Chapter 1
The special theory of relativity

1.1 Historical background

1905 is often described as Einstein's *annus mirabilis*: a wonderful year in which he came up with three remarkable ideas. These were the Brownian motion in fluids, the photoelectric effect and the special theory of relativity. Each of these was of a basic nature and also had a wide impact on physics. In this chapter we will be concerned with special relativity, which was arguably the most fundamental of the above three ideas.

It is perhaps a remarkable circumstance that, ever since the initiation of modern science with the works of Galileo, Kepler and Newton, there has emerged a feeling towards the end of each century that the end of physics is near: that is, most in-depth fundamental discoveries have been made and only detailed 'scratching at the surface' remains. This feeling emerged towards the end of the eighteenth century, when Newtonian laws of motion and gravitation, the studies in optics and acoustics, etc. had provided explanations of most observed phenomena. The nineteenth century saw the development of thermodynamics, the growth in understanding of electrodynamics, wave motion, etc., none of which had been expected in the previous century. So the feeling again grew that the end of physics was nigh. As we know, the twentieth century saw the emergence of two theories, fundamental but totally unexpected by the stalwarts of the nineteenth century, viz., relativity and quantum theory. Finally, the success of the attempt to unify electromagnetism with the weak interaction led many twentieth-century physicists to announce that the end of physics was not far off. That hope has not materialized even though the twenty-first century has begun.

While the above feeling of euphoria comes from the successes of the existing paradigm, the real hope of progress lies in those phenomena that seem anomalous, i.e., those that cannot be explained by the current paradigm. We begin our account with the notion of 'ether' or 'aether' (the extra 'a' for distinguishing the substance from the commonly used chemical fluid). Although Newton had (wrongly) resisted the notion that light travels as a wave, during the nineteenth century the concept of light travelling as a wave had become experimentally established through such phenomena as interference, diffraction and polarization. However, this understanding raised the next question: in what medium do these waves travel? For, conditioned by the mechanistic thinking of the Newtonian paradigm, physicists needed a medium whose disturbance would lead to the wave phenomenon. Water waves travel in water, sound waves propagate in a fluid, elastic waves move through an elastic substance ... so light waves also need a medium called aether in which to travel.

The fact that light seemed to propagate through almost a vacuum suggested that the proposed medium must be extremely 'non-intrusive' and so difficult to detect. Indeed, many unsuccessful attempts were made to detect it. The most important such experiment was conducted by Michelson and Morley.

1.2 The Michelson and Morley experiment

The basic idea behind the experiment conducted by A. A. Michelson and E. Morley in 1887 can be understood by invoking the example of a person rowing a boat in a river. Figure 1.1 shows a schematic diagram of a river flowing from left to right with speed v. A boatman who can row his boat at speed c in still water is trying to row along and across the river in different directions. In Figure 1.1(a) he rows in the direction of the current and finds that his net speed in that direction is $c + v$. Likewise (see Figure 1.1(b)), when he rows in the opposite direction his net speed is reduced to $c - v$. What is his speed when he rows across the river in the perpendicular direction as shown in Figure 1.1(c)? Clearly he must row in an oblique direction so that his velocity has a component v in a direction opposite to the current. This will compensate for the flow of the river. The remaining component $\sqrt{c^2 - v^2}$ will take him across the river in a perpendicular direction as shown in Figure 1.1(c).

Suppose now that he does this experiment of rowing down the river a distance d and back the same distance and then rows the same distance perpendicular to the current and back. What is the difference of time τ between the two round trips? The above details lead to the answer that

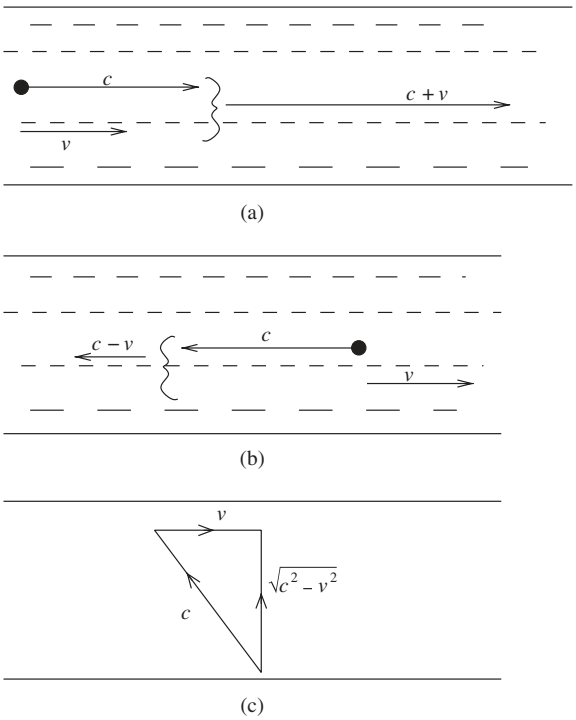

Fig. 1.1. The three cases of a boat being rowed in a river with an intrinsic speed c, the river flowing (from left to right) with speed v: (a) in the direction of the river flow, (b) opposite to that direction and (c) in a direction perpendicular to the flow of the river.

the time for the first trip exceeds that for the second by

$$\tau = \frac{d}{c-v} + \frac{d}{c+v} - \frac{2d}{\sqrt{(c^2 - v^2)}} \tag{1.1}$$

and, for small current speeds ($v \ll c$), we get the answer as

$$\tau \cong \frac{d}{c} \times \frac{v^2}{c^2}. \tag{1.2}$$

The Michelson–Morley experiment [1] used the Michelson interferometer and is schematically described by Figure 1.2. Light from a source S is made to pass through an inclined glass plate cum mirror P. The plate is inclined at an angle of 45° to the light path. Part of the light from the source passes through the transparent part of the plate and, travelling a distance d_1, falls on a plane mirror A, where it is reflected back. It then passes on to plate P and, getting reflected by the mirror part, it moves towards the viewing telescope. A second ray from the source first gets reflected by the mirror part of the plate P and then, after travelling a distance d_2, gets reflected again at the second mirror B. From there it passes through P and gets into the viewing telescope.

Now consider the apparatus set up so that the first path (length d_1) is in the E–W direction. In a stationary aether the surface of the Earth will

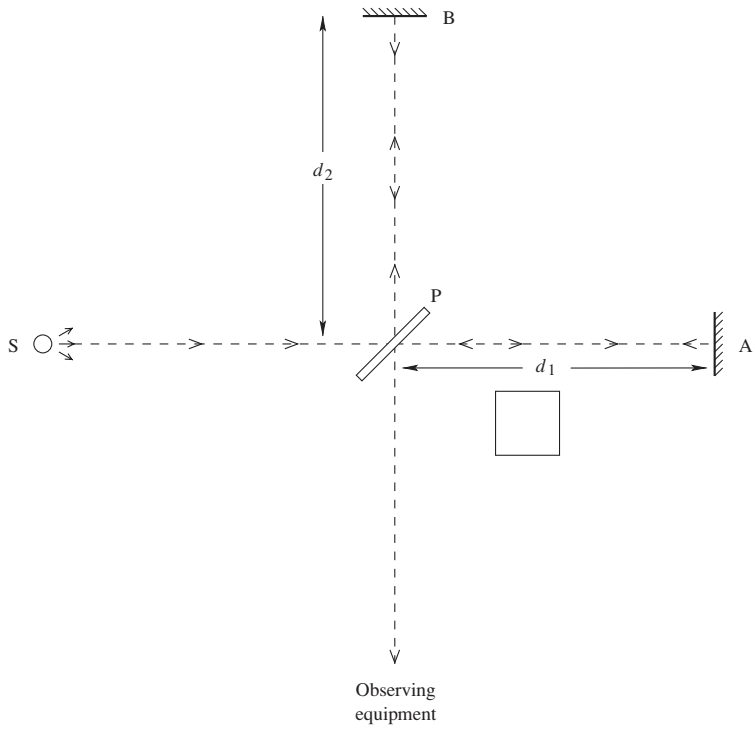

Fig. 1.2. The schematic arrangement of Michelson's interferometer, as described in the text.

have a velocity approximately equal to its orbital velocity of 30 km/s. Thus $(v/c)^2$ is of the order of 10^{-8}. In the actual experiment the apparatus was turned by a right angle so that the E–W and N–S directions of the arms were interchanged. So the calculation for the river-boat crossing can be repeated for both cases and the two times added to give the expected time difference as

$$\tau \cong \frac{d_1 + d_2}{c} \times \frac{v^2}{c^2}. \tag{1.3}$$

Although the effect expected looks very small, the actual sensitivity of the instrument was very good and it was certainly capable of detecting the effect *if indeed it were present.* The experiment was repeated several times. In the case that the Earth was at rest relative to the aether at the time of the experiment, six months later its velocity would be maximum relative to the aether. But an experment performed six months later also gave a null result.

The Michelson–Morley experiment generated a lot of discussion. Did it imply that there was no medium like aether present after all? Physicists not prepared to accept this radical conclusion came up with novel ideas to explain the null result. The most popular of these was the

notion of contraction proposed by George Fitzgerald and later worked on by Hendrik Lorentz. Their conclusion as summarized by J. Larmour in a contemporary (pre-relativity) text on electromagnetic theory reads as follows:

> . . . if the internal forces of a material system arise wholly from electromagnetic actions between the systems of electrons which constitute the atoms, then the effect of imparting to a steady material system a uniform velocity [**v**] of translation is to produce a uniform contraction of the system in the direction of motion, of amount $(1 - v^2/c^2)$ . . .

It is clear that a factor of this kind would resolve the problem posed by the Michelson–Morley experiment. For, by reducing the length travelled in the E–W direction by the above factor, we arrive at the same time of travel for both directions and hence a null result. Lorentz went further to give an elaborate physical theory to explain why the Fitzgerald contraction takes place.

The Michelson–Morley experiment was explained much more elegantly when Einstein proposed his special theory of relativity. We will return to this point after decribing what ideas led Einstein to propose the theory. As we will see, the Michelson–Morley experiment played no role whatsoever in leading him to relativity.

1.3 The invariance of Maxwell's equations

We now turn to Einstein's own approach to relativity [2], which was motivated by considerations of symmetry of the basic equations of physics, in particular the electromagnetic theory. For he discovered a conflict between Newtonian ideas of space and time and Maxwell's equations, which, since the mid 1860s, had been regarded as the fundamental equations of the electromagnetic theory. An elegant conclusion derived from them was that the electromagnetic fields propagated in space with the speed of light, which we shall henceforth denote by c. It was how this fundamental speed should transform, when seen by two observers in uniform relative motion, that led to the conceptual problems.

The Newtonian dynamics, with all its successes on the Earth and in the Cosmos, relied on what is known as the *Galilean transformation* of space and time as measured by two inertial observers. Let us clarify this notion further. Let O and O′ be two inertial observers, i.e., two observers on whom *no force acts*. By Newton's first law of motion both are travelling with uniform velocites in straight lines. Let the speed of O′ relative to O be **v**. Without losing the essential physical information we take parallel Cartesian axes centred at O and O′ with the X, X' axes parallel to the direction of **v**. We also assume that the respective time

coordinates of the two observers were so set that $t = t' = 0$ when O and O′ coincided.

Under these conditions the transformation law for spacetime variables for O and O′ is given by

$$t' = t, \qquad x' = x - vt, \qquad y' = y, \qquad z' = z. \tag{1.4}$$

Since v is a constant, the frames of reference move uniformly relative to each other. Laws of physics were expected to be invariant relative to such frames of reference. For example, because of constancy of v, we have equality of the accelerations $\ddot{x}$ and $\ddot{x}'$. Thus Newton's second law of motion is invariant under the Galilean transformation. Indeed, we may state a general expectation that the basic laws of physics should turn out to be invariant under the Galilean transformations. This may be called the *principle of relativity*.

Paving the way to a mechanistic philosophy, Newtonian dynamics nurtured the belief that the basic laws of physics will turn out to be mechanics-based and as such the Galilean transformation would play a key role in them. This belief seemed destined for a setback when applied to Maxwell's equations. Maxwell's equations in Gaussian units and in vacuum (with isolated charges and currents) may be written as follows:

$$\begin{aligned} \nabla \cdot \mathbf{B} = 0; \qquad & \nabla \times \mathbf{E} = -\frac{1}{c}\frac{\partial \mathbf{B}}{\partial t}; \\ \nabla \cdot \mathbf{D} = 4\pi\rho; \qquad & \nabla \times \mathbf{H} = \frac{1}{c}\frac{\partial \mathbf{D}}{\partial t} + \frac{4\pi}{c}\mathbf{j}. \end{aligned} \tag{1.5}$$

Here the fields $\mathbf{B}$, $\mathbf{E}$, $\mathbf{D}$ and $\mathbf{H}$ have their usual meaning and ρ and $\mathbf{j}$ are the charge and current density. We may set $\mathbf{D} = \mathbf{E}$ and $\mathbf{B} = \mathbf{H}$ in this situation. Then we get by a simple manipulation, in the absence of charges and currents,

$$\begin{aligned} \nabla \times \nabla \times \mathbf{H} &\equiv \nabla\, \nabla \cdot \mathbf{H} - \nabla^2 \mathbf{H} \\ &= \frac{1}{c}\frac{\partial}{\partial t}\nabla \times \mathbf{E} = -\frac{1}{c^2}\frac{\partial^2 \mathbf{H}}{\partial t^2}. \end{aligned} \tag{1.6}$$

From this we see that $\mathbf{H}$ satisfies the wave equation

$$\Box \mathbf{H} = \mathbf{0}. \tag{1.7}$$

Similarly $\mathbf{E}$ will also satisfy the wave equation, the operator $\Box$ standing for

$$\Box \equiv \frac{1}{c^2}\frac{\partial^2}{\partial t^2} - \nabla^2.$$

The conclusion drawn from this derivation is this: Maxwell's equations imply that the **E** and **H** fields propagate as waves with the speed c. Unless explicitly stated otherwise, we shall take $c = 1$.[1]

However, this innocent-looking conclusion leads to problems when we compare the experiences of two typical inertial observers, having a uniform relative velocity **v**. Suppose observer O sends out a wave towards observer O′ receding from him at velocity v directed along OO′. Our understanding of Newtonian kinematics will convince us that O′ will see the wave coming towards him with velocity $c - v$. But then we run foul of the *principle of relativity*: that the basic laws of physics are invariant under Galilean transformations. So Maxwell's equations should have the same formal structure for O and O′, with the conclusion that both these observers should see their respective vectors **E** and **H** propagate across space with speed c.

This was the problem Einstein worried about and to exacerbate it he took up the imaginary example of an observer travelling with the speed of the wave. What would such an observer see?

Let us look at the equations from a Galilean standpoint first. The Galilean transformation is given by

$$\mathbf{r}' = \mathbf{r} - \mathbf{v}t, \qquad t' = t. \tag{1.8}$$

Although the general transformation above can be handled, we will take its simplifed version in which O′ is moving away from O along the x-axis and O and O′ coincided when $t' = t = 0$. It is easy to see that the partial derivatives are related as follows:

$$\frac{\partial}{\partial x} = \frac{\partial}{\partial x'}, \qquad \frac{\partial}{\partial y} = \frac{\partial}{\partial y'}, \qquad \frac{\partial}{\partial z} = \frac{\partial}{\partial z'}, \qquad \frac{\partial}{\partial t} = \frac{\partial}{\partial t'} - v\,\frac{\partial}{\partial x'}.$$

If we apply these transformation formulae to the wave equation (1.7), we find that the form of the equation is changed to

$$\left(\frac{\partial}{\partial t'} - v\,\frac{\partial}{\partial x'}\right)^2 \mathbf{H} - \nabla'^2\mathbf{H} = \mathbf{0}. \tag{1.9}$$

Clearly Maxwell's equations are not invariant with respect to Galilean transformation. Indeed, if we want the equations to be invariant for all inertial observers, then we need, for example, the speed of light to be invariant for them, as seen from the above example of the wave equation. Can we think of some *other* transformation that will guarantee the above invariances?

In particular, let us ask this question: what is the simplest modification we can make to the Galilean transformation in order to preserve

[1] In this book, as a rule, we will choose units such that the speed of light is unity when measured in them.

the form of the wave equation? We consider the answer to this question for the situation of the two inertial observers O and O′ described above. We try linear transformations between their respective space and time coordinates (t, x, y, z) and (t', x', y', z') so as to get the desired answer. So we begin with

$$t' = a_{00}t + a_{01}x, \qquad x' = a_{10}t + a_{11}x, \qquad y' = y, \qquad z' = z. \tag{1.10}$$

With this transformation, it is not difficult to verify that the wave operator $\Box$ transforms as

$$\Box \equiv \left(a_{00}\frac{\partial}{\partial t'} + a_{10}\frac{\partial}{\partial x'}\right)^2 - \left(a_{01}\frac{\partial}{\partial t'} + a_{11}\frac{\partial}{\partial x'}\right)^2 - \frac{\partial^2}{\partial y'^2} - \frac{\partial^2}{\partial z'^2}. \tag{1.11}$$

A little algebra tells us that the right-hand side will reduce to the wave operator in the primed coordinates, provided that

$$a_{00}^2 - a_{01}^2 = 1, \qquad a_{10}^2 - a_{11}^2 = -1, \qquad a_{11}a_{01} = a_{10}a_{00}. \tag{1.12}$$

Now, if we assume that the origin of the frame of reference of O′ is moving with speed v with respect to the frame of O, then setting $x' = 0$ we get $va_{11} = -a_{10}$. Then from (1.12) we get $a_{01} = -va_{00}$. Finally we get the solution to these equations as

$$a_{11} = \gamma, \qquad a_{10} = -v\gamma = a_{01}, \qquad a_{00} = \gamma, \tag{1.13}$$

where

$$\gamma = (1 - v^2)^{-1/2}. \tag{1.14}$$

Thus the transformation that preserves the form of the wave equation is made up of the following relations between (t, x, y, z) and (t', x', y', z'), the coordinates of O and O′, respectively:

$$t' = \gamma(t - vx), \qquad x' = \gamma(x - vt), \qquad y' = y, \qquad z' = z. \tag{1.15}$$

It is easy to invert these relations so as to express the unprimed coordinates in terms of the primed ones. In that case we would find that the relations look formally the same but with $+v$ replacing $-v$:

$$t = \gamma(t' + vx'), \qquad x = \gamma(x' + vt'), \qquad y = y', \qquad z = z'. \tag{1.16}$$

Physically it means that, if O′ is moving with speed v relative to O, then O is moving with speed $-v$ relative to O′.

A more elaborate algebra will also show that the Maxwell equations are also invariant under the above transformation.

Einstein arrived at this result while considering the hypothetical observer travelling with the light wavefront. He found that such an observer could not exist. (This can be seen in our example below by

letting v go to $c = 1$.) In the process he arrived at the above transformation. As we will shortly see, this transformation has echoes of the work Lorentz had done in his attempts to explain the null result of the Michelson–Morley experiment. We will refer to such transformations by the name *Lorentz transformations*, the name given by Henri Poincaré to honour Lorentz for his original ideas in this field.

We also see that the space coordinates and the time coordinate get mixed up in a Lorentz transformation. Thus, for a family of inertial observers moving with different relative velocities, we cannot compartmentalize space and time as separate units. Rather they together form a four-dimensional structure, which we will henceforth call 'spacetime'.

Example 1.3.1 Consider (1.15) with the following definition of θ:

$$v = c \tanh \theta.$$

Then trigonometry leads us to the following transformation laws:

$$t' = t \cosh \theta - x \sinh \theta, \qquad x' = x \cosh \theta - t \sinh \theta, \qquad y' = y,$$
$$z' = z.$$

Compare the first two relations with the rotation of Cartesian axes x, y in two (space) dimensions:

$$x' = x \cos \theta - y \sin \theta, \qquad y' = y \cos \theta + x \sin \theta.$$

We may therefore consider the Lorentz transformation as a rotation through an imaginary angle $\mathrm{i}\theta$, if we define an imaginary time coordinate as $T = \mathrm{i}t$.

1.4 The origin of special relativity

Einstein thus found himself at a crossroads: the Newtonian mechanics was invariant under the Galilean transformation, whereas Maxwell's equations were invariant under the Lorentz transformation. One could try to modify the Maxwell equations and look for invariance of the new equations under the Galilean transformation. Alternatively, one could modify the Newtonian mechanics and make it invariant under the Lorentz transformation. Einstein chose the latter course. We will now highlight his development of the special theory of relativity.

We begin with the introduction of a special class of observers, the *inertial observers* in whose rest frame Newton's first law of motion holds. That is, these observers are under no forces and so move relative to one another with uniform velocities. Notice that there is no explicitly defined frame that could be considered as providing a frame of 'absolute rest'. Thus all inertial observers have equal status and so do their frames,

which are the inertial frames. This is in contrast with the Newtonian concept of absolute space, whose rest frame enjoyed a special status. We will comment on it further in Chapter 18 when we discuss Mach's principle.

The *principle of relativity* states that all basic laws of physics are the same for all inertial observers. Notice that this principle has not changed from its Newtonian form; but the inertial observers are now linked by Lorentz rather than Galilean transformations.

When applied to electricity and magnetism this principle tells us that Maxwell's equations are the same for all inertial observers: in particular, the speed of light c, which appears as the wave velocity in these equations, must be the same in all inertial reference frames. We also see that this requirement leads us to the Lorentz transformation. The transformation described by the equations (1.15) is called a 'special Lorentz transformation'. It can be easily generalized to the case in which the observer O′ moves with a constant velocity $\mathbf{v}$ in any arbitrary direction. The relevant relations are

$$t' = \gamma[t - (\mathbf{v}\cdot\mathbf{r})], \qquad \mathbf{r}' = \gamma(\mathbf{r}^* - \mathbf{v}t), \tag{1.17}$$

where

$$\mathbf{r}^* = \mathbf{r}/\gamma + (\gamma - 1)\mathbf{v}(\mathbf{v}\cdot\mathbf{r})/\gamma v^2. \tag{1.18}$$

We next look at some of the observable effects of this transformation on some measurements of events in space and time. For it is these effects that tell us what the special theory of relativity is all about.

Example 1.4.1 *Problem.* Show that (1.17) reduces to (1.15) for a special Lorentz transformation.

Solution. In the special Lorentz transformation, $\mathbf{v}$ is in the x-direction. So, if $\mathbf{e}$ is a unit vector in that direction,

$$\mathbf{v}\cdot\mathbf{r} = vx, \qquad \mathbf{r}^* = \mathbf{r}\sqrt{1-v^2} + \left(\frac{1}{\sqrt{1-v^2}} - 1\right)v^2 x\frac{1}{\gamma v^2}\mathbf{e},$$

where we have used (1.17) and (1.18). Thus $t' = \gamma(t - vx)$, which is as per (1.15). For the $\mathbf{r}'$ relation, note that the y–y' relation is $y' = y$. Similarly we have $z' = z$. The x–x' relation is

$$\begin{aligned} x' &= x + \gamma(\gamma - 1)v^2 x \cdot \frac{1}{\gamma v^2} - \gamma vt \\ &= x(1 + \gamma - 1) - \gamma vt = \gamma(x - vt). \end{aligned}$$

Thus we recover the special Lorentz transformation.