SPECIAL RELATIVITY. MATH2410

KOMISSAROV S.S
School of Mathematics
University of Leeds

2018

Contents

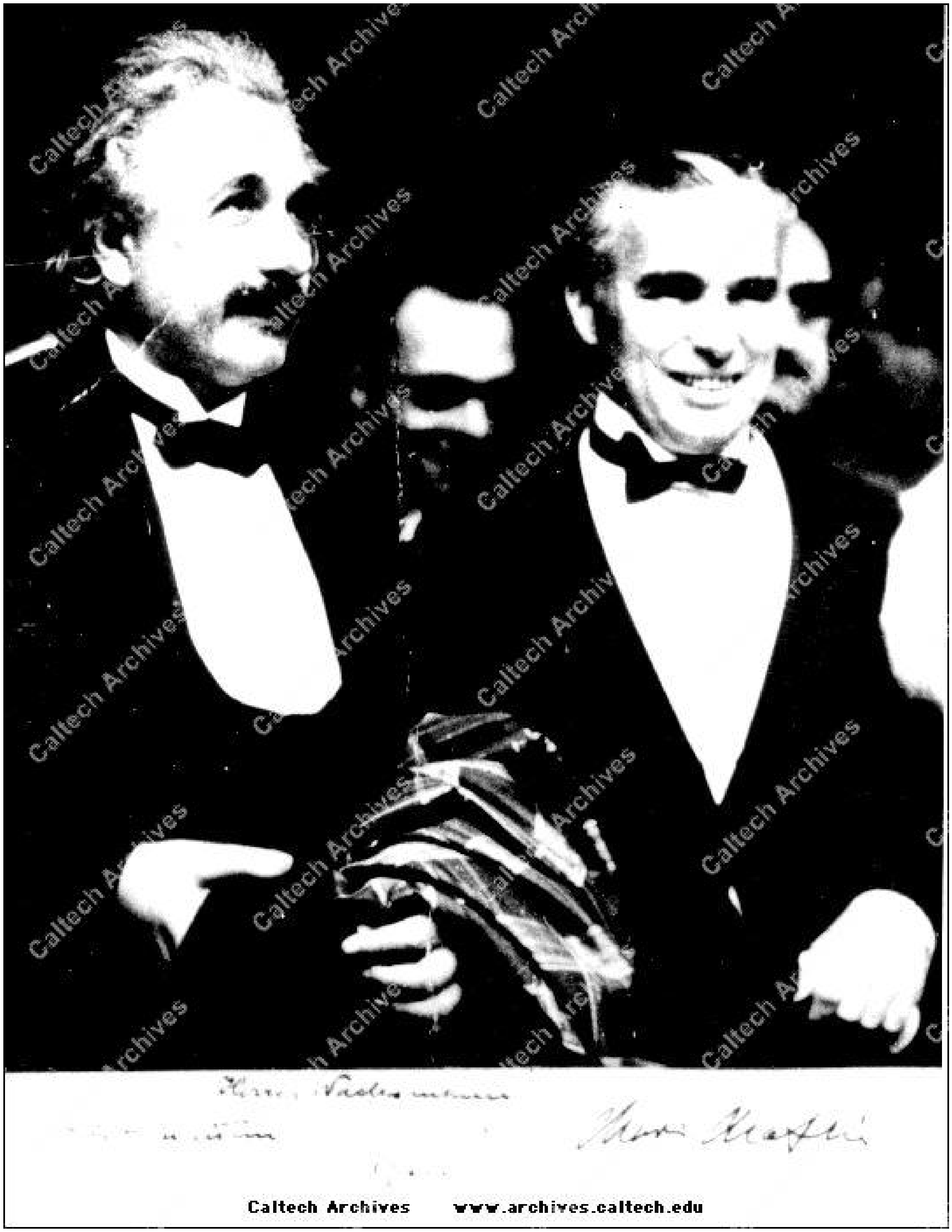

Figure 1: Charlie Chaplin: "They cheer me because they all understand me, and they cheer you because no one can understand you."

Figure 2: Arthur Stanley Eddington, the great English astrophysicist. From the conversation that took place in the lobby of The Royal Society: Silverstein - "... only three scientists in the world understand theory of relativity. I was told that you are one of them." Eddington - "Emm ... ." Silverstein - "Don't be so modest, Eddington!" Eddington - "On the contrary. I am just wondering who this third person might be."

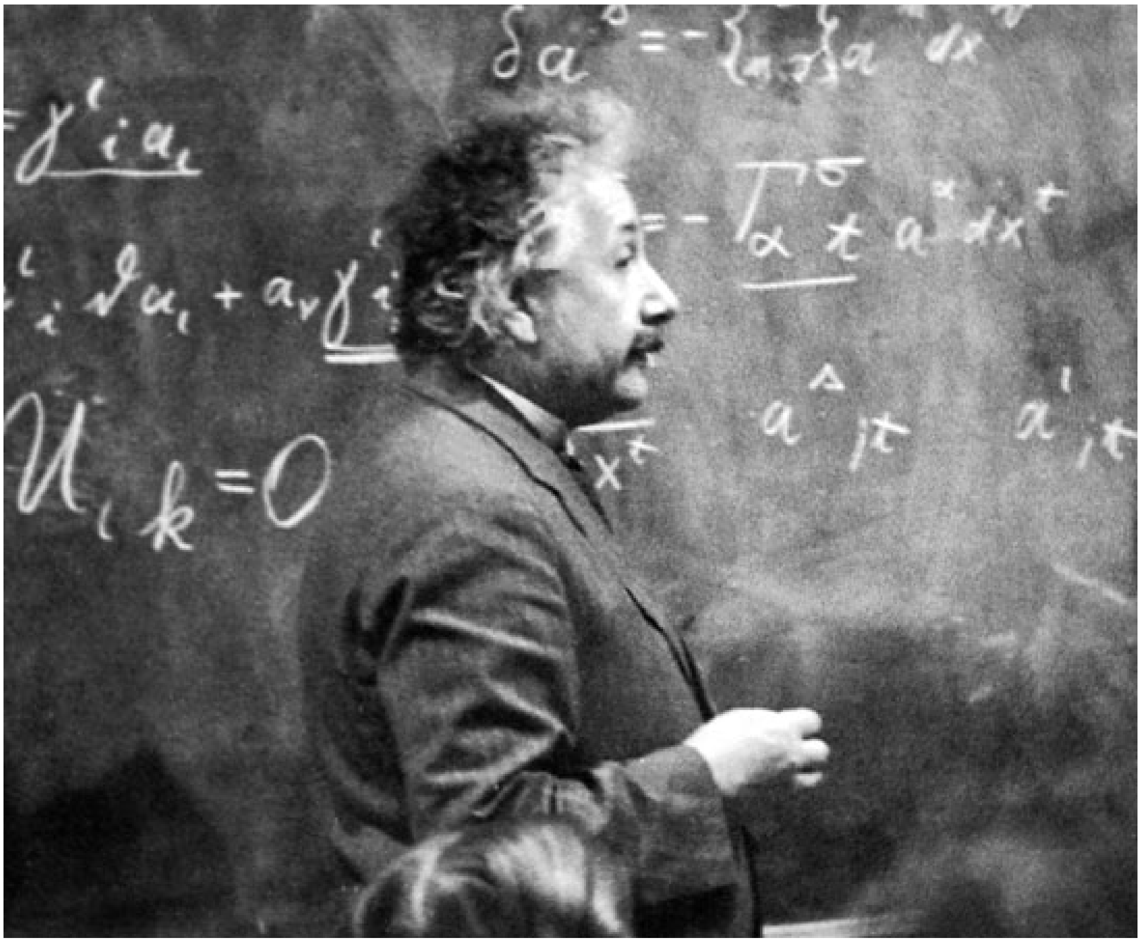

Figure 3: Einstein: “Since the mathematicians took over the theory of relativity I do no longer understand it.”

Chapter 1

Space and Time in Newtonian Physics

1.1 Space

The abstract concept of physical space reflects the properties of physical objects to have sizes, shapes and be located at different places relative to each other. In the pre-relativistic physics, the space is understood as a "room" which contains all physical objects. Hence it is considered as an independent physical reality. One can unambiguously determine whether an object is at rest or in motion in this space. The distances between physical objects can be reduced to the distances between points of this space where these bodies are located. This concept is called "the absolute space" and the motion in this space is "the absolute motion". As to the geometry of the absolute space, it is assumed to be Euclidean.

1.2 Time

The notion of time reflects our everyday-life observation that the world around us is continuously changing. That it can be viewed as a collection of events. These events may be instantaneous or have finite duration. In the pre-relativistic physics, it is assumed that all events are arranged in one unambiguous order. Hence a single periodic process, e.g. the Earth motion round the Sun, the motion of a pendulum etc., can be used as a reference to record this order.

This understanding of time leads naturally to the meaning of simultaneity as an absolute time relation. That is if somebody records two events as simultaneous, the same will be true for anyone else no matter where he is located and how he moves trough space, provided all sources of time-measurement errors are taken care of. Similarly, any event could be described by only one duration. It is easy to see why this concept of time is called the “absolute time”.

Both in theoretical and in practical terms, a unique temporal order of events can only be established if there exists a type of signals that propagate with infinite speed. In this case, when an event occurs in a remote place everyone can become aware of it instantaneously (!) by means of such “super-signals”. Hence, all events immediately divide into three groups with respect to this event: (i) The events simultaneous with it (signals from them arrive simultaneously with the event signal); (ii) The events preceding it (signals from them arrive earlier); (iii) The events following it (signals from them arrive later). If, however, there are no such super-signals, things become highly complicated as one needs to know not only the distances to the events but also the motion of the observer and the exact speed with which signals propagate through space.

Newtonian physics assumes that such infinite speed signals do exist and that they play a fundamental role in interactions between physical bodies. For example, in the Newtonian theory of gravity the gravitational force depends only on the current distance between two interacting bodies. So if one body changes its position then the other one “feels” this immediately, no matter how great the distance between them is.

1.3 Absolute Motion and Galilean relativity

It is relatively easy to detect and measure the motion of one physical body relative to another, the so-called “relative motion”. The same is true for the motion through some medium, provided one can interact with it. For example, a body moving through air experiences the air resistance, drag and lift forces as a result of this motion. This motion can also be considered as a kind of relative motion, the motion relative to this medium. What is about the absolute motion? How to detect it? If the absolute space was indeed like a room with walls then one would be able to detect the motion relative to these walls and call it the absolute motion but where are these walls and if they exist then what is beyond these walls? One can imagine that the whole space is filled with some hypothetical substance, e.g. "ether" or "plenum", and that one can detect the motion relative to it but will this be the absolute motion? Why would this substance be stationary in the space and hence serve as its "fabric"?

Our ancient predecessors used to believe that the Earth was at rest in the centre of the Universe (and hence in the absolute space). Hence the motion relative to the Earth could be considered as the absolute motion. The great Italian Galileo Galilei, who is regarded as the first true natural scientist, once made an observation which turned out to have far reaching consequences for modern physics. He noticed that it was impossible to tell whether a ship was anchored or coasting by means of mechanical experiments carried out inside the cabin of this ship. Nowadays, we use the generalised version of this conclusion known as the *Galilean Principle of Relativity*:

"It is impossible to detect the absolute motion by means of mechanical experiments."

In this formulation the focus is on the experiments. Theoretical physicists prefer a different formulation, which dictates the general form of equations in Theoretical Mechanics. It will be given a little bit later.

Obviously, the Galilean Principle of Relativity castes doubts over the whole idea of the absolute space. Yet, it is limited to the realm of mechanics only and hence leaves open the possibility of detecting the absolute motion by non-mechanical means.

1.4 Inertial frames of Newtonian mechanics

Newton (1643-1727) founded the classical mechanics - a basic set of mathematical laws of motion based on the ideas of absolute space and time described above. The First Law of Newtonian mechanics essentially states that

"The motion of a physical body which does not interact with other bodies (a free body) remains unchanged. It moves with constant speed along straight line in the absolute space."

If we choose a point O in the absolute space to serve as a reference point and denote as $\underline{r}$ the position vector (an arrow) whose base is anchored to O and the tip anchored to some physical body than the velocity of this body in the absolute space is defined as

$$\underline{w} = \frac{d\underline{r}}{dt} \tag{1.1}$$

and its acceleration as

$$\underline{a} = \frac{d\underline{w}}{dt} \,. \tag{1.2}$$

If we introduce Cartesian coordinates x^1, x^2 and x^3 centred on O then the components of the velocity aligned with the axes of this coordinate system of the absolute space are $w^i = dx^i/dt$ and the corresponding acceleration components are $a^i = dw^i/dt$. According to the First Law, for a free body $\underline{w}$ =const and $\underline{a} = 0$ which reads in the Cartesian components as w^i =const and $a^i = 0$.

Given the difficulties with detecting the absolute motion and hence finding a body at rest in the absolute space which can serve as the reference point "O" the First Law appears too abstract and impractical. In order to resolve this issue, the Newtonian mechanics comes up with the notion of "inertial reference frames". Each such frame can be considered as a solid coordinate grid (usually Cartesian but not always) moving with constant velocity as a whole (no rotation) in the absolute space. What is most important here is that the motion of free bodies relative to an inertial frames is the same as their absolute motion – they move with constant velocity.

1.5 Galilean transformation

Consider two Cartesian frames, S and $\tilde{S}$, with coordinates $\{x^i\}$ and $\{x^{\tilde{i}}\}$ respectively. Assume that (i) their corresponding axes are parallel, (ii) their origins coincide at time $t = 0$, (iii) frame $\tilde{S}$ is moving relative to S along the x^1 axis with constant speed v, as shown in Figure 1.1. This will be called the *standard configuration.*

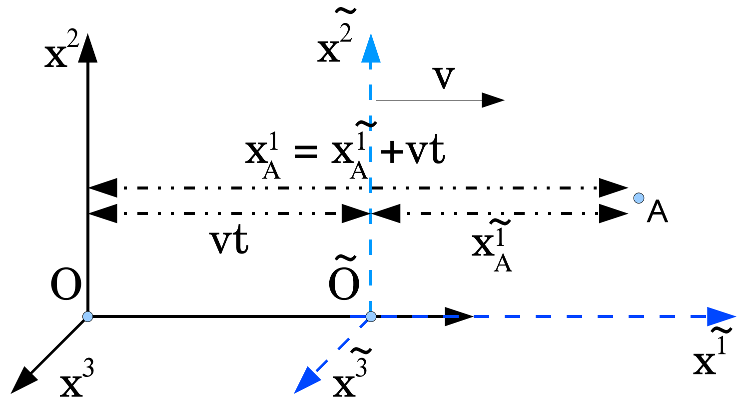

Figure 1.1: Measuring the x^1 coordinate of event A in two reference frames in the standard configuration. For simplicity, we show the case where this event occurs in the plane $x^3 = 0$.

The Galilean transformation relates the coordinates of events as measured in both frames. Given the absolute nature of time Newtonian physics, it is the same for both frames. So this may look over-elaborate if we write

$$t = \tilde{t}\,. \tag{1.3}$$

However, this make sense if we wish to stress that both fiducial observers, who ride these frames and make the measurements, use clocks made to the same standard and synchronised with each other.

Next let us consider the spatial coordinates of some event A. The x^2 coordinate of this event is its distance in the absolute space from the plane given by the equation $x^2 = 0$. Similarly, the $x^{\tilde{2}}$ coordinate of this event is its distance in the absolute space from the plane given by the equation $x^{\tilde{2}} = 0$. Since these two planes coincide both these distances are the same and hence $x^2 = x^{\tilde{2}}$. Similarly, we conclude that $x^3 = x^{\tilde{3}}$. As to the remaining coordinate, the distance between planes $x^1 = 0$ and $x^{\tilde{1}} = 0$ at the time of the event is vt, and hence $x^1 = x^{\tilde{1}} + vt$ (see Figure 1.1). Summarising,

$$
\begin{aligned}
x^1 &= x^{\tilde{1}} + vt, \\
x^2 &= x^{\tilde{2}}, \\
x^3 &= x^{\tilde{3}}.
\end{aligned}
\tag{1.4}
$$

This is the Galilean transformation for the standard configuration. If we allow the frame $\tilde{S}$ to move in arbitrary direction with velocity v^i then a more general result follows,

$$
x^i = x^{\tilde{i}} + v^i t. \tag{1.5}
$$

From this we derive the following two important conclusions:

- The velocity transformation law:

  $$
  w^i = w^{\tilde{i}} + v^i, \tag{1.6}
  $$

  where $w^i = dx^i/dt$ and $w^{\tilde{i}} = dx^{\tilde{i}}/dt$ are the velocities of a body as measured in frames S and $\tilde{S}$ respectively.

- The acceleration transformation law:

  $$
  a^i = a^{\tilde{i}}, \tag{1.7}
  $$

  where $a^i = d^2x^i/dt^2$ and $a^{\tilde{i}} = d^2x^{\tilde{i}}/dt^2$ are the accelerations of a body as measured in frames S and $\tilde{S}$ respectively. Thus, in both frames the acceleration is exactly the same.

Here we need to stress that these results are as fundamental as it gets, because they follow directly from the notions of absolute space and absolute time.

Now suppose that the frame S is at rest in the absolute space and hence the frame $\tilde{S}$ is an inertial frame moving in the absolute space and the fiducial observers riding these frames observe the motion of a free body. According to the First Law, the first observer obtains $a^i = 0$ and then according to Eq.(1.7) the inertial observer finds that $a^{\tilde{i}} = 0$ as well. Thus we reach the important conclusion that the motion of free bodies relative to any inertial frame is the same motion with constant velocity as relative to the absolute space. Thus, as far as the motion of free bodies is concerned, Newtonian mechanics is in full agreement with the Galilean relativity.

Equation 1.7 tells us that the body acceleration is the same in all inertial frames and it equals to that relative to the absolute space. This seems to give some solidity to the notion of absolute space but still does not deliver complete confidence. There exist infinitely many inertial frames and they all move relative to each other with constant speed. Only one of these frames is at rest in the absolute space (here we do not differentiate between frames with different orientation of their axes or/and locations of their origins) but we cannot tell which one. This make the absolute space a rather elusive if not ghostly "object".

1.6 The Second and Third Laws

If the motion of a body relative to the absolute space is in fact accelerated then it is not free from interactions with other bodies. The strength of this interaction is described by the force vector $\underline{f}$ and the body's ability to resist it by its inertial mass m. The precise meaning of this resistance is given by *the Second Law of Newtonian mechanics*

$$m\underline{a} = \underline{f}\,. \tag{1.8}$$

It is understood here that m is a parameter of the body as such and hence it does not depend on the force. Moreover, it is important to stress that Eq.(1.8) is not a definition of force. Instead, it describes the reaction of a body to the given force. Each kind of interaction involves additional laws determining the force vector as a function of other parameters (e.g the law of gravity). With this understanding, Eq.(1.8) determines the body's acceleration and mathematically it can be classified as a second order ordinary differential equation for the body's position vector $\underline{r}(t)$ in the absolute space.

The third law of Newtonian mechanics deals with binary interactions, or interactions involving only two bodies, say A and B. It states that

$$\underline{f}_a = -\underline{f}_b, \tag{1.9}$$

where $\underline{f}_a$ is the force acting on body A and $\underline{f}_b$ is the force acting on body B. Ultimately, this law leads to the conservation of momentum and mechanical energy.

So far we discussed these laws as applied in a reference frame at rest in the absolute space. Hence $\underline{f}$ is a vector in defined in the absolute space. How should these laws appear in an arbitrary inertial frame coasting through the absolute space? To be consistent with the Galilean relativity principle, they must have exactly the same form! Because both m and $\underline{a}$ are the same in all inertial frames, this means that $\underline{f}$ must be invariant as well and has to be described by equations which give the same vector in all inertial frames. This leads to the theoretician's version of the *Galilean Relativity principle*:

"All laws of mechanics must be invariant with respect to the Galilean transformation".

Let us consider the motion of a particle governed by equation (1.8). It is the same differential equation for any inertial frame. If we apply the same initial conditions (position and velocity at the initial time) then we will obtain exactly the same solution. In the context of mechanical experiments, this implies that if we setup the same experiment we obtain the same result, no matter how the laboratory is moving through the absolute space, provided this motion is not accelerated.

At this point, one can almost completely ignore the absolute space and consider the Newtonian mechanics as formulated for inertial frames. Some vectors in this formulation vary from frame to frame, like $\underline{r}$ and $\underline{v}$, others do not, like $\underline{a}$ and $\underline{f}$.

1.7 The lack of speed limit

Is there any speed limit a physical body can have in Newtonian mechanics? The answer to this question is No. To see this consider a particle of mass m under the action of constant force f. According to the second law of Newton its speed grows linearly

$$w = w_0 + \frac{f}{m}t$$

without a limit.

We can arrive to this conclusion using the velocity transformation law (1.6) as well. Suppose that in the frame $\tilde{S}$ of the standard configuration a body is moving along the x axis with the speed $\tilde{w} = 0.7w_{max} < w_{max}$, where w_{max} is the maximum possible speed, and that this frame moves relative to the frame S with the speed

$v = 0.7w_{max} < w_{max}$. Then according to Eq.(1.6) the speed of this body in the frame S is $w = 1.4w_{max} > w_{max}$ in conflict with the assumption that w_{max} is the maximum possible speed.

1.8 Light

The nature of light was a big mystery in Newtonian physics and a subject of heated debates. By analogy with sound waves, one point of view was that light was made by waves propagating in ether, the substance making the fabric of space. The speed of light was also of great interest to scientists. In contrast to sound, there is no phenomena in our everyday life which would clearly indicate that its speed is finite and as the result it was firmly believed that the it was infinite. There was also a strong philosophical reason for this assumption – as we have discussed above, signals with infinite speed would support very nicely the concept of absolute time. These ideas made light a fundamentally important phenomenon, intricately connected to the concepts of absolute space and time.

Some elaborate experiments have been made in attempts to measure the speed of light. One of the earliest is attributed to Galileo Galilei. It involves two people equipped with lanterns and placed far away from each other. They agree that initially they keep the lanterns covered until one of them rapidly uncovers his. In the meantime the other one waits till she sees the flash of light and then immediately uncovers hers. If the speed of light is finite then the first person sees the light of the second lantern only after some wait. Given the known distance between the two lanterns and the waiting time one can then than calculate the speed of light. Unfortunately, the waiting time was all too short to measure.

The first positive result came from astronomical observations. Dutch astronomer Roemer noticed that the motion of Jupiter's moons had systematic variation, which had a natural explanation only if light had a finite, though very large, speed. This phenomenon is related to the Doppler effect, which we will study later on. Since then many other measurements were made and they all agreed on the value for the speed of light

$$c \simeq 3 \times 10^{10} \text{cm/s}\,,$$

the value which did not seem to depend on the direction of light propagation. Within the framework of Newtonian physics, this could only be explained if in the ether the speed of light was the same in all direction and the Earth's speed through the ether (i.e. relative to the absolute space) was very small, below the accuracy of the speed of light measurements.

The development of the mathematical theory of electromagnetism resulted in the notions of electric and magnetic fields, which exist around electrically charged bodies. These fields do not manifest themselves in any other ways but via forces acting on other electrically charged bodies. Attempts do describe the properties of these fields mathematically resulted in Maxwell's equations, which agreed with the experiments most perfectly. What is the nature of electric and magnetic field? They could just reflect some internal properties of matter, like air, surrounding the electrically charged bodies. Indeed, it was found that the electric and magnetic fields depended on the chemical and physical state of surrounding matter. However, the experiments clearly indicated that the electromagnetic fields could also happily exist in vacuum (empty space). This fact prompted suggestions that in electromagnetism we are dealing with ether.

Analysis of Maxwell equations shows that electric and magnetic fields change via waves propagating with finite speed. In vacuum the speed of these waves is the same in all direction and equal to the known speed of light! When this had been discovered, Maxwell immediately interpreted light as electromagnetic waves or ether waves. Since according to the Galilean transformation the result of any speed measurement depends on the selection of inertial frame, the fact that Maxwell equations yielded a single speed could only mean that they are valid only in one particular frame, the rest frame of ether and hence absolute space.

However, Newtonian mechanics clearly shows that Earth cannot be exactly at rest in absolute space all the time. Indeed, it orbits the Sun and even if at one point of this orbit the speed of Earth's absolute motion is exactly zero it must be nonzero at all other points, reaching the maximum value equal to twice the orbital speed at the opposite point of the orbit. This simple argument shows that during one calendar year the speed of light should show variation of the order of the Earth orbital speed and that the speed of light should be different in different directions by at least the orbital speed. Provided the speed measurements

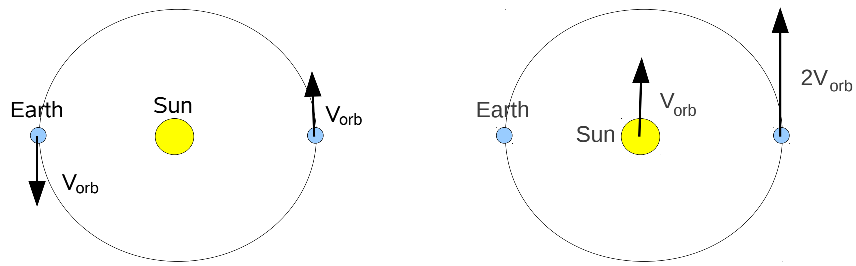

Figure 1.2: Left panel: Earth's velocity relative to the Sun at two opposite points of its orbit. Right panel: Earth's absolute velocity at two opposite points of its orbit and the Sun absolute velocity, assuming that at the left point the Earth velocity vanishes.

are sufficiently accurate we must be able to see these effects. American physicists Mickelson and Morley were first to design experiments of such accuracy (by the year 1887) and to everyone's amazement and disbelief their results were negative. Within their experimental errors, the speed of light was the same in all directions all the time! Since then, the accuracy of experiments has improved dramatically but the result is still the same, clearly indicating shortcomings of Newtonian physics with its absolute space and time. Moreover, no object has shown speed exceeding the speed of light. In his ground-braking work "On the electrodynamics of moving bodies", published in 1905, Albert Einstein paved way to new physics with completely new ideas on the nature of physical space and time, the Theory of Relativity, which accommodates these remarkable experimental findings.

1.9 Advanced material: Maxwell equations, electromagnetic waves, and Galilean invariance

1.9.1 Maxwell equations

Maxwell (1831-1879) completed the mathematical theory of electrodynamics. After his work, the evolution of electromagnetic field in vacuum is described by

$$\underline{\nabla}\cdot\underline{B} = 0, \tag{1.10}$$

$$\frac{1}{c}\frac{\partial \underline{B}}{\partial t} + \underline{\nabla}\times\underline{E} = 0, \tag{1.11}$$

$$\underline{\nabla}\cdot\underline{E} = 0, \tag{1.12}$$

$$-\frac{1}{c}\frac{\partial \underline{E}}{\partial t} + \underline{\nabla}\times\underline{B} = 0, \tag{1.13}$$

where c is a constant with dimension of speed. Laboratory experiments with electromagnetic materials allowed to measure this constant – it turned out to be equal to the speed of light!

Comment: Later, the works by Planck(1858-1947) and Einstein(1879-1955) lead to the conclusion that electromagnetic energy is emitted, absorbed, and propagate in discrete quantities, or photons. Thus, Newton's ideas have been partially confirmed as well. Such particle-wave duality is a common property of micro-particles that is accounted for in quantum theory.

1.9.2 Some relevant results from vector calculus

Notation: $\{x^k\}$ - Cartesian coordinates ($k = 1, 2, 3$); $\{\hat{\underline{e}}_k\}$ are the unit vectors along the x^k axes; $\underline{r} = x^k\hat{\underline{e}}_k$ is the position vector (radius vector) of the point with coordinates $\{x^k\}$; $\underline{A}(\underline{r}) = A^k(\underline{r})\hat{\underline{e}}_k$ is a vector field in Euclidean space (vector function); A^k are the components of vector $\underline{A}$ in the basis $\{\hat{\underline{e}}_k\}$; $f(\underline{r})$ is a scalar field in Euclidean space (scalar function).

The divergence of vector field $\underline{A}$ is defined as

$$\underline{\nabla}\cdot\underline{A} = \frac{\partial A^k}{\partial x^k}. \tag{1.14}$$

(Notice use of Einstein summation convention in this equation!). This is a scalar field. One can think of $\underline{\nabla}$ as a vector with components $\partial/\partial x^k$ and consider $\underline{\nabla}\cdot\underline{A}$ as a scalar product of $\underline{\nabla}$ and $\underline{A}$. The curl of vector field $\underline{A}$ is defined via the determinant rule for vector product

$$\underline{\nabla}\times\underline{A} = \begin{vmatrix} \hat{\underline{e}}_1 & \hat{\underline{e}}_2 & \hat{\underline{e}}_3 \\ \frac{\partial}{\partial x^1} & \frac{\partial}{\partial x^2} & \frac{\partial}{\partial x^3} \\ A^1 & A^3 & A^3 \end{vmatrix} = \tag{1.15}$$

$$= \underline{\hat{e}}_1(\frac{\partial A^3}{\partial x^2} - \frac{\partial A^2}{\partial x^3}) + \underline{\hat{e}}_2(\frac{\partial A^1}{\partial x^3} - \frac{\partial A^3}{\partial x^1}) + \underline{\hat{e}}_3(\frac{\partial A^2}{\partial x^1} - \frac{\partial A^1}{\partial x^2}). \tag{1.16}$$

This is a vector field. The gradient of a scalar field is defined as

$$\underline{\nabla} f = \frac{\partial f}{\partial x^1}\underline{\hat{e}}_1 + \frac{\partial f}{\partial x^2}\underline{\hat{e}}_2 + \frac{\partial f}{\partial x^3}\underline{\hat{e}}_3. \tag{1.17}$$

This is a vector field. The Laplacian of scalar field f is defined as

$$\nabla^2 f = \underline{\nabla}\cdot\underline{\nabla} f. \tag{1.18}$$

From this definition and eqs.(1.14,1.17) one finds that

$$\nabla^2 f = \frac{\partial^2 f}{\partial (x^1)^2} + \frac{\partial^2 f}{\partial (x^2)^2} + \frac{\partial^2 f}{\partial (x^3)^2}. \tag{1.19}$$

This is a scalar field. The Laplacian of vector field $\underline{A}$ is defined as

$$\nabla^2 \underline{A} = \underline{\hat{e}}_k \nabla^2 A^k. \tag{1.20}$$

(Notice use of Einstein summation convention in this equation!). This is a vector field. The following vector identity is very handy

$$\underline{\nabla}\times(\underline{\nabla}\times\underline{A}) = \underline{\nabla}(\underline{\nabla}\cdot\underline{A}) - \nabla^2\underline{A}. \tag{1.21}$$

1.9.3 Wave equation in electromagnetism

Apply $\underline{\nabla}\times$ to Eq.(1.11) to obtain

$$\frac{1}{c}\underline{\nabla}\times\frac{\partial \underline{B}}{\partial t} + \underline{\nabla}\times(\underline{\nabla}\times\underline{E}) = 0.$$

Using Eq.(1.21) and commutation of partial derivatives this reduces to

$$\frac{1}{c}\frac{\partial}{\partial t}\underline{\nabla}\times\underline{B} + \underline{\nabla}(\underline{\nabla}\cdot\underline{E}) - \nabla^2\underline{E} = 0.$$

The second term vanishes due to Eq.(1.12) and Eq.(1.13) allows us to replace $\underline{\nabla}\times\underline{B}$ with $\frac{\partial \underline{E}}{\partial t}$ in the first term. This gives us the final result

$$\frac{1}{c^2}\frac{\partial^2 \underline{E}}{\partial t^2} - \nabla^2\underline{E} = 0. \tag{1.22}$$

In a similar fashion one can show that

$$\frac{1}{c^2}\frac{\partial^2 \underline{B}}{\partial t^2} - \nabla^2 \underline{B} = 0. \tag{1.23}$$

These are examples of the canonical *wave equation*

$$\frac{1}{c^2}\frac{\partial^2 \psi}{\partial t^2} - \nabla^2 \psi = 0, \tag{1.24}$$

where $\psi(\underline{r}, t)$ is some function of space and time.

1.9.4 Plane waves

Look for solutions of Eq.(1.24) that depend only on t and x^1. Then

$$\frac{\partial \psi}{\partial x^2} = \frac{\partial \psi}{\partial x^3} = 0$$

and Eq.(1.24) reduces to

$$\frac{1}{c^2}\frac{\partial^2 \psi}{\partial t^2} - \frac{\partial^2 \psi}{\partial (x^1)^2} = 0. \tag{1.25}$$

This is a one-dimensional wave equation. It is easy to verify by substitution that it has solutions of the form

$$\psi_{\pm}(t, x^1) = f(x^1 \pm ct), \tag{1.26}$$

where $f(x)$ is an arbitrary twice differentiable function. $\psi_+(t, x^1) = f(x^1 + ct)$ describes waves propagating with speed c in the negative direction of the x^1 axis and $\psi_-(t, x^1) = f(x^1 - ct)$ describes waves propagating with speed c in the positive direction of the x^1 axis. Thus, equations (1.22,1.23), tell us straight away that Maxwell equations imply electromagnetic waves propagating with speed c, the speed of light.

1.9.5 Wave equation is not Galilean invariant

Can the electromagnetic phenomena be used to determine the absolute motion, that is motion relative to the absolute space. If like the equations of Newtonian mechanics the equations of electrodynamics are the same in all inertial frames then they cannot. Thus, it is important to see how the Maxwell equations transform under the Galilean transformation. However, it is sufficient to consider only the wave equation, Eq.(1.24), which is a derivative of Maxwell's equations. For simplicity sake, one can deal only with its one-dimensional version, Eq.(1.25). Denoting x^1 as simply x we have

$$-\frac{1}{c^2}\frac{\partial^2\psi}{\partial t^2}+\frac{\partial^2\psi}{\partial x^2}=0. \tag{1.27}$$

The Galilean transformation reads

$$\tilde{x}=x-vt.$$

It is easy to see that in new variables, $\{t,\tilde{x}\}$, equation 1.27 becomes

$$-\frac{1}{c^2}\frac{\partial^2\psi}{\partial t^2}+\frac{2v}{c^2}\frac{\partial^2\psi}{\partial t\partial\tilde{x}}+\left(1-\frac{v^2}{c^2}\right)\frac{\partial^2\psi}{\partial\tilde{x}^2}=0. \tag{1.28}$$

Since Eq.(1.28) has a different form compared to Eq.(1.27) we conclude that *the wave equation, and hence the Maxwell equations, are not invariant under Galilean transformation! One should be able to detect motion relative to the absolute space!*

In order to elucidate this result consider the wave solutions of Eq.(1.28). Direct substitution shows that this equation is satisfied by

$$\Psi_\pm(t,\tilde{x})=f(\tilde{x}+a_\pm t), \tag{1.29}$$

where $f(x)$ is again an arbitrary twice differentiable function and

$$a_\pm=v\pm c. \tag{1.30}$$

These are waves propagating with speeds $a_\pm$. In fact, $\Psi_+(t,\tilde{x})=f(\tilde{x}+(v+c)t)$ describes wave propagating with speed

$$\tilde{c}=-v-c$$

and $\Psi_-(t,\tilde{x})=f(\tilde{x}+(v-c)t)$ describes wave propagating with speed

$$\tilde{c}=-v+c$$

Comparing these results with Eq.(1.6 shows us that what we have got here is simply the Galilean velocity transformation for electromagnetic waves. Thus, we arrive to the following conclusions

1. Electromagnetic waves can propagate with speed c in all directions only in one very special inertial frame, namely the frame that is in rest in absolute space. In any other frame it will be different in different directions, as dictated by the Galilean transformation, and equal to c only in the directions normal to the frame velocity relative to the absolute space.

2. The Maxwell equations are not general. They hold only in the frame at rest in the absolute space.

In spite of looking very convincing these conclusions however do not comply with physical experiments which show beyond any doubt that in all frames the electromagnetic waves propagate with the same speed c in all directions! These experimental results show that the Galilean transformation is not that general as thought before Einstein, and hence the notions of space and time as described of Newtonian physics are not correct.

Chapter 2

Basic Special Relativity

2.1 Einstein's postulates

Paper "On the electrodynamics of moving bodies" by Einstein (1905).

- *Postulate 1 (Principle of Relativity):*

  "*All physical laws are the same (invariant) in all inertial frames*".

- *Postulate 2:*

  "*The speed of light (in vacuum) is the same in all inertial frames* ".

Postulate 1 implies that no physical experiment can be used to measure the absolute motion. In other words the notions of absolute motion and absolute space become redundant. As far as physics is concerned the absolute space does not exist! The inertial frames of Newtonian are defined using the notion of absolute space. They are still used in Special Relativity but they are defined in a different way where the absolute space is not involved. Namely, *a reference frame is called inertial if any particle which is not subjected to an external force moves with constant velocity relative to this frame.*

Postulate 2 is fully consistent with Postulate 1 (and hence may be considered as a derivative of Postulate 1). Indeed, if Maxwell's equations (1.10-1.13) are the same in all inertial frames then the electromagnetic waves propagate with the same speed which is given by the constant c in these equations. Since Postulate 2 is in conflict with the Galilean velocity addition it shows that the very basic properties of physical space and time have to be reconsidered. In the next section we carry out a number of very simple “thought experiments” which show how dramatic the required modifications are.

2.2 Einstein’s thought experiments

In these experiments we adopt the Speed of Light postulate and agree that whenever and wherever we encounter light, it invariable has the same speed, which we will denote as c.

2.2.1 Experiment 1. Relativity of simultaneity

A carriage is moving with speed v past a platform. A conductor, who stands in the middle of the carriage, sends simultaneously two light pulses in the opposite directions along the track. Both passengers and the crowd waiting on the platform observe how the pulses hit the ends of the carriage.

The carriage passengers agree that both pulses reach the ends simultaneously as they propagate with the same speed and have to cover the same distance (see left panel of figure 2.1). If $\tilde{L}$ is the carriage length as measured by its passengers then the required time is

$$\tilde{t}_{left} = \tilde{t}_{right} = \tilde{L}/2c\,.$$

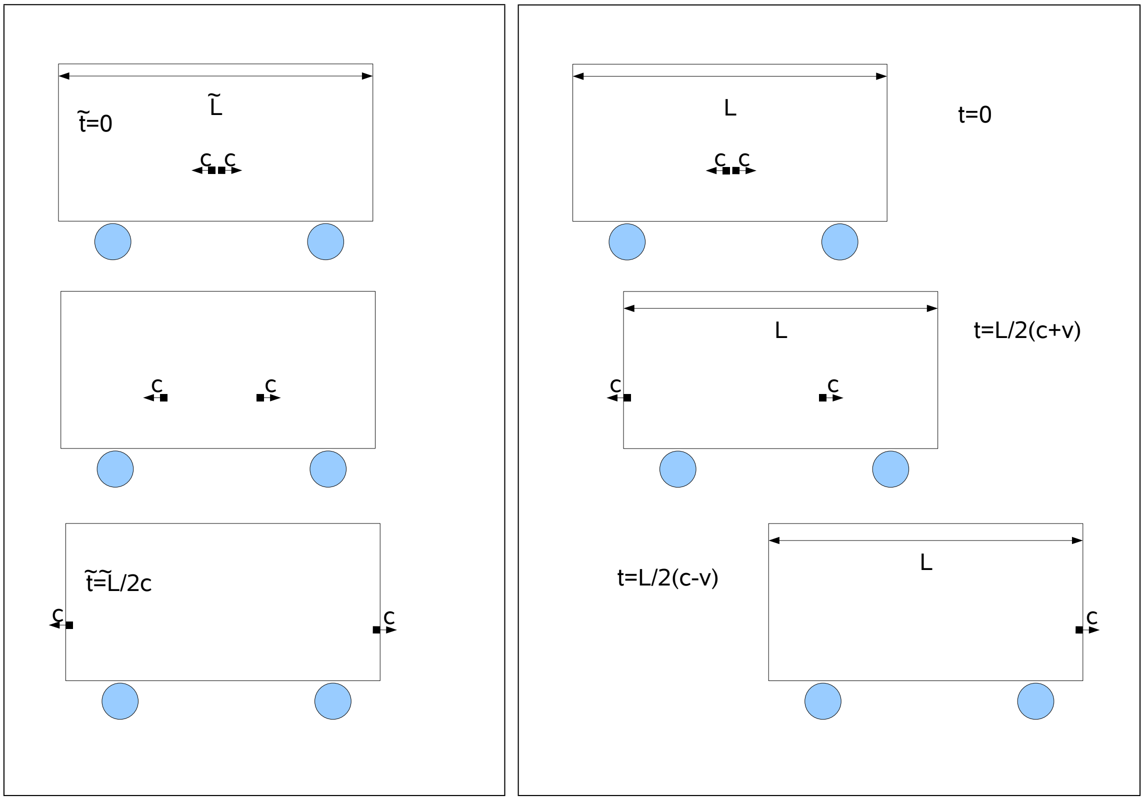

Figure 2.1: Thought experiment number 1. *Left panel:* Events as seen in the carriage frame. *Right panel:* Events as seen in the platform frame.

The crowd on the platform, however, see things differently. As both ends of the carriage move to the right the pulse sent to the left has to cover shorter distance and reaches its end earlier than the pulse sent to the right (right panel of figure 2.1). In fact, the required times are

$$t_{left} = \frac{L}{2}\left(\frac{1}{c+v}\right) \quad \text{and} \quad t_{right} = \frac{L}{2}\left(\frac{1}{c-v}\right),$$

where L is the carriage length as measured by the crowd (As we shell see later $L \neq \tilde{L}$ but this does not matter here.) Since $t_{left} \neq t_{right}$, these two events, which are simultaneous in the carriage frame, are not simultaneous in the platform frame. *This implies that the temporal order of events depends of the frame of reference and hence that the absolute time does no longer exists. Instead, each inertial frame must have its own time.* The next experiment supports this conclusion, showing that the same events may have different durations in different frames.

2.2.2 Experiment 2. Time dilation

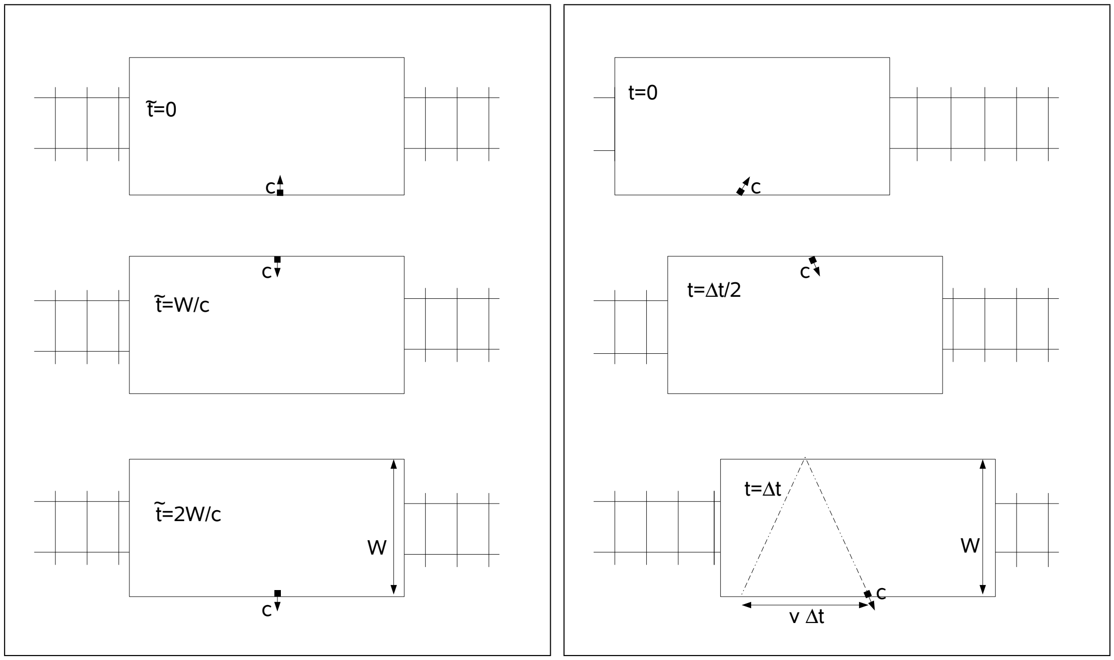

Figure 2.2: Thought experiment number 2. *Left panel:* Events as seen in the carriage frame. *Right panel:* Events as seen in the platform frame.

This time a single pulse is fired from one side of the carriage perpendicular to the track, reflects off the other side and returns back. In the carriage frame (left panel of figure 2.2) the pulse covers the distance $2W$, where W is the carriage width, and this takes time

$$\Delta\tilde{t} = \frac{2W}{c}.$$

In order to find Δt, the elapsed time as measured in the platform frame, we notice that we can write the distance covered by the pulse as $c\Delta t$ and as $2\sqrt{W^2 + (v\Delta t/2)^2}$ (right panel of figure 2.2). Thus,

$$(c\Delta t/2)^2 = W^2 + (v\Delta t/2)^2.$$

From this we find

$$\Delta t = \frac{2W}{c}\frac{1}{\sqrt{1 - v^2/c^2}} = \Delta\tilde{t}\frac{1}{\sqrt{1 - v^2/c^2}}.$$

Introducing the so-called Lorentz factor

$$\gamma = 1/\sqrt{1 - v^2/c^2}. \tag{2.1}$$

we can write this result as

$$\Delta t = \gamma \Delta \tilde{t}. \tag{2.2}$$

This shows us that not only simultaneity is relative but also the duration of events. (Notice that we assumed here that the carriage width is the same in both inertial frames. We will come back to this assumption later.)

Proper time and time of inertial frames.

The best way of measuring time is via some periodic process. Each standard clock is based on such a process. Its time is called the *proper time* of the clock. When instead of a clock we have some physical body it is also useful to introduce the proper time of this body – this can be defined as the time that would be measured by a standard clock moving with this body. One can also use proper time to describe time-separation between two events but only if there exists an inertial frame where these two events occur at the same place. In this case, it is the time that would be measured by a standard clock at rest in this location. We will denote the proper time using the Greek letter τ.

As to the *time of inertial frame*, it is measured by standard clocks which are at rest in this frame (co-moving with this it). Although in principle just one clock could be sufficient, it is best to think of the time as based on an imaginary system of many synchronised clocks located in convenient places (or even at every point of the frame). Indeed, a single clock is fine for measuring time of events which occur at the same location as this clock. For remote events one could rely on this single clock only when it is supported by a communication system of some sort which reports the distances to these events so that the associated time delays between their actual occurrence and the time when they are seen at the clock can be accounted for. Alternatively, one could use a whole grid of closely spaced synchronised standard clocks at rest in this frame. When an event occurs one simply records the time of the clock which happens to be at the same location as the event.

In this thought experiment, $\Delta\tilde{t}$ could measured by means of conductor's standard clock only – the pulse is fired from its location and then comes back to its location. Thus, $\Delta\tilde{t}$ is the proper time of the conductors clock. In contrast, Δt is the time of the inertial frame associated with the platform and not with the proper time of any of its clocks. Indeed, two clocks, located at points A and B, are required to determine this time interval.

Using these definitions, we can generalise the result (2.2) and write

$$\Delta t = \gamma \Delta \tau, \tag{2.3}$$

where τ is the proper time of some standard clock and t is the time of the inertial frame where this clock moves with the Lorentz factor γ.

One can introduce a proper time for any physical body, defining it as the time of an imaginary standard clock co-moving with it. Moreover, one can also define proper time for two separate events, but only if there exists a frame where they occur at the same place. This will be the time as measured in this frame and hence measured by a standard clock of this frame at the same location as the events. In the experiment, the proper time between the firing of the pulse and its return is measured by the conductor's clock.

Since $\gamma \geq 1$ by definition, equation (2.3) implies that $\Delta t \geq \Delta\tau$, with $\Delta t = \Delta\tau$ when $v = 0$. What does this mean? One can say that according to the time system of any inertial frame, any clock moving relative to it slows down. Since this result does not depend on the physical nature of the clock mechanism it inevitably implies that all physical processes within a moving body slow down compared to the processes of a similar body at rest in this inertial frame when measured using the time of this frame. When a different frame is used a different result may be obtained. In particular, in the frame of the aforementioned moving body the opposite is observed. Indeed in this frame the originally moving body is at rest and the originally stationary body is moving.

This effect is called the *time dilation*.

2.2.3 Experiment 3. Length contraction

This time the light pulse is fired from one end of the carriage along the track, gets reflected of the other end and comes back. In the carriage frame the time of pulse journey in both directions is equal to $\tilde{L}/c$ where $\tilde{L}$ is the carriage length as measured in the carriage frame (left panel of figure 2.3). Thus, the total time of the pulse journey is

$$\Delta\tilde{t} = 2\tilde{L}/c.$$

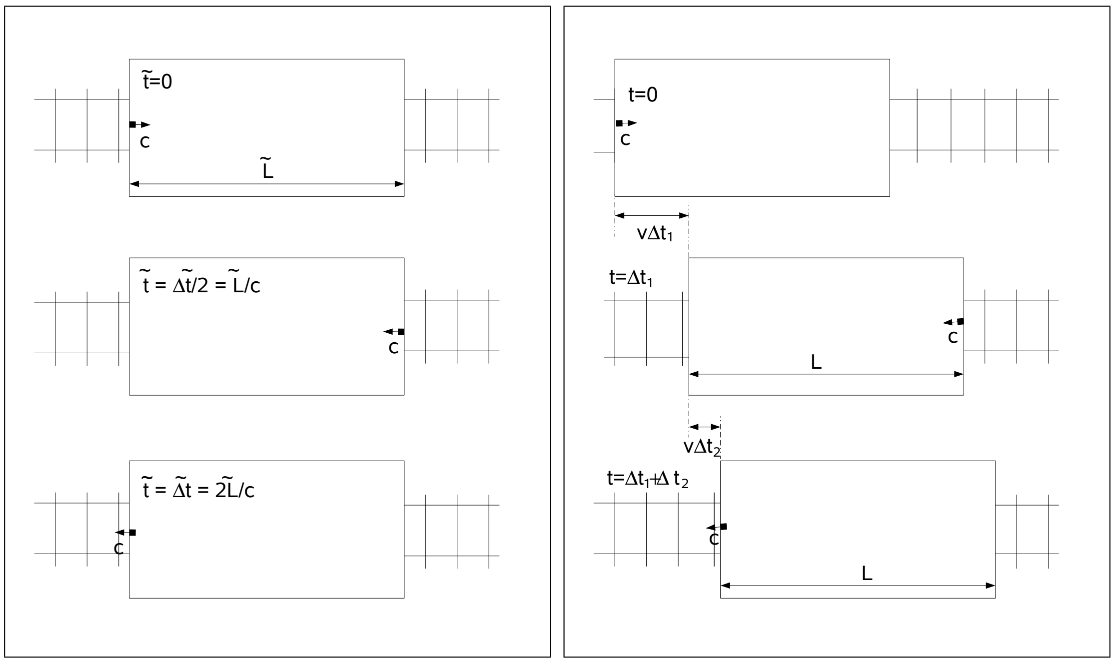

Figure 2.3: Thought experiment number 3. *Left panel:* Events as seen in the carriage frame. *Right panel:* Events as seen in the platform frame.

In the platform frame the first leg of the journey (before the reflection) takes some time Δt_1 and the second leg takes Δt_2 which is less then Δt_1 because of the carriage motion. To find Δt_1 we notice that the distance covered by the pulse during the first leg can be expressed as $c\Delta t_1$ and also as $L + v\Delta t_1$, where L is the length of the carriage as measured in the platform frame (see the right panel of figure 2.3). Thus,

$$\Delta t_1 = L/(c - v).$$

To find Δt_2 we notice that the distance covered by the pulse during the second leg can be expressed as $c\Delta t_2$ and also as $L - v\Delta t_2$, (see the right panel of figure 2.3). Thus,

$$\Delta t_2 = L/(c + v).$$

The total time is

$$\Delta t = \Delta t_1 + \Delta t_2 = \frac{L}{c-v} + \frac{L}{c+v} = \frac{2L}{c}\gamma^2.$$

Since $\Delta\tilde{t}$ is actually a proper time interval we can apply Eq.(2.3) and write

$$\Delta t = \gamma\Delta\tilde{t} = \gamma\frac{2\tilde{L}}{c}.$$

Combining the last two results we obtain

$$\frac{2L}{c}\gamma^2 = \gamma\frac{2\tilde{L}}{c}$$

which gives us

$$L = \tilde{L}/\gamma \tag{2.4}$$

Thus, the length of the carriage in the platform frame is different from that in the carriage frame. *This shows that if we accept that the speed of light is the same in all inertial frames then we have to get rid of the absolute space as well!*

Similarly to the definition of the proper time interval, the *proper length* of an object, which we will denote as L_0, is defined as the length measured in the frame where this object is at rest. In this experiment the proper length is $\tilde{L}$. Thus we can write

$$L = L_0/\gamma \tag{2.5}$$

In this equation the lengths are measured along the direction of relative motion of two inertial frames. Since, $\gamma > 1$ we conclude that L is always shorter than L_0. Hence the name of this effect – *length contraction.*

Consider two identical bars. When they are rested one alongside the other they have exactly the same length. Set them in relative motion in such a way that they are aligned with the direction of motion (see figure 2.4). In the frame where one bar is at rest the other bar is shorter, and the other way around. At first this may seem contradictory. However, this is in full agreement with the Principle of Relativity. In both frames we observe the same phenomenon – the moving bar becomes shorter. Moreover, the relativity of simultaneity explains how this can be actually possible. In order to measure the length of the

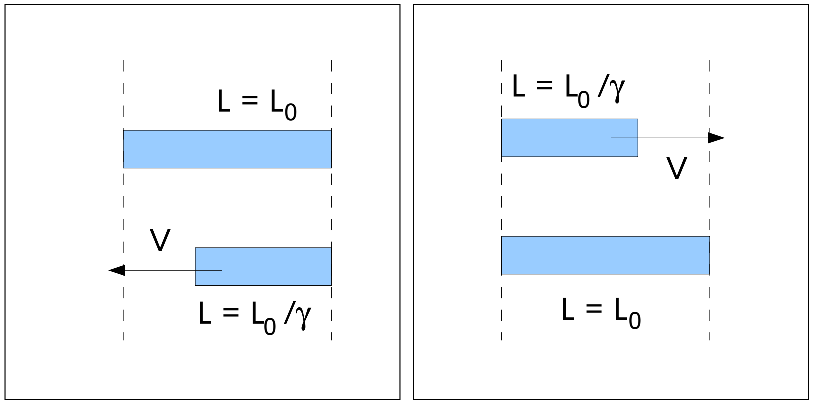

Figure 2.4: Two identical bars are aligned with the direction of their relative motion. In the rest frames of both bars the same phenomenon is observed – a moving bar is shorter.

moving bar the observer should mark the positions of its ends simultaneously and then to measure the distance between the marks. This way the observer in the rest frame of the upper bar in Figure 2.4 finds that the length of the lower bar is $L = L_0/\gamma$. However, the observer in the rest frame of the lower bar finds that the positions are not marked simultaneously, but the position of the left end is marked before the position of the right end (Recall the thought experiment 1 in order to verify this conclusion.). As the result, the distance between the marks, L' is even smaller than L_0/γ, in fact the actual calculations give $L' = L_0/\gamma^2 = L/\gamma$, in agreement with results obtained in the frame of the upper bar.

Lengths measured perpendicular to the direction of relative motion of two inertial frames must be the same in both frames. To show this, consider two identical bars perpendicular to the direction of motion (see Fig. 2.5). In this case there is in no need to know the simultaneous positions of the ends as they do not move in the direction along which the length is measured and, thus, the relativity of simultaneity is no longer important. For example, one could use two strings stretched parallel to the x axis so that the ends of one of the bars slide along these strings (Fig. 2.5). By observing whether the other bar fits between these strings or not one can decide if it is longer or shorter in the absolute sense – it does not matter which inertial observer makes this observation, the result will be the same. Let us say that the right bar

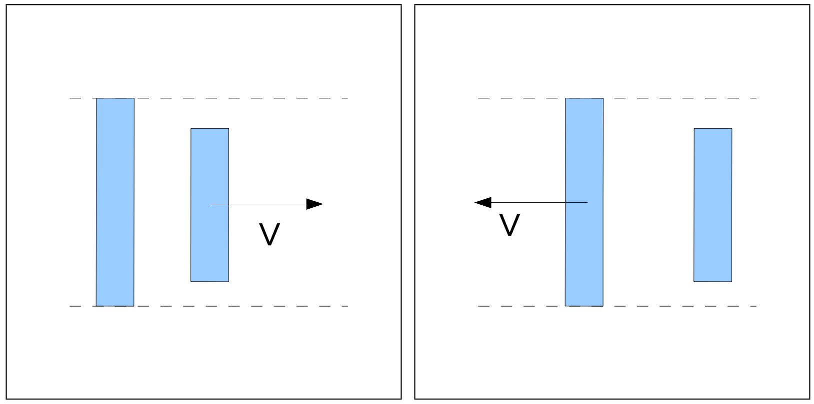

Figure 2.5: Two identical bars are aligned perpendicular to the direction of their relative motion. If the bars did not retain equal lengths then the equivalence of inertial frames would be broken. In one frame a moving bar gets shorter, whereas in the other it gets longer. This would contradict to the Principle of Relativity.

in Figure 2.5 is shorter. Then in the frame of the left bar moving bars contract, whereas in the frame of the right bar moving bars lengthen. This breaks the equivalence of inertial frames postulated in the Relativity Principle. Similarly, we show that the Relativity Principle does not allow the right bar to be longer then the left one. There is no conflict with this principle only if the bars have the same length.

2.2.4 Synchronisation of clocks

Consider a set of standard clocks placed on the carriage deck along the track. To make sure that all these clocks can be used for consistent time measurements the carriage passengers should synchronise them. This can be done by selecting the clock in the middle to be a reference clock and then by making sure that all other clocks show the same time simultaneously with the reference one. One way of doing this is by sending light signals at time t_0 from the reference clock to all the others. When a carriage clock receives this signal it should show time $t = t_0 + l/c$, where l is the distance to the reference clock. Now consider set of standard clocks placed on the platform along the track. This set can also be synchronised using the above procedure (now the reference clock will be in the middle of the platform). Recalling the result of the thought experiment 1 we are forced to conclude that to the platform crowd the carriage clocks will appear desynchronised (see fig.2.6) and the other way around – to the carriage passengers the platform clocks will appear desynchronised (see fig.2.7). That is if the passengers inside

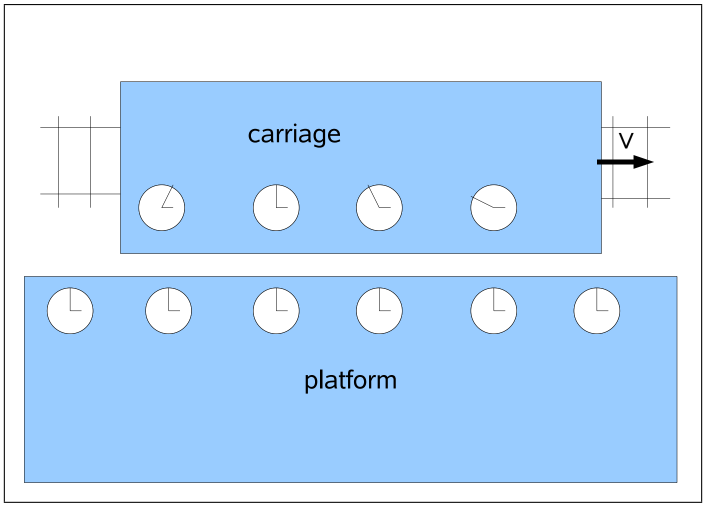

Figure 2.6: Clocks synchronised in the carriage frame appear desynchronised in the platform frame.

the carriage standing next to each of the carriage clocks are asked to report the time shown by the platform clock which is located right opposite to his clock when his clock shows time t_1 their reports will all have different readings, and the other way around. So in general, a set of clocks synchronised in one inertial frame will be appear as desynchronised in another inertial frame moving relative to the first one.

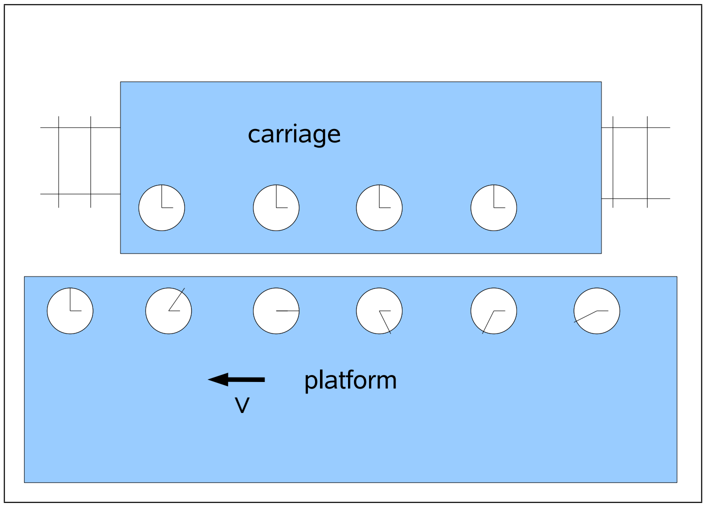

Figure 2.7: Clocks synchronised in the platform frame appear desynchronised in the carriage frame.

2.3 Lorentz transformation

In Special Relativity the transition from one inertial frame to another (in standard configuration) is no longer described by the Galilean transformation but by the Lorentz transformation. This transformation ensures that light propagates with the same speed in all inertial frames.

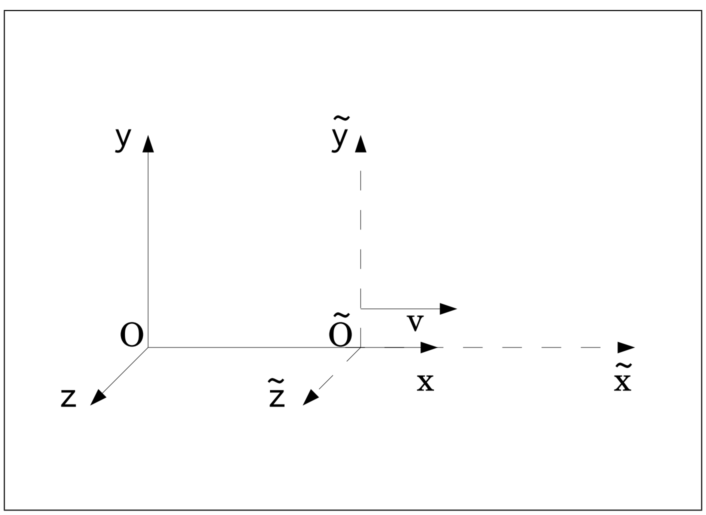

Figure 2.8: Two inertial frames in standard configuration

2.3.1 Derivation

The Galilean transformation

$$\begin{cases} t=\tilde{t} \\ x=\tilde{x}+v\tilde{t} \\ y=\tilde{y} \\ z=\tilde{z} \end{cases} \qquad \begin{cases} \tilde{t}=t \\ \tilde{x}=x-vt \\ \tilde{y}=y \\ \tilde{z}=z \end{cases} \tag{2.6}$$

is inconsistent with the second postulate of Special Relativity. The new transformation should have the form

$$\begin{cases} t=f(\tilde{t},\tilde{x},v) \\ x=g(\tilde{t},\tilde{x},v) \\ y=\tilde{y} \\ z=\tilde{z} \end{cases} \qquad \begin{cases} \tilde{t}=f(t,x,-v) \\ \tilde{x}=g(t,x,-v) \\ \tilde{y}=y \\ \tilde{z}=z \end{cases} \tag{2.7}$$

just because of the symmetry between the two frames. Indeed, the only difference between the frames S and $\tilde{S}$ is the direction of relative motion: If $\tilde{S}$ moves with speed v relative to S then S moves with speed $-v$ relative to $\tilde{S}$. Hence comes the change in the sign of v in the equations of direct and inverse transformations (2.7). Y and z coordinates are invariant because lengths normal to the direction of motion are unchanged (see Sec.2.2.3). Now we need to find functions f and g.

(2.6) is a linear transformation. Assume that (2.7) is linear as well (if our derivation fall through we will come back and try something less restrictive). Then

$$x=\gamma(v)\tilde{x}+\delta(v)\tilde{t}+\eta(v). \tag{2.8}$$

Clearly, we should have $\tilde{x}=-v\tilde{t}$ for any $\tilde{t}$ if $x=0$. Thus,

$$(-v\gamma+\delta)\tilde{t}+\eta(v)=0$$

for any $\tilde{t}$. This requires

$$\eta=0 \quad \text{and} \quad \delta=v\gamma.$$

Thus,

$$x = \gamma(\tilde{x}+v\tilde{t}), \tag{2.9}$$

$$\tilde{x} = \gamma(x-vt). \tag{2.10}$$

In principle, the symmetry of direct and inverse transformation is preserved both if $\gamma(-v) = \gamma(v)$ and $\gamma(-v) = -\gamma(v)$. However, it is clear that for $\tilde{x} \to +\infty$ we should have $x \to +\infty$ as well. This condition selects $\gamma(-v) = \gamma(v) > 0$.

Now to the main condition that allows us to fully determine the transformation. Suppose that a light signal is fired at time $t = \tilde{t} = 0$ in the positive direction of the x axis. (Note that we can always ensure that the standard time-keeping clocks located at the origins of S and $\tilde{S}$ show the same time when the origins coincide.) Eventually, the signal location will be $x = ct$ and $\tilde{x} = c\tilde{t}$. Substitute these into Eq.(2.10,2.9) to find

$$ct = \gamma c\tilde{t}(1 + v/c), \tag{2.11}$$

$$c\tilde{t} = \gamma ct(1 - v/c). \tag{2.12}$$

Now we substitute $c\tilde{t}$ from Eq.(2.12) into Eq.(2.11)

$$ct = \gamma^2 ct(1 - v/c)(1 + v/c)$$

and derive

$$\gamma = 1/\sqrt{(1 - v^2/c^2)}. \tag{2.13}$$

We immediately recognise the Lorentz factor.

In order to find function g of Eq.2.7 we simply substitute x from Eq.2.9 into Eq.2.10 and then express t as a function of $\tilde{t}$ and $\tilde{x}$:

$$\tilde{x} = \gamma[\gamma\tilde{x} + v\gamma\tilde{t} - vt],$$

$$\gamma vt = \gamma^2 v\tilde{t} - \tilde{x}(1 - \gamma^2).$$

It is easy to show that

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

Thus,

$$\gamma vt = \gamma^2 v\tilde{t} + \frac{v^2\gamma^2}{c^2}\tilde{x}$$

and finally

$$t = \gamma\left(\tilde{t} + \frac{v}{c^2}\tilde{x}\right). \tag{2.14}$$

Summarising, the coordinate transformations that keep the speed of light unchanged are

$$\begin{cases} t = \gamma(\tilde{t} + (v/c^2)\tilde{x}) \\ x = \gamma(\tilde{x} + v\tilde{t}) \\ y = \tilde{y} \\ z = \tilde{z} \end{cases} \qquad \begin{cases} \tilde{t} = \gamma(t - (v/c^2)x) \\ \tilde{x} = \gamma(x - vt) \\ \tilde{y} = y \\ \tilde{z} = z \end{cases} \tag{2.15}$$

They are due to Lorentz(1853-1928) and Larmor(1857-1942).

2.3.2 Newtonian limit

Consider the Lorentz transformations in the case of $v \ll c$. This is the realm of our everyday life. In fact even the fastest rockets fly with speeds which are much less than the speed of light. In order to obtain the equations of this limit we simply set $v/c \to 0$. Hence,

$$\gamma = (1 - v^2/c^2)^{-1/2} \to 1$$

and the transformation law for the x coordinate reduces from

$$x = \gamma(\tilde{x} + v\tilde{t}) \quad \tilde{x} = \gamma(x - vt)$$

to the old good Galilean form

$$x = \tilde{x} + v\tilde{t} \quad \tilde{x} = x - vt.$$

This is why the Galilean transformation appears to work so well.

Similarly, we find that the time equation of the Lorentz transformation reduces to

$$t = \tilde{t} + (v/c^2)\tilde{x} = \tilde{t} + (v/c)(\tilde{x}/c) \to \tilde{t}\,.$$

(The Newtonian limit can be reached by letting $c \to +\infty$, or just replacing c with $+\infty$.)

In order to understand the magnitude of relativistic corrections to the Galilean transformation under the typical conditions of our life experiences, we can use Maclaurin expansion of the Lorentz equations where we keep the first two terms. For an airliner, the typical speed is about $v = 300\text{m/s} = 3 \times 10^4\text{cm/s}$. Since $c = 3 \times 10^3\text{cm/s}$ we have $v/c = 10^{-6}$. The corresponding Lorentz factor

$$\gamma = (1 - (v/c)^2)^{-1/2} \simeq 1 + \frac{1}{2}(v/c)^2 \approx 1 + 10^{-12}\,.$$

Hence,

$$x = \gamma(\tilde{x} + v\tilde{t}) \simeq \tilde{x} + v\tilde{t} + 10^{-12}(\tilde{x} + v\tilde{t})\,.$$

For the distances as large as 3000 km, the correction is about $10^{-12} \times 3000\,\text{km} = 3 \times 10^{-4}\text{cm}$ only. Similarly, we find the corresponding time correction

$$t = \gamma\left(\tilde{t} + (v/c)^2(\tilde{x}/v)\right) \simeq (1 + 10^{-12})(\tilde{t} + 10^{-12}(\tilde{x}/v)) \approx \tilde{t} + 10^{-8}\text{s}\,.$$

2.3.3 Basic relativistic effects

Relativity of simultaneity

From equation (2.15) it follows that for any two events

$$\Delta t = \gamma(\Delta\tilde{t} + \frac{v}{c^2}\Delta\tilde{x})\,. \tag{2.16}$$

Hence the fact that $\Delta\tilde{t} = 0$ does not mean that $\Delta t = 0$. In fact,

$$\Delta t = \gamma\frac{v}{c^2}\Delta\tilde{x}\,,$$

which vanishes only if $\Delta\tilde{x} = 0$ as well.

Time dilation

Consider a standard clock at rest in the frame $\tilde{S}$. Hence for any two events in "life" of this clock $\Delta\tilde{x} = 0$ and $\Delta\tilde{t} = \Delta\tau$, where τ is the proper time of the clock. For such events equation (2.16) reads

$$\Delta t = \gamma\Delta\tau\,.$$

This is the time dilation formula.

Length contraction

From equation (2.15) it follows that for any two events

$$\Delta x = \gamma(\Delta\tilde{x} + v\Delta\tilde{t})\,. \tag{2.17}$$

Consider a bar of proper length l_0 at rest in the frame S and aligned with the x axis. Now let $\tilde{x}_1$ to be the coordinate of the left end of the bar at the time $\tilde{t}_1$ and $\tilde{x}_2$ to be the coordinate of the left end of the bar at the time $\tilde{t}_2$, $\Delta\tilde{x} = \tilde{x}_2 - \tilde{x}_1$ and $\Delta\tilde{t} = \tilde{t}_2 - \tilde{t}_1$. Whatever they are the corresponding $\Delta x = l_0$, the proper length of the bar. However in the frame $\tilde{S}$ the bar is moving along the x axis and hence $\Delta\tilde{x}$ will be the length l of the bar in this frame only if $\Delta\tilde{t} = 0$ – we need simultaneous positions of the ends! Hence, we find

$$l_0 = \gamma l\,,$$

which is the length contraction formula.

2.4 Relativistic velocity ”addition”

For $|v| > c$ the Lorentz factor

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

becomes imaginary and the equations of Lorentz transformation, as well as the time dilation and Lorentz contraction equations, become meaningless. This suggests that c **is the maximum possible speed in nature.** How can this possibly be the case ? If in the frame $\tilde{S}$ we have a body moving with speed $\tilde{w} > c/2$ to the right and this frame moves relative to the frame S with speed $v > c/2$, then in the frame S this body should move with speed $w > c/2 + c/2 = c$. However, in this calculation we have used the velocity addition law of Newtonian mechanics, $w = \tilde{w} + v$, which is based on the Galilean transformation, not the Lorentz transformation! So what does the Lorentz transformation tells us in this regard?

2.4.1 One-dimensional velocity "addition"

Consider a particle moving in the frame $\tilde{S}$ with speed $\tilde{w}$ along the x axis. Then we can write

$$\frac{d\tilde{x}}{d\tilde{t}} = \tilde{w}, \quad \frac{d\tilde{y}}{d\tilde{t}} = \frac{d\tilde{z}}{d\tilde{t}} = 0,$$

and

$$\frac{dx}{dt} = w, \quad \frac{dy}{dt} = \frac{dz}{dt} = 0.$$

From the first two equations of the Lorentz transformation (2.15) one has

$$\begin{cases} dt = \gamma(d\tilde{t} + (v/c^2)d\tilde{x}) \\ dx = \gamma(d\tilde{x} + vd\tilde{t}) \end{cases} . \tag{2.18}$$

Thus,

$$w = \frac{dx}{dt} = \frac{d\tilde{x} + vd\tilde{t}}{d\tilde{t} + (v/c^2)d\tilde{x}} =$$

$$= \frac{d\tilde{x}/d\tilde{t} + v}{1 + (v/c^2)d\tilde{x}/d\tilde{t}} = \frac{\tilde{w} + v}{1 + (v\tilde{w}/c^2)}.$$

The result is the relativistic velocity "addition" law for one-dimensional motion

$$w = \frac{\tilde{w} + v}{1 + (v\tilde{w}/c^2)}. \tag{2.19}$$

Obviously, this reduces to the Galilean result in the Newtonian limit ($c \to \infty$ or v, $\tilde{w} \ll c$).

Let us try few examples. For $v = \tilde{w} = c/2$ we obtain not $w = c$ but

$$w = \frac{c/2 + c/2}{1 + 1/4} = \frac{4}{5}c < c\,!$$

For $\tilde{w} = c$

$$w = \frac{c + v}{1 + vc/c^2} = c\frac{c + v}{c + v} = c$$

for any v. For $\tilde{w} = -c$

$$w = \frac{-c + v}{1 - vc/c^2} = c\frac{-c + v}{c - v} = -c$$

for any v. This suggest that the speed of light can be reached but cannot be exceeded.

Let us show rigorously that for any $v, \tilde{w} < c$ equation (2.19) gives $w < c$.

1. First we check that this holds when $\tilde{w} = 0$. Indeed, in this case Eq.(2.19) yields $w = v < c$.

2. Second, we show that w is a monotonically increasing function of $\tilde{w}$ and hence reaches its maximum when $\tilde{w} \to c$. Differentiating Eq.(2.19) we find

$$\frac{dw}{d\tilde{w}} = \frac{1 - v^2/c^2}{(1 + v\tilde{w}/c^2)^2} > 0 \quad \text{for} \quad \tilde{w}, v < c\,.$$

3. Finally, we have already seen that at the maximum $w = c$.

Thus w is indeed limited by c from above.

The inverse to Eq.2.19 law is the same as Eq.(2.19) up to the sign of v:

$$\tilde{w} = \frac{w - v}{1 - (vw/c^2)}. \tag{2.20}$$

The change of sign of v in this equation follows directly from the change of sign of v in the inverse Lorentz transformation (2.15). (Another way of deriving Eq.(2.20) is via finding $\tilde{w}$ from Eq.(2.19)

2.4.2 Three-dimensional velocity "addition"

In Sec.2.4.1 we considered the motion only along the x axis. If instead we allow the particle to move in arbitrary directions then using the same method we can derive the following result

$$w^x = \frac{w^{\tilde{x}} + v}{1 + vw^{\tilde{x}}/c^2}, \quad w^y = \frac{w^{\tilde{y}}}{\gamma(1 + vw^{\tilde{x}}/c^2)}, \quad w^z = \frac{w^{\tilde{z}}}{\gamma(1 + vw^{\tilde{x}}/c^2)}. \tag{2.21}$$

for the direct transformation and

$$w^{\tilde{x}} = \frac{w^x - v}{1 - vw^x/c^2}, \quad w^{\tilde{y}} = \frac{w^y}{\gamma(1 - vw^x/c^2)}, \quad w^{\tilde{z}} = \frac{w^z}{\gamma(1 - vw^x/c^2)}. \tag{2.22}$$

for the inverse one. (Notice again that the only difference in the direct and the inverse equations is the sign of v.)

2.5 Aberration of light

A light signal propagates at the angle $\tilde{\theta}$ to the x axis of frame $\tilde{S}$. What is the corresponding angle measured in frame S.

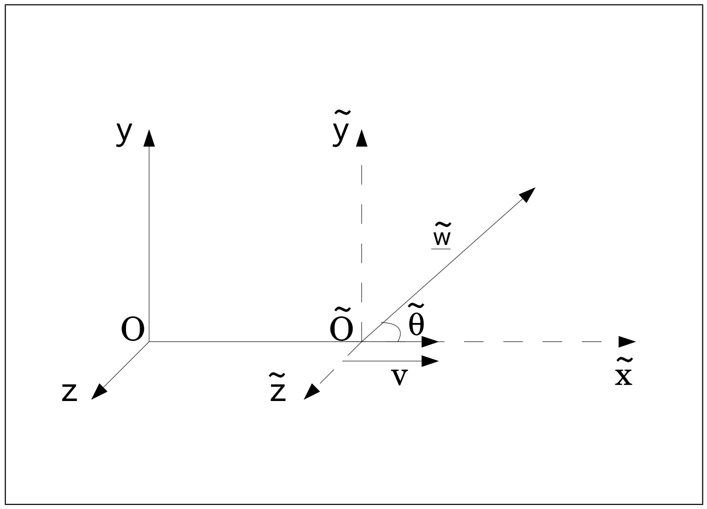

Figure 2.9:

We can always rotates both frames about the x-axis so that the signal is confined to the XOY plane as in figure 2.9. Then the velocity vector of the signal has the following components

$$w^{\tilde{x}} = c\cos\tilde{\theta}, \quad w^{\tilde{y}} = c\sin\tilde{\theta}, \quad w^{\tilde{z}} = 0,$$

in frame $\tilde{S}$ and

$$w^{x} = c\cos\theta, \quad w^{y} = c\sin\theta, \quad w^{z} = 0,$$

in frame S. Substitute these into the first two equations in (2.21) and find that

$$\cos\theta = \frac{\cos\tilde{\theta} + \beta}{1 + \beta\cos\tilde{\theta}} \tag{2.23}$$

$$\sin\theta = \frac{\sin\tilde{\theta}}{\gamma(1 + \beta\cos\tilde{\theta})}, \tag{2.24}$$

where $\beta = v/c$ From these one also finds

$$\tan\theta = \frac{\sin\tilde{\theta}}{\gamma(\beta + \cos\tilde{\theta})}. \tag{2.25}$$

Let us analyse the results. Differentiate Eq.(2.23) with respect to β

$$-\sin\theta\frac{d\theta}{d\beta} = \frac{\sin^2\tilde{\theta}}{(1+\beta\cos\tilde{\theta})^2}$$

and the substitute $\sin\theta$ from Eq.(2.24) to obtain

$$\frac{d\theta}{d\beta} = -\frac{\gamma\sin\tilde{\theta}}{1+\beta\cos\tilde{\theta}}.$$

This derivative is negative or zero ($\tilde{\theta} = 0, \pi$). Thus, θ decreases with β and the direction of light signal is closer to the x direction in frame S. To visualise this effect imagine an isotropic source of light moving with speed close to the speed of light. Its light emission will be beamed in the direction of motion (see fig.2.10) In fact, Eq.(2.23) shows that in the limit $v \to c$ ($\beta \to 1$) we have $\cos\theta \to 1$ and hence $\theta \to 0$ (with exception of $\tilde{\theta} = \pi$ which always gives $\theta = \pi$).

In the Newtonian limit ($c \to \infty$) Equations (2.23-2.25) reduce to

$$\theta = \tilde{\theta},$$

showing that there is no aberration of light effect.

2.6 Doppler effect

The Doppler effect describes the variation of periodic light signal due to the motion of emitter relative to receiver.

2.6.1 Transverse Doppler effect

Problem setup: A source (emitter) of periodic light signal (e.g. monochromatic wave) moves perpendicular to the light of sight of some observer with speed v. Find the frequency ν of the signal received by the observer if in the source frame the frequency of emitted light is ν_0.

Solution: $T_0 = 1/\nu_0$ is the period of the emitter in its rest frame. Suppose, that in the frame of the emitter this signal is emitted during the time $\Delta t_0 = NT_0$, where N is the number of produced wave crests. By definition, both these times are the proper times of the emitter.

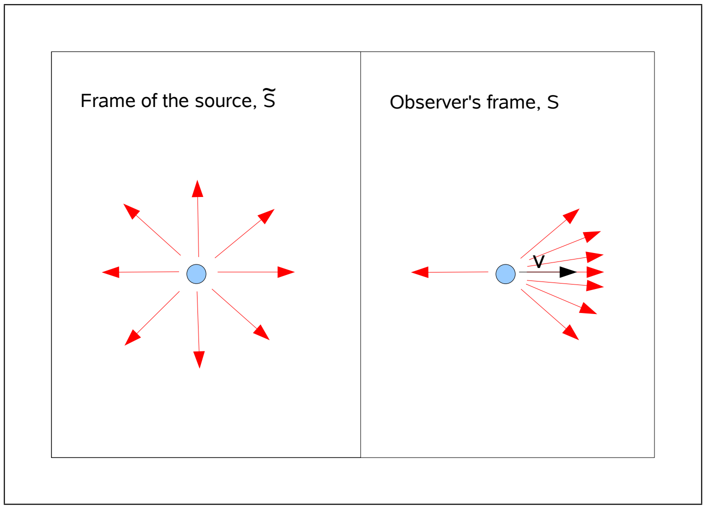

Figure 2.10: The effect of aberration of light on the radiation of moving sources. If in the rest frame of some light source its radiation is isotropic (left panel) then in the frame where the source is moving with relativistic speed its radiation is beamed in the direction of motion (right panel).

In the frame of receiver the emission process takes a longer time, Δt_e, given by the time dilation formula

$$\Delta t_e = \gamma \Delta t_0.$$

Assuming that the distance between the emitter and the receiver is so large that its change during the emission is negligible, the signal is received during the same time interval of the same length, $\Delta t_r = \Delta t_e$. Since the number of crests remains the same, the period of the received light is

$$T_r = \Delta t_r / N = \gamma T_0 \,. \tag{2.26}$$

Thus, the frequency of the received signal is

$$\nu = \frac{1}{T_r} = \frac{1}{\gamma T_0} = \frac{\nu_0}{\gamma}. \tag{2.27}$$

Since $\gamma > 1$ we have $\nu < \nu_0$. One can see that the transverse Doppler effect is entirely due to the time dilation.

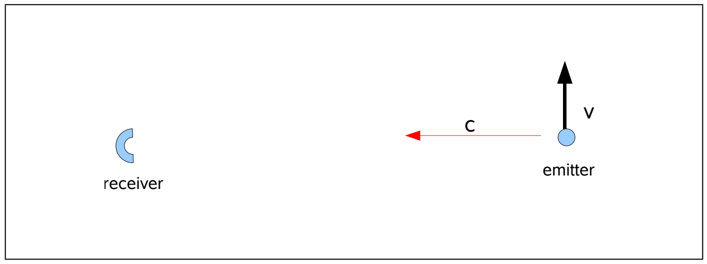

Figure 2.11:

2.6.2 Radial Doppler effect

Problem setup: Now we consider the case where the source is moving along the line of sight and so the distance between the emitter and the receiver is increasing during the process of emission. This has an additional effect on the frequency of the received signal.

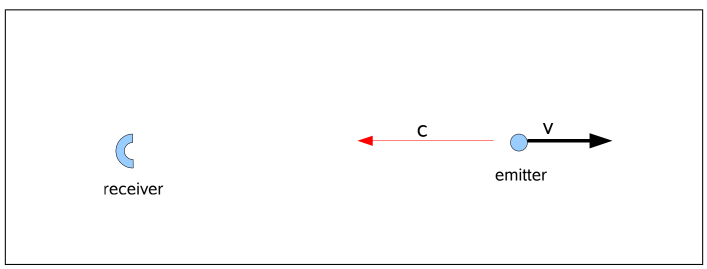

Figure 2.12:

Solution: The emission time as measured in the frame of the observer is still given by the time dilation formula

$$\Delta t_e = \gamma \Delta t_0 \,.$$

During the time Δt_e the emitter moves away from the observer by the distance

$$\Delta l = \Delta t_e v \,.$$

This leads to the increase in the period of reception time by $\Delta l / c$. Thus, the total reception time in this case is

$$\Delta t_r = \Delta t_e + \frac{\Delta t_e v}{c} = \Delta t_0 \gamma (1 + \beta) = \Delta t_0 \sqrt{\frac{1+\beta}{1-\beta}} \,.$$

This yields the observed period of the received signal

$$T_r = \frac{\Delta t_r}{N} = \frac{\Delta t_0}{N}\sqrt{\frac{1+\beta}{1-\beta}} = T_0\sqrt{\frac{1+\beta}{1-\beta}}\,,$$

and the observed frequency

$$\nu = \nu_0\sqrt{\frac{1-\beta}{1+\beta}}\,. \tag{2.28}$$

Notice that in this case $\nu < \nu_0$ if $v > 0$ (the emitter is moving away from the receiver) and $\nu > \nu_0$ if $v < 0$ (the emitter is approaching the receiver).

2.6.3 General case

Problem setup: In this case, the source is moving at angle θ with the line of sight. This includes the transverse ($\theta = \pi/2$) and the radial cases ($\theta = 0, \pi$).

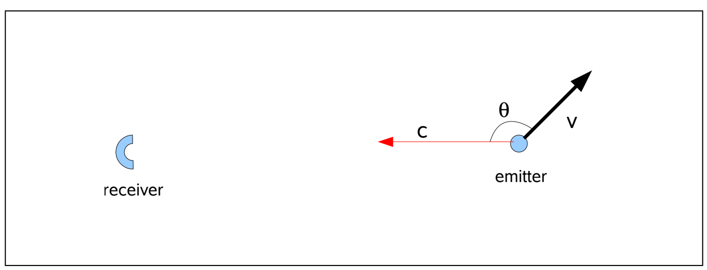

Figure 2.13:

Solution: The only difference compared to the radial case is in the rate of increase of the distance between the emitter and the receiver. Now

$$\Delta l = -\Delta t_e v \cos\theta\,,$$

where we assume that $v > 0$ and that the emitter is far away from the receiver. In order to see this, consider the triangle with sides l_0, l, and $v\Delta t_e$, where l_0 is the distance between the emitter and the receiver at the beginning of emission, l is this distance at the end of emission and

$v\Delta t_e$ is the distance covered by the emitter during the time of emission. From this triangle,

$$l^2 = l_0^2 + (v\Delta t_e)^2 - 2l_0 v\Delta t_e \cos\theta,$$

or

$$l^2 = l_0^2\left(1 - 2\cos\theta\left(\frac{v\Delta t_e}{l_0}\right) + \left(\frac{v\Delta t_e}{l_0}\right)^2\right).$$

For a distant emitter $v\Delta t_e/l_0 \ll 1$ and we have

$$l^2 \simeq l_0^2\left(1 - 2\cos\theta\left(\frac{v\Delta t_e}{l_0}\right)\right).$$

Moreover,

$$l \simeq l_0\left(1 - 2\cos\theta\left(\frac{v\Delta t_e}{l_0}\right)\right)^{1/2} \simeq l_0\left(1 - \cos\theta\left(\frac{v\Delta t_e}{l_0}\right)\right) = l_0 - \cos\theta v\Delta t_e\,.$$

Repeating the calculations of the radial case we then find that

$$T_r = T_0\gamma\left(1 - \beta\cos\theta\right)\,,$$

and

$$\nu = \frac{\nu_0}{\gamma\left(1 - \beta\cos\theta\right)}. \tag{2.29}$$

Now it is less clear whether $\nu > \nu_0$ or otherwise, and further analysis is required.

In the Newtonian limit ($c \to \infty$) one has $\nu = \nu_0$, so no effect is seen. In the case of $\beta \ll 1$, the Taylor expansion of Eq.2.29 gives

$$\nu = \nu_0(1 + \beta\cos\theta + O(\beta^2))\,.$$

Chapter 3

Space-time

3.1 Metric form of Euclidean space

By definition, in Euclidean space one can construct cuboids (rectangular parallelepipeds) such that the lengths of their edges, a, b, and c, and the diagonal l satisfy the following equation

$$l^2 = a^2 + b^2 + c^2, \tag{3.1}$$

no matter how big the cuboid is. This was strongly supported by the results of practical geometry.

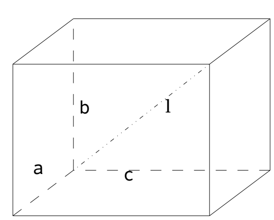

Figure 3.1: A cuboid

Given this property on can construct a set of Cartesian coordinates, $\{x^1, x^2, x^3\}$ (the same meaning as $\{x, y, z\}$). It is based on one reference point (the origin) and three mutually orthogonal reference directions (the axes). For any point of the space, its coordinates are essentially the lengths of the edges of the cuboid which is aligned with the reference directions, has one vertex fixed at the origin and the opposite vertex fixed and the point. These length are taken with the sign plus or minus depending on the cuboid position relative to the axes.

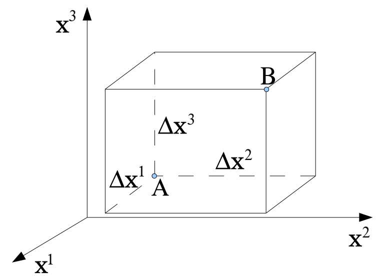

Figure 3.2: Cartesian coordinates

In Cartesian coordinates, the distance between point A and point B with coordinates $\{x_A^1, x_A^2, x_A^3\}$ and $\{x_B^1, x_B^2, x_B^3\}$ respectively is

$$\Delta l_{AB}^2 = (\Delta x^1)^2 + (\Delta x^2)^2 + (\Delta x^3)^2, \tag{3.2}$$

where $\Delta x^i = x_A^i - x_B^i$.

For infinitesimally close points this becomes

$$dl^2 = (dx^1)^2 + (dx^2)^2 + (dx^3)^2, \tag{3.3}$$

where dx^i are infinitesimally small differences between Cartesian coordinates of these points. This is called the metric form of Euclidean space in Cartesian coordinates.

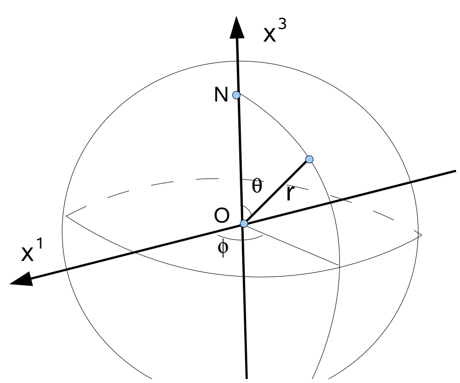

Figure 3.3: Spherical coordinates

For any system of coordinates one can introduce *coordinate lines and coordinate surfaces*. An x^k coordinate line is a line along which all coordinates but x^k are fixed. An x^k coordinate surface is a surface on which x^k coordinate is fixed. Obviously, the coordinate lines of Cartesian coordinates are straight and their coordinate surfaces are planes.

Other types of coordinates can used in Euclidean space as well. One example is the spherical coordinates, $\{r, \theta\,\phi\}$. Figure 3.3 shows how these coordinates are defined: they involve a reference point (the origin) and two reference directions. r is the distance from the origin to the point as measured along the straight line connecting them, ϕ is the angle between one of the reference directions (the azimuthal axis) and the plane containing the point and the other reference direction (the polar axis) and θ is the angle between the direction to the point and the polar axis. If $\{x^1, x^2, x^3\}$ are the Cartesian coordinates with the same origin, the x^1 axis aligned with the azimuthal axis and the x^3 axis aligned with the polar axis (as shown in Fig.3.3) then

$$x^1 = r \sin\theta \cos\phi; \tag{3.4}$$

$$x^2 = r \sin\theta \sin\phi; \tag{3.5}$$

$$x^3 = r \cos\theta. \tag{3.6}$$

Not all coordinate lines of the spherical coordinates are straight. Indeed, the θ coordinate lines are circles centred on the origin and the ϕ coordinate lines are circles centred on the polar axis. Such coordinate systems are called *curvilinear*. At their points of intersection θ, ϕ and r coordinate lines are perpendicular to each other. Such coordinate systems are called *orthogonal*.

In spherical coordinates, the distance between infinitesimally close points is

$$dl^2 = dr^2 + r^2 d\theta^2 + r^2 \sin^2\theta d\phi^2. \tag{3.7}$$

This is called the metric form of Euclidean space in spherical coordinates.

Since, the coefficients of $d\theta^2$ and $d\phi^2$ vary in space

$$\Delta l^2_{AB} \neq (\Delta r)^2 + r^2 \Delta\theta^2 + r^2 \sin^2\theta \Delta\phi^2 \tag{3.8}$$

for the distance between points A and B with finite separations $\Delta r, \Delta\theta, \Delta\phi$. However, one can find the distance between such points along any curve connecting them via integration along this curve

$$l_{ab} = \int_A^B dl \tag{3.9}$$

from A to B. For example the circumference of a circle of radius r_0 is

$$\Delta l = \int dl = 2\int_0^\pi r_0 d\theta = 2\pi r_0\,. \tag{3.10}$$

(Notice that we selected such coordinates that the circle is centred on the origin, $r = r_0$, and it is in a meridional plane, $\phi =$const. As the result, along the circle $dl = r_0 d\theta$.)

In the general case of coordinates which are both curvilinear and non-orthogonal, the distance between infinitesimally close points is given by the positive-definite quadratic form

$$dl^2 = \sum_{i=1}^{3}\sum_{j=1}^{3} g_{ij} dx^i dx^j, \tag{3.11}$$

where $g_{ij} = g_{ji}$ are coefficients that are functions of coordinates. (In fact, g_{ij} are the components of the so-called *metric tensor* in the coordinate basis of coordinates $\{x^i\}$. It is easy to see that in the case of Cartesian coordinates

$$g_{ij} = \delta_{ij} = \begin{cases} 1 & \text{if} \quad i = j \\ 0 & \text{if} \quad i \neq j \end{cases} \tag{3.12}$$

Not all positive definite metric forms correspond to Euclidean space. If there does not exist a coordinate transformation which reduces a given metric form to that of Eq.3.3 then the space with such metric form is not Euclidean. This mathematical result leads to the so-called Riemann geometry (and ultimately the curved spacetime of General Relativity).

3.1.1 Einstein summation rule

In the modern mathematical formulation of the Theory of Relativity it is important to distinguish between upper and lower indexes as the index position determines the mathematical nature of the indexed quantity. E.g. a single upper index indicates a vector (or a contravariant vector), as in b^i, whereas b_i stands for a mathematical object of a different type, the so-called one-form (or covariant vector). In the case of a metric space, one can establish a one-to-one connection between the covariant and contravariant vectors

In the case of coordinates, we use upper indices because $\{dx^i\}$ are components of a contravariant vector. Our syllabus is rather limited and we will not be able to explore the difference between covariant and contravariant vectors in full, as well as the difference between various types of tensors. However, we will keep using this proper notation.

The Einstein summation rule is a convention on the notation for summation over indexes. Namely, any index appearing once as a lower index and once as an upper index of the same indexed object or in the product of a number of indexed objects stands for summation over all allowed values of this index. Such index is called a dummy, or summation index. Indexes which are not dummy are called free indexes. Their role is to give a correct equation for any allowed index value, e.g. 1,2, and 3 in three-dimensional space.

According to this rule we can rewrite Eq.3.11 in a more concise form:

$$dl^2 = g_{ij} dx^i dx^j. \tag{3.13}$$

This rule allows to simplify expressions involving multiple summations. Here are some more examples:

1. $a_i b^i$ stands for $\sum_{i=1}^n a_i b^i$; here i is a dummy index;
2. In $a_i b_i$ i is a free index.
3. $a_i b^{kij}$ stands for $\sum_{i=1}^n a_i b^{kij}$; here k and j are free indexes and i is a dummy index;
4. $a^i \frac{\partial f}{\partial x^i}$ stands for $\sum_{i=1}^n a^i \frac{\partial f}{\partial x^i}$. Notice that index i in the partial derivative $\frac{\partial}{\partial x^i}$ is treated as a lower index.

3.2 Minkowski diagrams

Minkowski diagrams is the first step towards the development of the concept of space-time. Each instantaneous point-like physical event can be identified by stating its three spatial coordinates and time as measured in some inertial frame. In mathematical terms, this is a mapping of the events onto a four dimensional real space. Since it is impossible to illustrate the whole 4-dimensional and even 3-dimensional maps as plots, in the Minkowski diagrams we deal only with a section of the maps defined by the equations $y = y_0$ and $z = z_0$. This means that we show events with the same y and z coordinates.

Let us first to demonstrate some use of such diagram in Newtonian physics. Along the horizontal axis we show the x coordinate of the inertial frame S of the standard configuration and along the vertical axis its time (see the left panel of Fig.3.4). In fact, t can be multiplied by any constant. Mirroring the relativistic diagrams, we multiply it by the speed of light. It is significant that ct has the dimension of length, just like x. Any point on the plot corresponds to a particular real or possible physical event.

Any particle with finite life-time is represented by a line, called the *world line* of the particle. For example, the time axis is the world line of the frame's origin (or that of the standard clock located at the origin). The world line of a particle moving in space along the x axis with constant speed is also straight on the diagram but now it is inclined to the time axis. The inclination angle varies between 0 and $\pi/2$, depending on the particle speed. Simultaneous events reside on lines parallel to the x axis. Events with the same spacial position reside on lines parallel to the time axis. These are the x and t coordinate lines corresponding to the frame S respectively.

Now consider the frame $\tilde{S}$ of the standard configuration and figure out its coordinate lines on the same plot. Since in the Galilean transformation $t = \tilde{t}$, the $\tilde{x}$ coordinate lines coincide with the x coordinate lines. However, $\tilde{t}$ coordinate lines do not coincide with the t coordinate lines. For example, the line along which $\tilde{x} = 0$ satisfies the equation

$$x = \beta ct, \quad \text{where} \quad \beta = v/c\,.$$

This is a straight line inclined to the t axis by the angle

$$\tan\theta = \frac{x}{ct} = \beta. \tag{3.14}$$

This line is the world-line of the origin of the frame $\tilde{S}$ and at the same time the $\tilde{t}$ axis of its coordinates. All other $\tilde{t}$ coordinate lines are parallel to it.

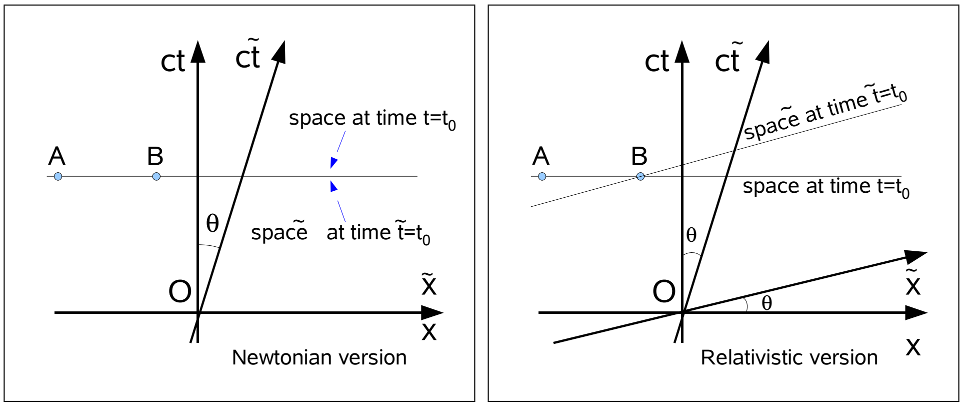

Figure 3.4: Newtonian (left panel) and Minkowskian (right panel) visions of space and time.

Now let us repeat the process using the relativistic framework. The coordinate lines of of corresponding to the inertial frame S are the same as before. To figure out the coordinate lines of the frame $\tilde{S}$ of the standard configuration we use the Lorentz transformation. The $\tilde{t}$ axis satisfies the condition $\tilde{x} = 0$. Since

$$\tilde{x} = \gamma(x - \beta ct)$$

this yield as before

$$x = \beta ct\,.$$

Hence, as before the $\tilde{t}$ coordinate lines are straight lines making the angle θ to the t axis, defined by the Eq.3.14.

The $\tilde{x}$ axis satisfies the condition $\tilde{t} = 0$. Since

$$c\tilde{t} = \gamma(ct - \beta x)\,,$$

this yields the line

$$ct = \beta x\,.$$

This is a straight line which make the angle θ to the x axis (see the right panel of Fig.3.4). All other $\tilde{x}$ coordinate lines are parallel to the $\tilde{x}$. Along each such line the time $\tilde{t}$ is constant but the time t is not.

One may say that the x axis corresponds to the space of the frame S at the time $t = 0$ and the $\tilde{x}$ axis corresponds to the space of the frame $\tilde{S}$ at the time $\tilde{t} = 0$. Similarly, the x coordinate line defined by the equation $t = t_0$ corresponds to the space of the frame S at the time $t = t_0$ and the $\tilde{x}$ coordinate line defined by the equation $\tilde{t} = \tilde{t}_0$ corresponds to the space of the frame $\tilde{S}$ at the time $\tilde{t} = \tilde{t}_0$.

Similarly, the t axis corresponds to the time of the frame S as measured at $x = 0$ and the $\tilde{t}$ axis corresponds to the time of the frame $\tilde{S}$ as measured at $\tilde{x} = 0$. The t coordinate line defined by the equation $x = x_0$ corresponds to the time of the frame S as measured at the point $x = x_0$ and the $\tilde{t}$ coordinate line defined by the equation $\tilde{x} = \tilde{x}_0$ corresponds to the space of the frame $\tilde{S}$ at the point $\tilde{x} = \tilde{x}_0$ and

Since $\beta < 1$ we have $\theta < \pi/4$. Thus, the time and the x axis of frame $\tilde{S}$ never merge, with the time axis staying above and the x axis below the line $ct = x$ which is the world line of a light signal sent from the origin at time $t = 0$.

3.3 Space-time

In Newtonian physics the physical space Euclidean. By definition, a space is Euclidean if one can introduce such coordinates x, y, z that its metric form becomes

$$dl^2 = dx^2 + dy^2 + dz^2 \tag{3.15}$$

at every point of the space. Such coordinates are called Cartesian coordinates. There are infinitely many different Cartesian coordinates. For example, one can obtain one system of Cartesian coordinates from another via rotation about the z axis by angle θ:

$$\begin{aligned}
\tilde{x} &= x\cos\theta + y\sin\theta, \\
\tilde{y} &= -x\sin\theta + y\cos\theta, \\
\tilde{z} &= z.
\end{aligned}$$

This transformation keeps the metric form invariant;

$$dl^2 = d\tilde{x}^2 + d\tilde{y}^2 + d\tilde{z}^2. \tag{3.16}$$

In Newtonian Physics it is assumed that the lengths of objects and distances between them do not depend on the selection of inertial frame where the length measurements are carried out. The geometry of Physical Space is reassuringly the same for every one. In Special Relativity the lengths of objects and distances between them do depend on the inertial frame used for the length measurements. The metric form (3.15) is not invariant under the Lorentz transformation. The geometry of the Physical Space is no longer reassuringly fixed. In mathematical terms the Physical Space can no longer be modelled as a metric space. This almost amounts to saying that the Physical Space is not quite a proper physical reality.

Einstein's teacher at Zurich Polytechnic, professor Minkowski (1864-1909), discovered that the following "generalised metric form"

$$ds^2 = -c^2 dt^2 + dx^2 + dy^2 + dz^2 \tag{3.17}$$

is in fact invariant under the Lorentz transformations. Obviously $dy^2 + dz^2 = d\tilde{y}^2 + d\tilde{z}^2$, so we only need to show that

$$-c^2 dt^2 + dx^2 = -c^2 d\tilde{t}^2 + d\tilde{x}^2.$$

The calculations are most concise when we use the Lorentz transformation written in terms of hyperbolic functions

$$\begin{aligned} ct &= c\tilde{t}\cosh\theta + \tilde{x}\sinh\theta, \\ x &= \tilde{x}\cosh\theta + c\tilde{t}\sinh\theta. \end{aligned} \tag{3.18}$$

To see that the Lorentz transformation can indeed be written this way, we first write it as

$$\begin{aligned} ct &= \gamma(c\tilde{t} + \beta\tilde{x}) \\ x &= \gamma(\tilde{x} + \beta c\tilde{t}), \end{aligned} \tag{3.19}$$

and then introduce θ via $\sinh\theta = \gamma\beta$, from which it follows that $\gamma = \cosh\theta$.

From the first two equations in (3.18) we have

$$\begin{aligned} cdt &= cd\tilde{t}\cosh\theta + d\tilde{x}\sinh\theta, \\ dx &= d\tilde{x}\cosh\theta + cd\tilde{t}\sinh\theta. \end{aligned}$$

Then

$$-c^2 dt^2 + dx^2 = -(cd\tilde{t}\cosh\theta + d\tilde{x}\sinh\theta)^2 + (d\tilde{x}\cosh\theta + cd\tilde{t}\sinh\theta)^2 =$$

$$-c^2 d\tilde{t}^2\cosh^2\theta - 2cd\tilde{x}d\tilde{t}\cosh\theta\sinh\theta - d\tilde{x}^2\sinh^2\theta +$$

$$+c^2 d\tilde{t}^2 \sinh^2\theta + 2cd\tilde{x}d\tilde{t}\cosh\theta\sinh\theta + d\tilde{x}^2\cosh^2\theta =$$

$$-c^2 d\tilde{t}^2(\cosh^2\theta - \sinh^2\theta) + d\tilde{x}^2(\cosh^2\theta - \sinh^2\theta) =$$

$$= -c^2 d\tilde{t}^2 + d\tilde{x}^2.$$

To make the last step we use the well known result $\cosh^2\theta - \sinh^2\theta = 1$. Thus, the quadratic form (3.17) is indeed Lorentz invariant and this fact suggests that space and time should be united into a single 4-dimensional metric space with metric form (3.17)!!! Like in the case of Minkowski diagrams points of this space are called *events*. This space is called *the space-time or the Minkowski space-time*. ds^2 is the generalised distance between points of this space – it is called the *space-time interval*. Denoting

$$x^0 = ct,\ x^1 = x,\ x^2 = y,\ x^3 = z \tag{3.20}$$

we can rewrite Eq.(3.17) as

$$ds^2 = -(dx^0)^2 + (dx^1)^2 + (dx^2)^2 + (dx^3)^2. \tag{3.21}$$

Given the similarity of this metric form with the metric form of Euclidean space in Cartesian coordinates the space-time of Special relativity is often called *pseudo-Euclidean* (where "pseudo" refers to the sign "-" in front of the first term) and the coordinates (3.20) are often called *pseudo-Cartesian. Thus, there is a one-to-one correspondence between inertial frames with their time and Cartesian grid and systems of pseudo-Cartesian coordinates in space-time!* The Lorentz transformation can now be considered as a transformation from one system of pseudo-Cartesian coordinates to another system of pseudo-Cartesian coordinates in space-time. Minkowski diagrams are simply maps of space-time onto a 2D Euclidean plane of a paper sheet. Notice that the usual Euclidean distances between the points on these diagrams do not reflect the generalised distances between the corresponding events in space-time.

In arbitrary curvilinear coordinates the space-time metric becomes

$$ds^2 = g_{\mu\nu}dx^\mu dx^\nu, \tag{3.22}$$

where the coefficients $g_{\mu\nu}$ are quite arbitrary as well. They can be written as components of a 4×4 matrix

$$\begin{pmatrix} g_{00} & g_{01} & g_{02} & g_{03} \\ g_{10} & g_{11} & g_{12} & g_{13} \\ g_{20} & g_{21} & g_{22} & g_{23} \\ g_{30} & g_{31} & g_{32} & g_{33} \end{pmatrix}$$

Here and in the rest of the notes we will assume that Greek indexes run from from 0 to 3 whereas Latin indexes run from 1 to 3. One can always impose the symmetry condition on the metric form

$$g_{\mu\nu} = g_{\nu\mu}, \tag{3.23}$$

and this is what we will always assume here. In pseudo-Cartesian coordinates

$$\begin{array}{c} g_{00} = -1, \; g_{11} = g_{22} = g_{33} = 1, \\ \text{and} \quad g_{\mu\nu} = 0 \quad \text{if} \quad \mu \neq \nu, \end{array} \tag{3.24}$$

or in the matrix form

$$g_{\mu\nu} = \begin{pmatrix} -1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \tag{3.25}$$

3.4 Light cone

Select some event, say event A, in space-time. Let $\{x^\nu\}$ to be the pseudo-Cartesian coordinates corresponding to some inertial frame. Consider the 3-dimensional hyper-surface of Minkowski space-time determined by the equation

$$-(\Delta x^0)^2 + (\Delta x^1)^2 + (\Delta x^2)^2 + (\Delta x^3)^2 = 0\,, \tag{3.26}$$

where $\Delta x^\nu = x^\nu - x^\nu_A$ and $\{x^\nu_A\}$ are the coordinates of the event A. This hyper-surface is called the light cone of the event A as it is made out of all events that are connected to A by world lines of light signals. Indeed, in the corresponding frame $(\Delta x^1)^2 + (\Delta x^2)^2 + (\Delta x^3)^2 = \Delta l^2$, the distance between the event with the coordinates $\{x^\nu\}$ and A, and

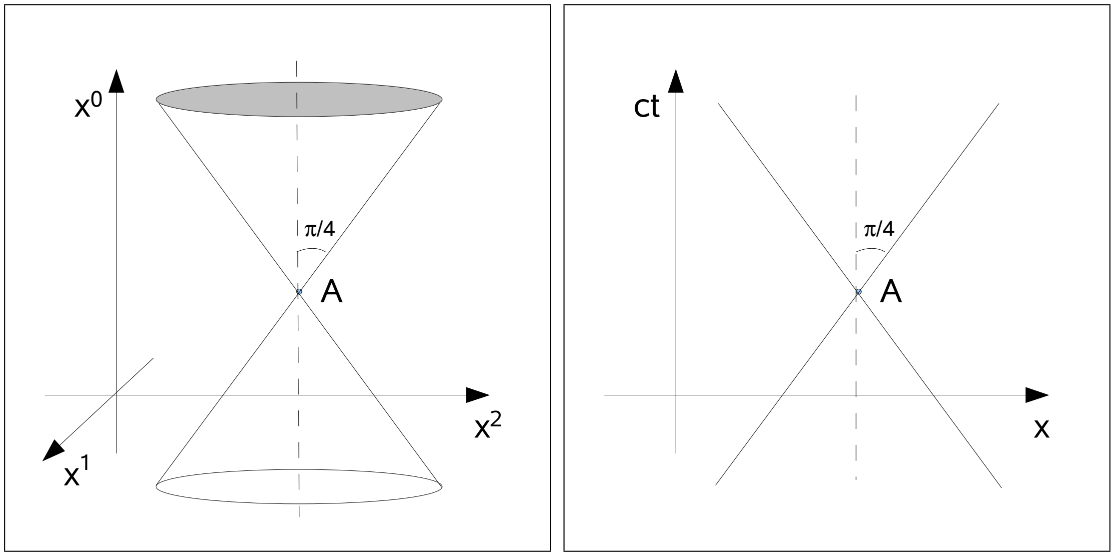

Figure 3.5: *Left panel:* The light cone of event A. *Right panel:* Appearance of the light cone in Minkowski diagram.

$\Delta x^0 = c\Delta t$ is the time interval between the event and A as measured in this frame. Hence $\Delta l/\Delta t = \pm c$.

In the hyper-plane $x^3 = x^3_A$ of space-time, the light cone define the surface

$$-(\Delta x^0)^2 + (\Delta x^1)^2 + (\Delta x^2)^2 = 0\,,$$

which is a cone of the half-opening angle $\pi/4$, centred on the x^0 coordinate line passing through A (see the left panel of fig.3.5). In the plane of Minkowski diagram defined by the equations $x^2 = x^2_A$ and $x^3 = x^3_A$, this equation reduces to

$$-(\Delta x^0)^2 + (\Delta x^1)^2 = 0\,,$$

which describes two straight lines passing through A at the angle of $\pi/4$ to the x^0 coordinate line (see the right panel of fig.3.5).

In the physical space of any inertial frame the bottom half of the light cone corresponds to a spherical light front converging onto the spacial location of A

$$c(t - t_A) = -\sqrt{(x - x_A)^2 + (y - y_A)^2 + (z - z_A)^2} \qquad (3.27)$$

and its upper half corresponds to a spherical light front diverging from the spacial location of A

$$c(t - t_A) = +\sqrt{(x - x_A)^2 + (y - y_A)^2 + (z - z_A)^2} \qquad (3.28)$$

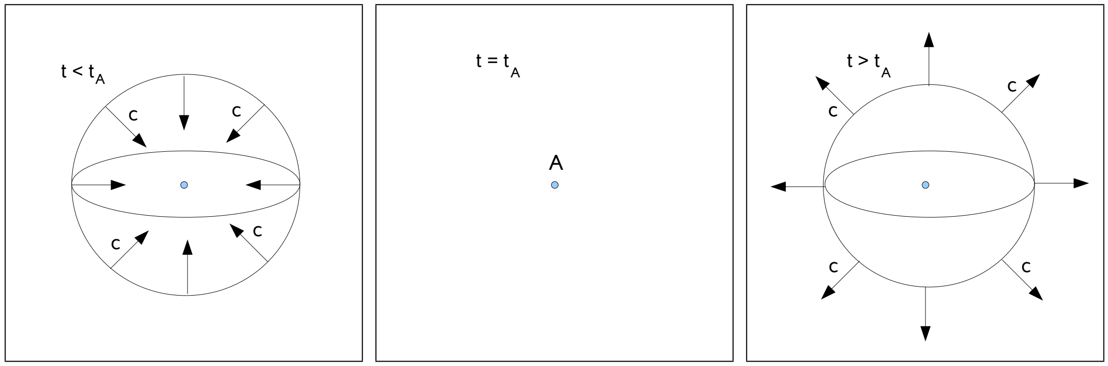

Figure 3.6: In 3-dimensional physical space of any inertial frame the lower half of the light cone of event A corresponds to a spherical light front converging of the location of event A (left panel, $t < t_A$), and its upper half corresponds to a spherical light front diverging from the location of this event (right panel, $t > t_A$).

(see fig.3.6). Due to the length contraction effect shapes and sizes of physical objects as well as distances between them differ in different inertial frames. However, the shape and size of these light fronts remain the same, being described by eqs.(3.27,3.28) in *all inertial frames*.

3.5 Causal structure of space-time

Consider a particle travelling with constant velocity in some inertial frame. Its world line is a straight line. Provided that it passes through the event A its equation is

$$(x - x_A)^2 + (y - y_A)^2 + (z - z_A)^2 = v^2(t - t_A)^2 \leq c^2(t - t_A)^2.$$

Thus, such a world line is located inside the light cone of A (see fig.3.7).

Physical objects interact with each other by means of waves or particles they produce. Let us call these interaction agents as *signals*. If an event A is caused (or triggered) by some other event B then this is arranged via a signal which B sends to A and which reaches the location of A before A occurs. Since the signal speed must be less or equal than the speed of light its world line must lie inside the lower half of the light cone of event A or on its surface if this is the light signal (see the left panel of fig.3.8).

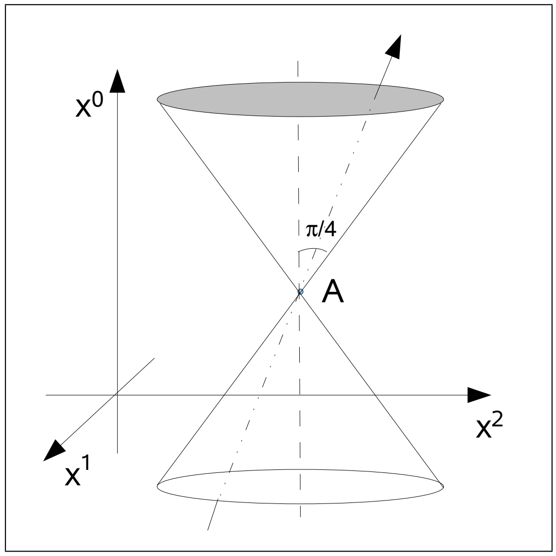

Figure 3.7: The world line of any physical object that passes through the event A lies inside the light cone of A.

On the other hand, the world lines of all possible signals sent by the event A fill the upper half of the light cone of A. Thus, the events that may be caused by A are located inside the upper half of the light cone of A (see the left panel of fig.3.8). As to the events that lie outside of the light cone of A they can neither cause not be caused by A – they are causally unrelated to A.

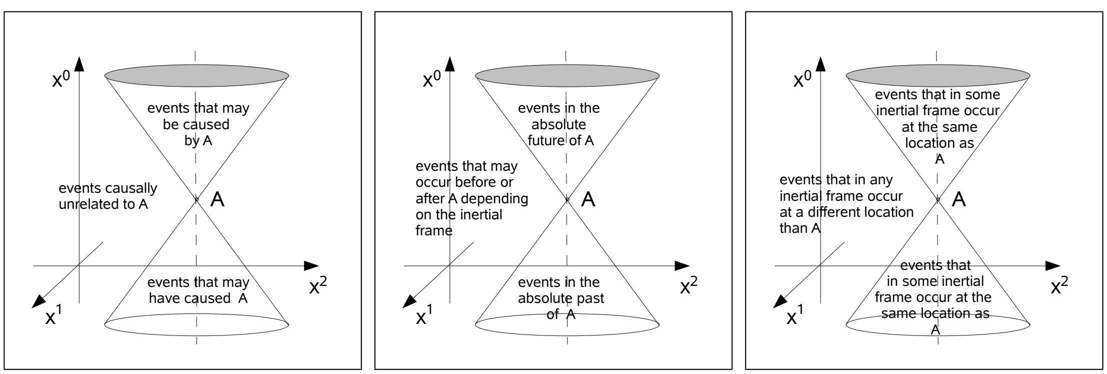

Figure 3.8: The light cone of event A, and the causal, spatial, and temporal connection of this event to all other events.

Consider an event B inside the light cone of A. Construct a straight line connecting A and B. Since this line is insight the light cone it can be the world line of some physical object. For example this can be a grid point of some inertial frame. In this frame A and B have the same spacial location. For any event C located outside the light cone of A the line connecting A and C cannot be a world line of any physical object as this would imply motion with speed exceeding the speed of light. Thus, in any inertial frame A and C will have different spacial locations (see the right panel of fig.3.8). Hence the events that can be casually connected are also the events that can have the same location in some inertial frame. On the contrary, the events that cannot be causally connected are spatially separated in all inertial frames.

The casual structure is also related with the temporal order of events. As we have seen, the events that can be caused by A are all in the future of A ($t > t_A$) in all inertial frames and the events that could cause A are in the past of A ($t < t_A$) in all inertial frames. Thus, the events that have unambiguous temporal location with respect to A are all inside the light cone of A. It is also easy to show that the events residing outside of the light cone of A can precede A in some frames and follow it in other frames.

3.6 Types of space-time intervals

The space-time metric in pseudo-Cartesian coordinates reads

$$ds^2 = -(dx^0)^2 + \sum_{i=1}^{3}(dx^i)^2 \,. \qquad (3.29)$$

This is the generalised distance between two infinitesimally close points in space-time. What is about the distance between points with finite separation? We should still measure them along straight lines. Consider the straight line defined by the parametric equation

$$x^\nu = a^\nu \lambda + b^\nu \,,$$

where λ is the line parameter. Hence, along the line

$$dx^\nu = a^\nu d\lambda$$

and obviously

$$\Delta x^\nu = a^\nu \Delta\lambda \,.$$

As to the distance along the line,

$$ds^2 = \left(-(a^0)^2 + \sum_{i=1}^{3}(a^i)^2\right) d\lambda^2 \,. \tag{3.30}$$

To find the generalised length of the line between two points with the parameter λ_0 and $\lambda_0 + \Delta\lambda$, it is found via the integration

$$\Delta s = \int_{\lambda_0}^{\lambda_0+\Delta\lambda} ds = \int_{\lambda_0}^{\lambda_0+\Delta\lambda} \left(-(a^0)^2 + \sum_{i=1}^{3}(a^i)^2\right)^{1/2} d\lambda$$

$$= \left(-(a^0)^2 + \sum_{i=1}^{3}(a^i)^2\right)^{1/2} \Delta\lambda \,.$$

Hence

$$\Delta s^2 = -(\Delta x^0)^2 + \sum_{i=1}^{3}(\Delta x^i)^2 \,. \tag{3.31}$$

This result has the same form as for infinitesimally close point. Like in the Euclidean space this similarity is specific to Cartesian coordinates, in Minkowskian space-time it is specific to pseudo-Cartesian coordinates.

For any event B that lies outside of the light cone of event A the space-time interval separating these two events is positive,

$$\Delta s^2_{AB} = -(\Delta x^0_{AB})^2 + \sum_{i=1}^{3}(\Delta x^i_{AB})^2 > 0, \tag{3.32}$$

where $\Delta x^\nu_{AB} = x^\nu_B - x^\nu_A$. Such intervals are called *space-like*. It is easy to show that there always exists an inertial frame where A and B are simultaneous and

$$\Delta s^2_{AB} = \sum_{i=1}^{3}(\Delta x^i_{AB})^2.$$

In some inertial frames moving relative to this one A occurs before B, in others B occurs before A. Thus, the temporal order of A and B depends on the choice of inertial frame.

For any event B on the surface of the light cone of event A the space-time interval separating these two events is zero,

$$\Delta s^2_{AB} = 0. \tag{3.33}$$

Such intervals are called *null.*

For any event B that lies inside the light cone of event A the space-time interval separating these two events is negative,

$$\Delta s^2_{AB} = -(\Delta x^0_{AB})^2 + \sum_{i=1}^{3}(\Delta x^i_{AB})^2 < 0. \tag{3.34}$$

Such intervals are called *time-like.* It is easy to show that there always exists an inertial frame where A and B are separated only in time but not in space and hence

$$\Delta s^2_{AB} = -(\Delta x^0_{AB})^2. \tag{3.35}$$

If B lies inside the upper half of the light cone of A then in any inertial frame A precedes B and otherwise if B lies inside the lower half of the light cone of A then in any inertial frame B precedes A. To see we use the Minkowski diagram (figure 3.9) that shows A, B and the axes of two inertial frames centred on the event A (this can always be done simply by resetting the zero time and changing the location of origin.). As we have already seen the angle θ is always less than $\pi/4$. Since the lines of simultaneity of frame $\tilde{S}$ also make angle θ with the x axis the event A precedes the event B in this frame as well as in frame S. One can say that B belongs to *the absolute future of A*. Similarly, we see that any event C that lies inside the lower half of the light cone of A always precedes A and thus belongs to *the absolute past of A*.

These temporal properties are summarised in the middle panel of fig.3.8.

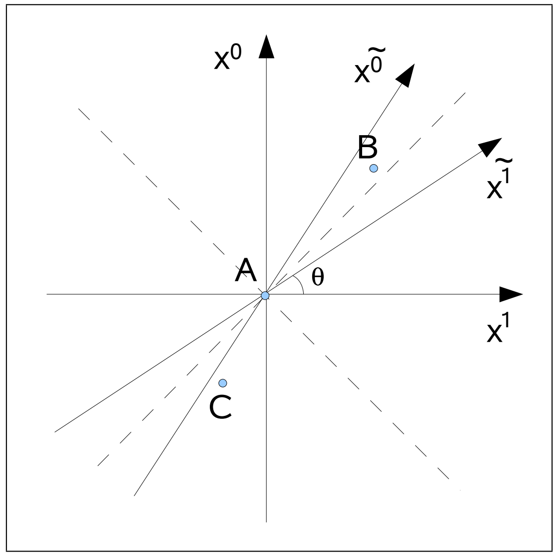

Figure 3.9:

3.7 Vectors

The visualisation/interpretation of vectors as straight arrows whose length describes the vector magnitude works fine for Euclidean space but is not fully suitable for curved manifolds (e.g. surfaces in Euclidean space) and space-time. To see the shortcomings of this interpretation in space-time consider the light cone of event A and a straight line on the surface of this cone (see fig.3.10). Take two points, B and C, on this line. They seem to define two different vectors $\overrightarrow{AB}$ and $\overrightarrow{AC}$. However, the generalised distances (squared) between these points are equal to zero

$$\Delta s^2_{AB} = \Delta s^2_{AC} = 0$$

and so are the generalised lengths of the arrows. Thus, arrows $\overrightarrow{AB}$ and $\overrightarrow{AC}$ have the same lengths and the same directions! Moreover, within this interpretation it is impossible to address the difference between vectors that connect points with space-like separation and vectors that connect points with time-like separation. Below we give the modern generalised definition of geometric vectors due to Cartan, which is very robust and suits very well the Theory of Relativity, both Special and General.

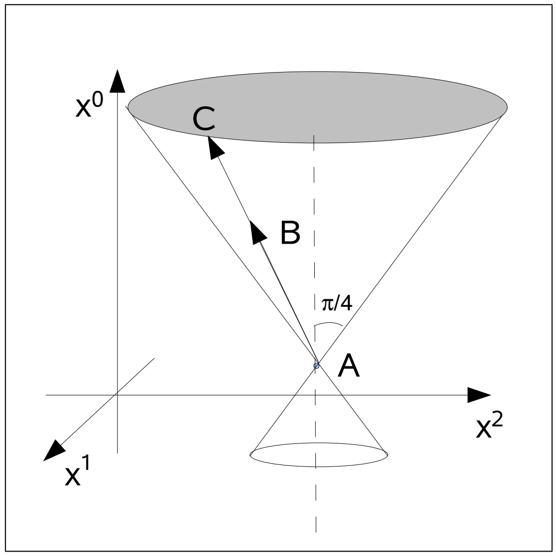

Figure 3.10:

3.7.1 Definition

Consider a point (we denote it as A) in space-time (in fact this theory applies to any other kind of space or manifold) and a curve (a one-dimensional continuous string of points) passing through this point. Let λ to be a parameter of this curve (coordinate of its points). This parameter defines a directional derivative $d/d\lambda$ at point A. Indeed, if F is a function defined on this space (space-time, or manifold) then on this curve it is a function of λ and can be differentiated with respect to it: $F = F(\lambda)$ and $F' = dF/d\lambda$.

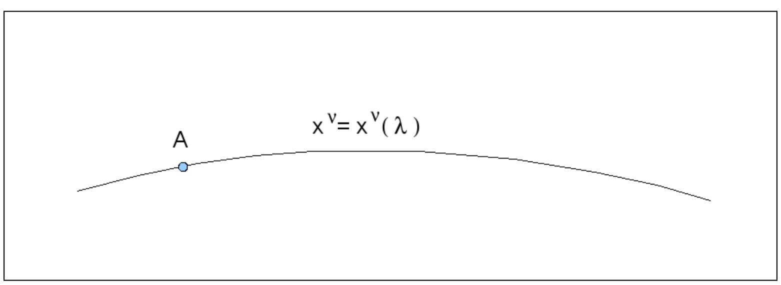

Figure 3.11:

Next introduce some coordinates in the space-time $\{x^\nu\}$. Now the curve can be described by the functions $x^\mu = x^\mu(\lambda)$. Moreover, $F = F(x^\nu)$ and can be differentiated with respect to x^ν as well. According to the chain rule

$$\frac{dF}{d\lambda} = \frac{dx^\nu}{d\lambda}\frac{\partial F}{\partial x^\nu},$$

where we applied the Einstein summation convention. Thus,

$$\frac{d}{d\lambda} = \frac{dx^\nu}{d\lambda}\frac{\partial}{\partial x^\nu}. \tag{3.36}$$

It is easy to verify that all directional derivatives defined in this way at point A form a 4-dimensional *abstract vector space*, which we will denote as T_A. The set of operators

$$\left\{\frac{\partial}{\partial x^\nu}\right\}, \quad \nu = 0, ..., 3$$

is called the coordinate basis in T_A, induced by the coordinates $\{x^\nu\}$, and

$$\frac{dx^\nu}{d\lambda}, \quad \nu = 0, ..., 3$$

are the components of the directional derivative $d/d\lambda$ in this basis. These directional derivatives are identified with vectors in modern geometry. They are often called *Cartan's vectors*. $d/d\lambda$ is also called the *tangent vector* to the curve with parameter λ at point A.

As it has been stressed this definition is very robust. It can be applied to the Euclidean space of Newtonian physics equally well. To see this consider the trajectory of a particle in Euclidean space and parametrise it using the Newtonian absolute time, t. The directional derivative d/dt is in fact the velocity vector of this particle,

$$\underline{v} = \frac{d}{dt}.$$

Indeed,

$$\frac{d}{dt} = \frac{dx^i}{dt}\frac{\partial}{\partial x^i}$$

and thus the components of $\underline{v}$ in the coordinate basis are

$$v^i = \frac{dx^i}{dt},$$

which is the familiar result of Particle Kinematics.

3.7.2 Operations of addition and multiplication

The operations of addition and multiplication by real number for vectors in T_A are defined in terms of their components. Let

$$\vec{a} = a^\nu \frac{\partial}{\partial x^\nu}, \quad \vec{b} = b^\nu \frac{\partial}{\partial x^\nu}, \quad \vec{c} = c^\nu \frac{\partial}{\partial x^\nu}, \quad \alpha \in R^1.$$

Then we say that

$$\vec{c} = \vec{a} + \vec{b} \quad \text{iff} \quad c^\nu = a^\nu + b^\nu \tag{3.37}$$

and

$$\vec{c} = \alpha \vec{a} \quad \text{iff} \quad c^\nu = \alpha a^\nu. \tag{3.38}$$

These definitions ensure all the familiar properties of vector addition and multiplication (commutative law, associative law etc.) Notice that we use "arrow" to indicate 4-dimensional vectors of space-time, e.g. $\vec{a}$, $\vec{b}$, whereas "underline" is reserved for 3-dimensional usual vectors of space, e.g. $\underline{a}$, $\underline{b}$.

3.7.3 Coordinate transformation

Introduce new coordinates, $\{x^{\tilde{\nu}}\}$, in space-time. Obviously,

$$x^{\tilde{\nu}} = x^{\tilde{\nu}}(x^\mu) \quad \text{as well as} \quad x^\mu = x^\mu(x^{\tilde{\nu}})$$

The transformation from $\{x^\nu\}$ to $\{x^{\tilde{\nu}}\}$ delivers the transformation matrix

$$A^{\tilde{\nu}}_{\mu} = \frac{\partial x^{\tilde{\nu}}}{\partial x^\mu}. \tag{3.39}$$

We agree to consider the upper index in this expression as the row index and the lower one as the column index. The inverse transformation, from $\{x^{\tilde{\nu}}\}$ to $\{x^\nu\}$, delivers another transformation matrix

$$A^{\mu}_{\tilde{\nu}} = \frac{\partial x^\mu}{\partial x^{\tilde{\nu}}}. \tag{3.40}$$

These two matrices are inverse to each other. Indeed their multiplication results in the unit matrix

$$A^{\tilde{\nu}}_{\mu} A^{\mu}_{\tilde{\beta}} = \frac{\partial x^{\tilde{\nu}}}{\partial x^\mu} \frac{\partial x^\mu}{\partial x^{\tilde{\beta}}} = \frac{\partial x^{\tilde{\nu}}}{\partial x^{\tilde{\beta}}} = \delta^{\tilde{\nu}}_{\tilde{\beta}}\,. \tag{3.41}$$

Here $\delta^{\tilde{\nu}}_{\tilde{\beta}}$ is Kronecker's delta symbol:

$$\delta^{\tilde{\nu}}_{\tilde{\beta}} = \begin{cases} 1 & \text{if} \quad \tilde{\nu} = \tilde{\beta} \\ 0 & \text{if} \quad \tilde{\nu} \neq \tilde{\beta} \end{cases} .$$

Notice that because one index of the transformation matrix is a lower index and the other one is an upper index, we can utilise the Einstein summation rule in Eq.(3.41). This is exactly the reason behind this indexation of the transformation matrix!

The new coordinates will introduce new coordinate basis at every point of space-time and in this new basis vectors will have different components. First we find the transformation of the coordinate basis of Cartan's vectors. According to the chain rule

$$\frac{\partial f}{\partial x^{\nu}} = \frac{\partial f}{\partial x^{\tilde{\mu}}}\frac{\partial x^{\tilde{\mu}}}{\partial x^{\nu}}.$$

Thus,

$$\frac{\partial}{\partial x^{\nu}} = A^{\tilde{\mu}}_{\nu}\frac{\partial}{\partial x^{\tilde{\mu}}}. \tag{3.42}$$

Similarly one finds

$$\frac{\partial}{\partial x^{\tilde{\mu}}} = A^{\nu}_{\tilde{\mu}}\frac{\partial}{\partial x^{\nu}}. \tag{3.43}$$

Notice that in these expressions we automatically select the correct transformation matrix when we apply the Einstein summation rule. Now we can find the transformation law for the components of vectors. On one hand

$$\vec{u} = u^{\nu}\frac{\partial}{\partial x^{\nu}} = u^{\nu}A^{\tilde{\mu}}_{\nu}\frac{\partial}{\partial x^{\tilde{\mu}}}.$$

On the other hand

$$\vec{u} = u^{\tilde{\mu}}\frac{\partial}{\partial x^{\tilde{\mu}}}.$$

Comparing these two results we immediately obtain

$$u^{\tilde{\mu}} = A^{\tilde{\mu}}_{\nu}u^{\nu}. \tag{3.44}$$

Similarly we find

$$u^{\nu} = A^{\nu}_{\tilde{\mu}}u^{\tilde{\mu}}. \tag{3.45}$$

In these equations, we automatically select the correct transformation matrix when we apply the Einstein summation rule. In the old fashion textbooks on the Theory of Relativity, it is the transformation laws 3.44 and 3.45 that define vectors (or to be more precise the so-called contravariant vectors).

Finally, we construct the transformation matrix of Lorentz transformations. In the case of pseudo-Cartesian coordinates the Lorentz transformation reads

$$\begin{cases} x^{\tilde{0}} = \gamma(x^0 - \beta x^1) \\ x^{\tilde{1}} = \gamma(x^1 - \beta x^0) \\ x^{\tilde{2}} = x^2 \\ x^{\tilde{3}} = x^3 \end{cases} . \tag{3.46}$$

From this we find that

$$\begin{cases} dx^{\tilde{0}} = \gamma(dx^0 - \beta dx^1) \\ dx^{\tilde{1}} = \gamma(dx^1 - \beta dx^0) \\ dx^{\tilde{2}} = dx^2 \\ dx^{\tilde{3}} = dx^3 \end{cases} \tag{3.47}$$

and thus

$$A^{\tilde{\mu}}_{\nu} = \begin{pmatrix} \gamma & -\beta\gamma & 0 & 0 \\ -\beta\gamma & \gamma & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} . \tag{3.48}$$

Thus, pseudo-Cartesian components of any 4-vector transform according to the same law as Eq.(3.46):

$$\begin{cases} a^{\tilde{0}} = \gamma(a^0 - \beta a^1) \\ a^{\tilde{1}} = \gamma(a^1 - \beta a^0) \\ a^{\tilde{2}} = a^2 \\ a^{\tilde{3}} = a^3 \end{cases} \tag{3.49}$$

In fact Eq.(3.47) is a particular example of this rule as dx^ν are the components of infinitesimal displacement vector $\overrightarrow{ds}$.

Similarly we find the inverse transformation matrix

$$A^{\mu}_{\tilde{\nu}} = \begin{pmatrix} \gamma & \beta\gamma & 0 & 0 \\ \beta\gamma & \gamma & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} . \tag{3.50}$$

and

$$\begin{cases} a^0 = \gamma(a^{\tilde{0}} + \beta a^{\tilde{1}}) \\ a^1 = \gamma(a^{\tilde{1}} + \beta a^{\tilde{0}}) \\ a^2 = a^{\tilde{2}} \\ a^3 = a^{\tilde{3}} \end{cases} \tag{3.51}$$

3.7.4 Infinitesimal displacement vectors

Consider a curve with parameter λ and its tangent vector $d/d\lambda$ at the point with coordinates x^ν. The infinitesimal increment of parameter due to infinitesimal displacement along the curve $d\lambda$ is a scalar by which we can multiply the tangent vector and obtain the *infinitesimal displacement vector*

$$\overrightarrow{ds} = d\lambda \frac{d}{d\lambda}. \tag{3.52}$$

The components of this vector in the coordinate basis can be found via

$$ds^\nu = d\lambda \frac{dx^\nu}{d\lambda} = dx^\nu. \tag{3.53}$$

Thus, the components of $\overrightarrow{ds}$ are the infinitesimal increments of coor-

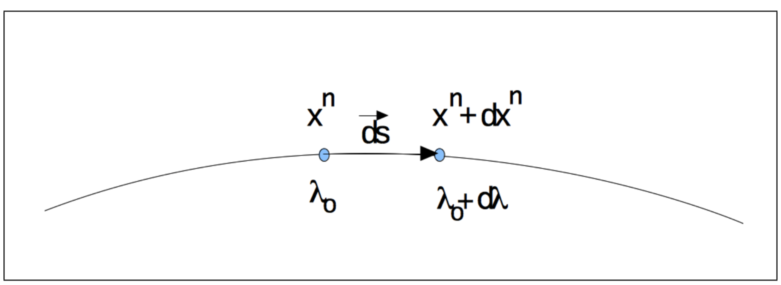

Figure 3.12:

dinates dx^ν corresponding to the infinitesimal increment of the curve parameter λ,

$$\overrightarrow{ds} = dx^\nu \frac{\partial}{\partial x^\nu}. \tag{3.54}$$

3.8 Tensors

3.8.1 Definition

In brief, tensors defined at point A are linear scalar operators acting on vectors from T_A. In other words, a tensor is a linear scalar function of, in general, several vector variables. The number of vectors the tensor needs to be "fed" with in order to produce a scalar is called its rank.

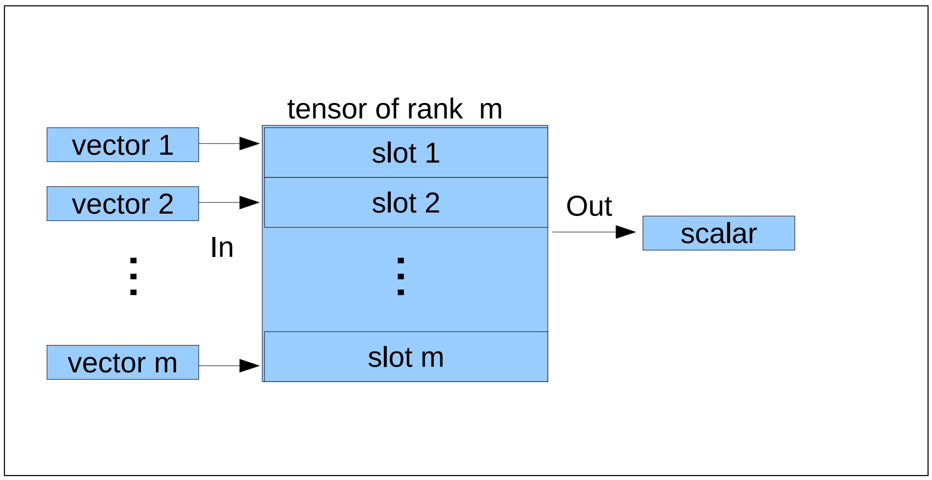

Figure 3.13:

Let, for example, T to be a second rank tensor defined at point A, $\vec{u}, \vec{v}, \vec{w} \in T_A$, and $\alpha, \beta \in R^1$. Then according to the definition

$$T(\vec{u}, \vec{v}) \in R^1; \tag{3.55}$$

$$T(\alpha\vec{u} + \beta\vec{w}, \vec{v}) = \alpha T(\vec{u}, \vec{v}) + \beta T(\vec{w}, \vec{v}); \tag{3.56}$$

$$T(\vec{v}, \alpha\vec{u} + \beta\vec{w}) = \alpha T(\vec{v}, \vec{u}) + \beta T(\vec{v}, \vec{w}). \tag{3.57}$$

3.8.2 Components of tensors

Components of tensors are defined via their action on the basis vectors. For example

$$T_{\nu\mu} = T\left(\frac{\partial}{\partial x^\nu}, \frac{\partial}{\partial x^\mu}\right)$$

are the components of the second-rank tensor T in the coordinate basis. The action of tensors on vectors can be fully described in terms of their components. For example,

$$T(\overrightarrow{u},\overrightarrow{v}) = T(u^{\nu}\frac{\partial}{\partial x^{\nu}}, v^{\mu}\frac{\partial}{\partial x^{\mu}}) = u^{\nu}v^{\mu}T(\frac{\partial}{\partial x^{\nu}}, \frac{\partial}{\partial x^{\mu}}) = u^{\nu}v^{\mu}T_{\nu\mu}.$$

Thus,

$$T(\overrightarrow{u},\overrightarrow{v}) = u^{\nu}v^{\mu}T_{\nu\mu}. \tag{3.58}$$

This is often described as a *contraction* of $T_{\nu\mu}$ with u^{ν} (over index ν) followed by a *contraction* with v^{μ} (over index μ).

3.8.3 Coordinate transformation

Introduce new coordinates $\{x^{\tilde{\mu}}\}$. The components $T_{\tilde{\nu}\tilde{\mu}}$ of tensor T in the new coordinate basis are

$$T_{\tilde{\nu}\tilde{\mu}} = T\left(\frac{\partial}{\partial x^{\tilde{\nu}}}, \frac{\partial}{\partial x^{\tilde{\mu}}}\right) = T\left(A^{\alpha}_{\tilde{\nu}}\frac{\partial}{\partial x^{\alpha}}, A^{\beta}_{\tilde{\mu}}\frac{\partial}{\partial x^{\beta}}\right) =$$

$$A^{\alpha}_{\tilde{\nu}}A^{\beta}_{\tilde{\mu}}\, T\left(\frac{\partial}{\partial x^{\alpha}}, \frac{\partial}{\partial x^{\beta}}\right) = A^{\alpha}_{\tilde{\nu}}A^{\beta}_{\tilde{\mu}}T_{\alpha\beta}$$

Thus,

$$T_{\tilde{\nu}\tilde{\mu}} = A^{\alpha}_{\tilde{\nu}}A^{\beta}_{\tilde{\mu}}T_{\alpha\beta} \tag{3.59}$$

Note that the transformation matrix A appears once per each index of T. One of the indexes of A is always the summation index (also known as *dummy* or *contraction index*) – this is α in the first transformation matrix in Eq.(3.59) and β in the second one. Whether this index is upper or lower one is determined by the Einstein summation convention – in general it must be the opposite case to that of the transformed tensor contracted with A. The other index (known as *free index* because it is not involved in the contraction) is the same both in case and letter as the corresponding index of T on the left-hand side – this is $\tilde{\nu}$ in the first transformation matrix in Eq.(3.59) and $\tilde{\mu}$ in the second one.

3.9 Metric tensor

3.9.1 Definition

Consider a curve with parameter λ and the infinitesimal displacement vector

$$\overrightarrow{ds} = d\lambda \frac{d}{d\lambda}$$

along this curve defined at the point where $\lambda = \lambda_0$.

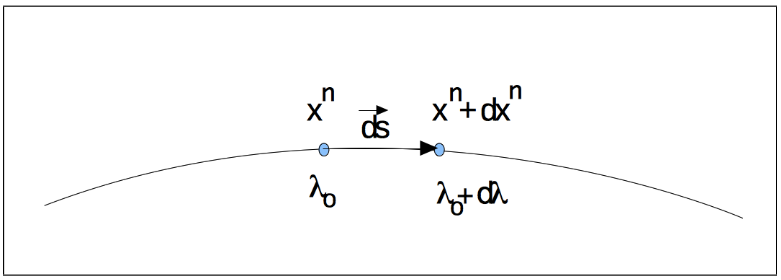

Figure 3.14:

A second rank tensor g which is symmetric,

$$g(\overrightarrow{u}, \overrightarrow{v}) = g(\overrightarrow{v}, \overrightarrow{u}), \tag{3.60}$$

and when acting on $\overrightarrow{ds}$ produces the space-time interval ds^2 between the points $\lambda = \lambda_0$ and $\lambda = \lambda_0 + d\lambda$ is called the *metric tensor*. Since the components of $\overrightarrow{ds}$ in the coordinate basis are dx^ν we have

$$ds^2 = g(\overrightarrow{ds}, \overrightarrow{ds}) = g_{\nu\mu} dx^\nu dx^\mu, \tag{3.61}$$

where $g_{\nu\mu}$ are the components of the metric tensor in the coordinate basis. We immediately recognise this as the metric form.

The metric tensor is also used to define the *scalar product of vectors* and their magnitudes,

$$(\overrightarrow{u} \cdot \overrightarrow{v}) = g(\overrightarrow{u}, \overrightarrow{v}) \tag{3.62}$$

and

$$|\overrightarrow{u}|^2 = g(\overrightarrow{u}, \overrightarrow{u}). \tag{3.63}$$

In components these results read

$$(\overrightarrow{u} \cdot \overrightarrow{v}) = g_{\nu\mu} u^\nu v^\mu \tag{3.64}$$

and

$$|\overrightarrow{u}|^2 = g_{\nu\mu} u^\nu u^\mu. \tag{3.65}$$

When we use pseudo-Cartesian components we have

$$g_{\nu\mu} = \begin{cases} -1 & \text{if} \quad \nu = \mu = 0; \\ +1 & \text{if} \quad \nu = \mu = 1, 2, 3; \\ 0 & \text{if} \quad \nu \neq \mu. \end{cases}$$

and hence the above expressions reduce to a simpler form, Namely

$$|\overrightarrow{u}|^2 = -(u^0)^2 + (u^1)^2 + (u^2)^2 + (u^3)^2 \tag{3.66}$$

and

$$(\overrightarrow{u} \cdot \overrightarrow{v}) = -u^0 v^0 + u^1 v^1 + u^2 v^2 + u^3 v^3\,. \tag{3.67}$$

When a non-pseudo-Cartesian basis is used the expressions can be more complex but the results stay the same. The same scalars of space-time.

3.9.2 Classification of space-time vectors

Since the space-time of Special Relativity is four dimensional, its scalars, vectors and tensors are often called 4-scalars, 4-vectors and 4-tensors. In contrast the usual 3-dimensional scalars, vectors and tensors are called 3-scalars, 3-vectors and 3-tensors.

A space-time vector $\overrightarrow{u}$ is called

- *time-like* if $|\overrightarrow{u}|^2 < 0$,
- *space-like* if $|\overrightarrow{u}|^2 > 0$
- *null* if $|\overrightarrow{u}|^2 = 0$

In general, we have to use Eq.3.65 to compute $|\overrightarrow{u}|^2$. However, the calculations are easer when we use pseudo-Cartesian components

For example, consider the space-time vector $\overrightarrow{u}$ which has the following components in some pseudo-Cartesian basis – $\overrightarrow{u} = (1, 1, 0, 0)$. Then

$$|\overrightarrow{u}|^2 = -1^2 + 1^2 + 0^2 + 0^2 = 0$$

and thus this vector is null.

Classification of the basis vectors of pseudo-Cartesian coordinates.

What are the magnitudes of pseudo-Cartesian basis vectors and the angles between them? This is easy to answer as

$$\left|\frac{\partial}{\partial x^\nu}\right|^2 = g\left(\frac{\partial}{\partial x^\nu}, \frac{\partial}{\partial x^\nu}\right) = g_{\nu\nu}.$$

Hence

$$\left|\frac{\partial}{\partial x^0}\right|^2 = g_{00} = -1 \quad , \tag{3.68}$$

and $\partial/\partial x^0$ is a unit time-like vector whereas

$$\left|\frac{\partial}{\partial x^i}\right|^2 = g_{ii} = 1 \tag{3.69}$$

and hence $\{\partial/\partial x^i\}$ $(i = 1, 2, 3)$ are unit space-like vectors. Similarly we find that

$$\left(\frac{\partial}{\partial x^\nu} \cdot \frac{\partial}{\partial x^\mu}\right) = g_{\nu\mu} = 0 \quad \text{if} \quad \nu \neq \mu \ ,$$

and thus these unit vectors are orthogonal to each other.

Chapter 4

Relativistic particle mechanics

4.1 Tensor equations and the Principle of Relativity

4.1.1 Lorentz invariant equations

The Einstein's Principle of Relativity requires all laws of Physics to be the same in all inertial frames and hence to be invariant under the Lorentz transformation. Since the laws of Newtonian physics are Galilean invariant they cannot be Lorentz invariant and have to be modified. How can this be done? Before, we do this we should try to answer the related question: "How to write Lorentz invariant equations?"

The great value of the concept of space-time and of the tensor calculus in the Theory of Relativity is that they suggest a straightforward way of constructing Lorentz invariant equations. Indeed, the Lorentz transformation, which in physical terms describes the transformation between two inertial frames in standard configuration, corresponds to a particular type of coordinate transformations in spacetime, namely a sub-class of transformations between two systems of pseudo-Cartesian coordinates. However, vector equations involving space-time vectors like

$$\vec{a} = m\vec{b}, \tag{4.1}$$

where m is a space-time scalar, are fully meaningful without any use of coordinates, and hence fully independent on the choice of coordinates in space-time. Once we introduce some coordinates we can write the representation of this equations in terms of components of vectors $\overrightarrow{a}$ and $\overrightarrow{b}$

$$a^\nu = mb^\nu \quad (\nu = 0, 1, 2, 3). \tag{4.2}$$

Clearly, the form of equation (4.2) is the same for any choice of coordinates in space-time and hence this equation is Lorentz invariant too (We also note that the left and the right hand sides of Eq.(4.2) transform in exactly the same way; see Eq.(3.44).). The same applies to all proper tensor equations in general, like

$$G_{\nu\mu} = aT_{\nu\mu}.$$

However, the detailed study of tensor operations is out of the scope of this course and we will deal only with vector equations.

4.1.2 3+1 split of space-time vectors

Thus, we now know how to construct Lorentz invariant equations. But we also need to figure out how to relate such equations with the corresponding equations of Newtonian physics. The Newtonian equations do not comply with the Einstein's Relativity Principle but they are very accurate when the involved speeds are much lower compared to the speed of light. Thus, the relativistic equations must reduce to the Newtonian ones in the limit of low speeds. In order, to compare the relativistic and Newtonian equation we need to find out how the 4-vector equations can be reduced to the 3-vector and 3-scalar equations characteristic for Newtonian Physics.

Equations like Eq.(4.2) differ from the vector equations of classical physics in that they have not 3 but 4 components. The following simple analysis shows that any 4-vector equation corresponds to a pair of traditional Newtonian-like equations, one them being a 3-scalar equation and the other one a 3-vector equation. Indeed, let us select an arbitrary *inertial frame*. This frame introduces hyper-surfaces ("hyper" because they are 3-dimensional) of simultaneity, $t = t_0$, in space-time. Each such hypersurface is the space of this inertial frame at a particular time. (Another inertial frame introduces different hypersurfaces of this sort and has different space and time.)

Now consider a transformation of spacial coordinates (x^i, $i = 1, 2, 3$) in the space of this frame. The transformation matrix of this transformation is

$$A^{\tilde{i}}_j = \frac{\partial x^{\tilde{i}}}{\partial x_j}, \quad (i, j = 1, 2, 3).$$

Quantities that remain invariants under such transformations are the usual spacial scalars (or 3-scalars). The components of usual spacial vectors (or 3-vectors) vary according to the vector transformation law

$$a^{\tilde{i}} = A^{\tilde{i}}_j a^j \quad (i = 1, 2, 3).$$

This transformation can also be viewed as a transformation of space-time coordinates but of a rather specific sort, namely

$$x^{\tilde{0}} = x^0, \quad x^{\tilde{i}} = x^{\tilde{i}}(x^j), \quad (i, j = 1, 2, 3).$$

Thus,

$$A^{\tilde{0}}_0 = 1, \quad A^{\tilde{0}}_i = 0, \quad A^{\tilde{i}}_0 = 0, \tag{4.3}$$

and so we have

$$A^{\tilde{\mu}}_\nu = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & A^{\tilde{1}}_1 & A^{\tilde{1}}_2 & A^{\tilde{1}}_3 \\ 0 & A^{\tilde{2}}_1 & A^{\tilde{2}}_2 & A^{\tilde{2}}_3 \\ 0 & A^{\tilde{3}}_1 & A^{\tilde{3}}_2 & A^{\tilde{3}}_3 \end{pmatrix}. \tag{4.4}$$

Next we apply this transformation law to a space-time vector. Whereas in general

$$a^{\tilde{\mu}} = A^{\tilde{\mu}}_\nu a^\nu$$

in our case we have

$$a^{\tilde{0}} = a^0 \tag{4.5}$$

and

$$a^{\tilde{i}} = A^{\tilde{i}}_j a^j \quad (i = 1, 2, 3). \tag{4.6}$$

Thus, the a^0 is invariant and hence behaves as a 3-scalar, whereas a^i behave as components of a 3-vector. This result is summarised in the following notation

$$\vec{a} = (a^0, \underline{a}). \tag{4.7}$$

Thus, any 4-vector equations splits into a 3-scalar equation and a 3-vector equation. For example, the 4-vector equation

$$\vec{a} = m\vec{b}$$

splits into the 3-scalar equation

$$a^0 = mb^0$$

and the 3-vector equation

$$\underline{a} = m\underline{b}.$$

This result shows that Newtonian physics missed important connections between various physical quantities and between various equations. These are uncovered in Special Relativity.

Tensors of space-time can be split into scalars, vectors and tensors of space but this is beyond the scope of out course. The space in question is the space of the selected inertial frame. For a different frame the space-time splits into time and space in a different way and so do its vectors and tensors.

4.2 4-velocity and 4-momentum

Before we attempt to write relativistic versions of the most important laws of particle dynamics, we need to introduce relevant 4-scalars and 4-vectors. The 3-velocity vector of Newtonian physics is the tangent vector along the particle trajectory with the absolute time as a parameter

$$\underline{v} = \frac{d}{dt} \quad (v^i = dx^i/dt)$$

(see the end of Sec.3.2.1).

In the same manner we can introduce the 4-velocity vector of a particle. Let us see how this is done. First instead of a curve in space, which is the particle trajectory, we need a curve in space-time. The only obvious choice is the world-line of the particle. Second, we need a time-like parameter along this curve. However, we no longer have the absolute time and each inertial frame has its own time. The only natural choice which gives us a unique parameter is the proper time of the particle, τ, which is the time as measured by a fiducial standard clock moving with this particle (the co-moving clock). From this definition, it follows that τ is a 4-scalar. Indeed, it is defined without any reference to a coordinate system in space-time. Thus, the 4-velocity of a particle can be defined as the directional derivative

$$\vec{u} = \frac{d}{d\tau} \tag{4.8}$$

along the world line of this particle.

Once a coordinate system is introduced for the space-time, we can find the components of $\vec{u}$ in the corresponding coordinate basis,

$$u^\nu = \frac{dx^\nu}{d\tau}. \tag{4.9}$$

Since dx^ν are also the components of the infinitesimal displacement 4-vector $\vec{ds}$ along the world line, in the same basis, we can also write

$$\vec{u} = \frac{\vec{ds}}{d\tau}. \tag{4.10}$$

This can be understood as the 4-vector $\vec{ds}$ divided by the 4-scalar $d\tau$, the corresponding increment of the proper time.

Consider an arbitrary inertial frame with arbitrary spatial coordinates x^i and the **normalised** time coordinate $x^0 = ct$, where t is the time of this inertial frame. Then all the basis 4-vectors $\partial/\partial x^i$ are orthogonal to $\partial/\partial x^0$ but in general they are not unit 4-vectors. Whereas the basis 4-vector

$$\frac{\partial}{\partial x^0} = \frac{1}{c}\frac{\partial}{\partial t}$$

is unit. In this basis,

$$u^0 = \frac{dx^0}{d\tau} = c\frac{dt}{d\tau}.$$

Since dt, $d\tau$, and the particle Lorentz factor Γ are related via the time dilation formula

$$dt = \Gamma d\tau,$$

we can write

$$u^0 = \Gamma c.$$

As to the other components,

$$u^i = \frac{dx^i}{d\tau} = \frac{dt}{d\tau}\frac{dx^i}{dt} = \Gamma v^i,$$

where $v^i = dx^i/dt$ are the components of the 3-velocity vector of this particle in this inertial frame in the coordinate basis of its space. These results are summarised using the following notation

$$\overrightarrow{u} = (\Gamma c, \Gamma \underline{v}). \tag{4.11}$$

Notice that Γ is a 3-scalar. (The most frequently used notation for the Lorentz factor is γ. However, in these notes we reserve γ specifically for the Lorentz factor of the relative motion of the two inertial frames involved in the Lorentz transformation and use Γ for the Lorentz factor associated with any other kind of motion.)

Given the above results, we can find the magnitude of the 4-velocity:

$$|\overrightarrow{u}|^2 = -\Gamma^2 c^2 + \Gamma^2 v^2 = -c^2\Gamma^2(1 - v^2/c^2) = -c^2\,.$$

Thus, 4-velocity is a time-like vector. Recall, that $|\overrightarrow{u}|^2$ is a 4-scalar and in any inertial frame and for any selection of space-time coordinates, the calculations of this quantity will yield exactly the same result, $-c^2$. For example, in the particle rest frame (the inertial frame where this particle is instantaneously at rest) $\Gamma = 1$, $\underline{v} = \underline{0}$ and

$$\overrightarrow{u} = (c, 0, 0, 0) \tag{4.12}$$

From this we immediately find that,

$$|\overrightarrow{u}|^2 = -c^2. \tag{4.13}$$

This example demonstrates how easy the calculations can become when we make a good choice of coordinate system/inertial frame.

Another important particle parameter in Newtonian mechanics is its inertial mass, which is assumed to be the same in all inertial frames. However, we cannot be sure that this assumption, which makes the inertial mass a 4-scalar, will be consistent with the relativistic mechanics. (In fact, as we will see very soon, it is not.) Yet, we need a 4-scalar to be used in 4-tensor equations. To make sure, we can use the same trick as with the proper time. We introduce the particle *rest mass* (or *proper mass*), m_0, which is defined as the inertial mass measured in the particle rest frame. This defines m_0 uniquely, yielding the same quantity no matter which inertial frame is actually used to describe the particle motion, or which coordinates we introduce in space-time for this purpose.

In Newtonian mechanics, the product of particle's inertial mass and its 3-velocity is another important 3-vector – the 3-momentum:

$$\underline{p} = m\underline{v} \,.$$

This suggests to introduce the 4-vector of momentum (or the 4-momentum vector) via

$$\overrightarrow{P} = m_0 \overrightarrow{u} \,. \tag{4.14}$$

Using Eq.(4.13) we immediately obtain

$$\left|\overrightarrow{P}\right|^2 = -m_0^2 c^2 \tag{4.15}$$

and thus the 4-momentum vector is also time-like. Using Eq.(4.11) we find that

$$\overrightarrow{P} = (m_0 \Gamma c, m_0 \Gamma \underline{v}). \tag{4.16}$$

The relativistic generalisation of the second law of Newton is also quite straightforward

$$\frac{d\overrightarrow{P}}{d\tau} = \overrightarrow{F}, \tag{4.17}$$

where $\overrightarrow{F}$ is a *4-vector of force* or just 4-force. We will come back to this law later after we consider the relativistic laws of conservation of energy and momentum.

4.3 Energy-momentum conservation

Here we will focus on isolated systems of particles. The term "isolated" means that these particles may interact with each other but not with their surrounding. Moreover, our analysis here will be restricted to only short-range interactions between particles (collisions) whereas the long-range interactions and the corresponding potential energy will be ignored.

The Newtonian mechanics states that the total energy and the total momentum of such a system are conserved (do not change in time):

$$\sum_{k=1} E_{(k)} = \text{const} \tag{4.18}$$

and

$$\sum_{k=1} \underline{p}_{(k)} = \text{const}, \tag{4.19}$$

where

$$E_{(k)} = m_{(k)} v^2_{(k)}/2$$

is the energy (kinetic energy) of the k-th particle and

$$\underline{p}_{(k)} = m_{(k)} \underline{v}_{(k)}$$

is its momentum ($m_{(k)}$ is its inertial mass and $\underline{v}_{(k)}$ is its 3-velocity). Note that we use round brackets here to indicate that k is the particle number (or name). This helps to avoid potential confusion with tensor indexes.

Now we need to find the Lorentz-invariant form of the momentum conservation law (4.19). This generalised law should be a 4-vector equation, involving vectors of space-time. The most obvious candidate is

$$\sum_{k=1} \overrightarrow{P}_{(k)} = \text{const}\,. \tag{4.20}$$

For this to be a good choice, we must be able to recover Eq.(4.19) in the limit of slow speeds ($v \ll c$). To see if this is the case, consider some arbitrary inertial frame and the corresponding 3+1 split of Eq.(4.20). This gives us the 3-scalar equation

$$\sum_{k=1} P^0_{(k)} = \text{const} \quad \text{or} \quad \sum_{k=1} m_{0,(k)} \Gamma_{(k)} c = \text{const} \tag{4.21}$$

and the 3-vector equation

$$\sum_{k=1} P^i_{(k)} = \text{const} \quad \text{or} \quad \sum_{k=1} m_{0,(k)} \Gamma_{(k)} \underline{v}_{(k)} = \text{const}. \tag{4.22}$$

In the Newtonian limit $\Gamma \to 1$ and the latter equation does reduce to the Newtonian momentum conservation law (4.19), with the rest mass playing the role of the inertial mass. Moreover, Eq.(4.22) will have exactly the same form as its Newtonian counterpart even for relativistic speeds if we define the particle's inertial mass as

$$m = m_0 \Gamma \tag{4.23}$$

and hence define the relativistic 3-momentum vector via the usual expression

$$\underline{p} = m\underline{v} \,. \tag{4.24}$$

In order to interpret the 3-scalar equation (4.21) we also consider its form in the Newtonian limit ($v \ll c$). Keeping the first two terms in the MacLaurin expansion of Γ

$$\Gamma = \left(1 - \left(\frac{v}{c}\right)^2\right)^{-1/2} = 1 + \frac{1}{2}\left(\frac{v}{c}\right)^2 + \frac{3}{4}\left(\frac{v}{c}\right)^4 + \dots \,,$$

we find that

$$m_0 \Gamma c \simeq m_0 c \left(1 + \frac{1}{2}\frac{v^2}{c^2}\right) = \frac{1}{c}\left(m_0 c^2 + \frac{1}{2} m_0 v^2\right) \tag{4.25}$$

and hence Eq.4.21 reads

$$\sum_{k=1} \left(m_{0,(k)} c^2 + \frac{1}{2} m_{0,(k)} v^2\right) = \text{const} \,. \tag{4.26}$$

Since in Newtonian mechanics the total mass is conserved,

$$\sum_{k=1} m_{0,(k)} = \text{const}$$

this equation is equivalent to the energy conservation law (4.18) of Newtonian mechanics. This implies that Eq.(4.21) is the relativistic version of the energy conservation and

$$E = m_0 \Gamma c^2 = mc^2 \tag{4.27}$$

is the relativistic version of the particle energy. Thus, the 4-vector equation splits into the Lorentz-invariant laws of energy and 3-momentum conservation. For this reason it is called *the energy-momentum conservation law.*

When $v = 0$, the particle energy E

$$E_0 = m_0 c^2, \tag{4.28}$$

which is called the *rest mass-energy* of the particle. If $v \neq 0$ then $E > E_0$ and the difference

$$K = E - E_0 \tag{4.29}$$

is called the kinetic energy. The interpretation of E_0 as energy becomes most convincing when we consider collisions (reactions) between atoms and subatomic particles. In such collision the rest mass and hence the rest-mass energy is generally not conserved

$$\sum_{k=1} m_{0,(k)} c^2 \neq \text{const}\,,$$

whereas the total energy is conserved and hence the kinetic energy may get converted into the rest-mass energy and the other way around. In fact it is the reduction of the total rest mass-energy of nuclear fuel in nuclear reactors during the process of radioactive decay that is the source of heat in nuclear power stations.

The definitions (4.24,4.27) allow us to write the 4-momentum vector as

$$\overrightarrow{P} = (E/c, \underline{p})\,, \tag{4.30}$$

showing its 3+1 splitting corresponding to a given inertial frame yields the energy and 3-momentum as measured in this frame. For this reason the 4-momentum is also called the energy-momentum vector. Thus, Special Relativity discovers the deep connection between the energy and 3-momentum.

Using pseudo-Cartesian components of $\overrightarrow{P}$ we find that

$$|\overrightarrow{P}|^2 = -\frac{E^2}{c^2} + (p^1)^2 + (p^2)^2 + (p^3)^2 = -\frac{E^2}{c^2} + p^2\,.$$

Combining this result with Eq.(4.15), we find the following useful identity

$$E^2 = p^2c^2 + m_0^2c^4 \quad \text{or} \quad E^2 = E_0^2 + p^2c^2\,. \tag{4.31}$$

4.4 Photons

How much energy can be associated with an electromagnetic wave (wave pack) of the given frequency ν ? In Classical Physics, this energy is unlimited and depends on the volume occupied by the wave – the larger the volume the higher the energy. However in 1900, German physicist Max Planck concluded that in the microphysical (quantum) processes of emission, the electromagnetic energy is emitted not in arbitrary amounts but in *quanta of energy*

$$E = h\nu, \tag{4.32}$$

where ν is the frequency of radiation and h is a constant, known as the Planck constant. Later, in 1905, Einstein concluded that the electromagnetic energy is not only emitted but also absorbed in quanta and proposed that it actually propagates in discrete portions. Thus, in some respects light can be considered as a collection of particles, called *photons*. This is an example of the so-called wave-particle duality of Quantum Physics.

Since photons move with speed of light, their world lines are null and we cannot introduce 4-velocity of photons in the same fashion as we have done for massive particles. Indeed, the 4-velocity is defined as $\overrightarrow{ds}/d\tau$ where τ is the proper time of the particle. This time is measured by the standard clock moving with the same speed as the particle. However, no clock can be forced to move with the speed of light. Indeed, if the rest mass of the clock is m_0 and it is moving with the Lorentz factor Γ, then according to Eq.(4.27) the total energy of this clock is $E = m_0\Gamma c^2$. Thus, as $v \to c$ we have $\Gamma \to \infty$ and hence $E \to \infty$ and given the finite energy resources available to man the speed of light cannot be reached by the clock.

How can photons move with the speed of light and still have the finite energy given by Eq.(4.32)? For fixed E, $\Gamma \to \infty$ implies $m_0 \to 0$. It is often said that the rest mass of photons is zero $m_0 = 0$ and hence they can be described as massless particles. However, this should not be taken literary. Indeed, there is no inertial frame where a photon can be rest – it moves with the speed of light in any frame. Hence a photon cannot be attributed with the rest mass just like it cannot be attributed with the proper time. As to the inertial mass, Einstein's $E = mc^2$ shows that any particle with non-vanishing energy E has the non-vanishing mass $m = E/c^2$. This applies to photons as well. Hence the 4-momentum of any particle can be written as

$$\overrightarrow{P} = (\frac{E}{c}, \frac{E}{c^2}\underline{v}) \,. \tag{4.33}$$

For a photon, $\underline{v} = c\underline{n}$ where $\underline{n}$ is the unit vector in the direction of motion. Hence

$$\overrightarrow{P} = (\frac{E}{c}, \frac{E}{c^2}c\underline{n}) = \frac{E}{c}(1, \underline{n}) = \frac{h\nu}{c}(1, \underline{n}) \,. \tag{4.34}$$

even if we cannot introduce 4-velocity for a particle of this sort. From the last result it follows that

$$|\overrightarrow{P}|^2 = 0 \,. \tag{4.35}$$

4.5 Particle collisions

The simplest and yet very important class of problems in mechanics is the particle collisions. In such problems a number of particles interact with each other by means of some short-range forces which can be rather complex and the actual interaction can be difficult to describe. However, the conservation of energy and momentum allows to answer a number of important questions on the final outcome of collisions. Here we consider few simple examples.

4.5.1 Nuclear recoil

Problem: If not excited via collisions with other particles atomic nuclei reside in the ground states. However, via collisions they can be moved to states with higher rest-mass energy, the excited states, and than spontaneously return to the ground state. The energy excess is then passed to another particle emitted by the nucleus during the transition. This particle can be a photon.

Denote the rest mass-energy of the ground state as E_g and the rest mass-energy of excited state $E_g + \delta E$. Find the energy of emitted photon in the frame where the excited nucleus is initially at rest.

Solution:

In the rest frame of the excited nucleus its 4-momentum is

$$\overrightarrow{Q}_0 = \left(\frac{E_g + \delta E}{c}, \underline{0} \right). \tag{4.36}$$

In the same frame its 4-momentum after the transition to the ground state is

$$\overrightarrow{Q}_1 = \left(\frac{E}{c}, \underline{p} \right), \tag{4.37}$$

where $\underline{p}$ is its 3-momentum and $E > E_g$ is its energy which includes the kinetic energy as well. In the same frame the 4-momentum of the emitted photon is

$$\overrightarrow{P} = \left(\frac{E_p}{c}, \frac{E_p}{c} \underline{n} \right), \tag{4.38}$$

where $\underline{n}$ is the direction of motion of the photon. The energy-momentum conservation law requires

$$\overrightarrow{Q}_0 = \overrightarrow{Q}_1 + \overrightarrow{P}. \tag{4.39}$$

From this point there are two ways to proceed, one is more elegant than the other.

The less elegant way: Here we simply use the fact that the corresponding components of 4-vectors on the left and right hand sides of Eq.4.39 must be the same:

$$\begin{cases} E_g + \delta E = E + E_p, \\ \underline{0} = \underline{p} + \frac{E_p}{c} \underline{n} \end{cases} \tag{4.40}$$

Form this we find

$$E = E_g + \delta E - E_p,$$

and

$$p^2 = \frac{E_p^2}{c^2}.$$

The next step is to utilise Eq.4.31, which in our case reads

$$E^2 = E_g^2 + p^2c^2.$$

Substituting the above expressions for E and p into this equation we obtain

$$(E_g + \delta E - E_p)^2 = E_g^2 + \frac{E_p^2}{c^2}c^2.$$

From this we can find E_p as a function of E_g and δE only. Simple calculations yield

$$E_p = \delta E \frac{\delta E + 2E_g}{2\delta E + 2E_g}.$$

Notice that $E_p < \delta E$. This is because a fraction of δE has been transformed into the kinetic energy of the nucleus. When the photon is emitted the nucleus receives a kick in the opposite direction (nuclear recoil) so that the total momentum of the system "nucleus+photon" remains zero.

The more elegant way: One can rewrite Eq.4.39 as

$$\overrightarrow{Q}_0 - \overrightarrow{P} = \overrightarrow{Q}_1.$$

This may seems not much different but in fact this significantly simplifies calculations because when we square both sides of this equations the result will include only the quantities which are given and the quantity we need to find.

Let us show this:

$$|\overrightarrow{Q}_0 - \overrightarrow{P}|^2 = |\overrightarrow{Q}_1|^2 \tag{4.41}$$

means that

$$|\overrightarrow{Q}_0|^2 + |\overrightarrow{P}|^2 - 2\overrightarrow{Q}_0 \cdot \overrightarrow{P} = |\overrightarrow{Q}_1|^2. \tag{4.42}$$

According to the general results (4.15) and (4.35), for any massive particle

$$|\overrightarrow{Q}|^2 = -m_0^2c^2 = -E_0^2/c^2, \tag{4.43}$$

where $E_0 = m_0c^2$ is the rest mass-energy of the particle. Thus, we have

$$|\overrightarrow{Q}_1|^2 = -E_g^2/c^2, \quad |\overrightarrow{Q}_0|^2 = -(E_g{+}\delta E)^2/c^2, \quad \text{and} \quad |\overrightarrow{P}|^2 = 0.$$

Moreover,

$$\overrightarrow{Q}_0 \cdot \overrightarrow{P} = -\frac{(E_g + \delta E)}{c}\frac{E_p}{c}.$$

Substituting these into Eq.4.42 we find

$$-(E_g + \delta E)^2 + 2(E_g + \delta E)E_p = -E_g^2\,,$$

which includes only one unknown quantity, E_p. Solving this for E_p we find the same answer as before

$$E_p = \delta E\frac{\delta E + 2E_g}{2\delta E + 2E_g}. \tag{4.44}$$

4.5.2 Absorption of neutrons

Problem: A neutron of rest mass-energy E_n and kinetic energy $2E_n$ is absorbed by a stationary B^{10} (Boron-10) nucleus of rest mass-energy $E_{10} = 10E_n$ and becomes B^{11} (Boron-11) nucleus. Find the rest mass-energy E_{11} of B^{11} immediately after the absorption (and hence before it moves from the excited state to the ground state).

Solution: Use the rest frame of B^{10} for the calculations. The 4-momentum of B^{10} and the neutron before the collision are

$$\overrightarrow{Q}_{10} = \left(\frac{10E_n}{c}, \underline{0}\right)\,,$$

$$\overrightarrow{Q}_n = \left(\frac{3E_n}{c}, \underline{p}\right)\,,$$

$$\overrightarrow{Q}_{11} = \left(\frac{E_{11}}{c}, \underline{p}_{11}\right)\,.$$

E_{11}, $\underline{p}$ and $\underline{p}_{11}$ are not known. The 4-momentum of B^{11} after the collision denote as $\overrightarrow{Q}_{11}$. The energy-momentum conservation reads

$$\overrightarrow{Q}_{10} + \overrightarrow{Q}_n = \overrightarrow{Q}_{11}.$$

From this equation we have

$$|\overrightarrow{Q}_{10}|^2 + |\overrightarrow{Q}_n|^2 + 2\overrightarrow{Q}_{10} \cdot \overrightarrow{Q}_n = |\overrightarrow{Q}_{11}|^2. \qquad (4.45)$$

Next we use the relation between the rest mass-energy of a particle and the magnitude of its 4-momentum vector:

$$|\overrightarrow{Q}_{10}|^2 = -\frac{(10E_n)^2}{c^2}, \quad |\overrightarrow{Q}_n|^2 = -\frac{(E_n)^2}{c^2}, \quad |\overrightarrow{Q}_{11}|^2 = -\frac{(E_{11})^2}{c^2},$$

and calculate

$$\overrightarrow{Q}_{10} \cdot \overrightarrow{Q}_n = -30\frac{E_n^2}{c^2}.$$

Finally we substitute these results into Eq.4.45

$$-E_n^2 - 100E_n^2 - 60E_n^2 = -E_{11}^2$$

and obtain

$$E_{11} = \sqrt{161}E_n \approx 12.7E_n.$$

4.6 4-acceleration and 4-force

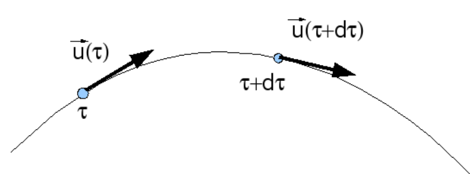

Figure 4.1:

The 4-acceleration of a massive particle is defined as

$$\overrightarrow{a} = \frac{d\overrightarrow{u}}{d\tau}. \qquad (4.46)$$

In this definition $d\overrightarrow{u} = \overrightarrow{u}(\tau + d\tau) - \overrightarrow{u}(\tau)$ and thus we subtract vectors defined at two different points. We can do this only if we can compare

4-vectors defined at different locations in space-time. In Special Relativity we say that two such vectors are identical iff they have the same pseudo-Cartesian components. It is easy to see that

$$(\vec{a} \cdot \vec{u}) = 0. \tag{4.47}$$

Indeed,

$$(\vec{a} \cdot \vec{u}) = \vec{u} \cdot \frac{d}{d\tau}\vec{u} = \frac{1}{2}\frac{d}{d\tau}(\vec{u} \cdot \vec{u}) = -\frac{1}{2}\frac{dc^2}{d\tau} = 0\,.$$

In the 3+1 form, Eq.(4.46) reads

$$\vec{a} = \frac{d}{d\tau}(\Gamma c, \Gamma \underline{v}) = \Gamma \frac{d}{dt}(\Gamma c, \Gamma \underline{v}) = \left(c\Gamma \frac{d\Gamma}{dt}, \Gamma \frac{d\Gamma}{dt}\underline{v} + \Gamma^2 \frac{d\underline{v}}{dt} \right).$$

Thus,

$$\vec{a} = \left(c\Gamma \frac{d\Gamma}{dt}, \Gamma \frac{d\Gamma}{dt}\underline{v} + \Gamma^2 \underline{a} \right), \tag{4.48}$$

where $\underline{a} = d\underline{v}/dt$ is the usual 3-acceleration vector.

The 4-force vector is defined as

$$\vec{F} = \frac{d\vec{P}}{d\tau}. \tag{4.49}$$

Since for massive particle $\vec{P} = m_0 \vec{u}$ we obtain that

$$\vec{F} = \frac{dm_0}{d\tau}\vec{u} + m_0 \vec{a} \tag{4.50}$$

and thus in general the 4-force is not parallel to the 4-acceleration. This fact gives rise to the following classification of forces is Special Relativity.

- A 4-force is called *pure* if

  $$dm_0/d\tau = 0.$$

  Eq.4.50 tells us that a pure 4-force is parallel to the 4-acceleration of particle

  $$\vec{F} = m_0 \vec{a}. \tag{4.51}$$

  Combining this result with Eq.(4.47) we also find that

  $$\vec{F} \cdot \vec{u} = 0. \tag{4.52}$$

- A 4-force is called *heat-like* if

$$\overrightarrow{a} = 0.$$

Eq.4.50 tells us that a heat-like 4-force is parallel to the 4-velocity of particle

$$\overrightarrow{F} = \frac{dm_0}{d\tau}\overrightarrow{u}. \tag{4.53}$$

Given the splitting (4.30) of 4-momentum we can split the 4-force vector of space-time into a 3-scalar and a 3-vector as well. We obtain

$$\overrightarrow{F} = \frac{d}{d\tau}(E/c, \underline{p}) = \Gamma\frac{d}{dt}(E/c, \underline{p}) = \left(\frac{\Gamma}{c}\frac{dE}{dt}, \Gamma\frac{d\underline{p}}{dt}\right).$$

Using the same definition of 3-force as in Newtonian mechanics,

$$\underline{f} = \frac{d\underline{p}}{dt}, \tag{4.54}$$

we can write the above result as

$$\overrightarrow{F} = \left(\frac{\Gamma}{c}\frac{dE}{dt}, \Gamma\underline{f}\right). \tag{4.55}$$

This equation tells us that the 4-force combines the effect on the particle 3-momentum with the effect on its energy.

Since $\overrightarrow{u} = (\Gamma cm, \Gamma\underline{v})$ we have

$$(\overrightarrow{F} \cdot \overrightarrow{u}) = \gamma^2(-\frac{dE}{dt} + \underline{f} \cdot \underline{v}).$$

On the other hand, using eqs.(4.47,4.50) we also obtain

$$(\overrightarrow{F} \cdot \overrightarrow{u}) = -c^2\frac{dm_0}{d\tau} = -c^2\gamma\frac{dm_0}{dt}.$$

Combining these two result we find

$$\frac{dE}{dt} = \frac{c^2}{\gamma}\frac{dm_0}{dt} + (\underline{f} \cdot \underline{v}). \tag{4.56}$$

Similar to Newtonian mechanics each kind of interaction should be described by additional laws specifying the 4-force. This is the subject for more advanced course on Special Relativity.