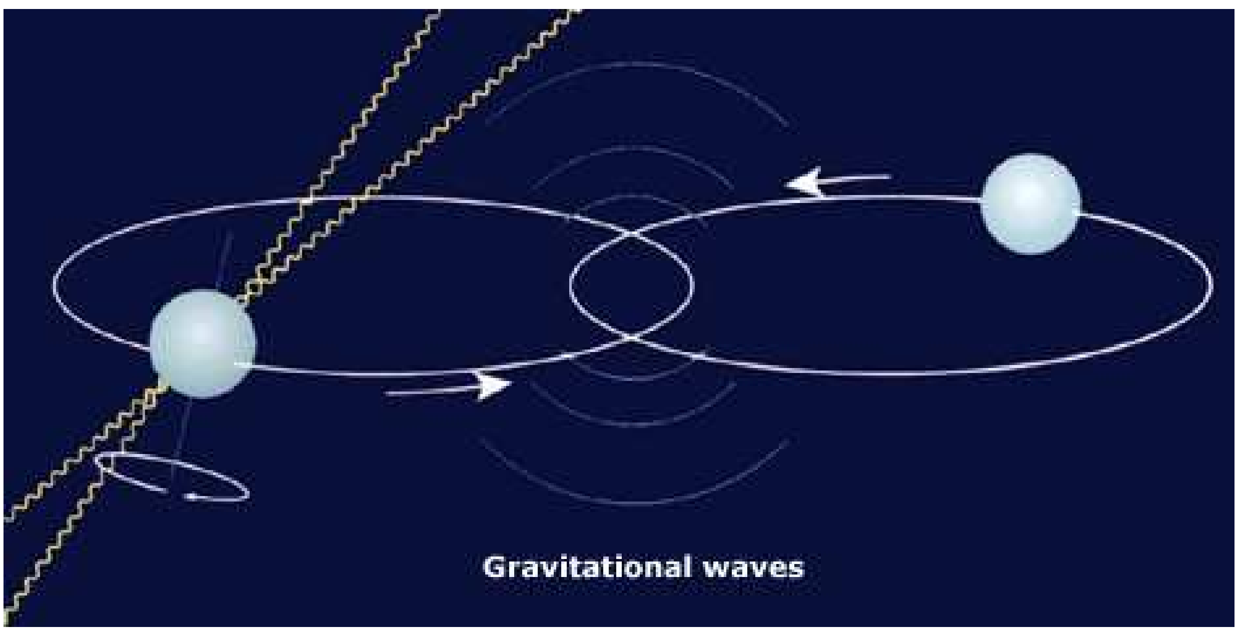

Fig. 26: The binary pulsar PSR1913+16

Hulse and Taylor were awarded the 1993 physics Nobel prize for this work which is regarded as excellent indirect evidence for the existence of gravitational radiation and of course provides yet another piece of experimental evidence for the correctness of general relativity.

[17] Given enough time—about 350 million years in this case—the orbit will have shrunk so much that the two stars should coalesce.

The LIGO gravitational radiation detection project

Direct observation of gravitational radiation may be possible with a special laser interferometry experiment involving lasers thousands of kilometres apart; one such planned experiment is called *LIGO* which stands for *Laser Interferometry Gravitational Wave Observatory.*

§ 13. Cosmology, the big bang and black holes

Black holes

General relativity also predicts that some stars may collapse to form what are called *black holes*. Black holes have very intriguing properties such as their *event horizon:* a mathematical surface surrounding the black hole which if crossed from the outside can *never be recrossed* thus trapping the object that did the crossing. Also nothing inside the event horizon—whether it was there when the black hole formed or came in afterwards—can ever get out not even light. The gravitational field is also *singular* at the centre of the black hole.

Black holes contain a singularity

Here is a brief sketch of some of the salient features of gravitational collapse.

Gravitational collapse is something that is worth investigating for very massive objects such as stars. This simple sounding idea is that, for a sufficiently massive body, the attractive force of gravity may be strong enough to cause it to start to *implode*.

To find something massive enough we have to choose a stellar object such as a star. Now, for a young active star, the burning of the nuclear fuel causes enough outward pressure to counteract all its gravitational inward pressure. However, since the nuclear fuel will eventually be used up this line of thought suggests that one calculate what gravity can do once it is not opposed by the nuclear reactions.

Black hole formation

When a large star has used up all its nuclear fuel it explodes into a supernova. The remains after the supernova explosion then collapse down to form the extremely dense object that is a neutron star. However if the mass M of these remains is bigger than about twice the mass of our Sun the collapse *does not stop* at the neutron star stage—nothing can overcome the gravitational forces—and a black hole is formed.

Evidence for black holes

Experimental evidence is accumulating very much in favour of some astronomical objects being black holes; candidates include certain X-rays sources such as Cygnus X-1, whose X rays are thought to be emitted by matter falling into a black hole; and incredibly massive objects weighing as much as 1.2×10^9 solar masses but occupying a volume only the size of our solar system—an example, found by the *Hubble space telescope*, being the core of galaxy $NGC4261$ which is thought to be too dense to be anything other than a black hole.

The big bang

The entire present day Universe may have originated in a past singularity known as the *big bang*; a possibility for which there is considerable experimental evidence nowadays. This has resulted in the big bang being taken very seriously.

Both the big bang and black holes are natural products of general relativity; and so the mounting experimental evidence for them is yet more experimental evidence for general relativity.

CHAPTER VIII

Einstein's two papers of 1905

§ 1. Einstein's first 1905 paper

In this section we provide an English translation of Einstein's original paper which appeared in German, the precise German reference being: *Zur Elektrodynamik bewegter Körper (On the electrodynamics of moving bodies),* Annalen der Physik, **322**, 891–921, (1905).

ON THE ELECTRODYNAMICS OF MOVING BODIES

By A. EINSTEIN

June 30, 1905

It is known that Maxwell's electrodynamics—as usually understood at the present time—when applied to moving bodies, leads to asymmetries which do not appear to be inherent in the phenomena. Take, for example, the reciprocal electrodynamic action of a magnet and a conductor. The observable phenomenon here depends only on the relative motion of the conductor and the magnet, whereas the customary view draws a sharp distinction between the two cases in which either the one or the other of these bodies is in motion. For if the magnet is in motion and the conductor at rest, there arises in the neighbourhood of the magnet an electric field with a certain definite energy, producing a current at the places where parts of the conductor are situated. But if the magnet is stationary and the conductor in motion, no electric field arises in the neighbourhood of the magnet. In the conductor, however, we find an electromotive force, to which in itself there is no corresponding energy, but which gives rise—assuming equality of relative motion in the two cases discussed—to electric currents of the same path and intensity as those produced by the electric forces in the former case.

Examples of this sort, together with the unsuccessful attempts to discover any motion of the earth relatively to the "light medium," suggest that the phenomena of electrodynamics as well as of mechanics possess no properties corresponding to the idea of absolute rest. They suggest rather that, as has already been shown to the first order of small quantities, the same laws of electrodynamics and optics will be valid for all frames of reference for which the equations of mechanics hold good. [1] We will raise this conjecture (the purport of which will hereafter be called the "Principle of Relativity") to the status of a postulate, and also introduce another postulate, which is only apparently irreconcilable with the former, namely, that light is always propagated in empty space with a definite velocity c which is independent of the state of motion of the emitting body. These two postulates suffice for the attainment of a simple and consistent theory of the electrodynamics of moving bodies based on Maxwell's theory for stationary bodies. The introduction of a "luminiferous ether" will prove to be superfluous inasmuch as the view here to be developed will not require an "absolutely stationary space" provided with special properties, nor assign a velocity-vector to a point of the empty space in which electromagnetic processes take place.

The theory to be developed is based—like all electrodynamics—on the kinematics of the rigid body, since the assertions of any such theory have to do with the relationships between rigid bodies (systems of co-ordinates), clocks, and electromagnetic processes. Insufficient consideration of this circumstance lies at the root of the difficulties which the electrodynamics of moving bodies at present encounters.

I. KINEMATICAL PART

§ 1. Definition of Simultaneity

Let us take a system of co-ordinates in which the equations of Newtonian mechanics hold good. [2] In order to render our presentation more precise and to distinguish this system of co-ordinates verbally from others which will be introduced hereafter, we call it the "stationary system."

If a material point is at rest relatively to this system of co-ordinates, its position can be defined relatively thereto by the employment of rigid standards of measurement and the methods of Euclidean geometry, and can be expressed in Cartesian co-ordinates.

If we wish to describe the *motion* of a material point, we give the values of its co-ordinates as functions of the time. Now we must bear carefully in mind that a mathematical description of this kind has no physical meaning unless we are quite clear as to what we understand by "time." We have to take into account that all our judgments in which time plays a part are always judgments of *simultaneous events*. If, for instance, I say, "That train arrives here at 7

[1] The preceding memoir by Lorentz was not at this time known to the author.

[2] i.e. to the first approximation.

o'clock," I mean something like this: "The pointing of the small hand of my watch to 7 and the arrival of the train are simultaneous events." [3]

It might appear possible to overcome all the difficulties attending the definition of "time" by substituting "the position of the small hand of my watch" for "time." And in fact such a definition is satisfactory when we are concerned with defining a time exclusively for the place where the watch is located; but it is no longer satisfactory when we have to connect in time series of events occurring at different places, or—what comes to the same thing—to evaluate the times of events occurring at places remote from the watch.

We might, of course, content ourselves with time values determined by an observer stationed together with the watch at the origin of the co-ordinates, and co-ordinating the corresponding positions of the hands with light signals, given out by every event to be timed, and reaching him through empty space. But this co-ordination has the disadvantage that it is not independent of the standpoint of the observer with the watch or clock, as we know from experience. We arrive at a much more practical determination along the following line of thought.

If at the point A of space there is a clock, an observer at A can determine the time values of events in the immediate proximity of A by finding the positions of the hands which are simultaneous with these events. If there is at the point B of space another clock in all respects resembling the one at A, it is possible for an observer at B to determine the time values of events in the immediate neighbourhood of B. But it is not possible without further assumption to compare, in respect of time, an event at A with an event at B. We have so far defined only an "A time" and a "B time." We have not defined a common "time" for A and B, for the latter cannot be defined at all unless we establish *by definition* that the "time" required by light to travel from A to B equals the "time" it requires to travel from B to A. Let a ray of light start at the "A time" t_{A} from A towards B, let it at the "B time" t_{B} be reflected at B in the direction of A, and arrive again at A at the "A time" t'_{A}.

In accordance with definition the two clocks synchronize if

$$t_{\mathrm{B}} - t_{\mathrm{A}} = t'_{\mathrm{A}} - t_{\mathrm{B}}.$$

We assume that this definition of synchronism is free from contradictions, and possible for any number of points; and that the following relations are universally valid:—

1. If the clock at B synchronizes with the clock at A, the clock at A synchronizes with the clock at B.

2. If the clock at A synchronizes with the clock at B and also with the clock at C, the clocks at B and C also synchronize with each other.

Thus with the help of certain imaginary physical experiments we have settled what is to be understood by synchronous stationary clocks located at different places, and have evidently obtained a definition of "simultaneous," or

[3] We shall not here discuss the inexactitude which lurks in the concept of simultaneity of two events at approximately the same place, which can only be removed by an abstraction.

"synchronous," and of "time." The "time" of an event is that which is given simultaneously with the event by a stationary clock located at the place of the event, this clock being synchronous, and indeed synchronous for all time determinations, with a specified stationary clock.

In agreement with experience we further assume the quantity

$$\frac{2\mathrm{AB}}{t'_A - t_A} = c,$$

to be a universal constant—the velocity of light in empty space.

It is essential to have time defined by means of stationary clocks in the stationary system, and the time now defined being appropriate to the stationary system we call it "the time of the stationary system."

§ 2. On the Relativity of Lengths and Times

The following reflexions are based on the principle of relativity and on the principle of the constancy of the velocity of light. These two principles we define as follows:—

1. The laws by which the states of physical systems undergo change are not affected, whether these changes of state be referred to the one or the other of two systems of co-ordinates in uniform translatory motion.

2. Any ray of light moves in the "stationary" system of co-ordinates with the determined velocity c, whether the ray be emitted by a stationary or by a moving body. Hence

$$\text{velocity} = \frac{\text{light path}}{\text{time interval}}$$

where time interval is to be taken in the sense of the definition in § 1.

Let there be given a stationary rigid rod; and let its length be l as measured by a measuring-rod which is also stationary. We now imagine the axis of the rod lying along the axis of x of the stationary system of co-ordinates, and that a uniform motion of parallel translation with velocity v along the axis of x in the direction of increasing x is then imparted to the rod. We now inquire as to the length of the moving rod, and imagine its length to be ascertained by the following two operations:—

(a) The observer moves together with the given measuring-rod and the rod to be measured, and measures the length of the rod directly by superposing the measuring-rod, in just the same way as if all three were at rest.

(b) By means of stationary clocks set up in the stationary system and synchronizing in accordance with § 1, the observer ascertains at what points of the stationary system the two ends of the rod to be measured are located at a definite time. The distance between these two points, measured by the measuring-rod already employed, which in this case is at rest, is also a length which may be designated "the length of the rod."

In accordance with the principle of relativity the length to be discovered by the operation (a)—we will call it "the length of the rod in the moving system"—must be equal to the length l of the stationary rod.

The length to be discovered by the operation (b) we will call "the length of the (moving) rod in the stationary system." This we shall determine on the basis of our two principles, and we shall find that it differs from l.

Current kinematics tacitly assumes that the lengths determined by these two operations are precisely equal, or in other words, that a moving rigid body at the epoch t may in geometrical respects be perfectly represented by *the same* body *at rest* in a definite position.

We imagine further that at the two ends A and B of the rod, clocks are placed which synchronize with the clocks of the stationary system, that is to say that their indications correspond at any instant to the "time of the stationary system" at the places where they happen to be. These clocks are therefore "synchronous in the stationary system."

We imagine further that with each clock there is a moving observer, and that these observers apply to both clocks the criterion established in § 1 for the synchronization of two clocks. Let a ray of light depart from A at the time[4] t_{A}, let it be reflected at B at the time t_{B}, and reach A again at the time t'_{A}. Taking into consideration the principle of the constancy of the velocity of light we find that

$$t_{\mathrm{B}} - t_{\mathrm{A}} = \frac{r_{\mathrm{AB}}}{c-v} \text{ and } t'_{\mathrm{A}} - t_{\mathrm{B}} = \frac{r_{\mathrm{AB}}}{c+v}$$

where r_{AB} denotes the length of the moving rod—measured in the stationary system. Observers moving with the moving rod would thus find that the two clocks were not synchronous, while observers in the stationary system would declare the clocks to be synchronous.

So we see that we cannot attach any *absolute* signification to the concept of simultaneity, but that two events which, viewed from a system of co-ordinates, are simultaneous, can no longer be looked upon as simultaneous events when envisaged from a system which is in motion relatively to that system.

§ 3. Theory of the Transformation of Co-ordinates and Times from a Stationary System to another System in Uniform Motion of Translation Relatively to the Former

Let us in "stationary" space take two systems of co-ordinates, i.e. two systems, each of three rigid material lines, perpendicular to one another, and issuing from a point. Let the axes of X of the two systems coincide, and their axes of Y and Z respectively be parallel. Let each system be provided with a rigid measuring-rod and a number of clocks, and let the two measuring-rods, and likewise all the clocks of the two systems, be in all respects alike.

[4] "Time" here denotes "time of the stationary system" and also "position of hands of the moving clock situated at the place under discussion."

Now to the origin of one of the two systems (k) let a constant velocity v be imparted in the direction of the increasing x of the other stationary system (K), and let this velocity be communicated to the axes of the co-ordinates, the relevant measuring-rod, and the clocks. To any time of the stationary system K there then will correspond a definite position of the axes of the moving system, and from reasons of symmetry we are entitled to assume that the motion of k may be such that the axes of the moving system are at the time t (this "t" always denotes a time of the stationary system) parallel to the axes of the stationary system.

We now imagine space to be measured from the stationary system K by means of the stationary measuring-rod, and also from the moving system k by means of the measuring-rod moving with it; and that we thus obtain the co-ordinates x, y, z, and ξ, η, ζ respectively. Further, let the time t of the stationary system be determined for all points thereof at which there are clocks by means of light signals in the manner indicated in § 1; similarly let the time τ of the moving system be determined for all points of the moving system at which there are clocks at rest relatively to that system by applying the method, given in § 1, of light signals between the points at which the latter clocks are located.

To any system of values x, y, z, t, which completely defines the place and time of an event in the stationary system, there belongs a system of values ξ, η, ζ τ, determining that event relatively to the system k, and our task is now to find the system of equations connecting these quantities.

In the first place it is clear that the equations must be *linear* on account of the properties of homogeneity which we attribute to space and time.

If we place $x' = x - vt$, it is clear that a point at rest in the system k must have a system of values x', y, z, independent of time. We first define τ as a function of x', y, z, and t. To do this we have to express in equations that τ is nothing else than the summary of the data of clocks at rest in system k, which have been synchronized according to the rule given in § 1.

From the origin of system k let a ray be emitted at the time τ_0 along the X-axis to x', and at the time τ_1 be reflected thence to the origin of the co-ordinates, arriving there at the time τ_2; we then must have $\frac{1}{2}(\tau_0 + \tau_2) = \tau_1$, or, by inserting the arguments of the function τ and applying the principle of the constancy of the velocity of light in the stationary system:—

$$\frac{1}{2}\left[\tau(0,0,0,t) + \tau\left(0,0,0,t + \frac{x'}{c-v} + \frac{x'}{c+v}\right)\right] = \tau\left(x',0,0,t + \frac{x'}{c-v}\right).$$

Hence, if x' be chosen infinitesimally small,

$$\frac{1}{2}\left(\frac{1}{c-v} + \frac{1}{c+v}\right)\frac{\partial\tau}{\partial t} = \frac{\partial\tau}{\partial x'} + \frac{1}{c-v}\frac{\partial\tau}{\partial t},$$

or

$$\frac{\partial \tau}{\partial x'} + \frac{v}{c^2 - v^2} \frac{\partial \tau}{\partial t} = 0.$$

It is to be noted that instead of the origin of the co-ordinates we might have chosen any other point for the point of origin of the ray, and the equation just obtained is therefore valid for all values of x', y, z.

An analogous consideration—applied to the axes of Y and Z—it being borne in mind that light is always propagated along these axes, when viewed from the stationary system, with the velocity $\sqrt{c^2 - v^2}$ gives us

$$\frac{\partial \tau}{\partial y} = 0, \frac{\partial \tau}{\partial z} = 0.$$

Since τ is a *linear* function, it follows from these equations that

$$\tau = a \left(t - \frac{v}{c^2 - v^2} x' \right)$$

where a is a function $\phi(v)$ at present unknown, and where for brevity it is assumed that at the origin of k, $\tau = 0$, when $t = 0$.

With the help of this result we easily determine the quantities ξ, η, ζ by expressing in equations that light (as required by the principle of the constancy of the velocity of light, in combination with the principle of relativity) is also propagated with velocity c when measured in the moving system. For a ray of light emitted at the time $\tau = 0$ in the direction of the increasing ξ

$$\xi = c\tau \text{ or } \xi = ac \left(t - \frac{v}{c^2 - v^2} x' \right).$$

But the ray moves relatively to the initial point of k, when measured in the stationary system, with the velocity $c - v$, so that

$$\frac{x'}{c - v} = t.$$

If we insert this value of t in the equation for ξ, we obtain

$$\xi = a \frac{c^2}{c^2 - v^2} x'.$$

In an analogous manner we find, by considering rays moving along the two other axes, that

$$\eta = c\tau = ac \left(t - \frac{v}{c^2 - v^2} x' \right)$$

when

$$\frac{y}{\sqrt{c^2 - v^2}} = t, \ x' = 0.$$

Thus

$$\eta = a\frac{c}{\sqrt{c^2 - v^2}}y \text{ and } \zeta = a\frac{c}{\sqrt{c^2 - v^2}}z.$$

Substituting for x' its value, we obtain

$$\begin{aligned}
\tau &= \phi(v)\beta(t - vx/c^2),\\
\xi &= \phi(v)\beta(t - vt),\\
\eta &= \phi(v)y,\\
\zeta &= \phi(v)z,
\end{aligned}$$

where

$$\beta = \frac{1}{\sqrt{1 - v^2/c^2}},$$

and ϕ is an as yet unknown function of v. If no assumption whatever be made as to the initial position of the moving system and as to the zero point of τ, an additive constant is to be placed on the right side of each of these equations.

We now have to prove that any ray of light, measured in the moving system, is propagated with the velocity c, if, as we have assumed, this is the case in the stationary system; for we have not as yet furnished the proof that the principle of the constancy of the velocity of light is compatible with the principle of relativity.

At the time $t = \tau = 0$, when the origin of the co-ordinates is common to the two systems, let a spherical wave be emitted therefrom, and be propagated with the velocity c in system K. If (x, y, z) be a point just attained by this wave, then

$$x^2 + y^2 + z^2 = c^2t^2.$$

Transforming this equation with the aid of our equations of transformation we obtain after a simple calculation

$$\xi^2 + \eta^2 + \zeta^2 = c^2\tau^2.$$

The wave under consideration is therefore no less a spherical wave with velocity of propagation c when viewed in the moving system. This shows that our two fundamental principles are compatible.[5]

In the equations of transformation which have been developed there enters an unknown function ϕ of v, which we will now determine.

For this purpose we introduce a third system of co-ordinates K′, which relatively to the system k is in a state of parallel translatory motion parallel to the axis of X, such that the origin of co-ordinates of system k moves with velocity

[5] The equations of the Lorentz transformation may be more simply deduced directly from the condition that in virtue of those equations the relation $x^2 + y^2 + z^2 = c^2t^2$ shall have as its consequence the second relation $\xi^2 + \eta^2 + \zeta^2 = c^2\tau^2$.

$-v$ on the axis of X. At the time $t = 0$ let all three origins coincide, and when $t = x = y = z = 0$ let the time t' of the system K$'$ be zero. We call the co-ordinates, measured in the system K$'$, x', y', z', and by a twofold application of our equations of transformation we obtain

$$\begin{array}{rclcl} t' & = & \phi(-v)\beta(-v)(\tau + v\xi/c^2) & = & \phi(v)\phi(-v)t, \\ x' & = & \phi(-v)\beta(-v)(\xi + v\tau) & = & \phi(v)\phi(-v)x, \\ y' & = & \phi(-v)\eta & = & \phi(v)\phi(-v)y, \\ z' & = & \phi(-v)\zeta & = & \phi(v)\phi(-v)z. \end{array}$$

Since the relations between x', y', z' and x, y, z do not contain the time t, the systems K and K$'$ are at rest with respect to one another, and it is clear that the transformation from K to K$'$ must be the identical transformation. Thus

$$\phi(v)\phi(-v) = 1.$$

We now inquire into the signification of $\phi(v)$. We give our attention to that part of the axis of Y of system k which lies between $\xi = 0, \eta = 0, \zeta = 0$ and $\xi = 0, \eta = l, \zeta = 0$. This part of the axis of Y is a rod moving perpendicularly to its axis with velocity v relatively to system K. Its ends possess in K the co-ordinates

$$x_1 = vt,\ y_1 = \frac{l}{\phi(v)},\ z_1 = 0$$

and

$$x_2 = vt,\ y_2 = 0,\ z_2 = 0.$$

The length of the rod measured in K is therefore $l/\phi(v)$; and this gives us the meaning of the function $\phi(v)$. From reasons of symmetry it is now evident that the length of a given rod moving perpendicularly to its axis, measured in the stationary system, must depend only on the velocity and not on the direction and the sense of the motion. The length of the moving rod measured in the stationary system does not change, therefore, if v and $-v$ are interchanged. Hence follows that $l/\phi(v) = l/\phi(-v)$, or

$$\phi(v) = \phi(-v).$$

It follows from this relation and the one previously found that $\phi(v) = 1$, so that the transformation equations which have been found become

$$\begin{aligned} \tau &= \beta(t - vx/c^2), \\ \xi &= \beta(x - vt), \\ \eta &= y, \\ \zeta &= z, \end{aligned}$$

where

$$\beta = 1/\sqrt{1 - v^2/c^2}.$$

§ 4. Physical Meaning of the Equations Obtained in Respect to Moving Rigid Bodies and Moving Clocks

We envisage a rigid sphere[6] of radius R, at rest relatively to the moving system k, and with its centre at the origin of co-ordinates of k. The equation of the surface of this sphere moving relatively to the system K with velocity v is

$$\xi^2 + \eta^2 + \zeta^2 = \mathrm{R}^2.$$

The equation of this surface expressed in x, y, z at the time $t = 0$ is

$$\frac{x^2}{(\sqrt{1 - v^2/c^2})^2} + y^2 + z^2 = \mathrm{R}^2.$$

A rigid body which, measured in a state of rest, has the form of a sphere, therefore has in a state of motion—viewed from the stationary system—the form of an ellipsoid of revolution with the axes

$$\mathrm{R}\sqrt{1 - v^2/c^2},\ \mathrm{R},\ \mathrm{R}.$$

Thus, whereas the Y and Z dimensions of the sphere (and therefore of every rigid body of no matter what form) do not appear modified by the motion, the X dimension appears shortened in the ratio $1 : \sqrt{1 - v^2/c^2}$, i.e. the greater the value of v, the greater the shortening. For $v = c$ all moving objects—viewed from the "stationary" system—shrivel up into plane figures.† For velocities greater than that of light our deliberations become meaningless; we shall, however, find in what follows, that the velocity of light in our theory plays the part, physically, of an infinitely great velocity.

It is clear that the same results hold good of bodies at rest in the "stationary" system, viewed from a system in uniform motion.

Further, we imagine one of the clocks which are qualified to mark the time t when at rest relatively to the stationary system, and the time τ when at rest relatively to the moving system, to be located at the origin of the co-ordinates of k, and so adjusted that it marks the time τ. What is the rate of this clock, when viewed from the stationary system?

Between the quantities x, t, and τ, which refer to the position of the clock, we have, evidently, $x = vt$ and

$$\tau = \frac{1}{\sqrt{1 - v^2/c^2}}(t - vx/c^2).$$

Therefore,

[6] That is, a body possessing spherical form when examined at rest.

† **Editor's note: In the original 1923 English edition, this phrase was erroneously translated as "plain figures". I have used the correct "plane figures" in this edition.**

$$\tau = t\sqrt{1-v^2/c^2} = t - (1-\sqrt{1-v^2/c^2})t$$

whence it follows that the time marked by the clock (viewed in the stationary system) is slow by $1-\sqrt{1-v^2/c^2}$ seconds per second, or—neglecting magnitudes of fourth and higher order—by $\frac{1}{2}v^2/c^2$.

From this there ensues the following peculiar consequence. If at the points A and B of K there are stationary clocks which, viewed in the stationary system, are synchronous; and if the clock at A is moved with the velocity v along the line AB to B, then on its arrival at B the two clocks no longer synchronize, but the clock moved from A to B lags behind the other which has remained at B by $\frac{1}{2}tv^2/c^2$ (up to magnitudes of fourth and higher order), t being the time occupied in the journey from A to B.

It is at once apparent that this result still holds good if the clock moves from A to B in any polygonal line, and also when the points A and B coincide.

If we assume that the result proved for a polygonal line is also valid for a continuously curved line, we arrive at this result: If one of two synchronous clocks at A is moved in a closed curve with constant velocity until it returns to A, the journey lasting t seconds, then by the clock which has remained at rest the travelled clock on its arrival at A will be $\frac{1}{2}tv^2/c^2$ second slow. Thence we conclude that a balance-clock[7] at the equator must go more slowly, by a very small amount, than a precisely similar clock situated at one of the poles under otherwise identical conditions.

§ 5. The Composition of Velocities

In the system k moving along the axis of X of the system K with velocity v, let a point move in accordance with the equations

$$\xi = w_\xi \tau, \eta = w_\eta \tau, \zeta = 0,$$

where w_ξ and w_η denote constants.

Required: the motion of the point relatively to the system K. If with the help of the equations of transformation developed in § 3 we introduce the quantities x, y, z, t into the equations of motion of the point, we obtain

$$\begin{aligned} x &= \frac{w_\xi + v}{1+vw_\xi/c^2}t, \\ y &= \frac{\sqrt{1-v^2/c^2}}{1+vw_\xi/c^2}w_\eta t, \\ z &= 0. \end{aligned}$$

Thus the law of the parallelogram of velocities is valid according to our theory only to a first approximation. We set

[7] Not a pendulum-clock, which is physically a system to which the Earth belongs. This case had to be excluded.

$$V^2 = \left(\frac{dx}{dt}\right)^2 + \left(\frac{dy}{dt}\right)^2,$$
$$w^2 = w_\xi^2 + w_\eta^2,$$
$$a = \tan^{-1} w_y/w_x,$$

a is then to be looked upon as the angle between the velocities v and w. After a simple calculation we obtain

$$V = \frac{\sqrt{(v^2 + w^2 + 2vw\cos a) - (vw\sin a/c)^2}}{1 + vw\cos a/c^2}.$$

It is worthy of remark that v and w enter into the expression for the resultant velocity in a symmetrical manner. If w also has the direction of the axis of X, we get

$$V = \frac{v + w}{1 + vw/c^2}.$$

It follows from this equation that from a composition of two velocities which are less than c, there always results a velocity less than c. For if we set $v = c - \kappa, w = c - \lambda$, κ and λ being positive and less than c, then

$$V = c\frac{2c - \kappa - \lambda}{2c - \kappa - \lambda + \kappa\lambda/c} < c.$$

It follows, further, that the velocity of light c cannot be altered by composition with a velocity less than that of light. For this case we obtain

$$V = \frac{c + w}{1 + w/c} = c.$$

We might also have obtained the formula for V, for the case when v and w have the same direction, by compounding two transformations in accordance with § 3. If in addition to the systems K and k figuring in § 3 we introduce still another system of co-ordinates k' moving parallel to k, its initial point moving on the axis of X with the velocity w, we obtain equations between the quantities x, y, z, t and the corresponding quantities of k', which differ from the equations found in § 3 only in that the place of "v" is taken by the quantity

$$\frac{v + w}{1 + vw/c^2};$$

from which we see that such parallel transformations—necessarily—form a group.

We have now deduced the requisite laws of the theory of kinematics corresponding to our two principles, and we proceed to show their application to electrodynamics.

II. ELECTRODYNAMICAL PART

§ 6. Transformation of the Maxwell-Hertz Equations for Empty Space. On the Nature of the Electromotive Forces Occurring in a Magnetic Field During Motion

Let the Maxwell-Hertz equations for empty space hold good for the stationary system K, so that we have

$$\begin{aligned}
\frac{1}{c}\frac{\partial \mathrm{X}}{\partial t} &= \frac{\partial \mathrm{N}}{\partial y} - \frac{\partial \mathrm{M}}{\partial z}, & \frac{1}{c}\frac{\partial \mathrm{L}}{\partial t} &= \frac{\partial \mathrm{Y}}{\partial z} - \frac{\partial \mathrm{Z}}{\partial y}, \\
\frac{1}{c}\frac{\partial \mathrm{Y}}{\partial t} &= \frac{\partial \mathrm{L}}{\partial z} - \frac{\partial \mathrm{N}}{\partial x}, & \frac{1}{c}\frac{\partial \mathrm{M}}{\partial t} &= \frac{\partial \mathrm{Z}}{\partial x} - \frac{\partial \mathrm{X}}{\partial z}, \\
\frac{1}{c}\frac{\partial \mathrm{Z}}{\partial t} &= \frac{\partial \mathrm{M}}{\partial x} - \frac{\partial \mathrm{L}}{\partial y}, & \frac{1}{c}\frac{\partial \mathrm{N}}{\partial t} &= \frac{\partial \mathrm{X}}{\partial y} - \frac{\partial \mathrm{Y}}{\partial x},
\end{aligned}$$

where (X, Y, Z) denotes the vector of the electric force, and (L, M, N) that of the magnetic force.

If we apply to these equations the transformation developed in § 3, by referring the electromagnetic processes to the system of co-ordinates there introduced, moving with the velocity v, we obtain the equations

$$\begin{aligned}
\frac{1}{c}\frac{\partial \mathrm{X}}{\partial \tau} &= \frac{\partial}{\partial \eta}\left\{\beta\left(\mathrm{N} - \frac{v}{c}\mathrm{Y}\right)\right\} - \frac{\partial}{\partial \zeta}\left\{\beta\left(\mathrm{M} + \frac{v}{c}\mathrm{Z}\right)\right\}, \\
\frac{1}{c}\frac{\partial}{\partial \tau}\left\{\beta\left(\mathrm{Y} - \frac{v}{c}\mathrm{N}\right)\right\} &= \frac{\partial \mathrm{L}}{\partial \xi} - \frac{\partial}{\partial \zeta}\left\{\beta\left(\mathrm{N} - \frac{v}{c}\mathrm{Y}\right)\right\}, \\
\frac{1}{c}\frac{\partial}{\partial \tau}\left\{\beta\left(\mathrm{Z} - \frac{v}{c}\mathrm{M}\right)\right\} &= \frac{\partial}{\partial \xi}\left\{\beta\left(\mathrm{M} - \frac{v}{c}\mathrm{Z}\right)\right\} - \frac{\partial \mathrm{L}}{\partial \eta}, \\
\frac{1}{c}\frac{\partial \mathrm{L}}{\partial \tau} &= \frac{\partial}{\partial \zeta}\left\{\beta\left(\mathrm{Y} - \frac{v}{c}\mathrm{N}\right)\right\} - \frac{\partial}{\partial \eta}\left\{\beta\left(\mathrm{Z} - \frac{v}{c}\mathrm{M}\right)\right\}, \\
\frac{1}{c}\frac{\partial}{\partial \tau}\left\{\beta\left(\mathrm{M} - \frac{v}{c}\mathrm{Z}\right)\right\} &= \frac{\partial}{\partial \xi}\left\{\beta\left(\mathrm{Z} - \frac{v}{c}\mathrm{M}\right)\right\} - \frac{\partial \mathrm{X}}{\partial \zeta}, \\
\frac{1}{c}\frac{\partial}{\partial \tau}\left\{\beta\left(\mathrm{N} - \frac{v}{c}\mathrm{Y}\right)\right\} &= \frac{\partial \mathrm{X}}{\partial \eta} - \frac{\partial}{\partial \xi}\left\{\beta\left(\mathrm{Y} - \frac{v}{c}\mathrm{N}\right)\right\},
\end{aligned}$$

where

$$\beta = 1/\sqrt{1 - v^2/c^2}.$$

Now the principle of relativity requires that if the Maxwell-Hertz equations for empty space hold good in system K, they also hold good in system k; that is to say that the vectors of the electric and the magnetic force—(X′, Y′, Z′) and (L′, M′, N′)—of the moving system k, which are defined by their ponderomotive effects on electric or magnetic masses respectively, satisfy the following equations:—

$$\begin{aligned}
\frac{1}{c}\frac{\partial \mathrm{X}'}{\partial \tau} &= \frac{\partial \mathrm{N}'}{\partial \eta} - \frac{\partial \mathrm{M}'}{\partial \zeta}, & \frac{1}{c}\frac{\partial \mathrm{L}'}{\partial \tau} &= \frac{\partial \mathrm{Y}'}{\partial \zeta} - \frac{\partial \mathrm{Z}'}{\partial \eta}, \\
\frac{1}{c}\frac{\partial \mathrm{Y}'}{\partial \tau} &= \frac{\partial \mathrm{L}'}{\partial \zeta} - \frac{\partial \mathrm{N}'}{\partial \xi}, & \frac{1}{c}\frac{\partial \mathrm{M}'}{\partial \tau} &= \frac{\partial \mathrm{Z}'}{\partial \xi} - \frac{\partial \mathrm{X}'}{\partial \zeta}, \\
\frac{1}{c}\frac{\partial \mathrm{Z}'}{\partial \tau} &= \frac{\partial \mathrm{M}'}{\partial \xi} - \frac{\partial \mathrm{L}'}{\partial \eta}, & \frac{1}{c}\frac{\partial \mathrm{N}'}{\partial \tau} &= \frac{\partial \mathrm{X}'}{\partial \eta} - \frac{\partial \mathrm{Y}'}{\partial \xi}.
\end{aligned}$$

Evidently the two systems of equations found for system k must express exactly the same thing, since both systems of equations are equivalent to the Maxwell-Hertz equations for system K. Since, further, the equations of the two systems agree, with the exception of the symbols for the vectors, it follows that the functions occurring in the systems of equations at corresponding places must agree, with the exception of a factor $\psi(v)$, which is common for all functions of the one system of equations, and is independent of ξ, η, ζ and τ but depends upon v. Thus we have the relations

$$\begin{array}{rclrcl} \mathrm{X}' &=& \psi(v)\mathrm{X}, & \mathrm{L}' &=& \psi(v)\mathrm{L}, \\ \mathrm{Y}' &=& \psi(v)\beta\left(\mathrm{Y} - \frac{v}{c}\mathrm{N}\right), & \mathrm{M}' &=& \psi(v)\beta\left(\mathrm{M} - \frac{v}{c}\mathrm{Z}\right), \\ \mathrm{Z}' &=& \psi(v)\beta\left(\mathrm{Z} - \frac{v}{c}\mathrm{M}\right), & \mathrm{N}' &=& \psi(v)\beta\left(\mathrm{N} - \frac{v}{c}\mathrm{Y}\right). \end{array}$$

If we now form the reciprocal of this system of equations, firstly by solving the equations just obtained, and secondly by applying the equations to the inverse transformation (from k to K), which is characterized by the velocity $-v$, it follows, when we consider that the two systems of equations thus obtained must be identical, that $\psi(v)\psi(-v) = 1$. Further, from reasons of symmetry [8] and therefore

$$\psi(v) = 1,$$

and our equations assume the form

$$\begin{array}{rclrcl} \mathrm{X}' &=& \mathrm{X}, & \mathrm{L}' &=& \mathrm{L}, \\ \mathrm{Y}' &=& \beta\left(\mathrm{Y} - \frac{v}{c}\mathrm{N}\right), & \mathrm{M}' &=& \beta\left(\mathrm{M} + \frac{v}{c}\mathrm{Z}\right), \\ \mathrm{Z}' &=& \beta\left(\mathrm{Z} + \frac{v}{c}\mathrm{M}\right), & \mathrm{N}' &=& \beta\left(\mathrm{N} - \frac{v}{c}\mathrm{Y}\right). \end{array}$$

As to the interpretation of these equations we make the following remarks: Let a point charge of electricity have the magnitude "one" when measured in the stationary system K, i.e. let it when at rest in the stationary system exert a force of one dyne upon an equal quantity of electricity at a distance of one cm. By the principle of relativity this electric charge is also of the magnitude "one" when measured in the moving system. If this quantity of electricity is at rest relatively to the stationary system, then by definition the vector (X, Y, Z) is equal to the force acting upon it. If the quantity of electricity is at rest relatively to the moving system (at least at the relevant instant), then the force acting upon it, measured in the moving system, is equal to the vector (X′, Y′, Z′). Consequently the first three equations above allow themselves to be clothed in words in the two following ways:—

1. If a unit electric point charge is in motion in an electromagnetic field, there acts upon it, in addition to the electric force, an "electromotive force" which, if we neglect the terms multiplied by the second and higher powers of

[8] If, for example, X=Y=Z=L=M=0, and N $\neq$ 0, then from reasons of symmetry it is clear that when v changes sign without changing its numerical value, Y′ must also change sign without changing its numerical value.

v/c, is equal to the vector-product of the velocity of the charge and the magnetic force, divided by the velocity of light. (Old manner of expression.)

2. If a unit electric point charge is in motion in an electromagnetic field, the force acting upon it is equal to the electric force which is present at the locality of the charge, and which we ascertain by transformation of the field to a system of co-ordinates at rest relatively to the electrical charge. (New manner of expression.)

The analogy holds with "magnetomotive forces." We see that electromotive force plays in the developed theory merely the part of an auxiliary concept, which owes its introduction to the circumstance that electric and magnetic forces do not exist independently of the state of motion of the system of co-ordinates.

Furthermore it is clear that the asymmetry mentioned in the introduction as arising when we consider the currents produced by the relative motion of a magnet and a conductor, now disappears. Moreover, questions as to the "seat" of electrodynamic electromotive forces (unipolar machines) now have no point.

§ 7. Theory of Doppler's Principle and of Aberration

In the system K, very far from the origin of co-ordinates, let there be a source of electrodynamic waves, which in a part of space containing the origin of co-ordinates may be represented to a sufficient degree of approximation by the equations

$$\begin{aligned} \mathrm{X} &= \mathrm{X}_0 \sin\Phi, & \mathrm{L} &= \mathrm{L}_0 \sin\Phi, \\ \mathrm{Y} &= \mathrm{Y}_0 \sin\Phi, & \mathrm{M} &= \mathrm{M}_0 \sin\Phi, \\ \mathrm{Z} &= \mathrm{Z}_0 \sin\Phi, & \mathrm{N} &= \mathrm{N}_0 \sin\Phi, \end{aligned}$$

where

$$\Phi = \omega\left\{t - \frac{1}{c}(lx + my + nz)\right\}.$$

Here $(\mathrm{X}_0, \mathrm{Y}_0, \mathrm{Z}_0)$ and $(\mathrm{L}_0, \mathrm{M}_0, \mathrm{N}_0)$ are the vectors defining the amplitude of the wave-train, and l, m, n the direction-cosines of the wave-normals. We wish to know the constitution of these waves, when they are examined by an observer at rest in the moving system k.

Applying the equations of transformation found in § 6 for electric and magnetic forces, and those found in § 3 for the co-ordinates and the time, we obtain directly

$$\begin{aligned} \mathrm{X}' &= \mathrm{X}_0 \sin\Phi', & \mathrm{L}' &= \mathrm{L}_0 \sin\Phi', \\ \mathrm{Y}' &= \beta(\mathrm{Y}_0 - v\mathrm{N}_0/c)\sin\Phi', & \mathrm{M}' &= \beta(\mathrm{M}_0 + v\mathrm{Z}_0/c)\sin\Phi', \\ \mathrm{Z}' &= \beta(\mathrm{Z}_0 + v\mathrm{M}_0/c)\sin\Phi', & \mathrm{N}' &= \beta(\mathrm{N}_0 - v\mathrm{Y}_0/c)\sin\Phi', \\ \Phi' &= \omega'\left\{\tau - \tfrac{1}{c}(l'\xi + m'\eta + n'\zeta)\right\} \end{aligned}$$

where

$$\begin{aligned}
\omega' &= \omega\beta(1 - lv/c),\\
l' &= \frac{l - v/c}{1 - lv/c},\\
m' &= \frac{m}{\beta(1 - lv/c)},\\
n' &= \frac{n}{\beta(1 - lv/c)}.
\end{aligned}$$

From the equation for ω' it follows that if an observer is moving with velocity v relatively to an infinitely distant source of light of frequency ν, in such a way that the connecting line "source-observer" makes the angle ϕ with the velocity of the observer referred to a system of co-ordinates which is at rest relatively to the source of light, the frequency ν' of the light perceived by the observer is given by the equation

$$\nu' = \nu \frac{1 - \cos\phi \cdot v/c}{\sqrt{1 - v^2/c^2}}.$$

This is Doppler's principle for any velocities whatever. When $\phi = 0$ the equation assumes the perspicuous form

$$\nu' = \nu \sqrt{\frac{1 - v/c}{1 + v/c}}.$$

We see that, in contrast with the customary view, when $v = -c, \nu' = \infty$.

If we call the angle between the wave-normal (direction of the ray) in the moving system and the connecting line "source-observer" ϕ', the equation for l' assumes the form

$$\cos\phi' = \frac{\cos\phi - v/c}{1 - \cos\phi \cdot v/c}.$$

This equation expresses the law of aberration in its most general form. If $\phi = \frac{1}{2}\pi$, the equation becomes simply

$$\cos\phi' = -v/c.$$

We still have to find the amplitude of the waves, as it appears in the moving system. If we call the amplitude of the electric or magnetic force A or A′ respectively, accordingly as it is measured in the stationary system or in the moving system, we obtain

$$\mathrm{A}'^2 = \mathrm{A}^2 \frac{(1 - \cos\phi \cdot v/c)^2}{1 - v^2/c^2}$$

which equation, if $\phi = 0$, simplifies into

$$\mathrm{A}'^2 = \mathrm{A}^2 \frac{1 - v/c}{1 + v/c}.$$

It follows from these results that to an observer approaching a source of light with the velocity c, this source of light must appear of infinite intensity.

§ 8. Transformation of the Energy of Light Rays. Theory of the Pressure of Radiation Exerted on Perfect Reflectors

Since $\mathrm{A}^2/8\pi$ equals the energy of light per unit of volume, we have to regard ${\mathrm{A}'}^2/8\pi$, by the principle of relativity, as the energy of light in the moving system. Thus ${\mathrm{A}'}^2/\mathrm{A}^2$ would be the ratio of the "measured in motion" to the "measured at rest" energy of a given light complex, if the volume of a light complex were the same, whether measured in K or in k. But this is not the case. If l, m, n are the direction-cosines of the wave-normals of the light in the stationary system, no energy passes through the surface elements of a spherical surface moving with the velocity of light:—

$$(x - lct)^2 + (y - mct)^2 + (z - nct)^2 = \mathrm{R}^2.$$

We may therefore say that this surface permanently encloses the same light complex. We inquire as to the quantity of energy enclosed by this surface, viewed in system k, that is, as to the energy of the light complex relatively to the system k.

The spherical surface—viewed in the moving system—is an ellipsoidal surface, the equation for which, at the time $\tau = 0$, is

$$(\beta\xi - l\beta\xi v/c)^2 + (\eta - m\beta\xi v/c)^2 + (\zeta - n\beta\xi v/c)^2 = \mathrm{R}^2.$$

If S is the volume of the sphere, and S′ that of this ellipsoid, then by a simple calculation

$$\frac{\mathrm{S}'}{\mathrm{S}} = \frac{\sqrt{1 - v^2/c^2}}{1 - \cos\phi \cdot v/c}.$$

Thus, if we call the light energy enclosed by this surface E when it is measured in the stationary system, and E′ when measured in the moving system, we obtain

$$\frac{\mathrm{E}'}{\mathrm{E}} = \frac{{\mathrm{A}'}^2\mathrm{S}'}{\mathrm{A}^2\mathrm{S}} = \frac{1 - \cos\phi \cdot v/c}{\sqrt{1 - v^2/c^2}},$$

and this formula, when $\phi = 0$, simplifies into

$$\frac{\mathrm{E}'}{\mathrm{E}} = \sqrt{\frac{1 - v/c}{1 + v/c}}.$$

It is remarkable that the energy and the frequency of a light complex vary with the state of motion of the observer in accordance with the same law.

Now let the co-ordinate plane $\xi = 0$ be a perfectly reflecting surface, at which the plane waves considered in § 7 are reflected. We seek for the pressure of light exerted on the reflecting surface, and for the direction, frequency, and intensity of the light after reflexion.

Let the incidental light be defined by the quantities A, $\cos\phi$, ν (referred to system K). Viewed from k the corresponding quantities are

$$\begin{aligned} \mathrm{A}' =& \mathrm{A}\frac{1-\cos\phi\cdot v/c}{\sqrt{1-v^2/c^2}}, \\ \cos\phi' =& \frac{\cos\phi - v/c}{1-\cos\phi\cdot v/c}, \\ \nu' =& \nu\frac{1-\cos\phi\cdot v/c}{\sqrt{1-v^2/c^2}}. \end{aligned}$$

For the reflected light, referring the process to system k, we obtain

$$\begin{aligned} \mathrm{A}'' &= \mathrm{A}' \\ \cos\phi'' &= -\cos\phi' \\ \nu'' &= \nu' \end{aligned}$$

Finally, by transforming back to the stationary system K, we obtain for the reflected light

$$\begin{aligned} \mathrm{A}''' &= \mathrm{A}''\frac{1+cos\phi''\cdot v/c}{\sqrt{1-v^2/c^2}} = \mathrm{A}\frac{1-2\cos\phi\cdot v/c+v^2/c^2}{1-v^2/c^2}, \\ \cos\phi''' &= \frac{\cos\phi''+v/c}{1+\cos\phi''\cdot v/c} = -\frac{(1+v^2/c^2)\cos\phi - 2v/c}{1-2\cos\phi\cdot v/c+v^2/c^2}, \\ \nu''' &= \nu''\frac{1+\cos\phi''\cdot v/c}{\sqrt{1-v^2/c^2}} = \nu\frac{1-2\cos\phi\cdot v/c+v^2/c^2}{1-v^2/c^2}. \end{aligned}$$

The energy (measured in the stationary system) which is incident upon unit area of the mirror in unit time is evidently $\mathrm{A}^2(c\cos\phi - v)/8\pi$. The energy leaving the unit of surface of the mirror in the unit of time is $\mathrm{A}'''^2(-c\cos\phi''' + v)/8\pi$. The difference of these two expressions is, by the principle of energy, the work done by the pressure of light in the unit of time. If we set down this work as equal to the product Pv, where P is the pressure of light, we obtain

$$\mathrm{P} = 2\cdot\frac{\mathrm{A}^2}{8\pi}\frac{(\cos\phi - v/c)^2}{1-v^2/c^2}.$$

In agreement with experiment and with other theories, we obtain to a first approximation

$$\mathrm{P} = 2\cdot\frac{\mathrm{A}^2}{8\pi}\cos^2\phi.$$

All problems in the optics of moving bodies can be solved by the method here employed. What is essential is, that the electric and magnetic force of the light which is influenced by a moving body, be transformed into a system of co-ordinates at rest relatively to the body. By this means all problems in the optics of moving bodies will be reduced to a series of problems in the optics of stationary bodies.

§ 9. Transformation of the Maxwell-Hertz Equations when Convection-Currents are Taken into Account

We start from the equations

$$\begin{aligned}
\frac{1}{c}\left\{\frac{\partial \mathrm{X}}{\partial t}+u_x\rho\right\} &= \frac{\partial \mathrm{N}}{\partial y}-\frac{\partial \mathrm{M}}{\partial z}, & \frac{1}{c}\frac{\partial \mathrm{L}}{\partial t} &= \frac{\partial \mathrm{Y}}{\partial z}-\frac{\partial \mathrm{Z}}{\partial y},\\
\frac{1}{c}\left\{\frac{\partial \mathrm{Y}}{\partial t}+u_y\rho\right\} &= \frac{\partial \mathrm{L}}{\partial z}-\frac{\partial \mathrm{N}}{\partial x}, & \frac{1}{c}\frac{\partial \mathrm{M}}{\partial t} &= \frac{\partial \mathrm{Z}}{\partial x}-\frac{\partial \mathrm{X}}{\partial z},\\
\frac{1}{c}\left\{\frac{\partial \mathrm{Z}}{\partial t}+u_z\rho\right\} &= \frac{\partial \mathrm{M}}{\partial x}-\frac{\partial \mathrm{L}}{\partial y}, & \frac{1}{c}\frac{\partial \mathrm{N}}{\partial t} &= \frac{\partial \mathrm{X}}{\partial y}-\frac{\partial \mathrm{Y}}{\partial x},
\end{aligned}$$

where

$$\rho=\frac{\partial \mathrm{X}}{\partial x}+\frac{\partial \mathrm{Y}}{\partial y}+\frac{\partial \mathrm{Z}}{\partial z}$$

denotes 4π times the density of electricity, and (u_x, u_y, u_z) the velocity-vector of the charge. If we imagine the electric charges to be invariably coupled to small rigid bodies (ions, electrons), these equations are the electromagnetic basis of the Lorentzian electrodynamics and optics of moving bodies.

Let these equations be valid in the system K, and transform them, with the assistance of the equations of transformation given in §§ 3 and 6, to the system k. We then obtain the equations

$$\begin{aligned}
\frac{1}{c}\left\{\frac{\partial \mathrm{X}'}{\partial \tau}+u_\xi\rho'\right\} &= \frac{\partial \mathrm{N}'}{\partial \eta}-\frac{\partial \mathrm{M}'}{\partial \zeta}, & \frac{1}{c}\frac{\partial \mathrm{L}'}{\partial \tau} &= \frac{\partial \mathrm{Y}'}{\partial \zeta}-\frac{\partial \mathrm{Z}'}{\partial \eta},\\
\frac{1}{c}\left\{\frac{\partial \mathrm{Y}'}{\partial \tau}+u_\eta\rho'\right\} &= \frac{\partial \mathrm{L}'}{\partial \zeta}-\frac{\partial \mathrm{N}'}{\partial \xi}, & \frac{1}{c}\frac{\partial \mathrm{M}'}{\partial \tau} &= \frac{\partial \mathrm{Z}'}{\partial \xi}-\frac{\partial \mathrm{X}'}{\partial \zeta},\\
\frac{1}{c}\left\{\frac{\partial \mathrm{Z}'}{\partial \tau}+u_\zeta\rho'\right\} &= \frac{\partial \mathrm{M}'}{\partial \xi}-\frac{\partial \mathrm{L}'}{\partial \eta}, & \frac{1}{c}\frac{\partial \mathrm{N}'}{\partial \tau} &= \frac{\partial \mathrm{X}'}{\partial \eta}-\frac{\partial \mathrm{Y}'}{\partial \xi},
\end{aligned}$$

where

$$\begin{aligned}
u_\xi &= \frac{u_x-v}{1-u_xv/c^2}\\
u_\eta &= \frac{u_y}{\beta(1-u_xv/c^2)}\\
u_\zeta &= \frac{u_z}{\beta(1-u_xv/c^2)},
\end{aligned}$$

and

$$\begin{aligned}
\rho' &= \frac{\partial \mathrm{X}'}{\partial \xi}+\frac{\partial \mathrm{Y}'}{\partial \eta}+\frac{\partial Z'}{\partial \zeta}\\
&= \beta(1-u_xv/c^2)\rho.
\end{aligned}$$

Since—as follows from the theorem of addition of velocities (§ 5)—the vector (u_ξ, u_η, u_ζ) is nothing else than the velocity of the electric charge, measured in the system k, we have the proof that, on the basis of our kinematical principles, the electrodynamic foundation of Lorentz's theory of the electrodynamics of moving bodies is in agreement with the principle of relativity.

In addition I may briefly remark that the following important law may easily be deduced from the developed equations: If an electrically charged body is in motion anywhere in space without altering its charge when regarded from a system of co-ordinates moving with the body, its charge also remains—when regarded from the "stationary" system K—constant.

§ 10. Dynamics of the Slowly Accelerated Electron

Let there be in motion in an electromagnetic field an electrically charged particle (in the sequel called an "electron"), for the law of motion of which we assume as follows:—

If the electron is at rest at a given epoch, the motion of the electron ensues in the next instant of time according to the equations

$$m\frac{d^2x}{dt^2} = \epsilon\mathrm{X}$$
$$m\frac{d^2y}{dt^2} = \epsilon\mathrm{Y}$$
$$m\frac{d^2z}{dt^2} = \epsilon\mathrm{Z}$$

where x, y, z denote the co-ordinates of the electron, and m the mass of the electron, as long as its motion is slow.

Now, secondly, let the velocity of the electron at a given epoch be v. We seek the law of motion of the electron in the immediately ensuing instants of time.

Without affecting the general character of our considerations, we may and will assume that the electron, at the moment when we give it our attention, is at the origin of the co-ordinates, and moves with the velocity v along the axis of X of the system K. It is then clear that at the given moment ($t = 0$) the electron is at rest relatively to a system of co-ordinates which is in parallel motion with velocity v along the axis of X.

From the above assumption, in combination with the principle of relativity, it is clear that in the immediately ensuing time (for small values of t) the electron, viewed from the system k, moves in accordance with the equations

$$m\frac{d^2\xi}{d\tau^2} = \epsilon\mathrm{X}',$$
$$m\frac{d^2\eta}{d\tau^2} = \epsilon\mathrm{Y}',$$
$$m\frac{d^2\zeta}{d\tau^2} = \epsilon\mathrm{Z}',$$

in which the symbols ξ, η, ζ, X′, Y′, Z′ refer to the system k. If, further, we decide that when $t = x = y = z = 0$ then $\tau = \xi = \eta = \zeta = 0$, the transformation equations of §§ 3 and 6 hold good, so that we have

$$\xi = \beta(x - vt), \eta = y, \zeta = z, \tau = \beta(t - vx/c^2),$$
$$\mathrm{X}' = \mathrm{X}, \mathrm{Y}' = \beta(\mathrm{Y} - v\mathrm{N}/c), \mathrm{Z}' = \beta(\mathrm{Z} + v\mathrm{M}/c).$$

With the help of these equations we transform the above equations of motion from system k to system K, and obtain

$$\left.\begin{array}{rcl} \frac{d^2x}{dt^2} & = & \frac{\epsilon}{m\beta^3}\mathrm{X} \\ \frac{d^2y}{dt^2} & = & \frac{\epsilon}{m\beta}\left(\mathrm{Y} - \frac{v}{c}\mathrm{N}\right) \\ \frac{d^2z}{dt^2} & = & \frac{\epsilon}{m\beta}\left(\mathrm{Z} - \frac{v}{c}\mathrm{M}\right) \end{array}\right\} \qquad \text{(A)}$$

Taking the ordinary point of view we now inquire as to the "longitudinal" and the "transverse" mass of the moving electron. We write the equations (A) in the form

$$\begin{array}{rcccl} m\beta^3\frac{d^2x}{dt^2} & = & \epsilon\mathrm{X} & = & \epsilon\mathrm{X}', \\ m\beta^2\frac{d^2y}{dt^2} & = & \epsilon\beta\left(\mathrm{Y} - \frac{v}{c}\mathrm{N}\right) & = & \epsilon\mathrm{Y}', \\ m\beta^2\frac{d^2z}{dt^2} & = & \epsilon\beta\left(\mathrm{Z} - \frac{v}{c}\mathrm{M}\right) & = & \epsilon\mathrm{Z}', \end{array}$$

and remark firstly that ϵX′, ϵY′, ϵZ′ are the components of the ponderomotive force acting upon the electron, and are so indeed as viewed in a system moving at the moment with the electron, with the same velocity as the electron. (This force might be measured, for example, by a spring balance at rest in the last-mentioned system.) Now if we call this force simply "the force acting upon the electron," [9] and maintain the equation—mass × acceleration = force—and if we also decide that the accelerations are to be measured in the stationary system K, we derive from the above equations

$$\text{Longitudinal mass} = \frac{m}{(\sqrt{1 - v^2/c^2})^3}.$$
$$\text{Transverse mass} = \frac{m}{1 - v^2/c^2}.$$

With a different definition of force and acceleration we should naturally obtain other values for the masses. This shows us that in comparing different theories of the motion of the electron we must proceed very cautiously.

We remark that these results as to the mass are also valid for ponderable material points, because a ponderable material point can be made into an electron (in our sense of the word) by the addition of an electric charge, *no matter how small.*

[9] The definition of force here given is not advantageous, as was first shown by M. Planck. It is more to the point to define force in such a way that the laws of momentum and energy assume the simplest form.

We will now determine the kinetic energy of the electron. If an electron moves from rest at the origin of co-ordinates of the system K along the axis of X under the action of an electrostatic force X, it is clear that the energy withdrawn from the electrostatic field has the value $\int \epsilon \mathrm{X}\, dx$. As the electron is to be slowly accelerated, and consequently may not give off any energy in the form of radiation, the energy withdrawn from the electrostatic field must be put down as equal to the energy of motion W of the electron. Bearing in mind that during the whole process of motion which we are considering, the first of the equations (A) applies, we therefore obtain

$$\begin{aligned} \mathrm{W} &= \int \epsilon \mathrm{X}\, dx = m \int_0^v \beta^3 v\, dv \\ &= mc^2 \left\{ \frac{1}{\sqrt{1 - v^2/c^2}} - 1 \right\}. \end{aligned}$$

Thus, when $v = c$, W becomes infinite. Velocities greater than that of light have—as in our previous results—no possibility of existence.

This expression for the kinetic energy must also, by virtue of the argument stated above, apply to ponderable masses as well.

We will now enumerate the properties of the motion of the electron which result from the system of equations (A), and are accessible to experiment.

1. From the second equation of the system (A) it follows that an electric force Y and a magnetic force N have an equally strong deflective action on an electron moving with the velocity v, when $\mathrm{Y} = \mathrm{N}v/c$. Thus we see that it is possible by our theory to determine the velocity of the electron from the ratio of the magnetic power of deflexion A_m to the electric power of deflexion A_e, for any velocity, by applying the law

$$\frac{\mathrm{A}_m}{\mathrm{A}_e} = \frac{v}{c}.$$

This relationship may be tested experimentally, since the velocity of the electron can be directly measured, e.g. by means of rapidly oscillating electric and magnetic fields.

2. From the deduction for the kinetic energy of the electron it follows that between the potential difference, P, traversed and the acquired velocity v of the electron there must be the relationship

$$\mathrm{P} = \int \mathrm{X} dx = \frac{m}{\epsilon} c^2 \left\{ \frac{1}{\sqrt{1 - v^2/c^2}} - 1 \right\}.$$

3. We calculate the radius of curvature of the path of the electron when a magnetic force N is present (as the only deflective force), acting perpendicularly to the velocity of the electron. From the second of the equations (A) we obtain

$$-\frac{d^2 y}{dt^2} = \frac{v^2}{\mathrm{R}} = \frac{\epsilon}{m} \frac{v}{c} \mathrm{N} \sqrt{1 - \frac{v^2}{c^2}}$$

or

$$\mathrm{R} = \frac{mc^2}{\epsilon} \cdot \frac{v/c}{\sqrt{1 - v^2/c^2}} \cdot \frac{1}{\mathrm{N}}.$$

These three relationships are a complete expression for the laws according to which, by the theory here advanced, the electron must move.

In conclusion I wish to say that in working at the problem here dealt with I have had the loyal assistance of my friend and colleague M. Besso, and that I am indebted to him for several valuable suggestions.

About this Document

This edition of Einstein's On the Electrodynamics of Moving Bodies is based on the English translation of his original 1905 German-language paper (published as Zur Elektrodynamik bewegter Körper, in Annalen der Physik. **17***:891, 1905) which appeared in the book The Principle of Relativity, published in 1923 by Methuen and Company, Ltd. of London.*

Most of the papers in that collection are English translations from the German Das Relativatsprinzip, 4th ed., published by in 1922 by Tuebner.

Numbered footnotes are as they appeared in the 1923 edition; editor's notes are marked by a dagger (†). The 1923 English translation modified the notation used in Einstein's 1905 paper to conform to that in use by the 1920's; for example, c denotes the speed of light, as opposed the V *used by Einstein in 1905.*

This edition was originally prepared by John Walker in LaTEX. The present version was modified from LaTEX to TEX by Charles Nash.

This document is available in a variety of formats from the Web site:

http://www.fourmilab.ch/

§ 2. Einstein's $E = mc^2$ paper

In this section we provide a translation of Einstein's second paper in which he derives the famous equation $E = mc^2$. It is only three pages long. The reference for the original German article is *Ist die Trägheit eines Körpers von seinem Energiegehalt abhängig?*, *Annalen der Physik*, **323**, 639–641, (1905).

DOES THE INERTIA OF A BODY DEPEND UPON ITS ENERGY-CONTENT?

By A. EINSTEIN

September 27, 1905

The results of the previous investigation lead to a very interesting conclusion, which is here to be deduced.

I based that investigation on the Maxwell-Hertz equations for empty space, together with the Maxwellian expression for the electromagnetic energy of space, and in addition the principle that:—

The laws by which the states of physical systems alter are independent of the alternative, to which of two systems of coordinates, in uniform motion of parallel translation relatively to each other, these alterations of state are referred (principle of relativity).

With these principles [1] as my basis I deduced *inter alia* the following result (§ 8):—

Let a system of plane waves of light, referred to the system of co-ordinates (x, y, z), possess the energy l; let the direction of the ray (the wave-normal) make an angle ϕ with the axis of x of the system. If we introduce a new system of co-ordinates (ξ, η, ζ) moving in uniform parallel translation with respect to the system (x, y, z), and having its origin of co-ordinates in motion along the axis of x with the velocity v, then this quantity of light—measured in the system (ξ, η, ζ)—possesses the energy

$$l^* = l\frac{1 - \frac{v}{c} cos\phi}{\sqrt{1 - v^2/c^2}}$$

where c denotes the velocity of light. We shall make use of this result in what follows.

[1] The principle of the constancy of the velocity of light is of course contained in Maxwell's equations.

Let there be a stationary body in the system (x, y, z), and let its energy—referred to the system (x, y, z) be E_0. Let the energy of the body relative to the system (ξ, η, ζ) moving as above with the velocity v, be H_0.

Let this body send out, in a direction making an angle ϕ with the axis of x, plane waves of light, of energy $\frac{1}{2}\mathrm{L}$ measured relatively to (x, y, z), and simultaneously an equal quantity of light in the opposite direction. Meanwhile the body remains at rest with respect to the system (x, y, z). The principle of energy must apply to this process, and in fact (by the principle of relativity) with respect to both systems of co-ordinates. If we call the energy of the body after the emission of light E_1 or H_1 respectively, measured relatively to the system (x, y, z) or (ξ, η, ζ) respectively, then by employing the relation given above we obtain

$$\begin{aligned} \mathrm{E}_0 &= \mathrm{E}_1 + \frac{1}{2}\mathrm{L} + \frac{1}{2}\mathrm{L}, \\ \mathrm{H}_0 &= \mathrm{H}_1 + \frac{1}{2}\mathrm{L}\frac{1 - \frac{v}{c}cos\phi}{\sqrt{1 - v^2/c^2}} + \frac{1}{2}\mathrm{L}\frac{1 + \frac{v}{c}cos\phi}{\sqrt{1 - v^2/c^2}} \\ &= \mathrm{H}_1 + \frac{\mathrm{L}}{\sqrt{1 - v^2/c^2}} \end{aligned}$$

By subtraction we obtain from these equations

$$\mathrm{H}_0 - \mathrm{E}_0 - (\mathrm{H}_1 - \mathrm{E}_1) = \mathrm{L}\left\{\frac{1}{\sqrt{1 - v^2/c^2}} - 1\right\}.$$

The two differences of the form $\mathrm{H} - \mathrm{E}$ occurring in this expression have simple physical significations. H and E are energy values of the same body referred to two systems of co-ordinates which are in motion relatively to each other, the body being at rest in one of the two systems (system (x, y, z)). Thus it is clear that the difference $\mathrm{H} - \mathrm{E}$ can differ from the kinetic energy K of the body, with respect to the other system (ξ, η, ζ), only by an additive constant C, which depends on the choice of the arbitrary additive constants of the energies H and E. Thus we may place

$$\begin{aligned} \mathrm{H}_0 - \mathrm{E}_0 &= \mathrm{K}_0 + \mathrm{C}, \\ \mathrm{H}_1 - \mathrm{E}_1 &= \mathrm{K}_1 + \mathrm{C}, \end{aligned}$$

since C does not change during the emission of light. So we have

$$\mathrm{K}_0 - \mathrm{K}_1 = \mathrm{L}\left\{\frac{1}{\sqrt{1 - v^2/c^2}} - 1\right\}.$$

The kinetic energy of the body with respect to (ξ, η, ζ) diminishes as a result of the emission of light, and the amount of diminution is independent of the properties of the body. Moreover, the difference $\mathrm{K}_0 - \mathrm{K}_1$, like the kinetic energy of the electron (§ 10), depends on the velocity.

Neglecting magnitudes of fourth and higher orders we may place

$$\mathrm{K}_0 - \mathrm{K}_1 = \frac{1}{2}\frac{\mathrm{L}}{c^2}v^2.$$

From this equation it directly follows that:—

If a body gives off the energy L in the form of radiation, its mass diminishes by L/c^2*.* The fact that the energy withdrawn from the body becomes energy of radiation evidently makes no difference, so that we are led to the more general conclusion that

The mass of a body is a measure of its energy-content; if the energy changes by L, the mass changes in the same sense by $\mathrm{L}/9 \times 10^{20}$, the energy being measured in ergs, and the mass in grammes.

It is not impossible that with bodies whose energy-content is variable to a high degree (e.g. with radium salts) the theory may be successfully put to the test.

If the theory corresponds to the facts, radiation conveys inertia between the emitting and absorbing bodies.

About this Document

This edition of Einstein's Does the Inertia of a Body Depend upon its Energy-Content is based on the English translation of his original 1905 German-language paper (published as Ist die Trägheit eines Körpers von seinem Energiegehalt abhängig?, in Annalen der Physik. **18***, 639, 1905) which appeared in the book The Principle of Relativity, published in 1923 by Methuen and Company, Ltd. of London. Most of the papers in that collection are English translations by W. Perrett and G.B. Jeffery from the German Das Relativatsprinzip, 4th ed., published by in 1922 by Tuebner.*

The footnote is as it appeared in the 1923 edition. The 1923 English translation modified the notation used in Einstein's 1905 paper to conform to that in use by the 1920's; for example, c denotes the speed of light, as opposed the V used by Einstein in 1905. In this paper Einstein uses L to denote energy; the italicised sentence in the conclusion may be written as the equation "$m = \mathrm{L}/c^2$*" which, using the more modern E instead of L to denote energy, may be trivially rewritten as "*$\mathrm{E} = mc^2$*".*

This edition was prepared by John Walker in LaTEX . The present version was modified from LaTEX to TEX by Charles Nash.

This document is available in a variety of formats from the Web site:

http://www.fourmilab.ch/

Contents

CHAPTER IV

CHAPTER V

CHAPTER VI

CHAPTER VII

CHAPTER VIII