Chapter 3

The Lorentz transformation

In *The Wonderful World* and appendix 1, the reasoning is kept as direct as possible. Much use is made of graphical arguments to back up the mathematical results. Now we will introduce a more algebraic approach. This is needed in order to go further. In particular, it will save a lot of trouble in calculations involving a change of reference frame, and we will learn how to formulate laws of physics so that they obey the Main Postulates of the theory.

3.1 Introducing the Lorentz transformation

The Lorentz transformation, for which this chapter is named, is the coordinate transformation which replaces the Galilean transformation presented in eq. (2.1).

Let S and S′ be reference frames allowing coordinate systems (t, x, y, z) and (t', x', y', z') to be defined. Let their corresponding axes be aligned, with the x and x' axes along the line of relative motion, so that S′ has velocity v in the x direction in reference frame S. Also, let the origins of coordinates and time be chosen so that the origins of the two reference frames coincide at $t = t' = 0$. Hereafter we refer to this arrangement as the 'standard configuration' of a pair of reference frames. In such a standard configuration, if an event has coordinates (t, x, y, z) in S, then its coordinates in S′ are given by

$$
\begin{aligned}
t' &= \gamma(t - vx/c^2) && (3.1)\\
x' &= \gamma(-vt + x) && (3.2)\\
y' &= y && (3.3)\\
z' &= z && (3.4)
\end{aligned}
$$

where $\gamma = \gamma(v) = 1/(1 - v^2/c^2)^{1/2}$. This set of simultaneous equations is called the Lorentz transformation; we will derive it from the Main Postulates of Special Relativity in section 3.2.

By solving for (t, x, y, z) in terms of (t', x', y', z') you can easily derive the inverse Lorentz transformation:

$$
\begin{aligned}
t &= \gamma(t' + vx'/c^2) & (3.5)\\
x &= \gamma(vt' + x') & (3.6)\\
y &= y' & (3.7)\\
z &= z' & (3.8)
\end{aligned}
$$

This can also be obtained by replacing v by $-v$ and swapping primed and unprimed symbols in the first set of equations. This is how it must turn out, since if S$'$ has velocity $\mathbf{v}$ in S, then S has velocity $-\mathbf{v}$ in S$'$ and both are equally valid inertial frames.

Let us immediately extract from the Lorentz transformation the phenomena of time dilation and Lorentz contraction. For the former, simply pick two events at the same spatial location in S, separated by time τ. We may as well pick the origin, $x = y = z = 0$, and times $t = 0$ and $t = \tau$ in frame S. Now apply eq. (3.1) to the two events: we find the first event occurs at time $t' = 0$, and the second at time $t' = \gamma\tau$, so the time interval between them in frame S' is $\gamma\tau$, i.e. longer than in the first frame by the factor γ. This is time dilation.

For Lorentz contraction, one must consider not two events but two worldlines. These are the worldlines of the two ends, in the x direction, of some object fixed in S. Place the origin on one of these worldlines, and then the other end lies at $x = L_0$ for all t, where L_0 is the rest length. Now consider these worldlines in the frame S$'$ and pick the time $t' = 0$. At this moment, the worldline passing through the origin of S is also at the origin of S$'$, i.e. at $x' = 0$. Using the Lorentz transformation, the other worldline is found at

$$
t' = \gamma(t - vL_0/c^2), \qquad x' = \gamma(-vt + L_0). \tag{3.9}
$$

Since we are considering the situation at $t' = 0$ we deduce from the first equation that $t = vL_0/c^2$. Substituting this into the second equation we obtain $x' = \gamma L_0(1 - v^2/c^2) = L_0/\gamma$. Thus in the primed frame at a given instant the two ends of the object are at $x' = 0$ and $x' = L_0/\gamma$. Therefore the length of the object is reduced from L_0 by a factor γ. This is Lorentz contraction.

For relativistic addition of velocities, eq. (22.8), consider a particle moving along the x' axis with speed u in frame S$'$. Its worldline is given by $x' = ut'$. Substituting in (3.6) we obtain $x = \gamma(vt' + ut') = \gamma^2(v + u)(t - vx/c^2)$. Solve for x as a function of t and one obtains $x = wt$ with w as given by (22.8).

For the Doppler effect, consider a photon emitted from the origin of S at time t_0. Its worldline

$$\gamma$$

$$\beta = \sqrt{1 - 1/\gamma^2}, \qquad \frac{\gamma - 1}{\beta^2} = \frac{\gamma^2}{1+\gamma} \tag{3.10}$$

$$\frac{d\gamma}{dv} = \gamma^3 v/c^2, \qquad \frac{d}{dv}(\gamma v) = \gamma^3 \tag{3.11}$$

$$\frac{dt}{d\tau} = \gamma, \qquad \frac{dt'}{dt} = \gamma_v(1 - \mathbf{u}\cdot\mathbf{v}/c^2) \tag{3.12}$$

$$\gamma(w) = \gamma(u)\gamma(v)(1 - \mathbf{u}\cdot\mathbf{v}/c^2) \tag{3.13}$$

Table 3.1: Useful relations involving γ. $\beta = v/c$ is the speed in units of the speed of light. $dt/d\tau$ relates the time between events on a worldline to the proper time, for a particle of speed v. dt'/dt relates the time between events on a worldline for two reference frames of relative velocity $\mathbf{v}$, with $\mathbf{u}$ the particle velocity in the unprimed frame. If two particles have velocities $\mathbf{u}, \mathbf{v}$ in some reference frame then $\gamma(w)$ is the Lorentz factor for their relative velocity.

is $x = c(t - t_0)$. The worldline of the origin of S$'$ is $x = vt$. These two lines intersect at $x = vt = c(t - t_0)$, hence $t = t_0/(1 - v/c)$. Now use the Lorentz transformation eq. (3.1), then invert to convert times into frequencies, and one obtains eq. (22.7).

To summarize:

> The Postulates of relativity, taken together, lead to a description of spacetime in which the notions of simultaneity, time duration, and spatial distance are well-defined in each inertial reference frame, but their values, for a given pair of events, can vary from one reference frame to another. In particular, objects evolve more slowly and are contracted along their direction of motion when observed in a reference frame relative to which they are in motion.

A good way to think of the Lorentz transformation is to regard it as a kind of 'translation' from the t, x, y, z 'language' to the t', x', y', z' 'language'. The basic results given above serve as an introduction, to increase our confidence with the transformation and its use. In the rest of this chapter we will use it to treat more general situations, such as addition of non-parallel velocities, the Doppler effect for light emitted at a general angle to the direction of motion, and other phenomena.

Table 3.1 summarizes some useful formulae related to the Lorentz factor $\gamma(v)$. Derivations of (3.12), (3.13) will be presented in section 3.5; derivation of the others is left as an exercise for the reader.

Why not start with the Lorentz transformation?
Question: "The Lorentz transformation allows all the basic results of time dilation, Lorentz contraction, Doppler effect and addition of velocities to be derived quite readily. Why not start with it, and avoid all the trouble of the slow step-by-step arguments presented in *The Wonderful World*?"
Answer: The cautious step-by-step arguments are needed in order to understand the results, and the character of spacetime. Only then is the physical meaning of the Lorentz transformation clear. We can present things quickly now because spacetime, time dilation and space contraction were already discussed at length in *The Wonderful World* and appendix 1. Such a discussion has to take place somewhere. The derivation of the Lorentz transformation given in section 3.2 can seem like mere mathematical trickery unless we maintain a firm grasp on what it all means.

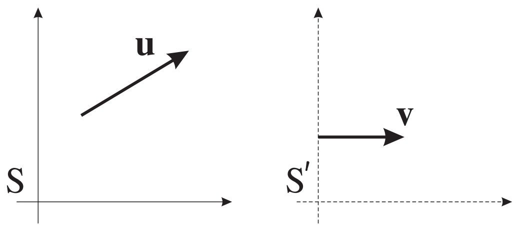


Figure 3.1: A particle has velocity $\mathbf{u}$ in frame S. Frame S′ moves at velocity $\mathbf{v}$ relative to S, with its spatial axes aligned with those of S.

3.2 Derivation of Lorentz transformation

[*Section omitted in lecture-note version.*]

3.3 Velocities

Let reference frames S, S′ be in standard configuration with relative velocity v, and suppose a particle moves with velocity $\mathbf{u}$ in S (see figure 3.1). What is the velocity $\mathbf{u}'$ of this particle in S′?

For the purpose of the calculation we can without loss of generality put the origin of coordinates on the worldline of the particle. Then the trajectory of the particle is $x = u_x t, y = u_y t, z = u_z t$.

Applying the Lorentz transformation, we have

$$\begin{aligned} x' &= \gamma(-vt + u_x t) \\ y' &= u_y t \\ z' &= u_z t \end{aligned} \tag{3.14}$$

for points on the trajectory, with

$$t' = \gamma(t - v u_x t/c^2). \tag{3.15}$$

This gives $t = t'/\gamma(1 - u_x v/c^2)$, which, when substituted into the equations for x', y', z' implies

$$u'_x = \frac{u_x - v}{1 - u_x v/c^2}, \tag{3.16}$$

$$u'_y = \frac{u_y}{\gamma(1 - u_x v/c^2)}, \tag{3.17}$$

$$u'_z = \frac{u_z}{\gamma(1 - u_x v/c^2)}. \tag{3.18}$$

Writing

$$\mathbf{u} = \mathbf{u}_{\parallel} + \mathbf{u}_{\perp} \tag{3.19}$$

where $\mathbf{u}_{\parallel}$ is the component of $\mathbf{u}$ in the direction of the relative motion of the reference frames, and $\mathbf{u}_{\perp}$ is the component perpendicular to it, the result is conveniently written in vector notation:

$$\mathbf{u}'_{\parallel} = \frac{\mathbf{u}_{\parallel} - \mathbf{v}}{1 - \mathbf{u}\cdot\mathbf{v}/c^2}, \qquad \mathbf{u}'_{\perp} = \frac{\mathbf{u}_{\perp}}{\gamma_v\,(1 - \mathbf{u}\cdot\mathbf{v}/c^2)}. \tag{3.20}$$

These equations are called the equations for the 'relativistic transformation of velocities' or 'relativistic addition of velocities'. The subscript on the γ symbol acts as a reminder that it refers to $\gamma(v)$ not $\gamma(u)$. If $\mathbf{u}$ and $\mathbf{v}$ are the velocities of two particles in any given reference frame, then $\mathbf{u}'$ is their relative velocity (think about it!).

When $\mathbf{u}$ is parallel to $\mathbf{v}$ we regain eq. (22.8).

When $\mathbf{u}$ is perpendicular to $\mathbf{v}$ we have $\mathbf{u}'_{\parallel} = -\mathbf{v}$ and $\mathbf{u}'_{\perp} = \mathbf{u}/\gamma$. The latter can be interpreted as an example of time dilation (in S$'$ the particle takes a longer time to cover a given distance). For this case $u'^2 = u^2 + v^2 - u^2v^2/c^2$.

Sometimes it is useful to express the results as a single vector equation. This is easily done using $\mathbf{u}_{\|} = (\mathbf{u}\cdot\mathbf{v})\mathbf{v}/v^2$ and $\mathbf{u}_{\perp} = \mathbf{u} - \mathbf{u}_{\|}$, giving:

$$\mathbf{u}' = \frac{1}{1-\mathbf{u}\cdot\mathbf{v}/c^2}\left[\frac{1}{\gamma_v}\mathbf{u} - \left(1 - \frac{\mathbf{u}\cdot\mathbf{v}}{c^2}\frac{\gamma_v}{1+\gamma_v}\right)\mathbf{v}\right]. \tag{3.21}$$

It will be useful to have the relationship between the gamma factors for $\mathbf{u}', \mathbf{u}$ and $\mathbf{v}$. One can obtain this by squaring (3.21) and simplifying, but the algebra is laborious. A much better way is to use an argument via proper time. This will be presented in section 3.5; the result is given in eq. (3.13). That equation also serves as a general proof that the velocity addition formulae never result in a speed $w > c$ when $u, v \le c$. For, if $u \le c$ and $v \le c$ then the right hand side of (3.13) is real and non-negative, and therefore $\gamma(w)$ is real, hence $w \le c$.

Let θ be the angle between $\mathbf{u}$ and $\mathbf{v}$, then $u_{\|} = u\cos\theta$, $u_{\perp} = u\sin\theta$, and from (3.20) we obtain

$$\tan\theta' = \frac{u'_{\perp}}{u'_{\|}} = \frac{u\sin\theta}{\gamma_v(u\cos\theta - v)}. \tag{3.22}$$

This is the way a direction of motion transforms between reference frames. In the formula $\mathbf{v}$ is the velocity of frame S′ relative to frame S. The classical (Galillean) result would give the same formula but with $\gamma = 1$. Therefore the distinctive effect of the Lorentz transformation is to 'throw' the velocity forward more than one might expect (as well as to prevent the speed exceeding c). See figure 3.6 for examples. (We shall present a quicker derivation of this formula in section 3.5.3 by using a 4-vector.)

3.4 Lorentz invariance and four-vectors

It is possible to continue by finding equations describing the transformation of acceleration, and then introducing force and its transformation. However, a much better insight into the whole subject is gained if we learn a new type of approach in which time and space are handled together.

First, let us arrange the coordinates t, x, y, z into a vector of four components. It is good practice to make all the elements of such a '4-vector' have the same physical dimensions, so we let the first component be ct, and define

$$\mathsf{X} \equiv \begin{pmatrix} ct \\ x \\ y \\ z \end{pmatrix}. \tag{3.23}$$

Is it ok to set $c = 1$?
It is a common practice to set $c = 1$ for convenience when doing mathematical manipulations in special relativity. Then one can leave c out of the equations, which reduces clutter and can make things easier. When you need to calculate a specific number for comparison with experiment, you must either put all the c's back into your final equations, or remember that the choice $c = 1$ is only consistent when the units of distance and time (and all other units that depend on them) are chosen appropriately. For example, one could work with seconds for time, and light-seconds for distance. (One light-second is equal to 299792458 metres). The only problem with this approach is that you must apply it consistently throughout. To identify the positions where c or a power of c appears in an equation, one can use dimensional analysis, but when one has further quantities also set equal to 1, this can require some careful thought. Alternatively you can make sure that all the units you use (including mass, energy etc.) are consistent with $c = 1$.
Some authors like to take this further, and argue that relativity teaches us that there is something basically wrong about giving different units to time and distance. We recognise that the height and width of any physical object are just different uses of essentially the same type of physical quantity, namely spatial distance, so the ratio of height to width is a dimensionless number. One might want to argue that, similarly, temporal and spatial separation are just different uses of essentially the same quantity, namely separation in spacetime, so the ratio of time to distance (what we call speed) should be regarded as dimensionless.
Ultimately this is a matter of taste. Clearly time and space are intimately related, but they are not quite the same: there is no way that a proper time could be mistaken for, or regarded as, a rest length, for example. My preference is to regard the statement 'set $c = 1$' as a shorthand for 'set $c = 1$ distance-unit per time-unit', in other words I don't regard speed as dimensionless, but I recognise that to choose 'natural units' can be convenient. 'Natural units' are units where c has the value '1 speed-unit'.

We will always use a capital letter and the plain font as in 'X' for 4-vector quantities. For the familiar '3-vectors' we use a bold Roman font as in '$\mathbf{x}$', and mostly but not always a small letter. You should think of 4-vectors as column vectors not row vectors, so that the Lorentz transformation equations can be written

$$\mathsf{X}' = \mathcal{L}\mathsf{X} \tag{3.24}$$

with

$$\mathcal{L} \equiv \begin{pmatrix} \gamma & -\gamma\beta & 0 & 0 \\ -\gamma\beta & \gamma & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \tag{3.25}$$

Question: Can we derive Special Relativity directly from the invariance of the interval? Do we have to prove that the interval is Lorentz-invariant first?
Answer: This question addresses an important technical point. It is good practice in physics to look at things in more than one way. A good way to learn Special Relativity is to take the Postulates as the starting point, and derive everything from there. This is approach adopted in *The Wonderful World of Relativity* and also in this book. Therefore you can regard the logical sequence as "postulates ⇒ Lorentz transformation ⇒ invariance of interval and other results." However, it turns out that the spacetime interval alone, if we *assume* its frame-independence, is sufficient to derive everything else! This more technical and mathematical argument is best assimilated after one is already familiar with Relativity. Therefore we are not adopting it at this stage, but some of the examples in this chapter serve to illustrate it. In order to proceed to General Relativity it turns out that the clearest line of attack is to assume by postulate that an invariant interval can be defined by combining the squares of coordinate separations, and then derive the nature of spacetime from that and some further assumptions about the impact of mass-energy on the interval. This leads to 'warping of spacetime', which we observe as a gravitational field.

where

$$\beta \equiv \frac{v}{c}. \tag{3.26}$$

The right hand side of equation (3.24) represents the product of a 4×4 matrix $\mathcal{L}$ with a 4×1 vector X, using the standard rules of matrix multiplication. You should check that eq. (3.24) correctly reproduces eqs. (3.1) to (3.4).

The inverse Lorentz transformation is obviously

$$\mathsf{X} = \mathcal{L}^{-1}\mathsf{X}' \tag{3.27}$$

(just multiply both sides of (3.24) by $\mathcal{L}^{-1}$), and one finds

$$\mathcal{L}^{-1} = \begin{pmatrix} \gamma & \gamma\beta & 0 & 0 \\ \gamma\beta & \gamma & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}. \tag{3.28}$$

It should not surprise us that this is simply $\mathcal{L}$ with a change of sign of β. You can confirm that $\mathcal{L}^{-1}\mathcal{L} = I$ where I is the identity matrix.

When we want to refer to the components of a 4-vector, we use the notation

$$\mathsf{X}^\mu = \mathsf{X}^0, \mathsf{X}^1, \mathsf{X}^2, \mathsf{X}^3, \qquad \text{or} \qquad \mathsf{X}^t, \mathsf{X}^x, \mathsf{X}^y, \mathsf{X}^z, \tag{3.29}$$

where the zeroth component is the 'time' component, ct for the case of X as defined by (3.23), and the other three components are the 'spatial' components, x, y, z for the case of (3.23). The reason to put the indices as superscipts rather than subscripts will emerge later.

3.4.1 Rapidity

Define a parameter ρ by

$$\tanh(\rho) = \frac{v}{c} = \beta, \tag{3.30}$$

then

$$\cosh(\rho) = \gamma, \;\; \sinh(\rho) = \beta\gamma, \;\; \exp(\rho) = \left(\frac{1+\beta}{1-\beta}\right)^{1/2}, \tag{3.31}$$

so the Lorentz transformation is

$$\mathcal{L} = \begin{pmatrix} \cosh\rho & -\sinh\rho & 0 & 0 \\ -\sinh\rho & \cosh\rho & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}. \tag{3.32}$$

The quantity ρ is called the *hyperbolic parameter* or the *rapidity*. The form (3.32) can be regarded as a 'rotation' through an imaginary angle $i\rho$. This form makes some types of calculation easy. For example, the addition of velocities formula $w = (u+v)/(1+uv/c^2)$ (for motions all in the same direction) becomes

$$\tanh\rho_w = \frac{\tanh\rho_u + \tanh\rho_v}{1+\tanh\rho_u \tanh\rho_v}$$

where $\tanh\rho_w = w/c$, $\tanh\rho_u = u/c$, $\tanh\rho_v = v/c$. I hope you are familiar with the formula for $\tanh(A+B)$, because if you are then you will see immediately that the result can be expressed as

$$\rho_w = \rho_u + \rho_v. \tag{3.33}$$

Thus, for the case of relative velocities all in the same direction, the rapidities add, a simple result. An example application to straight line motion is discussed in section 4.2.1.

Example. A rocket engine is programmed to fire in bursts such that each time it fires, the rocket achieves a velocity increment of u, meaning that in the inertial frame where the rocket is at rest before the engine fires, its speed is u after the engine stops. Calculate the speed w of the rocket relative to its starting rest frame after n such bursts, all collinear.

Answer. Define the rapidities ρ_u and ρ_w by $\tanh \rho_u = u/c$ and $\tanh \rho_w = w/c$, then by (3.33) we have that ρ_w is given by the sum of n increments of ρ_u, i.e. $\rho_w = n\rho_u$. Therefore $w = c \tanh(n\rho_u)$. (This can also be written $w = c(z^n - 1)/(z^n + 1)$ where $z = \exp(2\rho_u)$.)

You can readily show that the Lorentz transformation can also be written in the form

$$\begin{pmatrix} ct' + x' \\ ct' - x' \\ y' \\ z' \end{pmatrix} = \begin{pmatrix} e^{-\rho} & & & \\ & e^{\rho} & & \\ & & 1 & \\ & & & 1 \end{pmatrix} \begin{pmatrix} ct + x \\ ct - x \\ y \\ z \end{pmatrix}. \tag{3.34}$$

We shall mostly not adopt this form, but it is useful in some calculations.

3.4.2 Lorentz invariant quantities

Under a Lorentz transformation, a 4-vector changes, but not out of all recognition. In particular, a 4-vector has a size or 'length' that is not affected by Lorentz transformations. This is like 3-vectors, which preserve their length under rotations, but the 'length' has to be calculated in a specific way.

To find our way to the result we need, first recall how the length of a 3-vector is calculated. For $\mathbf{r} = (x, y, z)$ we would have $r \equiv |\mathbf{r}| \equiv \sqrt{x^2 + y^2 + z^2}$. In vector notation, this is

$$|\mathbf{r}|^2 = \mathbf{r} \cdot \mathbf{r} = \mathbf{r}^T \mathbf{r} \tag{3.35}$$

where the dot represents the scalar product, and in the last form we assumed $\mathbf{r}$ is a column vector, and $\mathbf{r}^T$ denotes its transpose, i.e. a row vector. Multiplying that 1×3 row vector onto the 3×1 column vector in the standard way results in a 1×1 'matrix', in other words a scalar, equal to $x^2 + y^2 + z^2$.

The 'length' of a 4-vector is calculated similarly, but with a crucial sign that enters in because time and space are not exactly the same as each other. For the 4-vector X given in eq. (3.23), you are invited to check that the combination

$$-(\mathsf{X}^0)^2 + (\mathsf{X}^1)^2 + (\mathsf{X}^2)^2 + (\mathsf{X}^3)^2 \tag{3.36}$$

is 'Lorentz-invariant'. That is,

$$-c^2 t'^2 + x'^2 + y'^2 + z'^2 = -c^2 t^2 + x^2 + y^2 + z^2, \tag{3.37}$$

c.f. eq. (2.7). In matrix notation, this quantity can be written

$$-c^2 t^2 + x^2 + y^2 + z^2 = \mathsf{X}^T g \mathsf{X} \tag{3.38}$$

where

$$g = \begin{pmatrix} -1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}. \tag{3.39}$$

More generally, if A is a 4-vector, and $\mathsf{A}' = \mathcal{L}\mathsf{A}$, then we have

$$\begin{aligned} \mathsf{A}'^T g \mathsf{A}' &= (\mathcal{L}\mathsf{A})^T g (\mathcal{L}A) \\ &= A^T (\mathcal{L}^T g \mathcal{L}) A, \end{aligned} \tag{3.40}$$

(where we used $(MN)^T = N^T M^T$ for any pair of matrices M, N). Therefore $\mathsf{A}'^T g \mathsf{A}' = \mathsf{A}^T g \mathsf{A}$ as long as

$$\boxed{\mathcal{L}^T g \mathcal{L} = g.} \tag{3.41}$$

You should now check that g as given in eq. (3.39) is indeed the solution to this matrix equation. This proves that for *any* quantity A that transforms in the same way as X, the scalar quantity $\mathsf{A}^T g \mathsf{A}$ is 'Lorentz-invariant', meaning that it does not matter which reference frame is picked for the purpose of calculating it, the answer will always come out the same.

g is called 'the metric' or 'the metric tensor'. A generalized form of it plays a central role in General Relativity.

symbol	definition	components	name(s)	invariant
X	X	$(ct, \mathbf{r})$	4-displacement, interval	$-c^2\tau^2$
U	$d\mathsf{X}/d\tau$	$(\gamma c, \gamma \mathbf{u})$	4-velocity	$-c^2$
P	$m_0\mathsf{U}$	$(E/c, \mathbf{p})$	energy-momentum, 4-momentum	$-m_0^2c^2$
F	$d\mathsf{P}/d\tau$	$(\gamma W/c, \gamma \mathbf{f})$	4-force, work-force	
J	$\rho_0\mathsf{U}$	$(c\rho, \mathbf{j})$	4-current density	$-c^2\rho_0^2$
A	A	$(\varphi/c, \mathbf{A})$	4-vector potential	
A	$d\mathsf{U}/d\tau$	$\gamma(\dot\gamma c, \dot\gamma \mathbf{u} + \gamma \mathbf{a})$	4-acceleration	a_0^2
K	$\Box\phi$	$(\omega/c, \mathbf{k})$	wave vector	

Table 3.2: A selection of useful 4-vectors. Some have more than one name. Their definition and use is developed in the text. The Lorentz factor γ is γ_u, i.e. it refers to the speed u of the particle in question in the given reference frame. $\dot\gamma$ is used for $d\gamma/dt$ and $W = dE/dt$. The last column gives the invariant squared 'length' of the 4-vector, but is omitted in those cases where it is less useful in analysis. Above the line are time-like 4-vectors; below the line the acceleration is space-like, the wave vector may be space-like or time-like.

In the case of the spacetime displacement (or 'interval') 4-vector X, the invariant 'length' we are discussing is the spacetime interval s previewed in eq. (2.7), taken between the origin and the event at X. As we mentioned in eq. (22.1), in the case of timelike intervals the invariant interval length is c times the proper time. To see this, calculate the length in the reference frame where the X has no spatial part, i.e. $x = y = z = 0$. Then it is obvious that $\mathsf{X}^T g \mathsf{X} = -c^2t^2$ and the time t is the *proper* time between the origin event 0 and the event at X, because it is the time in the frame where O and X occur at the same position.

Timelike intervals have a negative value for $s^2 \equiv -c^2t^2 + (x^2 + y^2 + z^2)$, so taking the square root would produce an imaginary number. However the significant quantity is the proper time given by $\tau = (-s^2)^{1/2}/c$; this is real not imaginary. In algebraic manipulations mostly it is not necessary to take the square root in any case. For intervals lying on the surface of a light cone the 'length' is zero and these are called **null** intervals.

Table 3.2 gives a selection of 4-vectors and their associated Lorentz-invariant 'length-squared'. These 4-vectors and the use of invariants in calculations will be developed as we proceed. The terminology 'timelike', 'null' and 'spacelike' is extended to all 4-vectors in an obvious way, according as $(\mathsf{A}^0)^2$ is greater than, equal to, or less than $(\mathsf{A}^1)^2 + (\mathsf{A}^2)^2 + (\mathsf{A}^3)^2$. N.B. a 'null' 4-vector is not necessarily zero; rather it is a 'balanced' 4-vector, poised on the edge between timelike and spacelike.

It is helpful to have a mathematical definition of what we mean in general by a 4-vector. The definition is: *a 4-vector is any set of four scalar quantities that transform in the same way as (ct, x, y, z) under a change of reference frame.* Such a definition is useful because it means that we can infer that the basic rules of vector algebra apply to 4-vectors. For example, the sum of two 4-vectors A and B, written $\mathsf{A} + \mathsf{B}$, is evaluated by summing the corresponding components, just as is done for 3-vectors. Standard rules of matrix multiplication apply, such as $\mathcal{L}(\mathsf{A} + \mathsf{B}) = \mathcal{L}\mathsf{A} + \mathcal{L}\mathsf{B}$. A small change in a 4-vector, written for example $d\mathsf{A}$, is itself a

4-vector.

You can easily show that (3.41) implies that $\mathsf{A}^T g \mathsf{B}$ is Lorentz-invariant for any pair of 4-vectors A, B. This combination is essentially a form of scalar product, so for 4-vectors we define

$$\mathsf{A} \cdot \mathsf{B} \equiv \mathsf{A}^T g \mathsf{B}. \tag{3.42}$$

That is, a central dot operator appearing between two 4-vector symbols is defined to be a shorthand notation for the combination $\mathsf{A}^T g \mathsf{B}$. The result is a scalar and it is referred to as the 'scalar product' of the 4-vectors. In terms of the components it is

$$-\mathsf{A}^0 \mathsf{B}^0 + (\mathsf{A}^1 \mathsf{B}^1 + \mathsf{A}^2 \mathsf{B}^2 + \mathsf{A}^3 \mathsf{B}^3).$$

A 'vector product' or 'cross product' can also be defined for 4-vectors, but it requires a 4×4 matrix to be introduced; this will be deferred until chapter 9.

3.5 Basic 4-vectors

3.5.1 Proper time

Consider a worldline, such as the one shown in figure 2.2. We would like to describe events along this line, and if possible we would like a description that does not depend on a choice of frame of reference. This is just like the desire to do classical (Newtonian) mechanics without picking any particular coordinate system: in Newtonian mechanics it is achieved by using 3-vectors. In Special Relativity, we use 4-vectors. We also need a parameter to indicate which event we are talking about, i.e. how 'how far' along the worldline it is. In Newtonian mechanics this job was done by the time, because that was a universal among reference frames connected by a Galilean transformation. In Special Relativity we use the *proper time* τ. By this we mean the integral of all the little infinitesimal bits of proper time 'experienced' by the particle along its history. This is a suitable choice because this proper time is Lorentz-invariant, i.e. agreed among all reference frames.

This basic role of proper time is a central idea of the subject.

In Newtonian mechanics a particle's motion is described by using a position 3-vector $\mathbf{r}$ that is a function of time, so $\mathbf{r}(t)$. This is a shorthand notation for three functions of t; the time t serves as a parameter. In relativity when we use a 4-vector to describe the worldline of some object, you should think of it as a function of the proper time along the worldline, so $\mathsf{X}(\tau)$. This is a shorthand notation for four functions of τ; the proper time τ serves as a parameter.

Let X be the displacement 4-vector describing a given worldline. This means its components in any reference frame S give $ct, x(t), y(t), z(t)$ for the trajectory relative to that frame. Two

4-vector notation; metric signature

Unfortunately there is more than one convention concerning notation for 4-vectors. There are two issues: the order of components, and the sign of the metric. For the former, the notation adopted in this book is the one that is most widely used now, but in the past authors have sometimes preferred to put the time component last instead of first, and then numbered the components 1 to 4 instead of 0 to 3. Also, sometimes you find $i = \sqrt{-1}$ attached to the time component. This is done merely to allow the invariant length-squared to be written $\sum_\mu (\mathsf{A}^\mu)^2$, the i^2 factor then takes care of the sign. One reason to prefer the introduction of the g matrix (eq. (3.39)) to the use of i is that it allows the transition to General Relativity to proceed more smoothly.

The second issue is the sign of g. When making the transition from Special to General Relativity, the almost universal practice in writing the Minkowski metric g is the one adopted in this book. However, within purely special relativistic treatments another convention is common, and is widely adopted in the particle physics community. This is to define g with the signs $1, -1, -1, -1$ down the diagonal, i.e. the negative of the version we adopt here. As long as one is consistent either convention is valid, but beware: changing convention will result in a change of sign of all scalar products. For example, we have $\mathsf{P}\cdot\mathsf{P} = -m^2c^2$ for the energy-momentum 4-vector, but the other choice of metric would give $\mathsf{P}\cdot\mathsf{P} = m^2c^2$. The trace of the metric (the sum of the diagonal elements) is called the *signature*. Our metric has signature $+2$, the other choice has signature -2. The reason that $1, -1, -1, -1$ is preferred by many authors is that it makes timelike vectors have positive 'size', and most of the important basic vectors are timelike (see table 3.2). However the reasons to prefer $-1, 1, 1, 1$ outweigh this in my opinion. They are

1. It is confusing to use $(+1, -1, -1, -1)$ in General Relativity.
2. Expressions like $\mathsf{U}\cdot\mathsf{P}$ ought to remind us of $\mathbf{u}\cdot\mathbf{p}$.
3. It is more natural to take the 4-gradient as $(-\partial/\partial ct, \partial/\partial x, \partial/\partial y, \partial/\partial z)$ since then it more closely resembles the familiar 3-gradient.

The 4-gradient (item 3) will be introduced in chapter 5 and its relation to the metric explained in chapter 9.

close together events on the worldline are (ct, x, y, z) and $(c(t+dt), x+dx, y+dy, z+dz)$. The proper time between these events is

$$\mathrm{d}\tau = \frac{(c^2dt^2 - dx^2 - dy^2 - dz^2)^{1/2}}{c} \tag{3.43}$$

$$= \mathrm{d}t\left(1 - u^2/c^2\right)^{1/2} \tag{3.44}$$

where $\mathbf{u} = (\mathrm{d}x/\mathrm{d}t, \mathrm{d}y/\mathrm{d}t, \mathrm{d}z/\mathrm{d}t)$ is the velocity of the particle in S. We thus obtain the important relation

$$\boxed{\frac{\mathrm{d}t}{\mathrm{d}\tau} = \gamma} \tag{3.45}$$

for neighbouring events on a worldline, where the γ factor is the one associated with the velocity of the particle in the reference frame in which t is calculated.

Eq. (3.12ii) concerns the time between events on a worldline as observed in two frames, neither of which is the rest frame. The worldline is that of a a particle having velocity $\mathbf{u}$ in the frame S, with $\mathbf{v}$ the velocity of S$'$ relative to S. To derive the result, let $(t, \mathbf{r}) = (t, \mathbf{u}t)$ be the coordinates in S of an event on the worldline of the first particle, then the Lorentz transformation gives

$$t' = \gamma_v(t - vx/c^2) = \gamma_v(t - \mathbf{u}\cdot\mathbf{v}t/c^2).$$

Differentiating with respect to t, with all the velocities held constant, gives eq. (3.12ii).

3.5.2 Velocity, acceleration

We have a 4-vector for spacetime displacement, so it is natural to ask whether there is a 4-vector for velocity, defined as a rate of change of the 4-displacement of a particle. To construct such a quantity, we note first of all that for 4-vector X, a small change $d\mathsf{X}$, is itself a 4-vector. To get a 'rate of change of X' we should take the ratio of $d\mathsf{X}$ to a small time interval, but take care: if we want the result to be a 4-vector then the small time interval had better be Lorentz invariant. Fortunately there is a Lorentz-invariant time interval that naturally presents itself: the proper time along the worldline. We thus arrive at the definition

$$\text{4-velocity} \qquad \mathsf{U} \equiv \frac{\mathrm{d}\mathsf{X}}{\mathrm{d}\tau}. \tag{3.46}$$

The 4-velocity four-vector has a direction in spacetime pointing along the worldline.

If we want to know the components of the 4-velocity in any particular frame, we use (3.45):

$$\mathsf{U} \quad\equiv\quad \frac{\mathrm{d}\mathsf{X}}{\mathrm{d}\tau} = \frac{\mathrm{d}\mathsf{X}}{\mathrm{d}t}\frac{\mathrm{d}t}{\mathrm{d}\tau} = (\gamma_u c,\, \gamma_u \mathbf{u}). \tag{3.47}$$

The invariant length or size of the 4-velocity is just c (this is obvious if you calculate it in the rest frame, but for practice you should do the calculation in a general reference frame too). This size is not only Lorentz invariant (that is, the same in all reference frames) but also constant (that is, not changing with time), even though U can change with time (it is the 4-velocity of a general particle undergoing any form of motion, not just inertial motion). In units where $c = 1$, a 4-velocity is a unit vector.

4-acceleration is defined as one would expect by $\mathsf{A} = d\mathsf{U}/d\tau = d^2\mathsf{X}/d\tau^2$, but now the relationship to a 3-vector is more complicated:

$$\mathsf{A} \equiv \frac{\mathrm{d}\mathsf{U}}{\mathrm{d}\tau} = \gamma\frac{\mathrm{d}\mathsf{U}}{\mathrm{d}t} = \gamma\left(\frac{\mathrm{d}\gamma}{\mathrm{d}t}c, \frac{\mathrm{d}\gamma}{\mathrm{d}t}\mathbf{u} + \gamma\mathbf{a}\right) \tag{3.48}$$

where, of course, $\gamma = \gamma(u)$ and $\mathbf{a}$ is the 3-acceleration. Using $\mathrm{d}\gamma/\mathrm{d}t = (\mathrm{d}\gamma/\mathrm{d}u)(\mathrm{d}u/\mathrm{d}t)$ with the γ relation (3.11) and[1] $\mathrm{d}u/\mathrm{d}t = (\mathbf{u}\cdot\mathbf{a})/u$, we find

$$\frac{\mathrm{d}\gamma}{\mathrm{d}t} = \gamma^3\frac{\mathbf{u}\cdot\mathbf{a}}{c^2}. \tag{3.49}$$

Therefore

$$\mathsf{A} = \gamma^2\left(\frac{\mathbf{u}\cdot\mathbf{a}}{c}\gamma^2, \frac{\mathbf{u}\cdot\mathbf{a}}{c^2}\gamma^2\mathbf{u} + \mathbf{a}\right). \tag{3.50}$$

In the rest frame of the particle this expression simplifies to

$$\mathsf{A} = (0, \mathbf{a}_0) \tag{3.51}$$

where we write $\mathbf{a}_0$ for the acceleration observed in the rest frame. If one takes an interest in the scalar product $\mathsf{U}\cdot\mathsf{A}$, one may as well evaluate it in the rest frame, and thus one finds that

$$\mathsf{U}\cdot\mathsf{A} = 0. \tag{3.52}$$

That is, *the 4-acceleration is always orthogonal to the 4-velocity.* This makes sense because the magnitude of the 4-velocity should not change: it remains a unit vector. 4-velocity is timelike and 4-acceleration is spacelike and orthogonal to it. This does not imply that 3-acceleration is orthogonal to 3-velocity, of course (it can be but usually is not).

Using the Lorentz-invariant length-squared of A one can relate the acceleration in any given reference frame to the acceleration in the rest frame $\mathbf{a}_0$:

$$\gamma^4\left(-\left(\frac{\mathbf{u}\cdot\mathbf{a}}{c}\right)^2\gamma^4 + \left(\frac{\mathbf{u}\cdot\mathbf{a}}{c^2}\gamma^2\mathbf{u} + \mathbf{a}\right)^2\right) = a_0^2. \tag{3.53}$$

This simplifies to

$$a_0^2 = \gamma^4 a^2 + \gamma^6(\mathbf{u}\cdot\mathbf{a})^2/c^2 \quad = \gamma^6(a^2 - (\mathbf{u}\wedge\mathbf{a})^2/c^2). \tag{3.54}$$

[1] $u = (\mathbf{u}\cdot\mathbf{u})^{1/2} \Rightarrow \mathrm{d}u/\mathrm{d}t = (1/2)(u^2)^{-1/2}(\mathbf{u}\cdot\mathbf{a} + \mathbf{a}\cdot\mathbf{u}) = \mathbf{u}\cdot\mathbf{a}/u$, or use $(\mathrm{d}/\mathrm{d}t)(u_x^2 + u_y^2 + u_z^2)^{1/2}$.

where we give two versions for the sake of convenience in later discussions. As a check, you can obtain the first version from the second by using the triple product rule.

When **u** and **a** are orthogonal, (3.54) gives $a_0 = \gamma^2 a$. For example, for circular motion the acceleration in the instantaneous rest frame is γ^2 times larger than the value in the rest frame of the circle, $a = u^2/r$.

When **u** and **a** are parallel, (3.54) gives $a_0 = \gamma^3 a$. Hence straight-line motion at constant a_0 is motion at constant $\gamma^3 a$. Using the gamma relation (3.11ii) this is motion at constant $(d/dt)(\gamma v)$, in other words constant rate of change of momentum, i.e. constant force. This will be discussed in detail in section 4.2.1. As γ increases, the acceleration in the original rest frame falls in proportion to $1/\gamma^3$, which is just enough to maintain a_0 at a constant value.

Addition of velocities: a comment

In section 3.5.2 we showed that the velocity 4-vector describing the motion of a particle has a constant magnitude or 'length', equal to c. It is a unit vector when $c = 1$ unit. This means that one should treat with caution the sum of two velocity 4-vectors:

$$\mathsf{U}_1 + \mathsf{U}_2 = ? \tag{3.55}$$

Although the sum on the left hand side is mathematically well-defined, the sum of two 4-velocities does *not* make another 4-velocity, because the sum of two timelike unit vectors is not a unit vector.

The idea of adding velocity vectors comes from classical physics, but if one pauses to reflect one soon realises that it is not the same sort of operation as, for example, adding two displacements. A displacement in spacetime added to another displacement in spacetime corresponds directly to another displacement. For the case of timelike displacements, for example, it could represent a journey from event A to event B, followed by a journey from event B to event C (where each journey has a definite start and finish time as well as position). Hence it makes sense to write

$$\mathsf{X}_1 + \mathsf{X}_2 = \mathsf{X}_3. \tag{3.56}$$

Adding velocity 4-vectors, however, gives a quantity with no ready physical interpretation. It is a bit like forming a sum of temperatures: one can add them up, but what does it mean? In the classical case the sum of 3-vector velocities makes sense because the velocity of an object C relative to another object A is given by the vector sum of the velocity of C relative to B and the velocity of B relative to A. In Special Relativity velocities don't sum like this: one must use instead the velocity transformation equations (3.20).

3.5.3 Momentum, energy

Supposing that we would like to develop a 4-vector quantity that behaves like momentum, the natural thing to do is to try multiplying a 4-velocity by a mass. We must make sure the mass we pick is Lorentz-invariant, which is easy: just use the rest mass. Thus we arrive at the definition

$$\text{4-momentum } \mathsf{P} \equiv m_0 \mathsf{U} = m_0 \frac{\mathrm{d}\mathsf{X}}{\mathrm{d}\tau}. \tag{3.57}$$

P, like U, points along the worldline. Using (3.12) we can write the components of P in any given reference frame as

$$\mathsf{P} = \gamma m_0 \frac{\mathrm{d}\mathsf{X}}{\mathrm{d}t} = (\gamma_u m_0 c,\, \gamma_u m_0 \mathbf{u}) \tag{3.58}$$

for a particle of velocity $\mathbf{u}$ in the reference frame.

In the next chapter (section 4.3), relativistic expressions for 3-momentum and energy will be developed. The argument can also be found in *The Wonderful World* and other references such as Feynman's lectures and the book by Taylor and Wheeler. One obtains the important expressions

$$E = \gamma m_0 c^2, \qquad \mathbf{p} = \gamma m_0 \mathbf{u} \tag{3.59}$$

for the energy and 3-momentum of a particle of rest mass m_0 and velocity $\mathbf{u}$. It follows that the 4-momentum can also be written

$$\mathsf{P} = (E/c, \mathbf{p})$$

and for this reason P is also called the energy-momentum 4-vector.

In the present chapter we have obtained this 4-vector quantity purely by mathematical argument, and we can call it 'momentum' if we chose. The step of claiming that this quantity has a conservation law associated with it is a further step, it is a statement of physical law. This will be presented in the next chapter.

The relationship

$$\frac{\mathbf{p}}{E} = \frac{\mathbf{u}}{c^2} \tag{3.60}$$

(which follows from (3.59)) can be useful for obtaining the velocity if the momentum and energy are known.

Invariant, covariant, conserved
Invariant or 'Lorentz-invariant' means *the same in all reference frames*
Covariant is, strictly, a technical term applied to four-vector quantities, but it is often used to mean 'invariant' when it is the mathemtical form of an equation (such as $\mathsf{F} = d\mathsf{P}/d\tau$) that is invariant
Conserved means 'not changing with time' or 'the same before and after'.
Rest mass is Lorentz-invariant but not conserved. Energy is conserved but not Lorentz-invariant.

We used the symbol m_0 for rest mass in the formulae above. This was for the avoidance of all doubt, so that it is clear that this is a rest mass and not some other quantity such as γm_0. Since rest mass is Lorentz invariant, however, it is by far the most important mass-related concept, and for this reason the practice of referring to γm_0 as 'relativistic mass' is mostly unhelpful. It is best avoided. Therefore we shall never use the symbol m to refer to γm_0. This frees us from the need to attach a subscript zero: throughout this book the symbol m will only ever refer to rest mass.

3.5.4 The direction change of a 4-vector under a boost

The simplicity of the components in $\mathsf{P} = (E/c, \mathbf{p})$ makes P a convenient 4-vector to work with in many situations. For example, to obtain the formula (3.22) for the transformation of a direction of travel, we can use the fact that P is a 4-vector. Suppose a particle has 4-momentum P in frame S. The 4-vector nature of P means that it transforms as $\mathsf{P}' = \mathcal{L}\mathsf{P}$ so

$$\begin{aligned} E'/c &= \gamma(E/c - \beta p_x), \\ p'_x &= \gamma(-\beta E/c + p_x), \\ p'_y &= p_y, \end{aligned}$$

and since the velocity is parallel to the momentum we can find the direction of travel in frame S$'$ by $\tan\theta' = p'_y/p'_x$:

$$\tan\theta' = \frac{p_y}{\gamma(-vE/c^2 + p_x)} = \frac{u_y}{\gamma_v(-v + u_x)} = \frac{u\sin\theta}{\gamma_v(u\cos\theta - v)},$$

where we used (3.60). This is valid for any 4-vector, if we take it that u refers to the ratio of the spatial to the temporal part of the 4-vector, multiplied by the speed of light.

Figure 3.2 gives a graphical insight into this result (see the caption for the argument). The diagram can be applied to any 4-vector, but since it can be useful when considering collision processes, an energy-momentum 4-vector is shown for illustrative purposes.

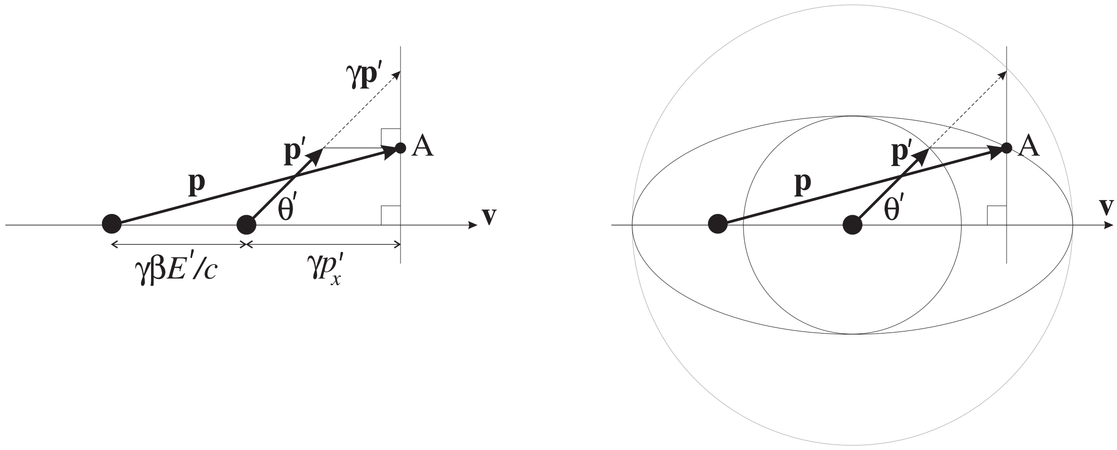


Figure 3.2: A graphical method for obtaining the direction in space of a 4-vector after a Lorentz 'boost', i.e. a change to another reference frame whose axes are aligned with the first. (N.B. this is neither a spacetime diagram nor a picture in space, it is purely a mathematical construction). Let frame S′ be in standard configuration with S. $\mathbf{p}'$ is a momentum vector in S′. The point A on the diagram is located such that its y position agrees with p'_y, and its x position is $\gamma p'_x$ from the foot of $\mathbf{p}'$. $\mathbf{p}$ is the momentum vector as observed in frame S. It is placed so that its foot is at a distance $\gamma\beta E'/c$ to the left of the foot of $\mathbf{p}$, and it extends from there to A. It is easy to check that it thus has the correct x and y components as given by Lorentz transformation of $\mathbf{p}'$. The interest is that one can show that when θ' varies while maintining p' fixed, the point A moves around an ellipse. Therefore the right hand diagram shows the general pattern of the relationship between $\mathbf{p}$ and $\mathbf{p}'$.

In the case of a null 4-vector (e.g. P for a zero-rest-mass particle) another form is often useful:

$$\cos\theta' = \frac{cp'_x}{E'} = \frac{\gamma(-\beta E/c + p\cos\theta)}{\gamma(E/c - \beta p\cos\theta)} = \frac{\cos\theta - \beta}{1 - \beta\cos\theta} \tag{3.61}$$

where we used $E = pc$.

3.5.5 Force

We now have at least two ways in which force could be introduced:

$$\mathsf{F} \stackrel{?}{=} m_0\mathsf{A} \qquad \text{or} \quad \mathsf{F} \stackrel{?}{=} \frac{\mathrm{d}\mathsf{P}}{\mathrm{d}\tau}. \tag{3.62}$$

Both of these are perfectly well-defined 4-vector equations, but they are not the same because the rest mass is not always constant. We are free to choose either because the relation is a *definition* of 4-force, and we can define things how we like. However, some definitions are more useful than others, and there is no doubt which one permits the most elegant theoretical description of the large quantity of available experimental data, it is the second:

$$\mathsf{F} \equiv \frac{\mathrm{d}\mathsf{P}}{\mathrm{d}\tau}. \tag{3.63}$$

The reason why this is the most useful way to define 4-force is related to the fact that P is conserved.

We have

$$\mathsf{F} = \frac{\mathrm{d}\mathsf{P}}{\mathrm{d}\tau} = \left(\frac{1}{c}\frac{\mathrm{d}E}{\mathrm{d}\tau}, \frac{\mathrm{d}\mathbf{p}}{\mathrm{d}\tau}\right),$$

where $\mathbf{p}$ is the relativistic 3-momentum $\gamma m_0 \mathbf{u}$. To work with F in practice it will often prove helpful to adopt a particular reference frame and study its spatial and temporal components separately. To this end we *define* a vector $\mathbf{f}$ by

$$\mathbf{f} \equiv \frac{\mathrm{d}\mathbf{p}}{\mathrm{d}t} \tag{3.64}$$

and this is called the force or 3-force. Then we have

$$\mathsf{F} = \frac{\mathrm{d}\mathsf{P}}{\mathrm{d}\tau} = \gamma\frac{\mathrm{d}\mathsf{P}}{\mathrm{d}t} = \gamma\frac{\mathrm{d}}{\mathrm{d}t}(E/c, \mathbf{p}) = (\gamma W/c,\, \gamma\mathbf{f}). \tag{3.65}$$

where $W = dE/dt$ can be recognised as the rate of doing work by the force.

3.5.6 Wave vector

Another 4-vector appears in the analysis of wave motion. It is the wave-4-vector (or "4-wave-vector")

$$\mathsf{K} = (\omega/c,\ \mathbf{k}) \tag{3.66}$$

where ω is the angular frequency of the wave, and $\mathbf{k}$ is the spatial wave-vector, which points in the direction of propagation and has size $k = 2\pi/\lambda$ for wavelength λ. We shall postpone the proof that K is a 4-vector till chapter 5. We introduce it here because it offers the most natural way to discuss the general form of the Doppler effect, for a source moving in an arbitrary direction. Note, the waves described by $(\omega/c,\ \mathbf{k})$ could be any sort of wave motion, not just light waves. They could be waves on water, or pressure waves, etc. The 4-wave-vector can refer to any quantity a whose behaviour in space and time takes the form

$$a = a_0 \cos(\mathbf{k}\cdot\mathbf{r} - \omega t)$$

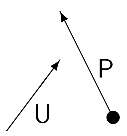
where the wave amplitude a_0 is a constant. The phase of the wave is

$$\phi = \mathbf{k}\cdot\mathbf{r} - \omega t = \mathsf{K}\cdot\mathsf{X}.$$

Since ϕ can be expressed as a dot product of 4-vectors, it is a Lorentz invariant quantity[2].

3.6 The joy of invariants

Suppose an observer moving with 4-velocity U observes a particle having 4-momentum P.



What is the energy E_O of the particle relative to the observer?

This is an eminently practical question, and we should like to answer it. One way (don't try it!) would be to express the P in component form in some arbitrary frame and Lorentz-transform to the rest frame of the observer. However you should learn to think in terms of 4-vectors, and not go to components if you don't need to.

[2] In chapter 5 we start by showing that ϕ is invariant without mentioning K, and then define K as its 4-gradient.

We know the quantity we are looking for must depend on both U and P, and it is a scalar. Therefore let's consider $\mathsf{U} \cdot \mathsf{P}$. This is such a scalar and has physical dimensions of energy. Evaluate it in the rest frame of the observer: there $\mathsf{U} = (c, 0, 0, 0)$ so we get minus c times the zeroth component of P in that frame, i.e. the particle's energy E in that frame, which is the very thing we wanted. In symbols, this is $\mathsf{U} \cdot \mathsf{P} = -E_{\rm O}$. Now bring in the fact that $\mathsf{U} \cdot \mathsf{P}$ is Lorentz invariant. This means that nothing was overlooked by evaluating it in one particular reference frame, it will always give $E_{\rm O}$. We are done: *the energy of the particle relative to the observer is* $-\mathsf{U} \cdot \mathsf{P}$.

This calculation illustrates a very important technique called **the method of invariants**. The idea has been stated beautifully by Hagedorn:

> *"If a question is of such a nature that its answer will always be the same, no matter in which inertial frame one starts, it must be possible to formulate the answer entirely with the help of those invariants which one can build with the available 4-vectors*[3]*. One then finds the answer in a particular inertial frame which one can choose freely and in such a way that the answer is there obvious or most easy. One looks then how the invariants appear in this particular system, expresses the answer to the problem by these same invariants, and one has found at the same time the general answer."*

He goes on to add that it is worthwhile to devote some time to thinking this through until one has understood that there is no hocus-pocus or guesswork and the method is completely safe. I agree!

Example. For any isolated system of particles, there exists a reference frame in which the total 3-momentum is zero. Such a frame is called the CM (centre of momentum) frame. For a system of two particles of 4-momenta P_1, P_2, what is the total energy in the CM frame?

Answer. We have three invariants to hand: $\mathsf{P}_1 \cdot \mathsf{P}_1 = -m_1^2 c^2$, $\mathsf{P}_2 \cdot \mathsf{P}_2 = -m_2^2 c^2$, and $\mathsf{P}_1 \cdot \mathsf{P}_2$. Other invariants, such as $(\mathsf{P}_1 + \mathsf{P}_2) \cdot (\mathsf{P}_1 + \mathsf{P}_2)$, can be expressed in terms of these three. Let S$'$ be the CM frame. In the CM frame the total energy is obviously $E'_1 + E'_2$. We want to write this in terms of invariants. In the CM frame we have, by definition, $\mathbf{p}'_1 + \mathbf{p}'_2 = 0$. This means that $(\mathsf{P}'_1 + \mathsf{P}'_2)$ has zero momentum part, and its energy part is the very thing we have been asked for. Therefore the answer can be written

$$E^{\rm CM}_{\rm tot} = E'_{\rm tot} = c\sqrt{-(\mathsf{P}'_1 + \mathsf{P}'_2) \cdot (\mathsf{P}'_1 + \mathsf{P}'_2)} = c\sqrt{-(\mathsf{P}_1 + \mathsf{P}_2) \cdot (\mathsf{P}_1 + \mathsf{P}_2)}, \tag{3.67}$$

where the last step used the invariant nature of the scalar product. We now have the answer we want in terms of the given 4-momenta, and it does not matter in what frame ('laboratory frame') they may have been specified.

We can now derive the eq. (3.13) relating the Lorentz factors for different 3-velocities. This is easily done by considering the quantity $\mathsf{U} \cdot \mathsf{V}$ where U and V are the 4-velocities of particles

[3] In a later chapter we shall generalise the use of invariants to tensors of any rank.

moving with velocities **u**, **v** in some frame. Then, using (3.47) twice,

$$\mathsf{U}\cdot\mathsf{V} = \gamma_u\gamma_v(-c^2 + \mathbf{u}\cdot\mathbf{v}).$$

Let **w** be the relativity 3-velocity of the particles, which is equal to the velocity of one particle in the rest frame of the other. In the rest frame of the first particle its velocity would be zero and that of the other particle would be **w**. Evaluating $\mathsf{U}\cdot\mathsf{V}$ in that frame gives

$$\mathsf{U}'\cdot\mathsf{V}' = -\gamma_w c^2.$$

Now use the fact that $\mathsf{U}\cdot\mathsf{V}$ is Lorentz-invariant. This means that evaluating it in any frame must give the same answer, so the above two expressions are equal:

$$\gamma_w c^2 = \gamma_u\gamma_v(c^2 - \mathbf{u}\cdot\mathbf{v}).$$

This is eq. (3.13). (See exercise **??** for another method).

3.7 Moving light sources

3.7.1 The Doppler effect

Suppose a wave source in frame S′ emits a plane wave of angular frequency ω_0 in a direction making angle θ_0 with the x' axis (we are using the subscript zero here to indicate the value in the frame where the source is at rest). Then the wave 4-vector in S' is $\mathsf{K}' = (\omega_0/c, k_0\cos\theta_0, k_0\sin\theta_0, 0)$.

Applying the inverse Lorentz transformation, the wave 4-vector in S is

$$\begin{pmatrix} \omega/c \\ k\cos\theta \\ k\sin\theta \\ 0 \end{pmatrix} = \begin{pmatrix} \gamma & \gamma\beta & 0 & 0 \\ \gamma\beta & \gamma & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} \omega_0/c \\ k_0\cos\theta_0 \\ k_0\sin\theta_0 \\ 0 \end{pmatrix} = \begin{pmatrix} \gamma(\omega_0/c + \beta k_0\cos\theta_0) \\ \gamma(\beta\omega_0/c + k_0\cos\theta_0) \\ k_0\sin\theta_0 \\ 0 \end{pmatrix}. \tag{3.68}$$

Therefore (extracting the first line, and the ratio of the next two):

$$\omega = \gamma\omega_0\left(1 + \frac{k_0}{\omega_0}v\cos\theta_0\right), \tag{3.69}$$

$$\tan\theta = \frac{\sin\theta_0}{\gamma(\cos\theta_0 + v(\omega_0/k_0)/c^2)}. \tag{3.70}$$

Eq. (3.69) is the Doppler effect. We did not make any assumption about the source, so this result describes waves of all kinds, not just light.

For light waves one has $\omega_0/k_0 = c$ so $\omega = \gamma\omega_0(1 + (v/c)\cos\theta_0)$. For $\theta_0 = 0$ we have the 'longitudinal Doppler effect' for light:

$$\frac{\omega}{\omega_0} = \gamma(1 + v/c) = \left(\frac{1 + v/c}{1 - v/c}\right)^{1/2}.$$

Another standard case is the 'transverse Doppler effect', observed when $\theta = \pi/2$, i.e. when the received light travels perpendicularly to the velocity of the source *in the reference frame of the receiver* (N.B. this is not the same as $\theta_0 = \pi/2$). From (3.70) this occurs when $\cos\theta_0 = -v/c$ so

$$\frac{\omega}{\omega_0} = \gamma(1 - v^2/c^2) = \frac{1}{\gamma}.$$

This can be interpreted as an example of time dilation: the process of oscillation in the source is slowed down by a factor γ. This is a qualitatively different prediction from the classical case (where there is no transverse effect) and so represents a direct test of Special Relativity. In practice the most accurate tests combine data from a variety of angles, and a comparison of the frequencies observed in the forward and back longitudinal directions allows the classical prediction to be ruled out, even when the source velocity is unknown.

It can be useful to have the complete Doppler effect formula in terms of the angle θ in the lab frame. This is most easily done[4] by considering the invariant $\mathsf{K} \cdot \mathsf{U}$ where U is the 4-velocity of the source. In the source rest frame this evaluates to $-(\omega_0/c)c = -\omega_0$. In the 'laboratory' frame S it evaluates to

$$(\omega/c,\ \mathbf{k}) \cdot (\gamma c, \gamma\mathbf{v}) = \gamma(-\omega + \mathbf{k}\cdot\mathbf{v}) = -\gamma\omega\left(1 - \frac{kv}{\omega}\cos\theta\right).$$

Therefore

$$\frac{\omega}{\omega_0} = \frac{1}{\gamma(1 - (v/v_p)\cos\theta)}. \tag{3.71}$$

where $v_p = \omega/k$ is the phase velocity in the lab frame. The transverse effect is easy to 'read off' from this formula (as is the effect at any θ).

The transverse Doppler effect has to be taken into account in high-precision atomic spectroscopy experiments. In an atomic vapour the thermal motion of the atoms results in 'Doppler broadening', a spread of observed frequencies, limiting the attainable precision. For atoms at room temperature, the speeds are of the order of a few hundred metres per second, giving rise to longitudinal Doppler shifts of the order of hundreds of MHz for visible light. To avoid this, a collimated atomic beam is used, and the transversely emitted light is detected. For a sufficiently well-collimated beam, the remaining contribution to the Doppler broadening is primarily from the transverse effect. In this way the experimental observation of time dilation has become commonplace in atomic spectroscopy laboratories, as well as in particle accelerators.

[4] Alternatively, first obtain a formula for $\cos\theta_0$ using the 2nd and 3rd lines of (3.68), see eq. (3.73).

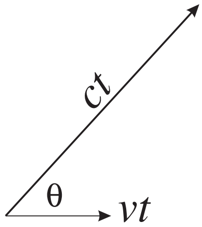


Figure 3.3:

3.7.2 Aberration and the headlight effect

The direction of travel of the waves is also interesting. Notice that eq. (3.70) is not the same as (3.22) when $\omega_0/k_0 \neq c$. This means that a particle emitted along the wave vector in the source frame does not in general travel in the direction of the wave vector in the receiver frame (if it is riding the crest of the wave, it still does so in the new frame but not in the normal direction). For a discussion of this in relation to group velocity and particle physics, see section 5.4.3.

The change in direction of travel of waves (especially light waves) when the same wave is observed in one of two different inertial frames is called *aberration*. The new name should not be taken to imply there is anything new here, however, beyond what we have already discussed. It is just an example of the change in direction of a 4-vector. The name arose historically because changes in the direction of rays in optics were referred to as 'aberration'.

The third line of (3.68) reads $k \sin\theta = k_0 \sin\theta_0$. For light waves the phase velocity is an invariant, so this can be converted into

$$\omega \sin\theta = \omega_0 \sin\theta_0. \tag{3.72}$$

This expresses the relation between Doppler shift and aberration.

Returning to (3.68) and taking the ratio of the first two lines one has, for the case $\omega_0/c = k_0$ (e.g. light waves):

$$\cos\theta = \frac{\cos\theta_0 + v/c}{1 + (v/c)\cos\theta_0}. \tag{3.73}$$

By solving this for $\cos\theta_0$ you can confirm that the formula for $\cos\theta_0$ in terms of $\cos\theta$ can be obtained as usual by swapping 'primed' for unprimed symbols and changing the sign of v (where here the 'primed' symbols are indicated by a subscript zero).

Consider light emitted by a point source fixed in S′. In any given time interval t in S, an

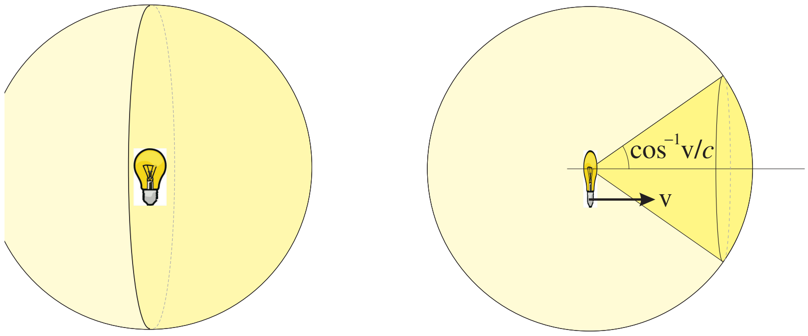


Figure 3.4: The headlight effect for photons. An ordinary incandescent light bulb is a good approximation to an isotropic emitter in its rest frame: half the power is emitted into each hemisphere. In any frame relative to which the light bulb moves at velocity $\mathbf{v}$, the emission is not isotropic but preferentially in the forward direction. The light appearing in the forward hemisphere of the rest frame is emitted in the general frame into a cone in the forward direction of half-angle $\cos^{-1} v/c$ (so $\sin\theta = 1/\gamma$). Its energy is also boosted. The remainder of the emitted light fills the rest of the full solid angle (the complete distribution is given in eqs. (3.80)), (3.81)).

emitted photon[5] moves through ct in the direction θ while the light source moves through vt in the x-direction, see figure 3.3. Consider the case $\theta_0 = \pi/2$, for example a photon emitted down the y' axis. For example, there might be a pipe layed along the y' axis and the photon travels down it. Observed in the other frame, such a pipe will be parallel to the y axis, and the photon still travels down it. In time t the photon travels through distance ct in a direction to be discovered, while the pipe travels through a distance vt in the x direction. Therefore for this case $c\cos\theta = v$, in agreement with (3.73). A source that emitted isotropically in its rest frame would emit half the light into the directions $\theta_0 \le \pi/2$. The receiver would then observe half the light to be directed into a cone with half-angle $\cos^{-1} v/c$, i.e. *less than* $\pi/2$; see figure 3.4. This 'forward beaming' is called the **headlight effect** or searchlight effect.

The full headlight effect involves both the direction and the intensity of the light. To understand the intensity (i.e. energy crossing unit area in unit time) consider figure 3.5 which shows a plane pulse of light propagating between two mirrors (such as in a laser cavity, for example). We consider a pulse which is rectangular in frame S$'$, and long enough so that it is monochromatic to good approximation, and wide enough so that diffraction can be neglected. Let the pulse length be n wavelengths, i.e. $n\lambda_0$ in frame S$'$. Imagine a small antenna which detects the pulse as it passes by. Such an antenna will register n oscillations. This number n must be frame-independent. It follows that the length of the pulse in frame S is $n\lambda$.

In frame S$'$ a given wavefront propagates as $x' = x'_0 + ct'\cos\theta_0$, $y' = y'_0 + ct'\sin\theta_0$. By taking a Lorentz transform one can find the location of the wavefronts in S at any given time t. One thus finds that in frame S the shape of the pulse, at any instant of time, is a parallelogram. The

[5] We use the word 'photon' for convenience here. It does not mean the results depend on a particle theory for light. It suffices that the waves travel in straight lines, i.e. along the direction of the wave vector. The 'photon' here serves as a convenient way to keep track of the motion of a given wavefront in vacuum.

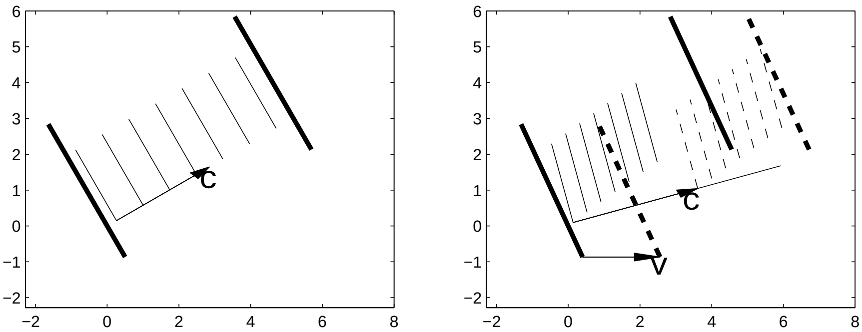


Figure 3.5: The effect of a change of reference frame on a plane wave. The diagrams show a pulse of light propagating between a pair of mirrors, for example the mirrors of a laser cavity. The left diagram shows the situation in S′, the rest frame of the mirrors. The right diagram shows the mirrors and wavefronts at two instants of time in frame S (full lines show the situation at $t = 0$, dashed lines show the situation at a later time t). In this frame the laser cavity suffers a Lorentz contraction and the pulse length is reduced by a larger factor. The wavefronts are no longer perpendicular to the mirror surfaces. The angles are such that the centre of each wavefront still arrives at the centre of the right mirror, and after reflection will meet the oncoming left mirror at its centre also. The width of the wavefronts is the same in the two frames.

direction of travel of each wavefront is given by (3.73), and the wavefront is perpendicular to this direction. One finds also (exercise **??**) that *the area of the wavefronts is Lorentz-invariant*. It follows that the volume of the pulse transforms in the same way as its wavelength. Now, the intensity I of a plane wave is proportional to the energy per unit volume u. We have, therefore:

$$\frac{I}{I_0} = \frac{u}{u_0} = \frac{E/\lambda}{E_0/\lambda_0} \tag{3.74}$$

where E is the energy of the pulse. Such a pulse of light can be regarded as an isolated system having zero rest mass and a well-defined energy-momentum 4-vector describing its total energy and momentum. This statement is non-trivial and will be reexamined in chapters 4 and 12. The zero rest mass, and the fact that the 3-momentum is in the direction of the 3-wave-vector, together mean that the energy-momentum must transform in the same way as the 4-wave-vector, and in particular $E/E_0 = \omega/\omega_0$. It follows that, for a plane wave, the intensity transforms as the square of the frequency:

$$\frac{I}{I_0} = \frac{u}{u_0} = \frac{\omega^2}{\omega_0^2}. \tag{3.75}$$

(This result can be obtained more directly by tensor methods.) This intensity increase even for a plane wave is the second part of the 'headlight effect'. It means that not only is there a steer towards forward directions, but also an increase in intensity of the plane wave components that are emitted in a forward direction.

Figure 3.6 presents the headlight effect along with some examples of equation (3.22), i.e. the transformation of particle velocities rather than wave vectors. If in an explosion in reference frame S′, particles are emitted in all directions with the same speed u', then in frame S the particle velocities are directed in a cone angled forwards along the direction of propagation of S′ in S, for $v > u'$, and mostly in such a cone for $v < u'$. This is not completely unlike the classical prediction (shown in the top two diagrams of figure 3.6, but the 'collimation' into a narrow beam is more pronounced in the relativistic case.

Here are some practical examples. When a fast-moving particle decays in flight, the products are emitted roughly isotropically in the rest frame of the decaying particle, and therefore in any other frame they move in a directed 'jet' along the line of motion of the original particle; these jets are commonly observed in particle accelerator experiments. They are a signature of the presence of a short-lived fast-moving particle that gave rise to the jet. Owing to the expansion of the universe, far off galaxies are moving away from us. The light emission from each galaxy is roughly isotropic in its rest frame, so owing to the headlight effect the light is mostly 'beamed' away from us, making the galaxies appear dimmer. This helps to resolve Olber's paradox (see exercises).

The headlight effect is put to good use in X-ray sources based on 'synchrotron radiation'. When a charged particle accelerates, its electric field must distort, with the result that it emits electromagnetic waves (see chapter 6). In the case of electrons moving in fast circular orbits, the centripetal acceleration results in radiation called synchrotron radiation. In the rest frame of the electron at any instant, the radiation is emitted symmetrically about an axis along the acceleration vector (i.e. about an axis along the radius vector from the centre of the orbit), and has maximum intensity in the plane perpendicular to this axis. However, in the laboratory frame two effects come into play: the Doppler effect and the headlight effect. The Doppler effect results in frequency shifts up to high frequency for light emitted in the forward direction, and the headlight effect ensures that most of the light appears in this direction. The result is a narrow beam, almost like a laser beam, of hard X-rays or gamma rays. This beam is continually swept around a circle, so a stationary detector will receive pulses of X-rays or gamma rays. (See section 6.6.1 for more information).

When one wants a bright source of X-rays, the synchrotron radiation is welcome. When one wants to accelerate particles to high velocities, on the other hand, the radiation is a problem. It represents a continuous energy loss that must be compensated by the accelerator. This limits the velocity that can be achieved in circular particle accelerators, and is a major reason why these accelerators have had to be made larger and larger: by increasing the radius of curvature, the acceleration and thus sychrotron radiation is reduced for any given particle energy. A quantitative calculation is presented in section 6.6.1.

So far we examined the headlight effect by finding the direction of any given particle or ray. Another important quantity is a measure of how much light is emitted into any given small range of directions. This is done by imagining a sphere around the light source, and asking how

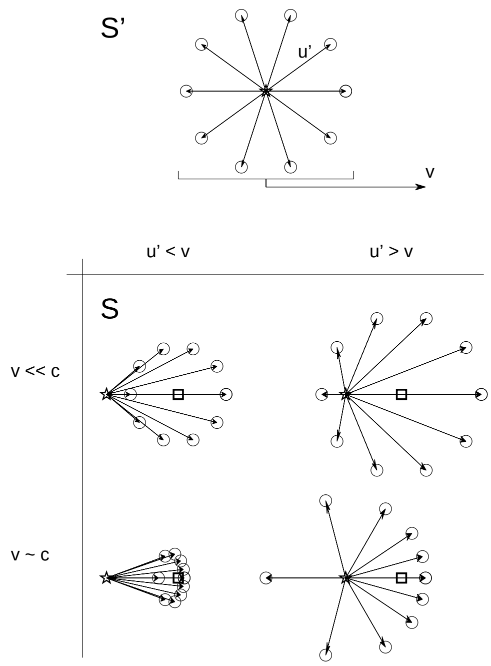


Figure 3.6: Transformation of velocities and the headlight effect. An isotropic explosion in frame S$'$ produces particles all moving at speed u' in S$'$, and a fragment is left at the centre of the explosion (top diagram). The fragment and frame S$'$ move to the right at speed v relative to frame S. The lower four diagrams show the situation in frame S. The $*$ shows the location of the explosion event. The square shows shows the present position of the central fragment; the circles show positions of the particles; the arrows show the velocities of the particles. The left diagrams show examples with $u' < v$, the right with $u' > v$. The top two diagrams show the case $u', v \ll c$. Here the particles lie on a circle centred at the fragment, as in classical physics. The bottom diagrams show examples with $v \sim c$, thus bringing out the difference between the relativistic and the classical predictions. The lower right shows $u' = c$: headlight effect for photons. The photons lie on a circle centred at the position of the explosion (not the fragment) but more of them move forward than backward.

much light falls onto a given region of the sphere.

Suppose N photons are emitted isotropically in frame S′. Then the number emitted into a ring at angle θ_0 with angular width $d\theta_0$ is equal to N multiplied by the surface area of the ring divided by the surface area of a sphere:

$$dN = N\frac{(2\pi r \sin\theta_0)(r d\theta_0)}{4\pi r^2}. \tag{3.76}$$

Here r is the radius of the sphere, so $r\sin\theta_0$ is the radius of the ring, and we used the fact that the surface area of such a narrow ring is simply its circumference multiplied by its width $r d\theta_0$. Hence

$$\frac{\mathrm{d}N}{\mathrm{d}\theta_0} = \frac{1}{2}\sin\theta_0. \tag{3.77}$$

We would like to find the corresponding quantity $dN/d\theta$ representing the number of photon velocities per unit range of angle in the other reference frame. This is obtained from $dN/d\theta = (dN/d\theta_0)(d\theta_0/d\theta)$. We invert (3.73) to obtain an expression for $\cos\theta_0$ in terms of $\cos\theta$, and then differentiate, which gives

$$\sin\theta_0 \frac{\mathrm{d}\theta_0}{\mathrm{d}\theta} = \sin\theta \frac{1 - v^2/c^2}{(1-(v/c)\cos\theta)^2} \tag{3.78}$$

and therefore

$$\frac{\mathrm{d}N}{\mathrm{d}\theta} = \frac{\mathrm{d}N}{\mathrm{d}\theta_0}\frac{\mathrm{d}\theta_0}{\mathrm{d}\theta} = \frac{1}{2}\sin\theta\frac{1 - v^2/c^2}{(1-(v/c)\cos\theta)^2}. \tag{3.79}$$

The solid angle subtended by the ring is $d\Omega = 2\pi\sin\theta d\theta$ in S and $d\Omega_0 = 2\pi\sin\theta_0 d\theta_0$ in S′. The conclusion for emission per unit range of solid angle is

$$\frac{\mathrm{d}N}{\mathrm{d}\Omega_0} = \frac{N}{4\pi}, \qquad \frac{\mathrm{d}N}{\mathrm{d}\Omega} = \frac{N}{4\pi}\frac{1 - v^2/c^2}{(1-(v/c)\cos\theta)^2} = \frac{N}{4\pi}\left(\frac{\omega}{\omega_0}\right)^2, \tag{3.80}$$

where the last step used the Doppler effect formula (3.71). Note that N, the total number of emitted particles, must be the same in both reference frames. The equation for $dN/d\Omega$ gives the enhancement (or reduction) factor for emission in forward (or backward) directions. For example, the enhancement factor for emission into a small solid angle in the directly forward direction (at $\theta = \theta_0 = 0$) is $(1-\beta^2)/(1-\beta)^2 = (1+\beta)/(1-\beta)$.

The simplicity of the final result on the right hand side of (3.80) is remarkable: the angles are so arranged that the number of photons per unit solid angle transforms in the same way as the square of the frequency. I have tried without success to find a simple reason for this. However, the case of emission in the forward or back direction can be argued as follows. Consider a single emission event, and two detectors. Let the detectors both present the same cross-sectional area, but move at different velocities towards (or away from) the source. They are positioned such that each detector finds itself at unit distance from the emission event, as observed in its own reference frame, when the emitted pulse arrives. By constructing an appropriate spacetime diagram, or otherwise, one can easily prove that these distances, when observed in the rest frame of the source, are proportional to λ, the wavelength observed by the detector. In other words, the detector receiving the more red-shifted light is further away, according to the source. Since the emission is isotropic in the source frame, it satisfies an inverse-square law, and therefore each such detector receives a number of photons in proportion to $1/\lambda^2$. This must be interpreted in the detector frame as a number of particles per unit solid angle in proportion to ω^2.

It should be possible to extend this argument to all angles, but then the area and angle of the detector aperture also has to be carefully considered.

Eq. (3.80) concerns the number of particle velocities or ray directions per unit solid angle, not the flux of energy per unit solid angle. To obtain the latter we need to combine eqs (3.80) and (3.75). The emission can always be expressed by Fourier analysis as a sum of plane waves; eq. (3.80) shows that for a point source the density (per unit solid angle) of plane wave components transforms as ω^2, and eq. (3.75) states that the intensity of each plane wave transforms as ω^2. It follows that, for a monochromatic source that emits isotropically in its rest frame, the flux of energy per unit solid angle transforms as

$$\frac{\mathrm{d}\mathcal{P}}{\mathrm{d}\Omega} = \left(\frac{\omega^4}{\omega_0^4}\right) \frac{\mathrm{d}\mathcal{P}_0}{\mathrm{d}\Omega_0}. \tag{3.81}$$

This fourth power relationship is a strong dependence. For example, for v close to c, eq. (3.69) gives $\omega \simeq 2\gamma\omega_0$ for emission in the forward direction. At $\gamma \simeq 100$, for example, the brightness in the forward direction is enhanced approximately a billion-fold.

3.7.3 Stellar aberration

'Stellar aberration' is the name for the change in direction of light arriving at Earth from a star, owing to the relative motion of the Earth and the star. Part of this relative motion is constant (over large timescales) so gives a fixed angle change: we can't tell it is there unless we have further information about the position or motion of the star. However, part of the angle change varies, owing to the changing direction of motion of the Earth in the course of a year, and this small part can be detected by sufficiently careful observations. Before carrying out a detailed calculation, let us note the expected order of magnitude of the effect: at $\theta' = \pi/2$ we have $\cos\theta = v/c$, therefore $\sin(\pi/2 - \theta) = v/c$. For $v \ll c$ this shows the angle $\pi/2 - \theta$ is

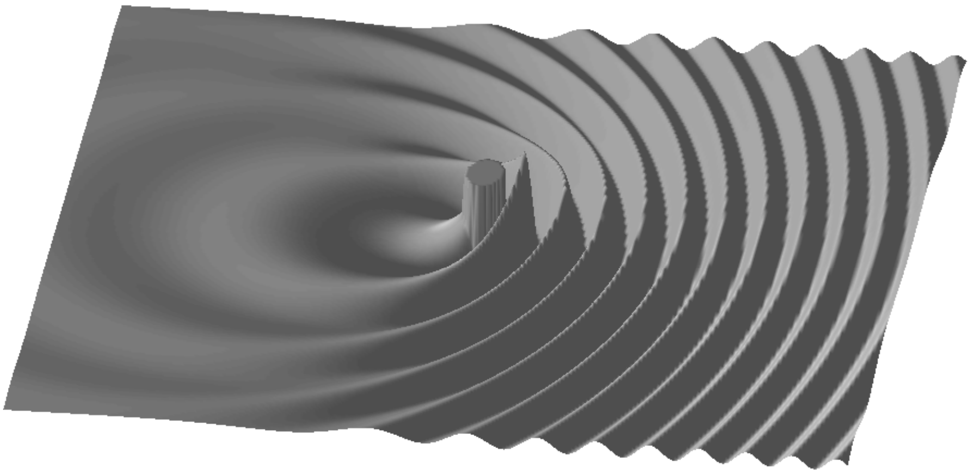


Figure 3.7: The Doppler effect and the headlight effect combine in this image of waves emitted by a moving oscillating source. The image shows an example where the emission is isotropic in the rest frame of the source, and the phase velocity is c. Each wavefront is circular, but more bunched up and brighter in the forward direction.

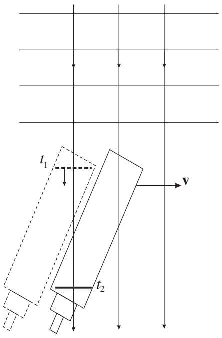


Figure 3.8: Stellar aberration pictured in the rest frame of the star. The light 'rains down' in the vertical direction, while the telescope fixed to planet Earth moves across. The horizontal lines show wavefronts. The thicker dashed wavefront shows the position at time t_1 of a portion of light that entered the telescope (dashed) a short time ago. In order that it can arrive at the bottom of the telescope, where the *same* bit of light is shown by a bold full line, it is clear that the telescope must be angled into the 'shower' of light. (To be precise, the bold line shows where the light would go if it were not focussed by the objective lens of the telescope. The ray passing through the centre of a thin lens is undeviated, so the focussed image appears centred on that ray.) This diagram suffices to show that a tilt of the telescope is needed, and in particular, if the telescope later moves in the opposite direction then its orientation must be changed if it is to be used to observe the same star.

small, so we can use the small angle approximation for the sin function, giving $\theta \simeq \pi/2 - v/c$. Indeed, since the velocities are small, one does not need relativity to calculate the effect. Over the course of six months the angle observed in the rest frame of the Earth is expected to change by about $2v/c \simeq 0.0002$ radians, which is $0.01°$ or about 40 seconds of arc. It is to his credit that in 1727 James Bradley achieved the required stability and precision in observations of the star γ-Draconis. In the course of a year he recorded angle changes in the light arriving down a telescope fixed with an accuracy of a few seconds of arc, and thus he clearly observed the aberration effect. In fact his original intention was to carry out triangulation using the Earth's orbit as baseline, and thus deduce the distance to the star. The triangulation or 'parallax' effect is also present, but it is much smaller than aberration for stars sufficiently far away. Bradley's observed angle changes were not consistent with parallax (the maxima and minima occured at the wrong points in the Earth's orbit), and he correctly inferred they were related to the velocity not the position of the Earth.

In the rest frame of the star, it is easy to picture the aberration effect: as the light 'rains down' on the Earth, the Earth with the telescope on it moves across. Clearly if a ray of light entering the top of the telescope is to reach the bottom of the telescope without hitting the sides, the telescope must not point straight at the star, it must be angled forward slightly into the 'shower' of light, see figure 3.8.

In the rest frame of the Earth, we apply eq. (3.73) supposing S$'$ to be the rest frame of the star. θ is the angle between the received ray and the velocity vector of the star in the rest frame of the Earth. First consider the case where the star does not move relative to the Sun, then v in the formula is the speed of the orbital motion of the Earth. Since this is small compared to c, one may use the binomial expansion $(1-(v/c)\cos\theta)^{-1} \simeq 1+(v/c)\cos\theta$ and then multiply out, retaining only terms linear in v/c, to obtain

$$\cos\theta' \simeq \cos\theta - \frac{v}{c}\sin^2\theta. \tag{3.82}$$

This shows that the largest difference between θ' and θ occurs when $\sin\theta = \pm 1$. This happens when Earth's velocity is at right angles to a line from the Earth to the star. For a star directly above the plane of Earth's orbit, the size of the aberration angle is constant and the star appears to move around a circle of angular diameter $2v/c$; for a star at some other inclination the star appears to move around an ellipse of (angular) major axis $2v/c$.

3.7.4 Visual appearances*

[*Section omitted in lecture-note version.*]

3.8 Summary

The main ideas of this chapter were the Lorentz transformation, 4-vectors and Lorentz invariant quantities, especially proper time. To help keep your thoughts on track, you should consider the spacetime displacement X and the energy-momentum P to be the 'primary' 4-vectors, the ones it is most important to remember. They have the simplest expression in terms of components (see table 3.2): their expressions do not involve γ. For wave motion, the 4-wave-vector is the primary quantity.

The next most simple 4-vectors are 4-velocity U and 4-force F.

Force, work, momentum and acceleration will be the subject of the next chapter.

Exercises

[*Section omitted in lecture-note version.*]

Chapter 4

Dynamics

We are now ready to carry out the sort of calculation one often meets in mechanics problems: the motion of a particle subject to a given force, and the study of collision problems through conservation laws.

Since the concept of force is familiar in classical mechanics, we shall start with that, treating problems where the force is assumed to be known, and we wish to derive the motion. However, since we are also interested in exploring the foundations of the subject, one should note that most physicists would agree that the notion of conservation of momentum is prior to, or underlies, the notion of force. In other words, force is to be understood as a useful way to keep track of the tendency of one body to influence the momentum of another when they interact. We *define* the 3-force $\mathbf{f}$ as equal to $d\mathbf{p}/dt$ where $\mathbf{p} = \gamma_v m_0 \mathbf{v}$ is the 3-momentum of the body it acts on. This proves to be a useful idea because there are many circumstances where the force can also be calculated in other ways. For example, for a spring satisfying Hooke's law we would have $\mathbf{f} = -k\mathbf{x}$ where $\mathbf{x}$ is the extension, and in electromagnetic fields we would have $\mathbf{f} = q(\mathbf{E}+\mathbf{v}\wedge\mathbf{B})$, etc. Therefore it makes sense to study cases where the force is given and the motion is to be deduced. However, the whole argument relies on the definition of momentum, and the reason momentum is defined as $\gamma_v m_0 \mathbf{v}$ is that this quantity satisfies a conservation law, which we shall discuss in section 4.3.

In the first section we introduce some general properties of the 4-force. We then treat various examples using the more familiar language of 3-vectors. This consists of various applications of the relativistic '2nd law of motion' $\mathbf{f} = d\mathbf{p}/dt$. In section 4.3 we then discuss the conservation of energy-momentum, and apply it to collision and scattering problems.

4.1 Force

Let us recall the definition of 4-force (eq. (3.63)):

$$\mathsf{F} \equiv \frac{d\mathsf{P}}{d\tau} = \left(\frac{1}{c}\frac{dE}{d\tau},\ \frac{d\mathbf{p}}{d\tau}\right) = \left(\frac{\gamma}{c}\frac{dE}{dt},\ \gamma\mathbf{f}\right). \tag{4.1}$$

where $\mathbf{f} \equiv d\mathbf{p}/dt$. Suppose a particle of 4-velocity U is subject to a 4-force F. Taking the scalar product, we obtain the Lorentz-invariant quantity

$$\mathsf{U} \cdot \mathsf{F} = \gamma^2 \left(-\frac{dE}{dt} + \mathbf{u} \cdot \mathbf{f}\right). \tag{4.2}$$

One expects that this should be something to do with the 'rate of doing work' by the force. Because the scalar product of two 4-vectors is Lorentz invariant, one can calculate it in any convenient reference frame and get an answer that applies in all reference frames. So let's calculate it in the rest frame of the particle ($\mathbf{u} = 0$), obtaining

$$\mathsf{U} \cdot \mathsf{F} = -c^2 \frac{dm_0}{d\tau}, \tag{4.3}$$

since in the rest frame $\gamma = 1$, $E = m_0 c^2$ and $dt = d\tau$. We now have the result in terms of all Lorentz-invariant quantities, and we obtain an important basic property of 4-force:

When $\mathsf{U} \cdot \mathsf{F} = 0$, *the rest mass is constant.*

A force which does not change the rest mass of the object it acts on is called a *pure force*. The work done by a pure force goes completely into changing the kinetic energy of the particle. In this case we can set (4.2) equal to zero, thus obtaining

$$\frac{dE}{dt} = \mathbf{f} \cdot \mathbf{u} \qquad [\text{ for pure force, } m_0 \text{ constant} \tag{4.4}$$

This is just like the classical relation between force and rate of doing work. An important example of a pure force is the force exerted on a charged particle by electric and magnetic fields. Fundamental forces that are non-pure include the strong and weak force of particle physics.

A 4-force which does not change a body's velocity is called *heat-like*. Such a force influences the rest-mass (for example by feeding energy into the internal degrees of freedom of a composite system such as a spring or a gas).

In this chapter we will study equations of motion only for the case of a pure force. The section on collision dynamics will include general forces (not necessarily pure), studied through their effects on momenta and energies.

4.1.1 Transformation of force

We introduced the 4-force on a particle by the sensible definition $\mathsf{F} = d\mathsf{P}/d\tau$. Note that this statement makes Newton's 2nd law a definition of force, rather than a statement about dynamics. Nonetheless, just as in classical physics, a physical claim is being made: we claim that there will exist cases where the size and direction of the 4-force can be established by other means, and then the equation can be used to find $d\mathsf{P}/d\tau$. We also make the equally natural definition $\mathbf{f} = d\mathbf{p}/dt$ for 3-force. However, we are then faced with the fact that a Lorentz factor γ appears in the relationship between F and $\mathbf{f}$: see eq. (4.1). This means that the transformation of 3-force, under a change of reference frame, depends not only on the 3-force $\mathbf{f}$ *but also on the velocity of the particle on which it acts.* The latter may also be called the velocity of the 'point of action of the force'.

Let $\mathbf{f}$ be a 3-force in reference frame S, and let $\mathbf{u}$ be the 3-velocity in S of the particle on which the force acts. Then, by applying the Lorentz transformation to $\mathsf{F} = (\gamma_u W/c, \gamma_u \mathbf{f})$, where $W = dE/dt$, one obtains

$$\begin{aligned} \frac{\gamma_{u'}}{c}\frac{dE'}{dt'} &= \gamma_v\gamma_u\left((dE/dt)/c - \beta \mathbf{f}_{\|}\right), \\ \gamma_{u'}\mathbf{f}_{\|} &= \gamma_v\gamma_u\left(-\beta(dE/dt)/c + \mathbf{f}_{\|}\right), \\ \gamma_{u'}\mathbf{f}_{\perp} &= \gamma_u \mathbf{f}_{\perp}, \end{aligned} \tag{4.5}$$

where $\mathbf{u}'$ is related to $\mathbf{u}$ by the velocity transformation formulae (3.20). With the help of eq. (3.13) relating the γ factors, one obtains

$$\mathbf{f}'_{\|} = \frac{\mathbf{f}_{\|} - (\mathbf{v}/c^2)dE/dt}{1 - \mathbf{u}\cdot\mathbf{v}/c^2}, \qquad \mathbf{f}'_{\perp} = \frac{\mathbf{f}_{\perp}}{\gamma_v(1 - \mathbf{u}\cdot\mathbf{v}/c^2)}. \tag{4.6}$$

These are the transformation equations for the components of $\mathbf{f}'$ parallel and perpendicular to the relative velocity of the reference frames, when in frame S the force $\mathbf{f}$ acts on a particle moving with velocity $\mathbf{u}$. (Note the similarity with the velocity transformation equations, owing to the similar relationship with the relevant 4-vector).

For the case of a pure force, it is useful to substitute (4.4) into (4.6)i, giving

$$\mathbf{f}'_{\|} = \frac{\mathbf{f}_{\|} - \mathbf{v}(\mathbf{f}\cdot\mathbf{u})/c^2}{1 - \mathbf{u}\cdot\mathbf{v}/c^2} \qquad [\text{ if } m_0\text{=const.} \tag{4.7}$$

Unlike in classical mechanics, $\mathbf{f}$ is not invariant between inertial reference frames. However, a special case arises when m_0 is constant and the force is parallel to the velocity $\mathbf{u}$. Then the force is the same in all reference frames whose motion is also parallel to $\mathbf{u}$. This is easily proved by using (4.7) with $\mathbf{f} \cdot \mathbf{u} = fu$, $\mathbf{u} \cdot \mathbf{v} = uv$ and $\mathbf{f}_\perp = 0$. Alternatively, simply choose S to be the rest frame ($\mathbf{u} = 0$) so one has $dE/dt = 0$, and then transform to any frame S$'$ with $\mathbf{v}$ parallel to $\mathbf{f}$. The result is $\mathbf{f}' = \mathbf{f}$ for all such S$'$.

The transformation equations also tell us some interesting things about forces in general. Consider for example the case $\mathbf{u} = 0$, i.e. $\mathbf{f}$ is the force in the rest frame of the object on which it acts. Then (4.6) says $\mathbf{f}'_\perp = \mathbf{f}_\perp/\gamma$, i.e. the transverse force in another frame is smaller than the transverse force in the rest frame. Since transverse area contracts by this same factor γ, we see that the force per unit area is independent of reference frame.

Suppose that an object is put in tension by forces that are just sufficient to break it in the rest frame. In frames moving perpendicular to the line of action of such forces, the tension force is reduced by a factor γ, and yet the object still breaks. Therefore the breaking strength of material objects is smaller when they move! We will see how this comes about for the case of electrostatic forces in chapter 6.

The **Trouton-Noble experiment** nicely illustrates the relativistic transformation of force—see figure 4.1.

Next, observe that if $\mathbf{f}$ is independent of $\mathbf{u}$, then $\mathbf{f}'$ does depend on $\mathbf{u}$. Therefore independence of velocity is not a Lorentz-invariant property. A force which does not depend on the particle velocity in one reference frame transforms into one that does in another reference frame. This is the case, for example, for electromagnetic forces. It is a problem for Newton's law of gravitation, however, which we deduce is not correct. To get the velocity-dependence of $\mathbf{f}'$ in terms of the velocity in the primed frame, i.e. $\mathbf{u}'$, use the velocity transformation equation (3.20) to write

$$\frac{1}{1 - \mathbf{u} \cdot \mathbf{v}/c^2} = \gamma_v^2 (1 + \mathbf{u}' \cdot \mathbf{v}/c^2). \tag{4.8}$$

4.2 Motion under a pure force

For a pure force we have $dm_0/dt = 0$ and so eq. (3.64) is

$$\mathbf{f} = \frac{d}{dt}(\gamma m_0 \mathbf{u}) = \gamma m_0 \mathbf{a} + m_0 \frac{d\gamma}{dt}\mathbf{u}, \tag{4.9}$$

$$\frac{dK}{dt} = \mathbf{f} \cdot \mathbf{u}. \tag{4.10}$$

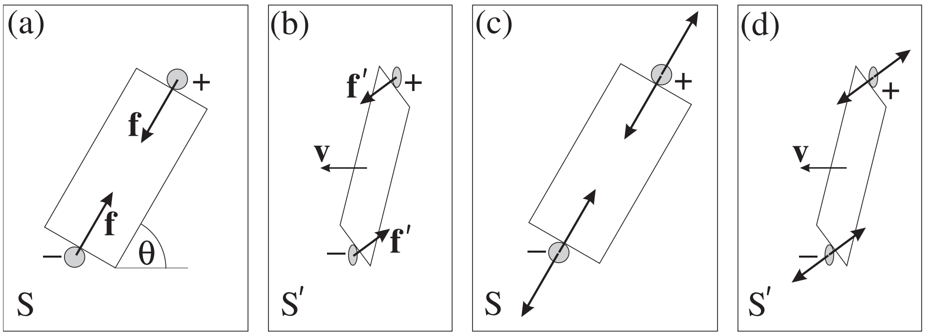


Figure 4.1: **The Trouton-Noble experiment** Suppose two opposite charges are attached to the ends of a non-conducting rod, so that they attract one another. Suppose that in frame S the rod is at rest, and oriented at angle θ to the horizontal axis. The forces exerted by each particle on the other are equal and opposite, directed along the line between them and of size f (fig. (a)). Now consider the situation in a reference frame S′ moving horizontally with speed v. The rod is Lorentz-contracted horizontally (the figure shows an example with $\gamma = 2.294$). The force transformation equations (4.6) state that in S′ the force is the same in the horizontal direction, but reduced in the vertical direction by a factor γ, as shown. Therefore the forces $\mathbf{f}'$ are not along the line between the particles in S′ (fig. (b)). Is there a net torque on the rod? This torque, if it existed, would allow the detection of an absolute velocity, in contradiction of the Principle of Relativity. The answer (supplied by Lorentz (1904)) is given by figures (c) and (d), which indicate the complete set of forces acting on each particle, including the reaction from the surface of the rod. These are balanced, in any frame, so there is no torque. (There are also balanced stresses in the material of the rod (not shown), placing it in compression.) In 1901 (i.e. before Special Relativity was properly understood) Fiztgerald noticed that the energy of the electromagnetic field in a capacitor carrying given charge would depend on its velocity and orientation (c.f. figure 6.1), implying that there would be a torque tending to orient the plates normal to the velocity through the 'aether'. The torque was sought experimentally by Trouton and Noble in 1903, with a null result. The underlying physics is essentially the same as for the rod with charged ends, but the argument in terms of field energy is more involved, because a static electromagnetic field cannot be treated as an isolated system, see exercise ???.

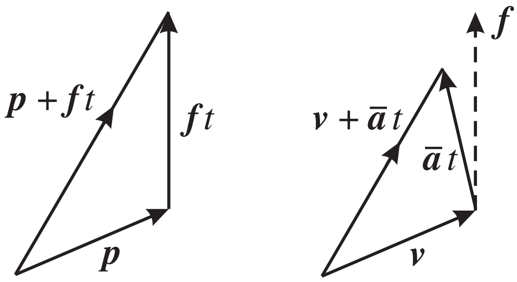


Figure 4.2: Force and acceleration are usually not parallel. The left diagram shows the change in momentum from $\mathbf{p}$ to $\mathbf{p}_f = \mathbf{p} + \mathbf{f}t$ when a constant force $\mathbf{f}$ acts for time t. The right diagram shows what happens to the velocity. The initial velocity is parallel to the initial momentum $\mathbf{p}$, and the final velocity is parallel to the final momentum $\mathbf{p}_f$, but the proportionality constant γ has changed, because the size of v changed. As a result the change in the velocity vector is not parallel to the line of action of the force. Thus the acceleration is not parallel to $\mathbf{f}$. (The figure shows $\bar{\mathbf{a}}t$ where $\bar{\mathbf{a}}$ is the mean acceleration during the time t; the acceleration is not constant in this example.)

We continue to use $\mathbf{u}$ for the velocity of the particle, so $\gamma = \gamma(u)$, and we rewrote eq. (4.4) in order to display all the main facts in one place, with $K \equiv E - m_0c^2$ the kinetic energy. The most important thing to notice is that the relationship between force and kinetic energy is the familiar one, but acceleration is not parallel to the force, except in special cases such as constant speed (leading to $d\gamma/dt = 0$) or $\mathbf{f}$ parallel to $\mathbf{u}$. Let us see why.

Force is defined as a quantity relating primarily to *momentum* not *velocity*. When a force pushes on a particle moving in some general direction, the particle is 'duty-bound' to increase its momentum components, each in proportion to the relevant force component. For example, the component of momentum perpendicular to the force, $\mathbf{p}_\perp$, should *not* change. Suppose the acceleration, and hence the velocity change, were parallel with the force. This would mean the component of velocity perpendicular to the force remains constant. However, in general the speed of the particle does change, leading to a change in γ, so this would result in a change in $\mathbf{p}_\perp$, which is not allowed. We deduce that when the particle speeds up it must redirect its velocity so as to reduce the component perpendicular to $\mathbf{f}$, and when the particle slows down it must redirect its velocity so as to increase the component perpendicular to $\mathbf{f}$. Figure 4.2 gives an example.

There are two interesting ways to write the $d\gamma/dt$ part. First, we have $E = \gamma m_0 c^2$ so when m_0 is constant we should recognise $d\gamma/dt$ as dE/dt up to constants:

$$\frac{d\gamma}{dt} = \frac{1}{m_0c^2}\frac{dE}{dt} = \frac{\mathbf{f}\cdot\mathbf{u}}{m_0c^2}, \tag{4.11}$$

using (4.4), so

$$\mathbf{f} = \gamma m_0 \mathbf{a} + \frac{\mathbf{f} \cdot \mathbf{u}}{c^2} \mathbf{u}. \tag{4.12}$$

This is a convenient form with which to examine the components of $\mathbf{f}$ parallel and perpendicular to the velocity $\mathbf{u}$. For the perpendicular component the second term vanishes: $f_\perp = \gamma m_0 a_\perp$. For the parallel component one has $\mathbf{f} \cdot \mathbf{u} = f_\| u$ and thus

$$\begin{aligned} f_\| &= \gamma m_0 a_\| + f_\| u^2/c^2 \\ \Rightarrow \quad f_\| &= \gamma^3 m_0 a_\|, \qquad f_\perp = \gamma m_0 a_\perp, \end{aligned} \tag{4.13}$$

where we restated the $f_\perp$ result in order to display them both together. Since any force can be resolved into longitudinal and transverse components, (4.13) provides one way to find the acceleration. Sometimes people like to use the terminology 'longitudinal mass' $\gamma^3 m_0$ and 'transverse mass' γm_0. This can be useful but we won't adopt it. The main point is that there is a greater inertial resistance to velocity changes (whether an increase or a decrease) along the direction of motion, compared to the inertial resistance to picking up a velocity component transverse to the current motion (and both excede the inertia of the rest mass).

One can also use (3.49) in (4.9), giving

$$\mathbf{f} = \gamma m_0 \left(\mathbf{a} + \gamma^2 \frac{\mathbf{u} \cdot \mathbf{a}}{c^2} \mathbf{u} \right) = \gamma^3 m_0 \left((1 - u^2/c^2) \mathbf{a} + \frac{\mathbf{u} \cdot \mathbf{a}}{c^2} \mathbf{u} \right). \tag{4.14}$$

This allows one to obtain the longitudinal and transverse acceleration without an appeal to work and energy.

4.2.1 Constant force (the 'relativistic rocket')

The phrase 'constant force' might have several meanings in a relativistic calculation. It could mean constant with respect to time in a given inertial frame or to proper time along a worldline, and it might refer to the 3-force or the 4-force. In this section we will study the case of motion of a particle subject to a 3-force whose size and direction is independent of time and position in a given reference frame.

The reader might wonder why we are not treating a constant 4-force. The reason is that this would be a somewhat unrealistic scenario. If the 4-force is independent of proper time then all parts of the energy-momentum 4-vector increase together, and this means the combination $E^2 - p^2 c^2$ must be changing, and we do not have a pure force. It is not impossible, but it represents a non-simple (and rather artificial) situation. If the 4-force on a particle is independent of

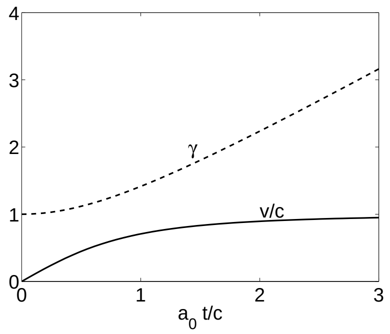


Figure 4.3: Speed (full curve) and Lorentz factor (dashed curve) as a function of time for straight-line motion under a constant force. The product of these two curves is a straight line.

reference frame time then its spatial part must be proportional to $1/\gamma_v$ where v is the speed of the particle in the reference frame. Again, it is not impossible but it is rather unusual or artificial.

The case of a 3-force $\mathbf{f}$ that is independent of time in a given reference frame, on the other hand, is quite common. It is obtained, for example, for a charged particle moving in a static uniform electric field.

Motion under a constant force, for the case of a particle starting from rest, is very easy to treat (the calculation is also presented in an early chapter of *The Wonderful World*):

$$\frac{d\mathbf{p}}{dt} = \mathbf{f} \qquad \Rightarrow \qquad \mathbf{p} = \mathbf{p}_0 + \mathbf{f}t$$

since $\mathbf{f}$ is constant. If $\mathbf{p}_0 = 0$ then the motion is in a straight line with $\mathbf{p}$ always parallel to $\mathbf{f}$, and by solving the equation $p = \gamma m_0 v = ft$ for v one finds

$$v = \frac{ft}{\sqrt{m_0^2 + f^2t^2/c^2}}. \tag{4.15}$$

(We are reverting to $\mathbf{v}$ rather than $\mathbf{u}$ for the particle velocity.) This result is plotted in figure 4.3. The case where $\mathbf{p}_0$ is not zero is treated in the exercises.

Example. An electron is accelerated from rest by a static uniform electric field of strength 1000 V/m. How long does it take (in the initial rest frame) for the electron's speed to reach 0.99c?

Answer. The equation $\mathbf{f} = q\mathbf{E}$ for the force due to an electric field is valid at all speeds. Therefore we have $f = 1.6 \times 10^{-13}$ N. The time is given by $t = \gamma mv/f \simeq 12$ μs.

In the previous section 4.1.1 (the transformation equations for force) we saw that in this case ($\mathbf{f}$ parallel to $\mathbf{v}$) the force is the same in all reference frames moving in the same direction as the particle. That is, if we were to evaluate the force in other reference frames moving parallel to the particle velocity, then we would find the same force. In particular, we might take an interest in the reference frame in which the particle is momentarily at rest at some given time. This is called the 'instantaneous rest frame' of the particle. N.B. this reference frame does not itself accelerate: it is an inertial frame. We would find that the force on the particle in this new reference frame is the same as in the first one, and therefore at the moment when the particle is at rest in the new reference frame, it has the very same acceleration that it had in the original rest frame when it started out! Such a particle always finds itself to have the same constant acceleration in its own rest frame, even though according to eq. (4.15) and figure 4.3 its acceleration falls to zero in the original reference frame as the particle speed approaches c. It is like the Alice and the Red Queen in Lewis Carroll's *Through the looking glass*, forever running to stand still. The particle accelerates and accelerates, and yet only approaches a constant velocity.

For a further comment on constant proper acceleration, see the end of section 3.5.2.

Let a_0 be the acceleration of the particle in in its instantaneous rest frame. $a_0 = f/m_0$ since $\gamma = 1$ at the moment when the particle is at rest. Therefore we can rewrite (4.15) as

$$v(t) = \frac{a_0 t}{(1 + a_0^2 t^2/c^2)^{1/2}}. \tag{4.16}$$

This can be integrated directly to give

$$\begin{aligned} x &= \frac{c^2}{a_0}\left(1 + \frac{a_0^2 t^2}{c^2}\right)^{1/2} + b \\ \Rightarrow \qquad (x-b)^2 - c^2 t^2 &= (c^2/a_0)^2. \end{aligned} \tag{4.17}$$

Here b is a constant of integration given by the initial conditions. At $t = 0$ the velocity dx/dt is zero and the position is $x(0) = c^2/a_0 + b$.

Equation (4.17) is the equation of a hyperbola, see figure 4.4. This type of motion is sometimes called 'hyperbolic motion.' It should be contrasted with the 'parabolic motion' (in spacetime) that is obtained for classical motion under a constant force. It is also useful to notice that $(x - b)^2 - c^2t^2$ is the invariant spacetime interval between the event $(t = 0, x = b)$ and the location (t, x) of the particle at any instant. Thus the motion maintains a constant interval from a certain event situated off the worldline. This event is singled out by the initial conditions and the size of the force.

Any hyperbolic curve can be usefully expressed in terms of hyperbolic functions. To this end,

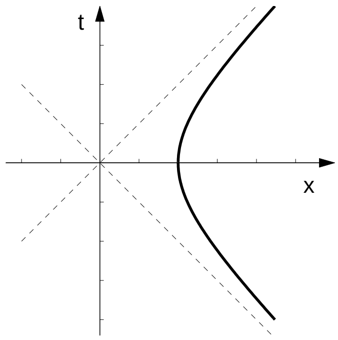


Figure 4.4: Spacetime diagram showing the worldline of a particle undergoing constant *proper acceleration*. That is, if at any event A on the worldline one picks the inertial reference frame S_A whose velocity matches that of the particle at A, then the acceleration of the particle at A, as observed in frame S_A, has a value a_0, independent of A. We say more succinctly that the acceleration is constant in the 'instantaneous rest frame' but strictly this phrase refers to a succession of inertial reference frames, not a single accelerating frame. The worldline is a hyperbola on the diagram, see eq. (4.17). The asymptotes are at the speed of light. The motion maintains a fixed spacetime interval from the event where the asymptotes cross (c.f. chapter 7). This type of motion can be produced by a constant force acting parallel to the velocity.

write (4.17) as

$$\left[\frac{a_0}{c^2}(x-b)\right]^2 - \left(\frac{a_0 t}{c}\right)^2 = 1 \tag{4.18}$$

and introduce a parameter θ defined by

$$\frac{a_0 t}{c} = \sinh\theta$$

so that the worldline can be expressed $\cosh^2\theta - \sinh^2\theta = 1$. One immediately obtains

$$a_0(x-b)/c^2 = \cosh\theta,$$

and eq. (4.16) is

$$v = c\tanh\theta. \tag{4.19}$$

It follows that $\gamma = \cosh\theta$, and therefore the Lorentz factor increases linearly with the distance covered.

By comparing (4.19) with (3.30) you can see that our parameter θ is the rapidity of the particle.

Now let's explore the proper time along the worldline. Using $dt/\tau = \gamma = \cosh\theta$ and $dt/d\theta = (c/a_0)\cosh\theta$ (from the definition of θ) we obtain

$$\frac{d\tau}{d\theta} = \frac{c}{a_0} \qquad \Rightarrow \qquad \tau = \frac{c\theta}{a_0}. \tag{4.20}$$

In other words, θ can also be understood as the proper time, in units of c/a_0, measured from the event where $v = 0$. This result can be used to make an exact calculation of the aging of the travelling twin in the twin paradox (see exercises).

The uniform increase of rapidity with proper time offers another way to think about constant acceleration. Let S_A be the instantaneous rest frame at some event A. At A the particle has zero velocity in the frame under consideration, and in the next small time interval $\mathrm{d}\tau$ it acquires a velocity $\mathrm{d}v = a_0\mathrm{d}\tau$, where we use τ since this is proper time. The rapidity ρ_0 increases from zero to $\tanh^{-1}(a_0\mathrm{d}\tau/c) \simeq a_0\mathrm{d}\tau/c$. Hence

$$\frac{\mathrm{d}\rho_0}{\mathrm{d}\tau} = \frac{a_0}{c} \tag{4.21}$$

where the equation applies in frame S_A for events in the vicinity of A. Now recall from the discussion in section 3.4.1 that, for velocity changes all in the same direction, rapidities add. It follows that the rapidity of the particle, as observed in any other frame S, is $\rho = \rho_A + \rho_0$, where ρ_A is the rapidity of frame S_A as observed in S. I insist again that S_A is an inertial frame, not an accelerating one, so ρ_A is constant. Hence

$$\frac{\mathrm{d}\rho}{\mathrm{d}\tau} = \frac{\mathrm{d}\rho_A}{\mathrm{d}\tau} + \frac{\mathrm{d}\rho_0}{\mathrm{d}\tau} = 0 + \frac{a_0}{c}. \tag{4.22}$$

This equation applies for events in the vicinity of A, but now we can argue that it doesn't matter what event A was chosen, we shall get the same result. Therefore the rapidity in frame S increases linearly with proper time. This is an alternative route to the derivation of (4.20), and gives a nice way to think about the whole process.

An important application of all the above results is to the design of linear particle accelerators, where a constant force is a reasonable first approximation to what can be achieved. Another application is to the study of a rocket whose engine is programmed in such a way as to maintain a constant proper acceleration. This means the rate of expulsion of rocket fuel should reduce in proportion to the remaining rest mass, so that the acceleration measured in the instantaneous rest frame stays constant. In an interstellar journey, a (not too large) constant proper acceleration might be desirable in order to offer the occupants of the rocket a constant 'artificial gravity'. For this reason, motion at constant proper acceleration is sometimes referred to as the case of a 'relativistic rocket'.

$$t = (c/a_0)\sinh\theta, \qquad (x - x_0) = (c^2/a_0)(\cosh\theta - 1) \tag{4.23}$$
$$v = c\tanh\theta, \qquad \gamma = \cosh\theta = 1 + a_0(x - x_0)/c^2 \tag{4.24}$$
$$\tau = c\theta/a_0, \qquad \gamma^3 a = a_0 \tag{4.25}$$

Table 4.1: A summary of results for straight-line motion at constant proper acceleration a_0 (sometimes called the 'relativistic rocket'). If the origin is chosen so that $x_0 = c^2/a_0$ then some further simplications are obtained, such as $x = \gamma x_0$, $v = c^2 t/x$.

The above formulae are gathered together in table 4.1. The situation of 'constant acceleration' (meaning constant proper acceleration) has many further fascinating properties and is discussed at length in chapter 7 as a prelude to General Relativity.

When the initial velocity is not along the line of the constant force, the proper acceleration is not constant (see exercises).

4.2.2 4-vector treatment of hyperbolic motion

If we make the most natural choice of origin, so that $b = 0$ in eq. (4.17), then the equations for x and t in terms of θ combine to make the 4-vector displacement

$$\mathsf{X} = (ct,\ x) = x_0(\sinh\theta,\ \cosh\theta) \tag{4.26}$$

where $x_0 = c^2/a_0$ and we suppressed the y and z components which remain zero throughout. We then obtain

$$\begin{aligned} \mathsf{U} &= \frac{d\mathsf{X}}{d\tau} = \frac{d\mathsf{X}}{d\theta}\frac{d\theta}{d\tau} = c(\cosh\theta,\ \sinh\theta) && (4.27)\\ \text{and}\quad \dot{\mathsf{U}} &= \frac{d\mathsf{U}}{d\tau} = a_0(\sinh\theta,\ \cosh\theta) = \frac{a_0^2}{c^2}\mathsf{X}. && (4.28)\\ \Rightarrow\quad \ddot{\mathsf{U}} &\propto \mathsf{U} && (4.29)\end{aligned}$$

where the dot signifies $d/d\tau$. We shall now show that this relationship between 4-velocity and rate of change of 4-acceleration can be regarded as the defining characteristic of hyperbolic motion.

Suppose we have motion that satisfies (4.29), i.e.

$$\dot{\mathsf{A}} = \alpha^2 \mathsf{U} \tag{4.30}$$

where α is a constant. Consider $\mathsf{A} \cdot \mathsf{A}$, and recall $\mathsf{A} \cdot \mathsf{A} = a_0^2$ (from eq. (3.51)). Differentiating with respect to τ gives

$$\frac{d}{d\tau}(a_0^2) = 2\dot{\mathsf{A}} \cdot \mathsf{A} = 2(\alpha^2 \mathsf{U}) \cdot \mathsf{A} = 0$$

where we used (4.30) and then the general fact that 4-velocity is perpendicular to 4-acceleration (eq. (3.52)). It follows that a_0 is constant. Hence (4.30) implies motion at constant proper acceleration.

The constant α is related to the proper acceleration. To find out how, consider $\mathsf{U} \cdot \mathsf{A} = 0$. Differentiating with respect to τ gives

$$\dot{\mathsf{A}} \cdot \mathsf{U} + \mathsf{A} \cdot \dot{\mathsf{U}} = 0 \qquad \Rightarrow \qquad \dot{\mathsf{A}} \cdot \mathsf{U} = -a_0^2 \tag{4.31}$$

(using eq. (3.51)). This is true for any motion, not just hyperbolic motion. Applying it to the case of hyperbolic motion, (4.30), we find $-\alpha^2 c^2 = -a_0^2$ hence $\alpha = a_0/c$.

Eq. (4.30) can be regarded as a 2nd order differential equation for U, and it can be solved straightforwardly using exponential functions. Upon substituting in the boundary condition $\mathsf{U} = (c, 0)$ at $\tau = 0$ one obtains the cosh function for U^0, and the boundary condition on $\dot{\mathsf{U}}$ leads to a sinh function for the spatial part, the same as we already obtained in the previous section.

To do the whole calculation starting from the 4-vector equation of motion

$$\mathsf{F} = m_0 \frac{d\mathsf{U}}{d\tau} \tag{4.32}$$

(valid for a pure force) we need to know what F gives the motion under consideration. Clearly it must be, in component form, $(\gamma \mathbf{f} \cdot \mathbf{v}/c, \gamma \mathbf{f})$ in the reference frame adopted in the previous section, but we would prefer a 4-vector notation which does not rely on any particular choice of frame. The most useful way to write it turns out to be

$$\mathsf{F} = \mathbb{F} g \mathsf{U}/c \tag{4.33}$$

where g is the metric and

$$\mathbb{F} = \begin{pmatrix} 0 & f_0 & 0 & 0 \\ -f_0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \tag{4.34}$$

(for a force along the x direction)[1], with constant f_0. Substituting this into (4.32) we obtain

$$\left.\begin{array}{rcl} f_0 \mathsf{U}^1/c &=& m_0 \dot{\mathsf{U}}^0 \\ f_0 \mathsf{U}^0/c &=& m_0 \dot{\mathsf{U}}^1 \end{array}\right\}$$

where the superscripts label the components of U. This pair of simultaneous first order differential equations may be solved in the usual way, by differentiating the second and substituting into the first, to find

$$\ddot{\mathsf{U}}^1 = \left(\frac{f_0}{m_0 c}\right)^2 \mathsf{U}^1.$$

This is one component of (4.30), whose solution we discussed above.

4.2.3 Circular motion

Another very simple case is obtained when $d\gamma/dt = 0$, i.e. motion at constant speed. From eq. (4.11) this happens when the force remains perpendicular to the velocity. An example is the force on a charged particle moving in a magnetic field: then

$$\mathbf{f} = q\mathbf{v} \wedge \mathbf{B} = \gamma m_0 \mathbf{a}. \tag{4.35}$$

The solution of the equation of motion proceeds exactly as in the classical (low velocity) case, except that a constant factor γ appears wherever the rest mass appears. For an initial velocity perpendicular to $\mathbf{B}$ the resulting motion is circular. The particle moves at speed v around a circle of radius

$$r = \frac{\gamma m_0 v}{qB} = \frac{p}{qB}. \tag{4.36}$$

In particle physics experiments, a standard diagnostic tool is to record the track of a particle in a uniform magnetic field of known strength. This equation shows that, if the charge q is also known, then the particle's momentum can be deduced directly from the radius of the track.

The equation is also crucial for the design of ring-shaped particle accelerators using magnetic confinement. It shows that, to maintain a given ring radius r, the strength of the magnetic field has to increase in proportion to the particle's momentum, not its speed. In modern accelerators the particles move at close to the speed of light anyway, so v is essentially fixed at $\simeq c$, but this

[1]The matrix $\mathbb{F}$ is introduced in (4.34) merely to show how to write the equation of motion we need. In chapter 9 we shall learn that $\mathbb{F}$ is a contravariant 2nd rank tensor, but you don't to worry about that for now.

does not free us from the need to build ever more powerful magnetic field coils if we want to confine particles of higher energy.

The period and angular frequency of the motion are

$$T = \frac{2\pi r}{v} = 2\pi \frac{\gamma m_0}{qB}, \qquad \omega = \frac{qB}{\gamma m_0}. \tag{4.37}$$

The classical result that the period is independent of the radius and speed is lost. This makes the task of synchronising applied electric field pulses with the motion of the particle (in order to accelerate the particle) more technically demanding. It required historically the development of the 'synchrotron' from the 'cyclotron'.

For helical motion, see exercise **??**

Combined electric and magnetic fields will be considered in chapter 12.

4.2.4 Motion under a central force

The case of a *central* force is that in which the force experienced by a particle is always directed towards or away from one point in space (in a given inertial frame). This is an important basic case partly because in the low-speed limit it arises in the 'two-body problem', where a pair of particles interact by a force directed along the line between them. In that case the equations can be simplified by separating them into one equation for the relative motion, and another for the motion of the centre of mass of the system. This simplification is possible because one can adopt the approximation that the field transmits cause and effect instantaneously between the particles, with the result that the force on one particle is always equal and opposite to the force on the other. In the case of high speeds this cannot be assumed. If two particles interact at a distance it must be because they both interact locally with a third party—for example the electromagnetic field—and the dynamics of the field cannot be ignored. We shall look into this more fully in chapter 12. The main conclusion for our present discussion is that the 'two-body' problem is really a 'two-body plus field' problem and has no simple solution.

Nevertheless, the idea of a central force remains important and can be a good model when one particle interacts with a very much heavier particle and energy loss by radiation is small—for example, a planet orbiting the sun. Then the acceleration of the heavy particle can be neglected, and in the rest frame of the heavy particle the other particle experiences, to good approximation, a central force. This can also be used to find out approximately how an electron orbiting an atomic nucleus would move if it did not emit electromagnetic waves.

Consider, then, a particle of rest mass m_0 and position vector $\mathbf{r}$ subject to a force

$$\mathbf{f} = f(r)\hat{\mathbf{r}}. \tag{4.38}$$

Introduce the **3-angular momentum**

$$\mathbf{L} \equiv \mathbf{r} \wedge \mathbf{p}. \tag{4.39}$$

By differentiating with respect to time one finds

$$\dot{\mathbf{L}} = \dot{\mathbf{r}} \wedge \mathbf{p} + \mathbf{r} \wedge \dot{\mathbf{p}} = \mathbf{r} \wedge \mathbf{f}, \tag{4.40}$$

(since $\mathbf{p}$ is parallel to $\dot{\mathbf{r}}$ and $\dot{\mathbf{p}} = \mathbf{f}$) which is true for motion under any force (and is just like the classical result). For the case of a central force one has conservation of angular-momentum:

$$\frac{\mathrm{d}\mathbf{L}}{\mathrm{d}t} = 0 \qquad \Longrightarrow \mathbf{L} = \text{const}. \tag{4.41}$$

It follows from this that the motion remains in a plane (the one containing the vectors $\mathbf{r}$ and $\mathbf{p}$), since if $\dot{\mathbf{r}}$ were ever directed out of that plane then $\mathbf{L}$ would necessarily point in a new direction. Adopting plane polar coordinates (r, ϕ) in this plane we have

$$\mathbf{p} = \gamma m_0 \mathbf{v} = \gamma m_0 (\dot{r},\, r\dot{\phi}) = (p_r,\, \gamma m_0 r\dot{\phi}). \tag{4.42}$$

Therefore

$$L = \gamma m_0 r^2 \dot{\phi}. \tag{4.43}$$

(The angular momentum vector being directed normal to the plane). Using $dt/d\tau = \gamma$ it is useful to convert this to the form

$$\frac{\mathrm{d}\phi}{\mathrm{d}\tau} = \frac{L}{m_0 r^2}. \tag{4.44}$$

Note also that $p^2 = p_r^2 + L^2/r^2$, which is like the classical result.

Let E be the energy of the particle, in the sense of its rest energy plus kinetic energy, then using $E^2 - p^2c^2 = m_0^2c^4$ we obtain

$$p_r^2 = \frac{E^2}{c^2} - \frac{L^2}{r^2} - m_0^2c^2. \tag{4.45}$$

To make further progress it is useful to introduce the concept of *potential energy* V. This is defined by

$$V \equiv -\int \mathbf{f} \cdot \mathrm{d}\mathbf{r}. \tag{4.46}$$

Such a definition is useful when the integral around any closed path is zero so that V is single-valued. When this happens the force is said to be *conservative.* Using (4.10) (valid for a pure force) we then find that during any small displacement $d\mathbf{r}$ the kinetic energy lost by the particle is equal to the change in V:

$$\begin{aligned} \mathrm{d}K &= (\mathbf{f} \cdot \mathbf{u})\mathrm{d}t = \mathbf{f} \cdot \mathrm{d}\mathbf{r} && (4.47) \\ &= -\mathrm{d}V. && (4.48) \end{aligned}$$

It follows that the quantity

$$\mathcal{E} \equiv E + V \tag{4.49}$$

is a constant of the motion. In classical mechanics V is often called ‘the potential energy of the particle’ and then $\mathcal{E}$ is called ‘the total energy of the particle’. However, strictly speaking V is not a property of the particle: it makes no contribution whatsoever to the energy possessed by the particle, which remains $E = \gamma m_0 c^2$. V is just a mathematical device introduced in order to identify a constant of the motion. Physically it could be regarded as the energy owned not by the particle but by the *other* system (such as an electric field) with which the particle is interacting.

We can write (4.49) in two useful ways:

$$\begin{aligned} \gamma m_0 c^2 + V &= \text{const} && (4.50) \\ \text{and} \quad p_r^2 c^2 + \frac{L^2 c^2}{r^2} + m_0^2 c^2 &= (\mathcal{E} - V)^2 && (4.51) \end{aligned}$$

(using (4.45)). Since for a given force, V is a known function of r, the first equation enables the Lorentz factor for the total speed to be obtained at any given r for given initial conditions. Using the angular momentum (also fixed by initial conditions) one can then also find $\dot{\phi}$ and hence $\dot{r}$.

Equation (4.51) is a differential equation for r as a function of time (since $p_r = \gamma m_0 \dot{r}$). It is easiest to seek a solution as a function of proper time τ, since

$$\frac{\mathrm{d}r}{\mathrm{d}\tau} = \frac{\mathrm{d}r}{\mathrm{d}t}\frac{\mathrm{d}t}{\mathrm{d}\tau} = \dot{r}\gamma = \frac{p_r}{m_0}$$

so we have

$$\frac{1}{2}m_0\left(\frac{\mathrm{d}r}{\mathrm{d}\tau}\right)^2 = \tilde{\mathcal{E}} - V_{\rm eff}(r) \tag{4.52}$$

where

$$\tilde{\mathcal{E}} \equiv \frac{\mathcal{E}^2 - m_0^2c^4}{2m_0c^2}, \tag{4.53}$$

$$V_{\rm eff}(r) \equiv \frac{V(r)(2\mathcal{E} - V(r))}{2m_0c^2} + \frac{L^2}{2m_0r^2}. \tag{4.54}$$

Equation (4.52) has precisely the same form as an equation for classical motion in one dimension in a potential $V_{\rm eff}(r)$. Therefore we can immediately deduce the main qualitative features of the motion. Consider for example an inverse-square-law force, such as that arising from Coulomb attraction between opposite charges. Writing $\mathbf{f} = -\alpha\hat{\mathbf{r}}/r^2$ and therefore $V = -\alpha/r$ we have

$$V_{\rm eff} = \frac{1}{2m_0c^2}\left(\frac{L^2c^2 - \alpha^2}{r^2} - \frac{2\alpha\mathcal{E}}{r}\right). \tag{4.55}$$

The second term gives an attractive $1/r$ potential well that dominates at large r. If the first term is non-zero then it dominates at small r and gives either a barrier or an attractive well, depending on the sign. Thus there are two cases to consider:

$$\text{(i) } L > L_{\rm c}, \qquad \text{(ii) } L \le L_{\rm c}; \qquad\qquad L_{\rm c} \equiv \frac{\alpha}{c}. \tag{4.56}$$

(i) For large angular momentum, the 'centrifugal barrier' is sufficient to prevent the particle approaching the origin, just as in the classical case. There are two types of motion: unbound motion (or 'scattering') when $\tilde{\mathcal{E}} > 0$, and bound motion when $\tilde{\mathcal{E}} < 0$, in which case r is constrained to stay in between turning points at $V_{\rm eff}(r) = \tilde{\mathcal{E}}$.

(ii) For small angular momentum, something qualitatively different from the classical behaviour occurs: when $L \le L_{\rm c}$ the motion has no inner turning point and the particle is 'sucked in' to the origin. The motion conserves L and therefore is a spiral in which $\gamma \to \infty$ as $r \to 0$. In this limit the approximation that the particle or system providing the central force does not itself accelerate is liable to break down; the main point is that a Coulomb-law scattering centre can result in a close collision even when the incident particle has finite angular momentum. In classical physics this type of behaviour would require an attractive force with a stronger dependence on distance. For a scattering process in which the incident particle has momentum $p_{\rm i}$ at infinity, and impact parameter b, the angular momentum is $L = bp_{\rm i}$. All particles with

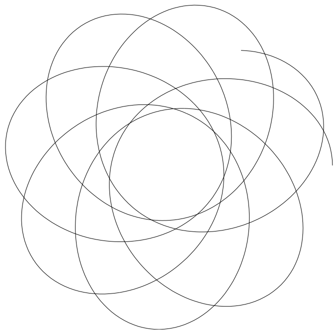


Figure 4.5: Example orbit of a fast-moving particle in a $1/r$ potential. Only part of the orbit is shown; its continuation follows the same pattern.

impact parameter below $b_c = L_c/p_{\rm i}$ will suffer a spiraling close collision. The collision cross-section for this process is

$$\pi b_c^2 = \frac{\pi\alpha^2}{c^2 p_{\rm i}^2} = \frac{\pi\alpha^2}{E^2 - m_0^2 c^4}.$$

This is very small in practice. For example, for an electron moving in the Coulomb potential of a proton $b_c \simeq 1.4 \times 10^{-12}$ m when the incident kinetic energy is 1 eV. Using the Newtonian formula for gravity to approximate conditions in the solar system, one obtains $b_c \simeq GM/vc$ where $M \simeq 2 \times 10^{30}$ kg is the solar mass and $v \ll c$ is the speed of an object such as a comet. b_c exceeds the radius of the sun when the incident velocity (far from the sun) is below 640 m/s.

In the case $L > L_{\rm c}$ and $\tilde{\mathcal{E}} < 0$, where there are stable bound orbits, a further difference from the classical motion arises. The classical $1/r$ potential leads to elliptical orbits, in which the orbit closes on itself after a single turn. This requires that the distance from the origin oscillates in step with the movement around the origin, so that after r completes one cycle between its turning points, ϕ has increased by 2π. There is no reason why this synchrony should be maintained when the equation of motion changes, and in fact it is not. The orbit has the form of a rosette—see figure 4.5. In order to deduce this, we can turn eq. (4.52) into an equation for the orbit, as follows. First differentiate with respect to τ, to obtain

$$m\frac{{\rm d}^2 r}{{\rm d}\tau^2} = -\frac{{\rm d}V_{\rm eff}}{{\rm d}r} \tag{4.57}$$

where we cancelled a factor of ${\rm d}r/{\rm d}\tau$ which is valid except at the stationary points. Apply this to the case of an inverse-square law force, for which the effective potential is given in (4.55):

$$\frac{{\rm d}^2 r}{{\rm d}\tau^2} = \frac{L^2 - \alpha^2/c^2}{m^2 r^3} - \frac{\alpha\mathcal{E}}{m^2 c^2 r^2}. \tag{4.58}$$

Although this equation can be tackled by direct integration, the best way to find the orbit is to make two changes of variable. Using (4.44) derivatives with respect to τ can be expressed in terms of derivatives with respect to ϕ. Then one changes variable from r to $u = 1/r$, obtaining

$$\frac{\mathrm{d}^2 r}{\mathrm{d}\tau^2} = -\left(\frac{L}{m}\right)^2 u^2 \frac{\mathrm{d}^2 u}{\mathrm{d}\phi^2} \tag{4.59}$$

and therefore (4.58) becomes

$$\frac{\mathrm{d}^2 u}{\mathrm{d}\phi^2} = -\left(1 - \frac{\alpha^2}{L^2 c^2}\right) u + \frac{\alpha \mathcal{E}}{L^2 c^2}. \tag{4.60}$$

This is the equation for simple harmonic motion. Hence the oribit is given by

$$r(\phi) = \frac{1}{u} = \frac{1}{A\cos(\tilde{\omega}(\phi - \phi_0)) + \alpha\mathcal{E}/(L^2c^2 - \alpha^2)} \tag{4.61}$$

where A and ϕ_0 are constants of integration, and

$$\tilde{\omega} = \sqrt{1 - \frac{\alpha^2}{L^2 c^2}}. \tag{4.62}$$

The radial motion completes one period when ϕ increases by $2\pi/\tilde{\omega}$. In the classical limit one has $\tilde{\omega} = 1$ which means the orbit closes (forming an ellipse). For the relativistic case far from the critical angular momentum, i.e. $L \gg \alpha/c$, one has $\tilde{\omega} \simeq 1 - \alpha^2/2L^2c^2$. Therefore when r returns to its minimum value (so-called *perihelion* in the case of planets orbiting the sun) ϕ has increased by 2π plus an extra bit equal to

$$\delta\phi = \frac{\pi\alpha^2}{L^2 c^2}. \tag{4.63}$$

The location of the innermost point of the orbit shifts around (or 'precesses') by this amount per orbit. For the case of an electron orbiting a proton, the combination α/Lc is equal to the fine structure constant when $L = \hbar$, and this motion was used by Sommerfeld to construct a semi-classical theory for the observed fine structure of hydrogen (subsequently replaced by the correct quantum treatment). For the case of gravitational attraction to a spherical mass, the result (4.63) is about 6 times smaller than the precession predicted by General Relativity.

4.2.5 (An)harmonic motion*

[*Section omitted in lecture-note version.*]

Invariant, conserved

		Lorentz invariant	conserved
energy	E	$\times$	✓
momentum	$\mathbf{p}$	$\times$	✓
rest mass	m	✓	$\times$
charge	q	✓	✓
charge density	ρ	$\times$	$\times$

4.3 The conservation of energy-momentum

So far we have discussed energy and momentum by introducing the definitions (3.59) without explaining where they come from (in this book, that is: an introduction is provided in *The Wonderful World*). Historically, in 1905 Einstein first approached the subject of force and acceleration by finding the equation of motion of a charged particle subject to electric and magnetic fields, assuming the charge remained constant and the Maxwell and Lorentz force equations were valid, and that Newton's 2nd law applied in the particle's rest frame. He could then use the theory he himself developed to understand what must happen in other frames, and hence derive the equation of motion for a general velocity of the particle. Subsequently Planck pointed out that the result could be made more transparent if one understood the 3-momentum to be given by $\gamma m_0 \mathbf{v}$. A significant further development took place in 1909 when Lewis and Tolman showed that this definition was consistent with momentum conservation in all reference frames. Nowadays, we can side-step these arguments by proceeding straight to the main result using 4-vector methods. However, when learning the subject the Lewis and Tolman argument remains a useful way in, so we shall present it first.

4.3.1 Elastic collision, following Lewis and Tolman

[*Section omitted in lecture-note version.*]

4.3.2 Energy-momentum conservation using 4-vectors

The Lewis and Tolman argument has the merit of being unsophisticated for the simplest case, but it is not easy to generalise it to all collisions. The use of 4-vectors makes the general argument much more straightforward.

By considering the worldline of a particle, we showed in chapter 3.1 that various 4-vectors, such as spacetime position X, 4-velocity $\mathsf{U} = d\mathsf{X}/d\tau$ and 4-momentum $\mathsf{P} = m_0\mathsf{U}$ could be associated with a single particle. In order to introduce a conservation law, we need to define first of all what we mean by the 4-momentum of a *collection* of particles. The definition is the obvious one:

$$\mathsf{P}_{\rm tot} \equiv \mathsf{P}_1 + \mathsf{P}_2 + \mathsf{P}_3 + \cdots + \mathsf{P}_n. \tag{4.64}$$

That is, we define the total 4-momentum of a collection of n particles to be the sum of the individual 4-momenta. Now we can state what we mean by the conservation of energy and momentum:

> **Law of conservation of energy and momentum**: the total energy-momemtum 4-vector of an isolated system is independent of time. In particular, it is not changed by internal interactions among the parts of the system.

In order to apply the insights of Special Relativity to dynamics, we state this conservation law as an axiom. Before going further we must check that it is consistent with the other axioms. We shall find that it is. Then one can use the conservation law to make predictions which must be compared with experiment. Further insight will be provided in chapter 11, where this conservation law is related to invariance of the *action* under translations in time and space.

Agreement with the Principle of Relativity

First we tackle the first stage, which is to that show energy-momentum conservation, as defined above, is consistent with the main Postulates (the Principle Principle of Relativity and the light speed Postulate). To show this we write down the conservation law in one reference frame, and then use the Lorentz transformation to find out how the same situation is described in another reference frame.

Let $\mathsf{P}_1, \mathsf{P}_2, \ldots \mathsf{P}_n$ be the 4-momenta of a set of particles, as observed in frame S. Then, by definition, the total 4-momentum is $\mathsf{P}_{\rm tot}$, given by (4.64). By calling the result of this sum a '4-momentum' and giving it a symbol $\mathsf{P}_{\rm tot}$ we are strongly implying that the sum total is itself a 4-vector. You might think that this is obvious, but in fact it requires further thought. After all, we already noted that adding up 4-velocities does not turn out to be a sensible thing to do—so why is 4-momentum any different? When we carry out the mathematical sum, summing the 4-momentum of one particle and the 4-momentum of a *different* particle, we are adding up things that are specified at different events in spacetime. When the terms in the sum can themselves change with time, we need to clarify at what moment each individual P_i term is to be taken. Therefore a more careful statement of the definition $\mathsf{P}_{\rm tot}$ would read:

$$\mathsf{P}_{\rm tot}(t = t_0) \equiv \mathsf{P}_1(t = t_0) + \mathsf{P}_2(t = t_0) + \mathsf{P}_3(t = t_0) + \cdots + \mathsf{P}_n(t = t_0) \tag{4.65}$$

where t_0 is the instant in some frame S at which the total 4-momentum is being defined.

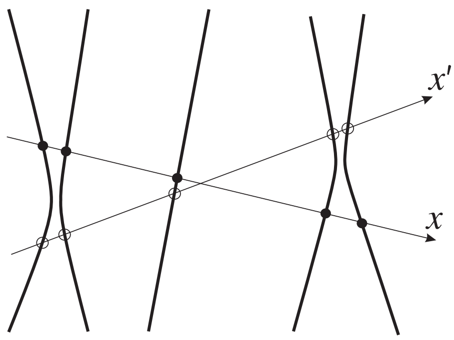


Figure 4.6: A set of worldlines is shown on a spacetime diagram, with lines of simultaneity for two different reference frames. The energy-momenta at some instant in frame S are defined at a different set of events (shown dotted) from those obtaining at some instant in frame S′ (circled). Therefore each term in the sum defining the total energy-momentum at some instant in S is not necessarily the Lorentz-transform of the corresponding term in the sum defining the total energy-momentum at some instant in S′. However, when the terms are added together, as long as 4-momentum conservation holds and the total system is isolated, the totals $\mathsf{P}_{\rm tot}$ and $\mathsf{P}'_{\rm tot}$ are Lorentz-transforms of each other. Proof: one can always choose to evaluate $\mathsf{P}_{\rm tot}$ by summing P_i at the circled events, i.e. those that are simultaneous in the *other* frame. This sum is the same as that at the dotted events, because the conservation law ensures that any collisions taking place do not change the total 4-momenta of the colliding partners, and between collisions each particle maintains a constant 4-momentum.

Now, if we apply the definition to the same set of particles, but now at some instant t'_0 in a different reference frame S′, we find the total 4-momentum in S′ is

$$\mathsf{P}'_{\rm tot}(t' = t'_0) \equiv \mathsf{P}'_1(t' = t'_0) + \mathsf{P}'_2(t' = t'_0) + \mathsf{P}'_3(t' = t'_0) + \cdots + \mathsf{P}'_n(t' = t'_0). \tag{4.66}$$

The problem is, the 4-momenta being summed in (4.65) are taken at a set of events simultaneous in S, while the 4-momenta being summed in (4.66) are being summed at a set of events simultaneous in S′. Owing to the relativity of simultaneity, these are two different sets of events. Therefore the individual terms are not necessarily Lorentz-transforms of each other:

$$\mathsf{P}'_i(t' = t'_0) \neq \mathcal{L}\mathsf{P}_i(t = t_0). \tag{4.67}$$

Therefore when we take the Lorentz-transform of $\mathsf{P}_{\rm tot}$ we will not obtain $\mathsf{P}'_{\rm tot}$, unless there is a physical constraint on the particles that makes their 4-momenta behave in such a way that $\mathcal{L}\mathsf{P}_{\rm tot}$ does equal $\mathsf{P}'_{\rm tot}$. Fortunately, the conservation law itself comes to the rescue, and provides precisely the constraint that is required! Proof (see figure 4.6): When forming the sum in one reference frame, one can always artificially choose a set of times t_i that lie in a plane of simultaneity for the other reference frame. Compared with the sum at t_0, the terms will either stay the same (for particles that move freely between t_0 and their t_i) or they will change (for particles that collide or interact in any way between t_0 and t_i), but if 4-momentum is conserved, such interactions do not change the total $\mathsf{P}_{\rm tot}$. QED.

We originally introduced P in section (3.5.3) as a purely mathematical quantity: a 4-vector related to 4-velocity and rest mass. That did not in itself tell us that P is conserved. However, if the natural world is mathematically consistent and Special Relativity describes it, then only certain types of quantity can be universally conserved (i.e. conserved in all reference frames). It makes sense to postulate a conservation law for something like $\gamma m_0 \mathbf{u}$ (3-momentum) because this is part of a 4-vector. The formalism of Lorentz transformations and 4-vectors enables us to take three further steps:

1. If a 4-vector is conserved in one reference frame then it is conserved in all reference frames.
2. "**Zero component lemma**": If one component of a 4-vector is conserved in all reference frames then the entire 4-vector is conserved.
3. A sum of 4-vectors, each evaluated at a different position (at some instant of time in a given reference frame), is itself a 4-vector if the sum is conserved.

Proof. We already dealt with item (3). For item (1) argue as follows. The word 'conserved' means 'constant in time' or 'the same before and after' any given process. For some chosen reference frame let P be the conserved quantity, with $\mathsf{P}_{\rm before}$ signifying its value before some process, and $\mathsf{P}_{\rm after}$. The conservation of P is then expressed by

$$\mathsf{P}_{\rm before} = \mathsf{P}_{\rm after}. \tag{4.68}$$

Now consider the situation in another reference frame. Since P is a 4-vector, we know how it transforms: we shall find

$$\mathsf{P}'_{\text{before}} = \mathcal{L}\mathsf{P}_{\text{before}}\,, \qquad \mathsf{P}'_{\text{after}} = \mathcal{L}\mathsf{P}_{\text{after}}.$$

By applying a Lorentz transformation to both sides of (4.68) we shall immediately find $\mathsf{P}'_{\text{before}} = \mathsf{P}'_{\text{after}}$, i.e. the quantity is also conserved in the new reference frame, QED. This illustrates how 4-vectors 'work': by expressing a physical law in 4-vector form we automatically take care of the requirements of the Principle of Relativity.

To prove the zero component lemma, consider the 4-vector $\Delta\mathsf{P} = \mathsf{P}_{\text{after}} - \mathsf{P}_{\text{before}}$. If one component of 4-momentum is conserved in all reference frames then one component of $\Delta\mathsf{P}$ is zero. Pick the x-component $\Delta\mathsf{P}^1$ for example. If there is a frame in which the y or z component is non-zero, then we can rotate axes to make the x component non-zero, contrary to the claim that it is zero in all reference frames. Therefore the y and z components are zero also. If there is a reference frame in which the time-component $\Delta\mathsf{P}^0$ is non-zero, then we can apply a Lorentz transformation to make $\Delta\mathsf{P}^1$ non-zero, contrary to the claim. Therefore $\Delta\mathsf{P}^0$ is zero. A similar argument can be made starting from any of the components, which concludes the proof.

4.3.3 Mass-energy equivalence

At first the zero component lemma might seem to be merely a piece of mathematics, but it is much more. It says that if we have conservation (in all reference frames) of a scalar quantity that is known to be one component of a 4-vector, then we have conservation of the whole 4-vector. This enables us to reduce the number of assumptions we need to make: instead of postulating conservation of 4-momentum, for example, we could postulate conservation of one of its components, say the x-component of momentum, in all reference frames, and we would immediately deduce not only conservation of 3-momentum but conservation of energy as well.

In classical physics the conservation laws of energy and momentum were separate: they do not necessarily imply one another. In Relativity they do. The conservation of the 3-vector quantity (momentum) is no longer separate from the conservation of the scalar quantity (energy). The unity of spacetime is here exhibited as a unity of energy and momentum. It is not that they are the same, but they are two parts of one thing.

Once we have found the formula relating the conserved 3-vector to velocity, i.e. $\mathbf{p} = \gamma m \mathbf{v}$ (the spatial part of $m\mathsf{U}$), we do not have any choice about the formula for the conserved scalar, up to a constant factor, it must be $E \propto \gamma m$ (the temporal part of $m\mathsf{U}$). Also, the constant factor must be c^2 in order to give the known formula for kinetic energy in the low-velocity limit and thus match with classical definition of what we call energy. *Thus the important relation "$E = mc^2$" follows from momentum conservation and the main Postulates.* This formula gives rise to a wonderful new insight—perhaps the most profound prediction of Special Relativity—namely the *equivalence of mass and energy.* By this we mean two things. First, in any process, kinetic energy of the reactants can contribute to rest mass of the products, and conversely. For example, in a collision where two particles approach and then stick together, there is a reference frame where the product is at rest. In that frame, we shall find $Mc^2 = \sum \gamma_i m_i c^2$ and therefore $M > \sum m_i$ where M is the rest mass of the product and m_i are the rest masses of the reactants.

The physical meaning of this rest mass M is *inertial.* It is "that which increases the momentum", i.e. the capacity of a body to make other things move when it hits them. It does not immediately follow that it is the same thing as gravitational mass. One of the foundational assumptions of General Relativity is that this inertial mass is indeed the same thing as gravitational mass, for a body at rest with no internal pressure.

The second part of the meaning of "equivalence of mass and energy" is that "rest mass" and "rest energy" are simply different words for the same thing (up to a multiplying constant, i.e. c^2). This is a strict equivalence. It is not that they are 'like' one another (as is sometimes asserted of space and time, where the likeness is incomplete), but they are strictly the same, just different words used by humans for the same underlying physical reality. In an exothermic reaction such as nuclear fission, therefore, rather than saying "mass is converted into energy" it is arguably more correct to say simply that energy is converted from one form to another. We have only ourselves to blame if we gave it a different name when it was located in the nucleus. The point can be emphasized by considering a more everyday example such as compression of an ordinary metal spring. When under compression, energy has been supplied to the spring, and we are taught to call it 'potential energy.' We may equally call it 'mass energy': it results in an increase in the rest mass of the spring (by the tiny amount of 10^{-17} kg per joule). When we enjoy the warmth from a wooden log fire, we are receiving benefit from a process of "conversion of mass to energy" just as surely as when we draw on the electrical power provided by a nuclear power station. The "binding energy" between the oxygen atoms and carbon atoms is another name for a rest mass deficit: each molecule has a smaller rest mass than the sum of the rest masses of the separate atoms. The tiny difference δm is enough to liberate noticeable amounts of energy ($\delta m c^2$) in another form such as heat.

4.4 Collisions

We will now apply the conservation laws to a variety of collision-type processes, starting with the most simple and gaining in complexity as we proceed. We will make repeated use of the formula $E^2 - p^2c^2 = m^2c^4$ which we can now recognise both as a statement about mass and energy, and also as a Lorentz invariant quantity associated with the energy-momentum 4-vector.

The quantities E_i, p_i, m_i will usually refer to the energy, momentum and rest mass of the i'th particle *after* the process. In particle physics experiments one typically gathers information on p and E (e.g. from curvature of particle tracks and from energy deposited in a detector, respectively), and some or all of the rest masses may be known. To extract a velocity one can use $\mathbf{v} = \mathbf{p}c^2/E$ (eq. (3.60)). However, not all the information is always available, and typically momenta can be obtained more precisely than energies. Even if one has a set of measurements that in principle gives complete information, it is still very useful to establish relations (constraints) that the data ought to obey, because this will allow the overall precision to be improved, consistency checks to be made, and systematic error uncovered. Also, it is crucial to have good systematic ways of looking for patterns in the data, because usually the interesting events are hidden in a great morass or background of more frequent but mundane processes.

1. *Spontaneous emission, radioactive decay.*
An atom at rest emits a photon and recoils. For a given energy level difference in the atom, what is the frequency of the emitted photon? A radioactive nucleus emits a single particle of given rest mass. For a given change in rest mass of the nucleus, what is the energy of the particle?

These are both examples of the same type of process. Before the process there is a single particle of rest mass M^* and zero momentum. The asterisk serves as a reminder that this is an excited particle that can decay. Afterwards there are two particles of rest mass m_1 and m_2. By conservation of momentum these move in opposite directions, so we only need to treat motion in one dimension. The conservation of energy and momentum gives

$$\begin{aligned} M^* c^2 &= E_1 + E_2, & (4.69) \\ p_1 &= p_2. & (4.70) \end{aligned}$$

The most important thing to notice is that, for given rest masses M^*, m_1, m_2, there is a *unique solution* for the energies and momenta (i.e. the sizes of the momenta; the directions must be opposed but otherwise they are unconstrained). This is because we have 4 unknowns E_1, E_2, p_1, p_2 and four equations—the above and $E_i^2 - p_i^2 c^2 = m_i^2 c^4$ for $i = 1, 2$.

Taking the square of the momentum equation, we have $E_1^2 - m_1^2 c^4 = E_2^2 - m_2^2 c^4$. After substituting for E_2 using (4.69), this is easily solved for E_1, giving

$$E_1 = \frac{M^{*2} + m_1^2 - m_2^2}{2M^*} c^2. \tag{4.71}$$

When the emitted particle is a photon, $m_1 = 0$ so this can be simplified. Let $E_0 = M^* c^2 - m_2 c^2$ be the gap between the energy levels of the decaying atom or nucleus in its rest frame. Then $M^{*2} - m_2^2 = (M^* + m_2)(M^* - m_2) = (2M^* - E_0/c^2)E_0/c^2$ so

$$E_1 = \left(1 - \frac{E_0}{2M^* c^2}\right) E_0. \tag{4.72}$$

The energy of the emitted photon is slightly smaller than the rest energy change of the atom. The difference $E_0^2/(2M^* c^2)$ is called the recoil energy.

2. *Absorption.*
[*Section omitted in lecture-note version.*]

3. *In-flight decay.*
It has not escaped our notice that absorption and emission are essentially the same process running in different directions, and therefore eq. (**??**) could be obtained from (4.71) by a change of reference frame. To treat the general case of a particle moving with any speed decaying into two or more products, it is better to learn some more general techniques employing 4-vectors.

Suppose a particle with 4-momentum P decays into various products. The conservation of 4-momentum reads

$$\mathsf{P} = \sum_i \mathsf{P}_i. \tag{4.73}$$

Therefore

$$M^2c^4 = E^2 - p^2c^2 = (\sum E_i)^2 - (\sum \mathbf{p}_i) \cdot (\sum \mathbf{p}_i)c^2. \tag{4.74}$$

Thus if all the products are detected and measured, one can deduce the rest mass M of the original particle.

In the case of just two decay products (a so-called *two body decay*), a useful simplification is available. We have

$$\mathsf{P} = \mathsf{P}_1 + \mathsf{P}_2. \tag{4.75}$$

Take the scalar product of each side with itself:

$$\mathsf{P} \cdot \mathsf{P} = \mathsf{P}^2 = \mathsf{P}_1^2 + \mathsf{P}_2^2 + 2\mathsf{P}_1 \cdot \mathsf{P}_2 \tag{4.76}$$

All these terms are Lorentz-invariant. By evaluating P^2 in any convenient reference frame, one finds $\mathsf{P}^2 = -M^2c^2$, and similarly $\mathsf{P}_1^2 = -m_1^2c^2$, $\mathsf{P}_2^2 = -m_2^2c^2$. Therefore

$$M^2 \;=\; m_1^2 + m_2^2 + \frac{2}{c^4}(E_1E_2 - \mathbf{p}_1 \cdot \mathbf{p}_2c^2) \tag{4.77}$$

(c.f. eq. (3.67)). This shows that to find M it is sufficient to measure the sizes of the momenta and the angle between them, if m_1 and m_2 are known.

The $\mathsf{P}_1 \cdot \mathsf{P}_2$ term in (4.76) can also be interpreted using eq. (**??**), giving

$$M^2 \;=\; m_1^2 + m_2^2 + 2m_1m_2\gamma(u) \tag{4.78}$$

e	0.510999 MeV
p	938.272 MeV
π_0	134.977 MeV
$\pi_\pm$	139.570 MeV
Z	(91.1876 ± 0.0021) GeV

Table 4.2: Some particles and their rest energies to six significant figures.

where u is the relative speed of the decay products.

Some further comments on the directions of the momenta are given in the discussion of elastic collisions below, in connection with figure **??** which applies to any 2-body process.

4. *Particle formation and centre of momentum frame*
A fast-moving particle of energy E, rest mass m, strikes a stationary one of rest mass M. One or more new particles are created. What are the energy requirements?

The most important idea in this type of collision is to consider the situation in the **centre of momentum frame**. This is the inertial frame of reference in which the total momentum is zero. The total energy of the system of particles in this reference frame is called the 'centre of momentum collision energy' $E_{\rm CM}$ or sometimes (by a loose use of language) the 'centre of mass energy'. The quickest way to calculate $E_{\rm CM}$ is to the use Lorentz invariant '$E^2 - p^2c^2$' applied to the total energy-momentum of the system. In the laboratory frame before the collision the total energy-momentum is $\mathsf{P} = (E/c + Mc,\ \mathbf{p})$ where $\mathbf{p}$ is the momentum of the incoming particle. In the centre of momentum frame the total energy-momentum is simply $(E_{\rm CM}/c, 0)$. Therefore by Lorentz invariance we have

$$\begin{aligned} E_{\rm CM}^2 &= (E + Mc^2)^2 - p^2c^2 \\ &= m^2c^4 + M^2c^4 + 2Mc^2E. \end{aligned} \tag{4.79}$$

If the intention is to create new particles by smashing existing ones together, then one needs to provide the incoming 'torpedo' particle with sufficient energy. In order to conserve momentum, the products of the collision must move in some way in the laboratory frame. This means that not all of the energy of the 'torpedo' can be devoted to providing the rest mass needed to create new particles. Some of it has to be used up furnishing the products with kinetic energy. The least kinetic energy in the centre of momentum frame is obviously obtained when all the products are motionless. This suggests that this is the optimal case, i.e. with the least kinetic energy in the laboratory frame also. To prove that this is so, apply eq. (4.78), in which M on the left hand side is the invariant associated with P, the total energy-momentum of the system. This shows that the minimum $\gamma(u)$ is attained at the minimum M. M can never be less than the sum of the post-collision rest masses, but it can attain that minimum if the products do not move in the centre of momentum frame. Therefore the threshold γ_u factor, and hence the threshold energy, is when

$$E_{\rm CM} = \sum_i m_i c^2 \tag{4.80}$$

where m_i are the rest masses of the collision products. Substituting this into (4.79) we obtain the general result:

$$E_{\rm th} = \frac{(\sum_i m_i)^2 - m^2 - M^2}{2M} c^2. \tag{4.81}$$

This gives the threshold energy in the laboratory frame for a particle m hitting a free stationary target M, such that collision products of total rest mass $\sum_i m_i$ can be produced.

Let us consider a few examples. Suppose we would like to create antiprotons by colliding a moving proton with a stationary proton. The process $\mathrm{p} + \mathrm{p} \to \bar{\mathrm{p}}$ does not exist in nature because it does not satisfy conservation laws associated with particle number, but the process $\mathrm{p} + \mathrm{p} \to \mathrm{p} + \mathrm{p} + \mathrm{p} + \bar{\mathrm{p}}$ is possible. Applying eq. (4.81) we find that the energy of the incident proton must be $7Mc^2$, i.e. 3.5 times larger than the minimum needed to create a proton/anti-proton pair.

In general, eq. (4.81) shows that there is an efficiency problem when the desired new particle is much heavier than the target particle. Suppose for example that we wanted to create Z bosons by smashing fast positrons into electrons at rest in the laboratory. Eq. (4.81) says the initial energy of the positrons must be approximately 90000 times larger than the rest-energy of a Z boson! Almost all the precious energy, provided to the incident particle using expensive accelerators, is 'wasted' on kinetic energy of the products. In Rindler's memorable phrase, "it is a little like trying to smash ping-pong balls floating in space with a hammer". This is the reason why the highest-energy particle accelerators now adopt a different approach, where two beams of particles with equal and opposite momenta are collided in the laboratory. In such a case the laboratory frame is the CM frame, so all the energy of the incident particles can in principle be converted into rest mass energy of the products. Getting a pair of narrow intense beams to hit each other presents a great technical challenge, but formidable as the task is, it is preferable to attempting to produce a single beam of particles with energies thousands of times larger. This is the way the Z boson was experimentally discovered in the 'SPS' proton-antiproton collider at CERN, Geneva in 1983, and subsequently produced in large numbers by that laboratory's large electron-positron collider ('LEP').

The process of creating particles through collisions is called *formation.* In practice the formed particle may be short-lived and never observed directly. The sequence of events may be, for example, $a + b \to X \to a + b$, or else X may be able to decay into other particles (in which case it is said to have more than one *decay channel*). The state consisting of X is a state of reasonably well-defined energy and momentum (broadened by the finite lifetime of the particle). It shows up in experiments as a large enhancement in the scattering cross section when a and b scatter off one another.

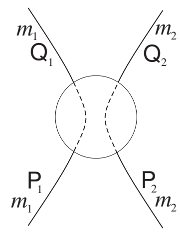


Figure 4.7: A generic elastic collision, in which the incoming 4-momenta are $\mathsf{P}_1, \mathsf{P}_2$, the outgoing 4-momenta are Q_1, Q_2. The rest masses m_1, m_2 are unchanged.

5. *3-body decay.* [*Section omitted in lecture-note version.*]

4.4.1 Elastic collisions

We term a collision *elastic* when the rest masses of the colliding particles are all preserved. Though less glamorous than inelastic processes and particle formation, elastic collisions are an important tool in particle physics for probing the structure of composite particles, and testing fundamental theories, for example of the strong and weak interactions.

A generic two-body elastic collision is shown in figure 4.7, in order to introduce notation. To conserve energy-momentum we have $\mathsf{P}_1+\mathsf{P}_2 = \mathsf{Q}_1+\mathsf{Q}_2$. Squaring this gives $\mathsf{P}_1^2+\mathsf{P}_2^2+2\mathsf{P}_1\cdot\mathsf{P}_2 = \mathsf{Q}_1^2+\mathsf{Q}_2^2+2\mathsf{Q}_1\cdot\mathsf{Q}_2$. But by hypothesis, $\mathsf{P}_1^2 = \mathsf{Q}_1^2$ and $\mathsf{P}_2^2 = \mathsf{Q}_2^2$. It follows that

$$\mathsf{P}_1\cdot\mathsf{P}_2 = \mathsf{Q}_1\cdot\mathsf{Q}_2. \tag{4.82}$$

Using (**??**) it is seen that this implies the relative speed of the particles is the same before and after the collision, just as occurs in classical mechanics.

In the centre of momentum (CM) frame, an elastic collision is so simple as to be almost trivial: the two particles approach one another along a line with equal and opposite momenta; after the collision they leave in opposite directions along another line, with the same relative speed and again equal and opposite momenta. The result in some other frame is most easily obtained by Lorentz transformation from this one.

Consider the case of identical particles ('relativistic billiards'). Let frame S′ be the CM frame, in which the initial and final speeds are all v. Choose the x' axis along the incident direction of one of the particles. If the final velocity in the CM frame of one particle is directed at some angle θ_0 to the x' axis in the anticlockwise direction, then the other is at $\theta_0-\pi$, i.e. $\pi-\theta_0$ in the clockwise direction.

We take an interest in the 'lab frame' S where one of the particles was initially at rest (see figure

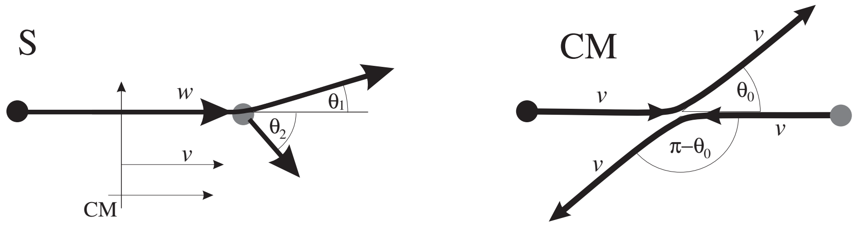


Figure 4.8: An elastic collision between particles of equal rest mass. The 'lab frame' S is taken to be that in which one of the particles is initially at rest. The CM moves at speed v relative to S. The incoming particle has speed $w = 2v/(1 + v^2/c^2)$ in S.

4.8). The post-collision angles θ_1 and θ_2 are related to θ_0 and $\theta_0 - \pi$ by the angle transformation equation for particle velocities (3.61), with the substitutions $\theta \to (\theta_0$ or $\theta_0 - \pi)$, $\theta' \to (\theta_1$ or $\theta_2)$, $u \to v$, $v \to -v$. Hence

$$\tan\theta_1 = \frac{\sin\theta_0}{\gamma_v(\cos\theta_0 + 1)}, \qquad \tan\theta_2 = \frac{\sin\theta_0}{\gamma_v(\cos\theta_0 - 1)},$$

where θ_1, θ_2 are both measured anticlockwise, with the result that they have opposite signs. Using these expressions we find, for $\theta_0 \neq 0$,

$$\tan(\theta_1 - \theta_2) = \frac{2\gamma_v}{(\gamma_v^2 - 1)\sin\theta_0}. \tag{4.83}$$

(The case $\theta_0 = 0$ has to be treated separately, but it is has an obvious answer). Hence at low speeds ($\gamma_v \to 1$) the opening angle $(\theta_1 - \theta_2)$ tends to 90°, the familiar classical result that the particles move at right angles. For higher speeds the opening angle is less than 90° because both particles are 'thrown forward' compared to the classical case, c.f. figure 3.6. Elastic collisions with opening angles below 90° are frequently seen in particle accelerators and in cosmic ray events in photographic emulsion detectors.

In terms of the relative speed w we have

$$\gamma_v^2 = \tfrac{1}{2}(\gamma(w) + 1)$$

by using the gamma relation (3.13). The relationship between θ_1 and θ_2 can also be written

$$\tan\theta_1 \tan\theta_2 \;=\; -1/\gamma_v^2 \qquad \Rightarrow \quad \tan\theta_2 = \gamma_v^{-2}\tan(\theta_1 - \pi/2), \tag{4.84}$$

using $\cot\theta = \tan(\pi/2 - \theta)$.

Compton scattering

'Compton scattering' is the scattering of light off particles, such that the recoil of the particles results in a change of wavelength of the light. When Arthur Compton (1892-1962) and others discovered changes in the wavelength of X-rays and γ-rays scattered by electrons, and especially changes that depended on scattering angle, it was very puzzling, because it is hard to see how a wave of given frequency can cause any oscillation at some other frequency when it drives a free particle. Compton's careful experimental observations gave him sufficiently accurate data to lend focus to his attempts to model the phenomenon theoretically. He hit upon a stunningly simple answer by combining the quantum theory of light, still in its infancy, with Special Relativity.

Let the initial and final properties of the photon be (E, p) and (E', p'), and let m be the rest mass of the target (assumed initially stationary). Then

$$E + mc^2 = E' + \sqrt{m^2c^4 + p_{\rm f}^2c^2}, \quad \mathbf{p} = \mathbf{p}' + \mathbf{p}_{\rm f} \tag{4.85}$$

where $p_{\rm f}$ is the final momentum of the target (such as an electron) whose rest mass is assumed unchanged. From the momentum equation we may obtain $p_{\rm f}^2 = p^2 + p'^2 - 2\mathbf{p}\cdot\mathbf{p}' = p^2 + p'^2 - 2pp'\cos\theta$ where θ is the angle between the incident and final directions of the scattered photon. Substituting this into the energy equation, and using $E = pc$, $E' = p'c$ for zero rest mass, one obtains after a little algebra

$$\begin{aligned} (E - E')mc^2 &= EE'(1 - \cos\theta) \\ \Rightarrow \quad \frac{1}{E'} - \frac{1}{E} &= \frac{1}{mc^2}(1 - \cos\theta). \end{aligned} \tag{4.86}$$

(See also exercise **??** for a neat method using 4-vectors.)

So far the calculation has concerned particles and their energies and momenta. If we now turn to quantum theory then we can relate the energy of a photon to its frequency, according to Planck's famous relation $E = h\nu$. Then eq. (4.86) becomes

$$\lambda' - \lambda = \frac{h}{mc}(1 - \cos\theta). \tag{4.87}$$

This is the Compton scattering formula.

A wave model of Compton scattering is not completely impossible to formulate, but the particle model presented above is much simpler. In a wave model, the change of wavelength arises from a Doppler effect owing to the motion of the target electron.

The quantity

$$\lambda_C \equiv \frac{h}{mc} \tag{4.88}$$

is called the **Compton wavelength**. For the electron its value is $2.4263102175(33) \times 10^{-12}$ m. It is poorly named because, although it may be related to wavelengths of photons, it is best understood as the distance scale below which quantum field theory (chapter 19) is required; both classical physics and non-relativistic quantum theory then break down. The Bohr radius can be written

$$a_0 = \frac{\lambda_C}{2\pi\alpha}$$

where α is the fine structure constant. Since $\alpha \ll 1$ we find that $a_0 \gg \lambda_C$, so quantum field theory is not required to treat the structure of atoms, at least in first approximation: Schrödinger's equation will do. The non-relativistic Schrödinger equation for the hydrogen atom can be written

$$-\frac{\lambda_C}{4\pi}\nabla^2\psi - \frac{\alpha}{r}\psi = \frac{i}{c}\frac{\partial\psi}{\partial t}.$$

Elastic terminology. Compton scattering appears here under the heading of 'elastic' processes because the rest masses do not change. However, the word 'elastic' can also be used to mean that the energies of the colliding parties are unchanged; Compton scattering is not elastic in that sense, except in the limit $m \to \infty$.

Inverse Compton scattering

The formula (4.86) shows that a photon scattering off a stationary particle always loses energy. A photon scattering off a moving particle can either lose or gain energy; the latter case is sometimes called 'inverse Compton scattering'. It is of course just another name for Compton scattering viewed from a different reference frame. In astrophysics such inverse Compton scattering is more important (because a more useful source of observational information) than Compton scattering.

Let $\mathsf{P}_1, \mathsf{Q}_1$ be the 4-momenta of the photon before and after the collision, and $\mathsf{P}_2, \mathsf{Q}_2$ be those of the other particle. Conservation of energy-momentum gives

$$\mathsf{P}_1 + \mathsf{P}_2 = \mathsf{Q}_1 + \mathsf{Q}_2.$$

Supposing that the initial conditions P_1 and P_2 are given, we would like to know the final properties of the photon, i.e. Q_1. To get rid of Q_2, isolated it and then square:

$$\begin{aligned}(\mathsf{P}_1 + \mathsf{P}_2 - \mathsf{Q}_1)^2 = \mathsf{Q}_2^2 \quad &\Rightarrow \quad \mathsf{P}_1^2 + \mathsf{P}_2^2 + 2(\mathsf{P}_1 \cdot \mathsf{P}_2 - \mathsf{P}_1 \cdot \mathsf{Q}_1 - \mathsf{P}_2 \cdot \mathsf{Q}_1) = 0 \\ &\Rightarrow \quad \mathsf{P}_1 \cdot \mathsf{Q}_1 = \mathsf{P}_2 \cdot (\mathsf{P}_1 - \mathsf{Q}_1)\end{aligned}$$

where we used $\mathsf{P}_1^2 = \mathsf{P}_2^2 = 0$. So far the result is true in general, for any angles. For the sake of simplicity, we now specialize to the case of a head on collision, i.e. $\mathsf{P}_1 = E_1(1,1)$, $\mathsf{Q}_1 = E_1'(1,-1)$, $\mathsf{P}_2 = \gamma m(1,-u)$ in one spatial dimension, and taking $c = 1$. We thus find

$$-2E_1E_1' = \gamma m[-E_1 + E_1' - u(E_1 + E_1')].$$

Solving for E_1' yields

$$E_1' = \frac{\gamma m(1+u)}{2 + \gamma m(1-u)/E_1}. \tag{4.89}$$

When $u \simeq 1$ (i.e. close to the speed of light) it is more useful to write $(1+u) \simeq 2$ and $(1-u) \simeq 1/2\gamma^2$, so

$$E_1' \simeq \frac{\gamma m}{1 + m/4\gamma E_1} \tag{4.90}$$

which further simplifies to $E_1' = 4\gamma^2 E_1$ (hence wave frequency $\nu' = 4\gamma^2\nu$) when $\gamma E_1 \ll m$.

This process is relevant in various astrophysical phenomena, such as X ray emission from active galactic nuclei, gamma ray emission in some quasars, and X ray emission in intergalactic space. For example, an electron with $\gamma \simeq 10^4$ colliding with a photon from the cosmic microwave background radiation (wavelength $\simeq 0.5$ cm) can result in a scattered X-ray photon. At higher energies, the incident particle loses a large fraction of its energy in a single collision.

Compton and inverse Compton scattering are also related to *bremsstrahlung* or 'breaking radiation,' which is the radiation emitted when charged particles are slowed, for example by elastic collisions with atomic nuclei.

More general treatment of elastic collisions*

[*Section omitted in lecture-note version.*]

4.5 Composite systems

In the discussion of Special Relativity in this book we have often referred to 'objects' or 'bodies' and not just to 'particles'. In other words we have taken it for granted that we can talk of a composite entity such as a brick or a plank of wood as a single 'thing', possessing a position, velocity and mass. The conservation laws are needed in order to make this logically coherent (the same is true in classical physics).

We use the word 'system' to refer to a collection of particles whose behaviour is going to be discussed. Such a system could consist of particles that are attached to one another, such as the atoms in a solid object, or it could be a loose collection of independent particles, such as the atoms in a low-density gas. In either case the particles do not 'know' that we have gathered them together into a 'system': the system is just our own selection, a notional 'bag' into which we have placed the particles, without actually doing anything to them. The idea of a system is usually invoked when the particles in question may interact with one another, but they are not interacting with anything else. Then we say we have an 'isolated system.' This terminology was already invoked in the previous section. We there talked about the total energy and total 3-momentum of such a system. Now we would like to enquire what it might mean to talk about the velocity and rest-mass of a composite system.

If a composite system can be discussed as a single object, then we should expect that its rest mass must be obtainable from its total energy-momentum in the standard way, i.e.

$$\mathsf{P}_{\text{tot}}^2 = -E_{\text{tot}}^2/c^2 + \mathbf{p}_{\text{tot}}^2 \equiv -m^2c^4. \tag{4.91}$$

This serves as the definition of the rest mass m of the composite system. It makes sense because the conservation law guarantees that P_{tot} is constant if the system is not subject to external forces.

One convenient way to calculate m is to work it out in the CM frame, where $\mathbf{p}_{\text{tot}} = 0$. Thus we find

$$m = E_{\text{CM}}/c^2 \tag{4.92}$$

where E_{CM} is the value of E_{tot} in the CM frame. Note that the *rest* mass of the composite system is equal to the *total energy* of the constituent particles (divided by c^2) in the CM frame, not the sum of their rest masses. For example, a system consisting of two photons propagating in different directions has a non-zero rest mass[2]. The photons propagating inside a hot oven or a bright star make a contribution to the rest mass of the respective system.

Relative to any other reference frame, the CM frame has some well-defined 3-velocity $\mathbf{u}_{\text{CM}}$, and therefore a 4-velocity $\mathsf{U}_{\text{CM}} = \gamma(u_{\text{CM}})(c, \mathbf{u}_{\text{CM}})$. You can now prove that

$$\mathsf{P}_{\text{tot}} = m\mathsf{U}_{\text{CM}} \tag{4.93}$$

(method: first prove that the direction of the spatial part agrees, then check the magnitudes of the 4-vectors.) This confirms that the composite system is behaving as we would expect for a single object of given rest mass and velocity. It also provides an easy way to find the velocity of the CM frame.

[2]For two or more photons all propagating in the same direction, there is no CM frame because reference frames cannot attain the speed of light.

4.6 Energy flux, momentum density, and force

There is an important general relationship between *flux* of energy $\mathbf{S}$ and *momentum per unit volume* $\mathbf{g}$. It is easily stated:

$$\boxed{\mathbf{S} = \mathbf{g}c^2. \qquad (4.94)}$$

$\mathbf{S}$ is the amount of energy crossing a surface (in the normal direction), per unit area per unit time, and $\mathbf{g}$ is the momentum per unit volume in the flow.

It would be natural to expect energy flux to be connected to *energy* density. For example, for a group of particles all having energy E and moving together at the same velocity $\mathbf{v}$, the energy density is $u = En$ where n is the number of particles per unit volume, and the number crossing a surface of area A in time t is $nA(vt)$, so $S = nvE = uv$: the energy flux is proportional to the energy density. However, if the particles are moving in some other way, for example isotropically, then the relationship changes. For particles effusing from a hole in a chamber of gas, for example, we find $S = (1/4)uv$.

Eq. (4.94) is more general. For the case of particles all moving along together, it is easy to prove by using the fact that $\mathbf{p} = E\mathbf{v}/c^2$ for each particle. The momentum density is then $\mathbf{g} = n\mathbf{p}$, and the energy flux is $\mathbf{S} = n\mathbf{v}E = n\mathbf{p}c^2 = \mathbf{g}c^2$. If we now consider more general scenarios, such as particles in a gas, we can apply this basic vector relationship to every small region and small range of velocities, and when we do the sum to find the two totals, the proportionality factor is c^2 for every term in the sum, so it remains c^2 in the total.

The particles we considered may or may not have had rest mass: the relationship $\mathbf{p} = E\mathbf{v}/c^2$ is valid for either, so (4.94) applies equally to light and to matter, and to the fields inside a material body. It is universal!

Another important idea is *momentum flow*.

We introduced force by *defining* it as the rate of change of momentum. We also established that momentum is conserved. These two facts, taken together, imply that another way to understand force is in terms of momentum flow. When more than one force acts, we can have a balance of forces, so the definition in terms of rate of change of momentum is no longer useful: there isn't any such rate of change. In a case like that, we know what we mean by the various forces in a given situation: we mean that we studied other cases and we claim that the momentum *would* change if the other forces were not present.

In view of the primacy of conservation laws over the notion of force, it can sometimes be helpful to adopt another physical intuition of what a force represents. A force per unit area, in any situation, can be understood as an 'offered' momentum flux, i.e. an amount of momentum flowing across a surface, per unit area per unit time. When a field or a body offers a pressure force to its environment, it is as if it is continually bringing up momentum to the boundary, like the molecules in a gas hitting the chamber walls, and 'offering' the momentum to the neighbouring system. If the neighbour wants to refuse the offer of acquiring momentum, it has to push back with a force: it makes a counter-offer of just enough momentum flow to prevent itself from acquiring any net momentum. In the case of a gas, such a picture of momentum flow is natural, but one could if one chose claim that precisely the same flow is taking place in a solid, or anywhere a force acts. The molecules don't have to move in order to transport momentum: they only need to push on their neighbours. It is a matter purely of taste whether one prefers the language of 'force' or 'momentum flow'.

These ideas will be important in the later chapters of the book, where we shall grapple with the important but tricky concept of the *stress-energy tensor*.

4.7 Exercises

[*Section omitted in lecture-note version.*]

Chapter 5

Further kinematics

[*Section omitted in lecture-note version.*]

5.1 The Principle of Most Proper Time

[*Section omitted in lecture-note version.*]

5.2 4-dimensional gradient

Now that we have got used to 4-vectors, it is natural to wonder whether we can develop 4-vector operators, the 'larger cousins', so to speak, of the gradient, divergence and curl. A first guess might be to propose a 4-gradient $((1/c)\partial/\partial t, \partial/\partial x, \partial/\partial y, \partial/\partial z)$. Although this quantity is clearly a sort of gradient operator, it is not the right choice because the gradient it produces is not a standard 4-vector. One can see this by a simple example.

Consider some potential function $V(t, x)$ whose gradient we would like to examine. We have in mind for V a scalar quantity that is itself Lorentz-invariant. This means, if we change reference frames, the value of V at any particular event in spacetime does not change. However, owing to time dilation and space contraction the rate of change of V with either of t' or x' is not necessarily the same as the rate of change with t or x.

Consider a very simple case: $V(t, x) = x$, i.e. a potential which in some reference frame S is independent of time and slopes upwards as a function of x, with unit gradient. For an observer S$'$ moving in the positive x-direction, the potential would be found to be time-dependent.

Because he is moving towards regions of higher V, at any fixed position in reference frame S$'$, V *increases* as a function of time. To get this increase right, clearly we need a plus sign not a minus sign in the transformation formula for the four-gradient of V. However, the Lorentz transformation (3.1) has a minus sign.

The answer to this problem is that we must define the 4-dimensional gradient operator as

$$\Box = \left(-\frac{1}{c}\frac{\partial}{\partial t}, \, \boldsymbol{\nabla} \right) = \left(-\frac{1}{c}\frac{\partial}{\partial t}, \frac{\partial}{\partial x}, \frac{\partial}{\partial y}, \frac{\partial}{\partial z} \right). \tag{5.1}$$

The idea is that with this definition, $\Box V$ is a 4-vector, as we shall now prove.

Consider two neighbouring events. In some reference frame S their coordinates are t, x, y, z and $t + dt, x + dx, y + dy, z + dz$. The change in the potential V between these events is

$$dV = \left(\frac{\partial V}{\partial t}\right)_x dt + \left(\frac{\partial V}{\partial x}\right)_t dx, \tag{5.2}$$

where for simplicity we have chosen a potential function that is independent of y and z. Therefore

$$\begin{aligned} \left(\frac{\partial V}{\partial t'}\right)_{x'} &= \left(\frac{\partial V}{\partial t}\right)_x \left(\frac{\partial t}{\partial t'}\right)_{x'} + \left(\frac{\partial V}{\partial x}\right)_t \left(\frac{\partial x}{\partial t'}\right)_{x'} \\ \text{and} \quad \left(\frac{\partial V}{\partial x'}\right)_{t'} &= \left(\frac{\partial V}{\partial t}\right)_x \left(\frac{\partial t}{\partial x'}\right)_{t'} + \left(\frac{\partial V}{\partial x}\right)_t \left(\frac{\partial x}{\partial x'}\right)_{t'}. \end{aligned} \tag{5.3}$$

where t', x' are coordinates in some other frame S$'$. The coordinate systems are related by the Lorentz transformation, so

$$t = \gamma(t' + (v/c^2)x'), \qquad x = \gamma(vt' + x')$$

from which

$$\begin{aligned} \left(\frac{\partial t}{\partial t'}\right)_{x'} &= \gamma, \qquad \left(\frac{\partial t}{\partial x'}\right)_{t'} = \gamma v/c^2 \\ \left(\frac{\partial x}{\partial t'}\right)_{x'} &= \gamma v, \qquad \left(\frac{\partial x}{\partial x'}\right)_{t'} = \gamma. \end{aligned}$$

Substituting these into (5.3) we have

$$\begin{aligned}
\left(\frac{\partial V}{\partial t'}\right)_{x'} &= \gamma\left(\left(\frac{\partial V}{\partial t}\right)_x + v\left(\frac{\partial V}{\partial x}\right)_t\right), \\
\left(\frac{\partial V}{\partial x'}\right)_{t'} &= \gamma\left(\frac{v}{c^2}\left(\frac{\partial V}{\partial t}\right)_x + \left(\frac{\partial V}{\partial x}\right)_t\right).
\end{aligned}$$

After multiplying the first equation by (-1), this pair of equations can be written which can be written

$$\begin{pmatrix} \frac{-1}{c}\frac{\partial}{\partial t'} \\ \frac{\partial}{\partial x'} \end{pmatrix} V = \begin{pmatrix} \gamma & -\beta\gamma \\ -\beta\gamma & \gamma \end{pmatrix} \begin{pmatrix} \frac{-1}{c}\frac{\partial}{\partial t} \\ \frac{\partial}{\partial x} \end{pmatrix} V,$$

which is

$$\Box' V = \mathcal{L}\,\Box V. \tag{5.4}$$

This proves that $\Box V$ is a 4-vector.

To gain some familiarity, let us examine what happens to the gradient of a function $V(t,x) = \phi(x)$ that depends only on x in reference frame S. In this case the slope $(\partial V/\partial x)$ in S and the slope $(\partial V/\partial x')$ in S$'$ are related by a factor γ:

$$\frac{\partial V}{\partial x'} = \gamma \frac{\partial V}{\partial x} \qquad [\text{ when } \tfrac{\partial V}{\partial t} = 0$$

This is a special relativistic effect, not predicted by the Galilean transformation. It can be understood in terms of space contraction. The observer S could pick two locations where the potential differs by some given amount $\Delta V = 1$ unit, say, and paint a red mark at each location, or place a stick extending from one location to the other. This is possible because V is independent of time in S. Suppose the marks are separated by 1 metre according to S (or the stick is 1 metre long in S). Any other observer S$'$ must agree that the potential at the first red mark differs from that at the other red mark by $\Delta V = 1$ unit, assuming that we are dealing with a Lorentz invariant scalar field. However, such an observer moving with respect to S must find that the two red marks are separated by a *smaller* distance (contracted by γ). He must conclude that the gradient is larger than 1 unit per metre by the Lorentz factor γ.

Similarly, when V depends on time but not position in S, then its rate of change in another reference frame is larger than $\partial V/\partial t$ owing to time dilation.

In classical mechanics, we often take an interest in the gradient of potential energy or of electric potential. You should beware however that potential energy is not Lorentz invariant, and neither is electric potential, so an attempt to calculate a 4-gradient of either of them is misconceived[1]. Instead they are each part of a 4-vector, and one may take an interest in the 4-divergence or 4-curl of the associated 4-vector. The definition of 4-divergence of a 4-vector field F is what one would expect:

$$\Box \cdot \mathsf{F} \equiv \Box^T (g\mathsf{F}) = \frac{1}{c}\frac{\partial \mathsf{F}^0}{\partial t} + \boldsymbol{\nabla} \cdot \mathbf{f} \tag{5.5}$$

where $\mathbf{f}$ is the spatial part of F (i.e. $\mathsf{F} = (\mathsf{F}^0, \mathbf{f})$). Note that the minus sign in the definition of $\Box$ combines with the minus sign in the scalar product (from the metric g) to produce plus signs in (5.5).

The 4-dimensional equivalent of curl is more complicated and will be discussed in chapter 9.

As an example, you should check that the 4-divergence of the spacetime displacement $\mathsf{X} = (ct, \mathbf{r})$ is simply

$$\Box \cdot \mathsf{X} = 4. \tag{5.6}$$

Example. (i) If ϕ and V are scalar fields (i.e. Lorentz scalar quantities that may depend on position and time), show that

$$\Box(\phi V) = V \Box \phi + \phi \Box V.$$

Answer. Consider first of all the time component:

$$\frac{1}{c}\frac{\partial}{\partial t}(\phi V) = \frac{1}{c}\left(\frac{\partial \phi}{\partial t} V + \phi \frac{\partial V}{\partial t}\right)$$

which is the time component of $V\Box\phi + \phi\Box V$. Proceeding similarly with all the other components (paying attention to the signs), the result is soon proved.

(ii) If ϕ is a scalar field and F is a 4-vector field (i.e. a 4-vector that may depend on position and time), prove that

$$\Box \cdot (\phi \mathsf{F}) = \mathsf{F} \cdot \Box \phi + \phi \Box \cdot \mathsf{F}.$$

Answer. This is just like the similar result for $\boldsymbol{\nabla} \cdot (\phi \mathbf{f})$ and may be proved similarly, by proceeding one partial derivative at a time (or by reference to chapter 9).

[1] This does not rule out that one could introduce a Lorentz scalar field Φ with the dimensions of energy, as a theoretical device, for example to model a 4-force by $-\Box\Phi$; such a force would be impure. An example is the scalar meson theory of the atomic nucleus, considered in chapter 18.

5.3 Current density, continuity

The general pattern with 4-vectors is that a scalar quantity appears with a 'partner' vector quantity. So far, examples have included time with spatial displacement, speed of light with particle velocity, energy with momentum. Once one has noticed the pattern it becomes possible to guess at further such 'partnerships'. Our next example is density and flux.

The density ρ of some quantity is the amount per unit volume, and the *flux* or current density $\mathbf{j}$ is a measure of flow, defined as 'amount crossing a small area, per unit area per unit time.' If the quantity under consideration is conserved (think of a flow of water, for example, or of electric charge), then the amount present in some closed region of space can only grow or shrink if there is a corresponding net flow in or out across the boundary of the region. The mathematical expression of this is

$$\frac{d}{dt}\int_R \rho\, dV = -\int_R \mathbf{j}\cdot d\mathbf{S} \tag{5.7}$$

where R signifies some closed region of space, the integral on the left is over the volume of the region, and the integral on the right is over the surface of the region. The minus sign is needed because by definition, in the surface integral, $d\mathbf{S}$ is taken to be an outward-pointing vector so the surface integral represents the net flow *out* of R. By applying Gauss's theorem, and arguing that the relation holds for all regions R, one obtains the *continuity equation*

$$\frac{d\rho}{dt} + \boldsymbol{\nabla}\cdot\mathbf{j} = 0. \tag{5.8}$$

This equation is reminiscent of the 4-divergence equation (5.5). Indeed if we tentatively conjecture that $(\rho c, \mathbf{j}) = \mathsf{J}$ is a 4-vector, then we can write the continuity equation in the covariant form

$$\Box\cdot\mathsf{J} = 0. \qquad [\text{ Continuity equation} \tag{5.9}$$

This is quite correct because $(\rho c, \mathbf{j})$ *is* a 4-vector. Let's see why.

We shall consider the question of flow for some arbitrary conserved quantity that we shall simply call 'particles'. The particles could be water molecules, in the case of a flow of water, or charge carriers in the case of electric charge, or the charge itself if the carriers are not conserved but the charge is (the question of two different signs for charge is easily kept in the account and will not be explicitly indicated in the following). We will allow ourselves to take the limit where the flow is continuous, like that of a continuous fluid, but using the word 'particles' helps to keep in mind that we want to be able to talk about the flow of a Lorentz-invariant quantity. For particles one can simply count the number of worldlines crossing some given 3-surface in spacetime; since this is merely a matter of counting it is obviously Lorentz invariant if 'particles' are not being created or destroyed.

Suppose some such particles are distributed throughout a region of space. In general the particles might move with different velocities, but suppose the velocities are smoothly distributed, not jumping abruptly from one value to another for neighbouring particles. Then in any small enough region, the particles in it all have the same velocity. Then we can speak of a rest frame for that small region. We define the rest number density ρ_0 to be the number of particles per unit volume in such a rest frame. ρ_0 can be a function of position and time, but note that by definition it is Lorentz invariant. It earns its Lorentz invariant status in just the same way that proper time does: it comes with reference frame 'pre-attached'. Now define

$$\mathsf{J} \equiv \rho_0 \mathsf{U} \tag{5.10}$$

where U is the four-velocity of the fluid at any time and position. Clearly J is a 4-vector because it is the product of an invariant and a 4-vector. We shall now show that, when defined this way, J will turn out to be equal to $(\rho c, \mathbf{j})$.

In the local rest frame, we have simply $\mathsf{J} = (\rho_0 c, \mathbf{0})$. If we pass from the rest frame to any other frame, then, by the Lorentz transformation, the zeroth component of J changes from $\rho_0 c$ to $\gamma \rho_0 c$. This is equal to ρc where ρ is the density in the new frame, because any given region of the rest frame (containing a fixed number of particles) will be Lorentz-contracted in the new frame, so that its volume is reduced by a factor γ, so the number per unit volume in the new frame is higher by that factor. Let $\mathbf{u}$ be the local flow velocity in the new frame. Then the flux is given by $\mathbf{j} = \rho \mathbf{u}$. It is obvious that this u is also the relative speed of the new frame and the local rest frame, so

$$\mathbf{j} = \rho \mathbf{u} = \gamma_u \rho_0 \mathbf{u}. \tag{5.11}$$

But this is just the spatial part of $\rho_0 \mathsf{U}$. Since we can use such a Lorentz transformation from the rest frame to connect ρ_0 and U to ρ and $\mathbf{j}$ for any part of the fluid, we have proved in complete generality that

$$\rho_0 \mathsf{U} = (\rho c, \mathbf{j}). \tag{5.12}$$

Hence $(\rho c, \mathbf{j})$ is a 4-vector as we suspected.

What we have gained from all this is some practice at identifying 4-vectors, and a useful insight into the continuity equation (5.9). Because the left hand side can be written as a scalar product of a 4-vector-operator and a 4-vector, it must be Lorentz invariant. Therefore the whole equation relates one invariant to another (zero). Therefore if the continuity equation is obeyed in one reference frame, then it is obeyed in all.

The continuity equation is a statement about conservation of particle number (or electric charge etc.). The 4-flux J is not itself conserved, but its null 4-divergence shows the conservation of the quantity whose flow it expresses. The conserved quantity is here a Lorentz scalar. This is in contrast to energy-momentum where the conserved quantity was the set of all components of a 4-vector. The latter can be treated by writing the divergence of a higher-order quantity called the stress-energy tensor—something we will do in chapter 12.

5.4 Wave motion

A plane wave (whether of light or of anything else, such as sound, or oscillations of a string, or waves at sea) has the general form

$$a = a_0 \cos(\mathbf{k} \cdot \mathbf{r} - \omega t) \tag{5.13}$$

where a is the displacement of the oscillating quantity (electric field component; pressure; height of a water wave; etc.), a_0 is the amplitude, ω the angular frequency and $\mathbf{k}$ the wave vector. As good relativists, we suspect that we may be dealing with a scalar product of two 4-vectors:

$$\mathsf{K} \cdot \mathsf{X} = (\omega/c, \mathbf{k}) \cdot (ct, \mathbf{r}) = \mathbf{k} \cdot \mathbf{r} - \omega t. \tag{5.14}$$

Let's see if this is right. That is, does the combination $(\omega/c,\, \mathbf{k})$ transform as a 4-vector under a change of reference frame?

A nice way to see that it does is simply to think about the phase of the wave,

$$\phi = \mathbf{k} \cdot \mathbf{r} - \omega t. \tag{5.15}$$

To this end we plot the wavefronts on a spacetime diagram. Figure 5.1 shows a set of wavefronts of a wave propagating along the positive x axis of some frame S. Be careful to read the diagram correctly: the whole wave appears 'static' on a spacetime diagram, and the lines represent the locus of a mathematically defined quantity. For example, if we plot the wave crests then we are plotting those events where the displacement a is at a maximum. For plane waves in one spatial dimension, each such locus is a line in spacetime. Note also that, because the phase velocity ω/k can be either smaller, equal to, or greater than the speed of light, a wavecrest locus (='ray') in spacetime can be either timelike, null, or spacelike.

One may plot the wavecrests in the first instance from the point of view of one particular reference frame (each line then has the equation $\omega t = kx + \phi$). However, a maximum excursion is a maximum excursion: all reference frames will agree on those events where the displacement is maximal, even though the amplitude (a_0 or a'_0) may be frame-dependent. It follows that the

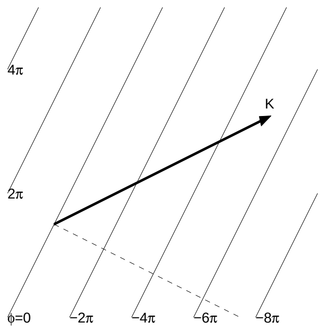


Figure 5.1: Wavefronts (surfaces of constant phase) in spacetime. It is easy to get confused by this picture, and imagine that it shows a snapshot of wavefronts in space. It does not. It shows the complete propagation history of a plane wave moving to the right in one spatial dimension. By sliding a spacelike slot up the diagram you can 'watch' the wavefronts march to the right as time goes on in your chosen reference frame (each wavefront will look like a dot in your slot). The purpose in showing this diagram is to press home the point that the set of events at some given value of phase, such as at a maximal displacement of the oscillating medium, is not frame-dependent. It is a given set of events in spacetime. Therefore ϕ is a scalar invariant. The direction of the wave 4-vector K may be constructed by drawing a vector in the direction down the phase gradient (shown dotted), and then changing the sign of the time component. The waves shown here have a phase velocity less than c. (For light waves in vacuum the wavefronts and the wave 4-vector are both null, i.e. sloping at 45° on such a diagram.) The wavelength λ in any given reference frame is indicated by the distance between events where successive wavecrest lines cross the position axis (line of simultaneity) of that reference frame. The period T is the time interval between events where successive wavecrest lines cross the time axis of the reference frame.

wavecrest locations are Lorentz invariant, and more generally so is the phase ϕ, because the Lorentz transformation is linear, so all frames agree on how far through the cycle the oscillation is between wavecrests.

We can now obtain K as the gradient of the phase:

$$\begin{aligned} \mathsf{K} &= \Box\phi = \left(-\frac{1}{c}\frac{\partial}{\partial t}, \boldsymbol{\nabla}\right)\phi \\ &= (\omega/c, \mathbf{k}), \end{aligned} \tag{5.16}$$

using (5.15). Since this is a 4-gradiant of a Lorentz scalar, it is a 4-vector.

Writing $v_{\rm p}$ for the phase velocity ω/k, we find the associated invariant

$$\mathsf{K}^2 = \omega^2\left(\frac{1}{v_{\rm p}^2} - \frac{1}{c^2}\right). \tag{5.17}$$

Therefore when $v_p < c$ the 4-wave-vector is spacelike, and when $v_p > c$ the 4-wave-vector is timelike. For light waves in vacuum the 4-wave-vector is null. The invariant also shows that a wave of any kind whose phase velocity is c in some reference frame will have that same phase velocity in all reference frames.

5.4.1 Wave equation

Wave motion such as that expressed in eq. (5.13) is understood mathematically as a solution of the wave equation

$$\frac{\partial^2 a}{\partial t^2} = v_{\rm p}^2\nabla^2 a. \tag{5.18}$$

Writing this

$$-\frac{1}{c^2}\frac{\partial^2 a}{\partial t^2} + \frac{v_{\rm p}^2}{c^2}\nabla^2 a = 0 \tag{5.19}$$

we observe that for the special case $v_p = c$ the wave equation takes the Lorentz covariant form

$$\Box^2 a = 0. \qquad [\text{ Wave equation!} \tag{5.20}$$

The operator is called the *d'Alembertian*[2]:

$$\Box^2 \equiv \Box \cdot \Box = -\frac{1}{c^2}\frac{\partial^2}{\partial t^2} + \nabla^2 \tag{5.21}$$

(a product of three minus signs made the minus sign here!). Hence the general idea of wave propagation can be very conveniently treated in Special Relaivity when the waves have phase velocity c. This will be used to great effect in the treatment of electromagnetism in chapter 6.

5.4.2 Particles and waves

While we are considering wave motion, let's briefly look at a related issue: the wave–particle duality. We will not try to introduce that idea with any great depth, that would be the job of another textbook, but it is worth noticing that the introduction of the photon model for light can be guided by Special Relativity.

Max Planck is associated with the concept of the photon, owing to his work on the Black Body radiation. However, when he introduced the idea of energy quantisation, he did not in fact have in mind that this should serve as a new model for the electromagnetic field. It was sufficient for his purpose merely to assert that energy was absorbed by matter in quantised 'lumps'. It was Einstein who extended the notion to the electromagnetic field itself, through his March 1905 paper. This paper is often mentioned in regard to the photoelectric effect, but this does not do justice to its full significance. It was a revolutionary re-thinking of the nature of electromagnetic radiation.

When teaching students about the photoelectric effect and its impact on the development of quantum theory, it makes sense, and it is the usual practice, to emphasize that the *energy* of the emitted electrons has no dependence on the *intensity* of the incident light. Rather, the energy depends linearly on the frequency of the light, while the light intensity influences the rate at which photo-electrons are generated. This leads one to propose the model $E = h\nu$ relating the energy of the light particles to the frequency of the waves.

However, this data was not available in 1905. There was evidence that the electron energy did not depend on the intensity of the light, and for the existence of a threshold frequency, but the linear relation between photoelectron energy and light frequency was *predicted* in Einstein's paper: it was not extracted from experimental data. Einstein's paper relied chiefly on arguments from thermodynamics and what we now call statistical mechanics: he calculated the entropy per unit volume of thermal radiation and showed that the thermodynamic behaviour of the

[2]Beware: many authors now use the symbol $\Box$ (without the ^{2}) for the d'Alembertian; they then have to use some other symbol for the 4-gradiant. Very confusing! To be fair, there is a reason: it is to make the index notation, to be introduced in chapter 9, more consistent. I have adopted a notation I believe to be the least confusing for learning purposes. Also, the d'Alembertian is often defined as $c^2\partial^2/\partial t^2 - \nabla^2$ (the negative of our $\Box^2$).

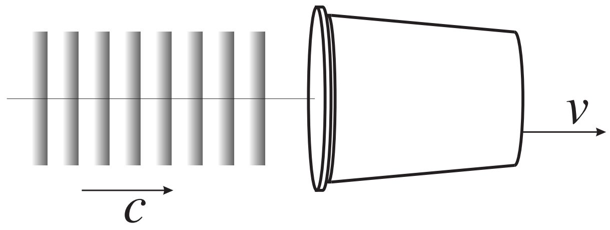


Figure 5.2: A parallel beam of light falls into a moving bucket.

radiation at a given frequency ν was the same as that of a gas of particles each carrying energy $h\nu$. The relationship $E = h\nu$ as applied to what we now call photons was thus first proposed by Einstein. However, his 1905 paper was still far short of a full model; it was not until Compton's experiments (1923) that the photon idea began to gain wide acceptance, and a thorough model required the development of quantum field theory, the work of many authors, with Dirac (1927) playing a prominent role.

In this section we shall merely point out one feature (which is not the one historically emphasized in 1905): *if one is going to attempt a particle model for electromagnetic waves, then Special Relativity can guide you on how to do it.* That is, we shall play the role of theoretical physicist, and assume merely that we know about classical electromagnetism and we would like to investigate what kind of photon model might be consistent with it.

Consider a parallel beam of light falling on a moving bucket (figure 5.2). We shall use this situation to learn about the way the energy and intensity of light transform between reference frames. In fact we already made a general observation about this in the discussion of the headlight effect in section 3.7.2, in connection with eq (3.74). The present discussion will proceed more cautiously, and thus exhibit the reasoning underlying (3.74).

Suppose that in frame S the light and the bucket move in the same direction, with speeds c and v respectively. Let u be the energy per unit volume in the light beam. The amount of energy flowing across a plane fixed is S of cross section A during time t is then $uA(ct)$. The 'intensity' I (or flux) is defined to be the power per unit area, so

$$I = uc. \tag{5.22}$$

We would like to calculate the amount of energy entering the bucket, and compare this between reference frames. To this end it is convenient to use the Lorentz invariance of the phase of the wave. We consider the energy and momentum that enters the bucket during a period when N wavefronts move into the bucket. In frame S these waves fill a total length $L = N\lambda$ where λ is the wavelength, so the energy entering the bucket is $E = N\lambda Au$. In the rest frame S$'$ of the bucket, between the events when the first and last of these wavefronts entered, the energy coming in must be

$$E' = N\lambda' A u' \tag{5.23}$$

since N is invariant and A is a transverse (therefore uncontracted) area. We argue that the portion or 'lump' of the light field now in the bucket (we can suppose the bucket is deep so the light has not been absorbed yet) can be considered to possess energy E and be propagating at speed c. It follows that its momentum must be $p = E/c$. Note, we have not invoked a particle model in order to assert this, we have merely claimed that the relation $p/E = v/c^2$, which we know to be valid for $v < c$, is also valid in the limit $v = c$. (We shall show in chapter 12 that electromagnetic field theory also confirms $p = E/c$ for light waves.) Applying a Lorentz transformation to the energy-momentum of the light, we obtain for the energy part:

$$E' = \gamma(1-\beta)E = \sqrt{\frac{1-\beta}{1+\beta}}E. \tag{5.24}$$

By Lorentz transforming the 4-wave-vector $(\omega/c, \mathbf{k})$, or by using the Doppler effect formula (22.7), we obtain for the wavelength

$$\lambda' = \sqrt{\frac{1+\beta}{1-\beta}}\lambda.$$

Substituting these results into (5.23) we find

$$\begin{aligned} u' &= \frac{E'}{N\lambda' A} = \frac{1-\beta}{1+\beta}\frac{E}{N\lambda A} \\ \Rightarrow \qquad I' &= \frac{1-\beta}{1+\beta} I \end{aligned} \tag{5.25}$$

where the last step uses (5.22).

Two things are striking in this argument. First, the energy of the light entering the bucket transforms in the same way as its frequency. Second, the energy does *not* transform in the same way as the intensity. When making an approach to a particle model, therefore, although one might naively have guessed that the particle energy should be connected to the intensity of the light, we see immediately that this will not work: it cannot be true in all reference frames for a given set of events. For, just as the number of wavefronts entering the bucket is a Lorentz invariant, so must the number of particles be: those particles could be detected and counted, after all, and the count displayed on the side of the bucket. Therefore the energies E and E' that we calculated must correspond to the *same* number of particles, so they are telling us about the energy per particle.

One will soon run into other difficulties with a guess that the particle energy is proportional to $\sqrt{I}$ or to $I\lambda$. It seems most natural to try $E \propto \nu$, the frequency. Indeed, with the further consideration that we need a complete energy-momentum 4-vector for our particle, not just a scalar energy, and we have to hand the 4-wave-vector of the light with just the right direction in spacetime (i.e. the null direction), it is completely natural to guess the right model, $E = h\nu$ and $\mathsf{P} = \hbar\mathsf{K}$.

5.4.3 Group velocity and particle velocity

Recall equations (3.61) and (3.70) for the angle change of the velocity of a particle and the wave-vector of a plane wave, respectively. We reproduce these here for convenience:

$$\tan\theta = \frac{\sin\theta_0}{\gamma(\cos\theta_0 + v v_{\rm p}/c^2)}, \tag{5.26}$$

$$\tan\theta = \frac{\sin\theta_0}{\gamma(\cos\theta_0 + v/u)}, \tag{5.27}$$

where the frames are labelled S and S_0, $v_{\rm p} \equiv \omega_0/k_0$ is the phase velocity of the waves in the frame S_0, and u is the speed of the particle in the frame S_0. These are both examples of a direction-change of a 4-vector, so they amount to the same formula: the first can be obtained from the second by the replacement $u \to (k_0/\omega_0)c^2$. However, the result is that a particle travelling along at the phase velocity of the waves (i.e. having the same speed and direction) in frame S_0 does not have in general have the same speed or direction as the phase velocity in frame S.

Something interesting emerges if we look at *group velocity*. The group velocity of a set of waves is defined

$$v_{\rm g} \equiv \frac{{\rm d}\omega}{{\rm d}k}. \tag{5.28}$$

Thus the group velocity depends on the way the frequency of the waves is related to their wavevector. There is no general formula for this, because it depends on the particular conditions, such as the behaviour of the refractive index for light waves in a transparent medium, or the dispersion relation for sound waves, etc. However, an interesting case to consider is waves that have the property that $\mathsf{K}\cdot\mathsf{K}$ is independent of k. Note, this does not necessarily have to happen: $\mathsf{K}\cdot\mathsf{K}$ is guaranteed to be Lorentz-invariant, but its value might in general be a function of frequency. However, if it does not depend on frequency then we have

$$-\omega^2/c^2 + k^2 = \text{const.}$$

After multiplying by c^2 and taking the derivative with respect to k, we obtain

$$v_{\rm g} = \frac{{\rm d}\omega}{{\rm d}k} = \frac{kc^2}{\omega} = \frac{c^2}{v_{\rm p}}. \tag{5.29}$$

The group velocity is in the same direction as the phase velocity, but has a different size (it is less than c if the phase velocity is greater than c).

Now consider a particle whose speed and direction u, θ_0 in frame S_0 matches that of the *group* velocity of a set of waves. Then we have $u = c^2/v_{\rm p}$. Substituting this into (5.27) we find that now the change in direction of the particle motion matches that of the wave motion. Also, by using eq. (3.60) one can show that the size of the speed follows the size of the group velocity as long as $p \propto k$ and $E \propto \omega$. This is the relationship between 4-momentum and 4-wave-vector that appears in the quantum mechanical treatment of particles in terms of de Broglie waves. Those waves satisfy the condition $\mathsf{K} \cdot \mathsf{K} =$ const, the constant in question being related to the rest-mass of the particle. Hence the wave-particle duality continues to make sense in a relativistic treatment, and we deduce that the speed of the particles should be understood as given by the group velocity (not the phase velocity) of the waves.

5.5 Acceleration and rigidity

Consider a stick that accelerates as it falls. For the sake of argument, suppose that in some reference frame S(x, y, z) a stick is extended along the x direction, and remains straight at all times. It accelerates in the y direction all as a piece (without bending) at constant acceleration a in S. The worldline of any particle of the stick is then given by

$$x = x_0 \tag{5.30}$$

$$y = \frac{1}{2}at^2 \tag{5.31}$$

during some interval for which $t < c/a$, where x_0 takes values in the range $-L_0/2$ to $L_0/2$ where L_0 is the rest length of the stick.

Now consider this stick from the point of view of a reference frame moving in the x direction (relative to S) at speed v. In the new frame the coordinates of a particle on the stick are given by

$$\begin{pmatrix} ct' \\ x' \\ y' \\ z' \end{pmatrix} = \begin{pmatrix} \gamma & -\gamma\beta & 0 & 0 \\ -\gamma\beta & \gamma & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} ct \\ x_0 \\ \frac{1}{2}at^2 \\ 0 \end{pmatrix} = \begin{pmatrix} \gamma(ct - \beta x_0) \\ \gamma(x_0 - vt) \\ \frac{1}{2}at^2 \\ 0 \end{pmatrix} \tag{5.32}$$

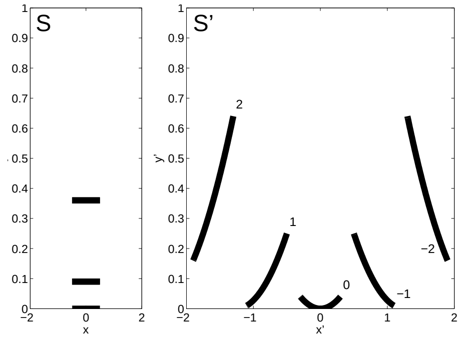


Figure 5.3: A rigid stick that remains straight and parallel to the x axis in frame S (left diagram), is here shown at five successive instants in frame S$'$ (right diagram). The stick has an initial velocity in the downwards (-ve y) direction and accelerates in the +ve y direction; frame S$'$ moves to the right (+ve x direction) at speed v relative to frame S. (In the example shown the stick has proper length 1, $v = 0.8$, and $a = 2$, all in units where $c = 1$.).

Figure 5.4: Spacetime diagram showing the worldsheet of the stick shown in figure 5.3.

Use the first line in this vector equation to express t in terms of t', obtaining $t = (t'/\gamma + \beta x_0/c)$, and substitute this into the rest:

$$\begin{pmatrix} ct' \\ x' \\ y' \\ z' \end{pmatrix} = \begin{pmatrix} ct' \\ x_0/\gamma - vt' \\ \frac{1}{2}a(\frac{t'}{\gamma} + \frac{\beta x_0}{c})^2 \\ 0 \end{pmatrix}. \tag{5.33}$$

When we allow x_0 to take values between 0 and L_0, this equation tells us the location in S' of the all the particles of the stick, at any given t'. It is seen that they lie along a parabola. Figure 5.3b shows the stick in frame S' at five successive values of t', and 5.4 shows the spacetime diagram.

This example shows that accelerated motion while maintaining a fixed shape in one reference frame will result in a changing shape for the object in other reference frames. This is because the worldlines of the particles of the object are curved, and the planes of simultaneity for most reference frames must intersect such a set of worldlines along a curve. This means that, for accelerating objects, the concept of 'rigid' behaviour is not Lorentz-invariant. The notion of 'remaining undeformed' cannot apply in all reference frames when a body is accelerating (see the exercises for further examples).

A related issue is the concept of a 'rigid body'. In classical physics this is a body which does not deform when a force is applied to it; it accelerates all of a piece. In Special Relativity this concept has to be abandoned. There is no such thing as a rigid body, if by 'rigid' we mean a body that does not deform when struck. This is because when a force is applied to one part of a body, only that part of the body is causally influenced by the force. Other parts, outside the future light cone of the event at which the force began to be applied, cannot possibly be influenced, whether to change their motion or whatever. It follows that the application of a force to one part of body *must* result in deformation of the body. Another way of stating this is to say that a rigid body is one for which the group velocity of sound goes to infinity, but this is ruled out by the Light Speed Postulate.

There can exist accelerated motion of a special kind, such that the different parts of a body move in synchrony so that proper distances are maintained. Such a body can be said to be 'rigid' while it accelerates. This is described in chapter 14.

5.5.1 The great train disaster[3]

Full fathom five thy father lies,
Of his bones are coral made:
Those are pearls that were his eyes,
Nothing of him that doth fade,
But doth suffer a sea-change
Into something rich and strange.
Sea-nymphs hourly ring his knell:
Hark! now I hear them, ding-dong, bell.

(Ariels's song from *The Tempest* by William Shakespeare)

The relativity of the shape of accelerated objects is nicely illustrated by a paradox in the general family of the stick and the hole (see for example *The Wonderful World*). Or perhaps, now that we understand relativity moderately well (let's hope), it is not a paradox so much as another fascinating example of the relativity of simultaneity and the transformation of force.

So, imagine a super train, 300 m long (rest length), that can travel at about 600 million miles per hour, or, to be precise, $\sqrt{8}c/3$. The train approaches a chasm of width 300 m (rest length) which is spanned by a bridge made of three suspended sections, each of rest length 100 m, see figure **??**. Owing to its Lorentz contraction by a factor $\gamma = 3$, the whole weight of the train has to be supported by just one section of the bridge. Unfortunately the architect has forgotten to take this into account: the cable snaps, the bridge section falls, and the train drops into the chasm.

At this point the architect arrives, both shocked and perplexed.

"But I did take Lorentz contraction into account," he says. "In fact, in the rest frame of the train, the chasm is contracted to 100 m, so the train easily extends right over it. Each section

[3] This is loosely based on the discussion by Fayngold.

of the bridge only ever has to support one ninth of the weight of the train. I can't understand why it failed, and I certainly can't understand how the train could fall down because it will not fit into such a short chasm."

Before we resolve the architect's questions, it is only fair to point out that the example is somewhat unrealistic in that, on all but the largest planets, the high speed of the train will exceed the escape velocity. Instead of pulling the train down any chasm, an ordinary planet's gravity would not even suffice to keep such a fast train on the planet surface: the train would continue in an almost straight line, moving off into space as the planet's surface curved away beneath it. A gravity strength such as that near the event horizon of a black hole would be needed to cause the crash. However, we could imagine that the train became electrically charged by rubbing against the planet's (thin) atmosphere and the force on it is electromagnetic in origin, then a more modest planet could suffice.

In any case, it is easy to see that whatever vertical force f' acted on each particle of the train in the rest frame of the planet, the force per particle in the rest frame of the train is considerably larger: $f = \gamma f' = 3f'$ (see eq. (4.6)). The breaking strength of the chain would appear to be something less than Nf' in the planet frame, where N is the number of particles in the whole of the train. In the train frame, this breaking strength will be reduced (says eq. (4.6)) to something less than $Nf'/\gamma = Nf'/3$. Hence, in the train frame, the number of particles n the bridge section can support is given by

$$n\gamma f' < Nf'/\gamma \qquad \Rightarrow \qquad n < N/\gamma^2. \tag{5.34}$$

It is not suprising, therefore, that the chain should break when only one ninth of the length of the train is on the bridge section (in the train's rest frame).

The architect's second comment is that the train in its rest frame will not fit horizontally into the chasm. This is of course true. However, by using the Lorentz transformation it is easy to construct the trajectories of all the particles of the train, and they are as shown in figure **??**. The vertical acceleration of the falling train is Lorentz-transformed into a bending downwards. The train which appeared 'rigid' in the planet frame is revealed in the horizontally moving frame to be as floppy as a snake as it plunges headlong through the narrow gap in the bridge. The spacetime story is encapsulated by the spacetime diagram shown in figure **??**.

5.5.2 Lorentz contraction and internal stress

The Lorentz contraction results in distortion of an object. The contraction is purely that: a contraction, not a rotation, but a contraction can change angles as well as distances in solid objects. For example, a picture frame that is square in its rest frame will be a parallelogram at any instant of time in most other inertial reference frames. The legs of a given ordinary table are not at right angles to its surface in most inertial reference frames.

For accelerating bodies, the change of shape associated with a change of inertial reference frame is more extreme. Examples are the twisted cylinder (see exercise **??**), and the falling stick or the train of the previous section. Things that accelerate can suffer a Lorentz-change into something rich and strange.

These observations invite the question, are these objects still in internal equilibrium, or are they subject to internal stresses? What is the difference between Lorentz contraction and the distortion that can be brought about by external forces?

John Bell proposed the following puzzle. Suppose two identical rockets are at rest relative to a space station, one behind the other, separated by $L = 100$ m. That is, in the rest frame S of the space station, the tail of the front rocket is $L = 100$ m in front of the tip of the back rocket. They are programmed to blast off simultaneously in the inertial reference frame S, and thereafter to burn fuel at the same rate. It is clear that the trajectory of either rocket will be identical in S, apart from the 100 m gap. In other words if the tail of the front rocket moves as $x(t)$ then the tip of the back rocket moves as $x(t) - L$. Therefore their separation remains 100 m in reference frame S.

Now suppose that before they blast off, a string of rest length $L_0 = 100$ m is connected between the rockets (as if one rocket were to tow the other, but they are still both provided with working engines), and suppose any forces exerted by the string are negligible compared to those provided by the rocket engines. Then, in frame S the string will suffer a Lorentz contraction to less than 100 m, but the rockets are still separated by 100 m. So what will happen? Does the string break?

I hope it is clear to you that the string will eventually break. It undergoes acceleration owing to the forces placed on it by the rockets. It will in turn exert a force on the rockets, and its Lorentz contraction means that that force will tend to pull the rockets together to a separation smaller than 100 m in frame S. This means that it begins to act as a tow rope. The fact that its length remains (very nearly) constant at 100 m in S, whereas it 'ought to' be L_0/γ, shows us that the engine of the rear rocket is not doing enough to leave the tow rope nothing to do: the tow rope is being stretched by the external forces. The combination of this stretching and the Lorentz contraction results in the observed constant string length in frame S.

Such a string is not in internal equilibrium. It will only be in internal equilibrium, exerting no outside forces, if it attains the length L_0/γ. As the rockets reach higher and higher speed relative to S, γ gets larger and larger, so the string is stretched more and more relative to its equilibrium length. If you need to be further convinced of this, then jump aboard the rest frame of the front rocket at some instant of time, and you will find the back rocket is trailing behind by considerably more than 100 m. At some point the material of the string cannot withstand further stretching, and the string breaks.

In the study of springs and Hooke's law, we say the length of a spring when it exerts zero force is called its 'natural' length. In Special Relativity, we call the length of a body in the rest frame of the body its 'proper' length: you might say this is the length that it 'thinks' it has. The proper length is, by definition, a Lorentz invariant. The natural length depends on reference frame however. The proper length does not have to be equal to the natural length.

A spring with no external forces acting on it, and for which any oscillations have damped away, will have its natural length. Suppose that length is $L_{\rm n}(0)$ in the rest frame of the spring. In inertial reference frames moving relative to the spring in a direction along its length, the natural length will be $L_{\rm n}(v) = L_{\rm n}(0)/\gamma$.

We now have three lengths to worry about: the length L that a body actually has in any given reference frame, its natural length $L_{\rm n}(v)$ in that reference frame, and its proper length L_0. The Lorentz contraction affects the length and the natural length. A stretched or compressed spring has a length in any given reference frame different from its natural length in that reference frame. Its proper length is $L_0 = \gamma L$. If $L \neq L_{\rm n}(v)$ then $L_0 \neq L_{\rm n}(0)$, i.e. a stretched or compressed spring has a proper length different from what the natural length would be in its rest frame.

In the example of the rockets joined by a string, in reference frame S the natural length of the string become shorter and shorter, but the string did not become shorter. In the sequence of rest frames of the centre of the string (i.e. each one is an inertial frame having the speed momentarily possessed by the string) the string's natural length was constant but its actual length became longer.

If a moving object is abruptly stopped, so that all of its parts stop at the same time in a reference frame other than the rest frame, then the length in that frame remains constant but the proper length gets shorter (it was γL, now it is L). If the object was previously moving freely with no internal stresses then now it will try to expand to its new natural length, but it has been prevented from doing so. Therefore it now has internal stresses: it is under compression.

Similarly, if an object having no internal stresses is set in motion so that all parts of the object get the same velocity increase at the same time in the initial rest frame S, then the length of the object in S stays constant while the proper length gets longer (it was L, now it is γL). Since the proper length increases, such a procedure results in internal stresses such that the object is now under tension.

More generally, to discover whether internal stresses are present, it suffices to discover whether the distance between neighbouring particles of a body is different from the natural distance. In the example of the great train disaster, the train is without internal stress as it bends during the free fall. The passengers too are without internal stress—except the pyschological kind, of course. If the natural (unstressed) shape of an accelerating object remains straight in some inertial reference frame, then in most other reference frames the natural (unstressed) shape will be bent.

5.6 General Lorentz boost

So far we have considered the Lorentz transformation only for a pair of reference frames in the standard configuration, where it has the simple form presented in eq. (3.25). More generally, inertial reference frames can have relative motion in a direction not aligned with their axes, and they can be rotated or suffer reflections with respect to one another. To distinguish these possibilities, the transformation for the case where the axes of two reference frames are mutually aligned, but they have a non-zero relative velocity, is called a *Lorentz boost*. A more general transformation, involving a rotation of coordinate axes as well as a relative velocity, is called a Lorentz transformation but not a boost.

The most general Lorentz boost, therefore, is for the case of two reference frames of aligned axes, whose relative velocity $\mathbf{v}$ is in some arbitrary direction relative to those axes. In order to obtain the matrix representing such a general boost, it is instructive to write the simpler case given in (3.25) in the vector form

$$\begin{aligned} ct' &= \gamma(ct - \boldsymbol{\beta}\cdot\mathbf{x}) \\ \mathbf{x}' &= \mathbf{x} + \left(-\gamma ct + \frac{\gamma^2}{1+\gamma}\boldsymbol{\beta}\cdot\mathbf{x}\right)\boldsymbol{\beta} \end{aligned} \tag{5.35}$$

This gives a strong hint that the general Lorentz boost is

$$\mathcal{L}(\mathbf{v}) = \begin{pmatrix} \gamma & -\gamma\beta_x & -\gamma\beta_y & -\gamma\beta_z \\ . & 1+\frac{\gamma^2}{1+\gamma}\beta_x^2 & \frac{\gamma^2}{1+\gamma}\beta_x\beta_y & \frac{\gamma^2}{1+\gamma}\beta_x\beta_z \\ . & . & 1+\frac{\gamma^2}{1+\gamma}\beta_y^2 & \frac{\gamma^2}{1+\gamma}\beta_y\beta_z \\ . & . & . & 1+\frac{\gamma^2}{1+\gamma}\beta_z^2 \end{pmatrix} \tag{5.36}$$

where the lower left part of the matrix can be filled in by using the fact that the whole matrix is symmetric. One can prove that this matrix is indeed the right one by a variety of dull but thorough methods, see exercises.

5.7 Lorentz boosts and rotations

Suppose a large regular polygon (e.g. 1 km to the side) is constructed out of wood and laid on the ground. A pilot then flies an aircraft around this polygon (at some fixed distance above it), see figure 5.5a.

Let N be the number of sides of the polygon. As the pilot approaches any given corner of the polygon, he observes that the polygon is Lorentz-contracted along his flight direction. If, for

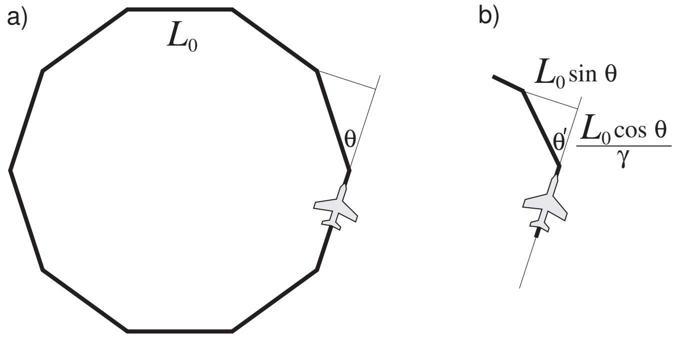


Figure 5.5: (a) A aircraft flies around a regular polygon. The polygon has N sides, each of rest length L_0. The angle between one side and the next, in the polygon rest frame, is $\theta = 2\pi/N$. (b) shows the local situation in the rest frame of the pilot as he approaches a corner and is about the make a turn through θ'. Since $\theta' > \theta$, the pilot considers that the sequence of angle turns he makes, in order to complete one circuit of the polygon, amount to more than 360°.

example, he considers the right-angled triangle formed by a continuation of the side he is on and a hypotenuse given by the next side of the polygon, then he will find the lengths of its sides to be $(L_0 \cos\theta)/\gamma$ and $L_0 \sin\theta$, where $\theta = 2\pi/n$, see figure 5.5b. He deduces that the angle he will have to turn through, in order to fly parallel to the next side, is θ' given by

$$\tan\theta' = \gamma \tan\theta. \tag{5.37}$$

Having made the turn, he can also consider the side receding from him and confirm that it makes this same angle θ' with the side he is now on.

For large N we have small angles, so

$$\theta' \simeq \gamma\theta. \tag{5.38}$$

After performing the manoeuvre N times, the aircraft has completed one circuit and is flying parallel to its original direction, and yet the pilot considers that he has steered through a total angle of

$$N\theta' = \gamma 2\pi. \tag{5.39}$$

Since $\gamma > 1$ we have a total steer by more than two pi radians, in order to go once around a circuit! The extra angle is given by

$$\Delta\theta = N\theta' - 2\pi = (\gamma - 1)2\pi. \tag{5.40}$$

This is a striking result. What is going on? Are the pilot's deductions faulty in some way? Perhaps something about the acceleration needed to change direction renders his argument invalid?

It will turn out that the pilot's reasoning is quite correct, but some care is required in the interpretation. The extra rotation angle is an example of a phenomenon called *Thomas rotation*. It is also often called *Thomas precession*, because it was first discovered in the context of a changing direction of an angular momentum vector. We will provide the interpretation and some more details in section 5.7.2. First we need a result concerning a simple family of three inertial reference frames.

5.7.1 Two boosts at right angles

Figure 5.6 shows a set of three reference frame axes, all aligned with one another at any instant of time in frame S'. Frame S'' is moving horizontally with respect to S' at speed v. Frame S is moving vertically with respect to S' at speed u. Let A be a particle at the origin of S, and B be a particle at the origin of S''.

We will calculate the angle between the line AB and the x-axis of S, and then the angle between AB and the x'-axis of S$''$. This will reveal an interesting phenomenon.

First consider the situation in S. Here A stays fixed at the origin, and B moves. We use the velocity transformation equations (3.20), noting that we have the simple case where the pair of velocities to be 'added' are mutually orthogonal. B has no vertical component of velocity in S$'$, so in S the vertical component of its velocity is $-u$. Its horizontal velocity in S$'$ is v, so in S its horizontal component of velocty is $v/\gamma(u)$. Therefore the angle θ between AB and the $x-$axis of frame S is given by

$$\tan\theta = \frac{u\gamma_u}{v}. \tag{5.41}$$

Now consider the situation in S$''$. Here B is fixed at the origin and A moves. The horizontal component of the velocity of A in S$''$ is $-v$, the vertical component is $u/\gamma(v)$. Therefore the angle θ'' between AB and the $x''-$axis of frame S$''$ is given by

$$\tan\theta'' = \frac{u}{\gamma_v v}. \tag{5.42}$$

Thus we find that $\theta \neq \theta''$. Since the three origins all coincide at time zero, the line AB is at all times parallel to the relative velocity of S and S''. This velocity is constant and it must be the

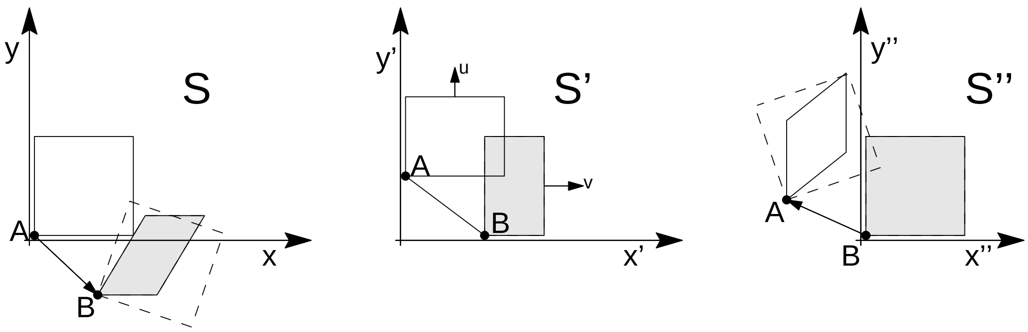


Figure 5.6: Two squares (i.e. each is a solid object that is square in its rest frame: it helps to think of them as physical bodies, not just abstract lines) of the same proper dimensions are in relative motion. In frame S′ the white square moves upwards at speed u and the grey square moves to the right at speed v. The central diagram shows the situation at some instant of time in S′: each square is contracted along its direction of motion. Frame S is the rest frame of the white square; S″ is the rest frame of the grey square. A and B are particles at the origins of S and S″ respectively. The left and right diagrams show the situation at some instant of time in S and S″ respectively. The reference frame axes of S and S″ have been chosen parallel to the sides of the fixed square in each case; those of S′ have been chosen parallel to the sides of both objects as they are observed in that frame. N.B. there are three diagrams here, not one! The diagrams have been oriented so as to bring out the fact that S and S′ are mutually aligned, and S′ and S″ are mutually aligned. However the fact that S and S″ are *not* mutually aligned is not directly indicated, it has to be inferred. The arrow AB on the left diagram indicates the velocity of S″ relative to S. The arrow BA on the right diagram indicates the velocity of S relative to S″. These two velocities are collinear (they are equal and opposite). The dashed squares in S and S″ show a shape that, if Lorentz contracted along the relative velocity AB, would give the observed parallelogram shape of the moving object in that reference frame. It is clear from this that the relationship between S and S″ is a boost combined with a rotation, not a boost alone. This rotation is the kinematic effect that gives rise to the Thomas precession. Take a long look at this figure: there is a lot here—it shows possibly the most mind-bending aspect of Special Relativity!

same (equal and opposite) when calculated in the two reference frames whose relative motion it describes. Therefore the interpretation of $\theta \neq \theta''$ must be that *the coordinate axes of S are* not *parallel to the coordinate axes of S''* (when examined either in reference frame S or in S'').

This is a remarkable result, because we started by stating that the axes of S and S'' are mutually aligned in reference frame S'. It is as if we attempted to line up three soldiers, with Private Smith aligned with Sergeant Smithers, and Sergeant Smithers aligned with Captain Smitherson, but somehow Private Smith is not aligned with Captain Smitherson. With soldiers, or lines purely in *space*, this would not be possible. What we have found is a property of constant-velocity motion in *spacetime*.

The sequence of passing from frame S to S' to S'' consists of two Lorentz boosts, but the overall result is not merely a Lorentz boost to the final velocity $\mathbf{w}$, but a boost combined with a rotation. Mathematically, this is

$$\mathcal{L}(-\mathbf{u})\mathcal{L}(\mathbf{v}) = \mathcal{L}(\mathbf{w})R(\Delta\theta) \tag{5.43}$$

where $\Delta\theta = \theta - \theta''$. We have proved the case where $\mathbf{u}$ and $\mathbf{v}$ are orthogonal. One can show that the pattern of this result holds more generally: a sequence of Lorentz boosts in different directions gives a net result that involves a rotation, even though each boost on its own produces no rotation. The rotation angle for orthogonal $\mathbf{u}$ and $\mathbf{v}$ can be obtained from (5.41) and (5.42) using the standard trigonometric formula $\tan(\theta - \theta'') = (\tan\theta - \tan\theta'')/(1 + \tan\theta\tan\theta'')$:

$$\tan\Delta\theta = \frac{uv(\gamma_u\gamma_v - 1)}{u^2\gamma_u + v^2\gamma_v}. \tag{5.44}$$

Note that the rotation effect is a purely kinematic result: it results purely from the geometry of spacetime. That is to say, the amount and sense of rotation is determined purely by the velocity changes involved, not by some further property of the forces which cause the velocity changes in any particular case. It is at the heart of the Thomas precession, which we will now discuss.

5.7.2 The Thomas precession

Let us return to the thought-experiment with which we began section 5.7: the aircraft flying around the polygon. This thought-experiment can be understood in terms of the rotation effect that results from a sequence of changes of inertial reference frame, as discussed in the previous section. The pilot's reasoning is valid, and it implies that a vector carried around a closed path by *parallel transport* will undergo a net rotation: it will finish pointing in a direction different to the one it started in.

Parallel transport is the type of transport when an object is translated as a whole, in some given direction, without rotating it. For example, if you pass someone a book, you will normally find that your action will rotate the book as your arm swings. However, with care you could adjust the angle between your hand and your arm, as your hand moves, so as to maintain the orientation of the book fixed. That would be a parallel transport.

In the aeroplane example, the aeroplane did not undergo a parallel transport, but if the pilot kept next to him a rod, initially parallel to the axis of the aircraft, and made it undergo a parallel transport, then after flying around the polygon he would find the rod was no longer parallel to the axis of the aircraft. He could make sure the rod had a parallel transport by attaching springs and feedback-controls to it, so that the forces at the two ends of the rod were always equal to one another (in size and direction) at each instant of time as defined in his instantaneous rest frame. His observations of his journey convince him that the angle between himself and such a rod increases by more than 360°, and he is right. On completing the circuit in an anticlockwise direction, the aircraft is on a final flight path parallel to its initial one, but the rod has undergone a net rotation clockwise, see fig. 5.7.

Parallel transport in everyday situations (in technical language, in flat Euclidean geometry) never results in a change of orientation of an object. However, it is possible to define parallel transport in more general scenarios, and then a net rotation can be obtained. To get a flavour of this idea, consider motion in two dimensions, but allow the 'two dimensional' surface to be curved in some way, such as the surface of a sphere. Define 'parallel transport' in this surface to mean the object has to lie in the surface, but it is not allowed to rotate relative to the nearby surface as it moves. For a specific example, think of carrying a metal bar over the surface of a non-rotating spherical planet. Hold the bar always horizontal (i.e. parallel to the ground at your location), and when you walk make sure the two ends of the bar move through the same distance relative to the ground: that is what we mean by parallel transport in this example. Start at the equator, facing north, so that the bar is oriented east-west. Walk due north to the north pole. Now, without rotating yourself or the bar, step to your right, and continue until you reach the equator again. You will find on reaching the equator that you are facing around the equator, and the bar is now oriented north-south. Next, again without turning, walk around the equator back to your starting point. You can take either the long route by walking forwards, or the short route by stepping backwards. In either case, when you reach your starting point, the bar, and your body, will have undergone a net rotation through 90°.

The example just given was intended merely to give you some general flavour of the idea of parallel transport. In special relativity, we do *not* have any curvature of spacetime, but we do have a geometry of spacetime rather than of space alone. In spacetime we define parallel transport to mean that an object is displaced without undergoing a rotation in its rest frame. This definition has to be clarified when an object undergoes acceleration, because then its rest frame is continuously changing. However, it is not hard to see what is needed. Let τ be the proper time at the center of mass of some object. At time τ, let the momentary rest frame of the object be called S. In the next instant of time $\tau + d\tau$, the object will be at rest in any one of an infinite number of frames S$'$, all having the same velocity but related to one another by a rotation of axes. Among all these frames, we pick the one whose axes are parallel to those of S, according to an observer at rest in S. If at time $\tau + d\tau$ the object has the same orientation

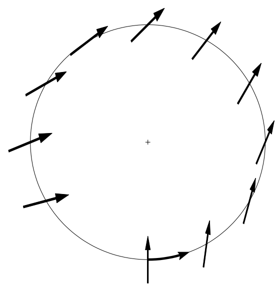


Figure 5.7: The evolution of an object (e.g. a wooden arrow) when it undergoes a parallel transport in spacetime, such that it is carried around a circle in some inertial reference frame. 'Parallel transport' means that at each moment, the evolution in the next small time interval can be described by a Lorentz boost, that is, an acceleration without rotation.

in this S′ as it had in S, then it is undergoing a parallel transport.

We also speak of parallel transport of a 'vector'. This is simply to liberate the definition from the need to talk about any particular physical object, but note that such a vector ultimately has to be defined in physical terms. It is a mathematical quantity behaving in the same way as a spatial displacement in the instantaneous rest frame, where, as always, spatial displacement is displacement relative to a reference body in uniform motion.

Is there a torque?

Students (and more experienced workers) are sometimes confused about the distinction between kinematic and dynamic effects. For example, the Lorentz contraction is a kinematic effect because it is the result of examining the same set of worldlines (those of the particles of a body) from the perspective of two different reference frames. Nonetheless, if a given object starts at rest and then is made to accelerate, then any change in its shape in a given frame (such as the initial rest frame) is caused by the forces acting on it—a dynamic effect. The insight obtained from Lorentz contraction in such a case is that it enables us to see what kind of dynamical contraction is the one that preserves the proper length. To be specific, if a rod starts at rest in S and then is accelerated to speed v in S by giving the same velocity-change to all the particles in the rod, then if the proper length is to remain unchanged, the particles must not be pushed simultaneously in S. Rather, the new velocity has to be acquired by the back of the rod first. No wonder then that it contracts.

In the case of Thomas rotation, a similar argument applies. Recall the example of the aeroplane, and suppose the aeroplane first approached the polygon in straight line flight along a tangent, and then flew around it. From the perspective of a reference frame fixed on the ground, the rod initially has a constant orientation (until the aircraft reaches the polygon), and then it begins to rotate. It must therefore be subject to a torque to set it rotating. It is not hard to see how the torque arises. Transverse forces on the rod are needed to make it accelerate with

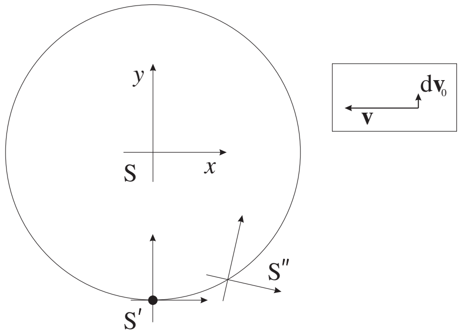


Figure 5.8: Analysis of motion around a circle. The frame S is that of the fixed circle. Frames S′ and S″ are successive rest frames of an object moving around the circle. The axes of S'' and S′ are arranged to be parallel in either of those frames. (Therefore they are not parallel in frame S which has been used to draw the diagram). The inset shows the velocities of S and S″ relative to S′.

the aeroplane around the polygon. If the application of these forces is simultaneous in the rest frame of the aeroplane, then in the rest frame of the polygon, the force at the back of the rod happens first, so there is a momentary torque about the centre of mass. This shows that the Thomas rotation is a companion to the Lorentz contraction.

5.7.3 Analysis of circular motion

We shall now analyze the case of motion around a circular trajectory. We already know the answer because the simple argument given at the start of this section for the aircraft flying above the polygon is completely valid, but to get a more complete picture it is useful to think about the sequence of rest frames of an object following a curved trajectory.

Figure 5.8 shows the case of a particle following a circular orbit. The axes xy are those of the reference frame S in which the circle is at rest. The particle is momentarily at rest in frame S′ at proper time τ and in frame S″ at the slighter later proper time $\tau + d\tau$. The axes of both S and of S″ are constructed to be parallel to those of S′ for an observer at rest in S′. Nevertheless, as we already showed in section 5.7, the axes of S and S″ are not parallel in S or S″.

In order to analyse the acceleration of the particle at proper time τ, we adopt its momentary rest frame S'. Let $\mathbf{a}_0$ be the acceleration of the particle in this frame. For a small enough time interval $d\tau$ the change in velocity is

$$d\mathbf{v}_0 = \mathbf{a}_0 d\tau. \tag{5.45}$$

This will be the velocity of S'' relative to S'. The subscript zero is to indicate that the quantity is as observed in the momentary rest frame. $\mathbf{a}_0$ is directed towards the center of the circle, which at the instant τ is in the direction of the positive y' axis, therefore $dv_{0x} = 0,\ dv_{0y} = dv_0$.

Let $\mathbf{v}$ be the velocity of S' relative to S. $\mathbf{v}$ and $d\mathbf{v}_0$ are mutually perpendicular, so we have a situation exactly as was discussed in section 5.7, with the speeds u and v now replaced by dv_0 and v. Angles θ and θ'' (equations (5.41) and (5.42)) are both small, and $\gamma(dv_0) \simeq 1$, so we have a rotation of the rest frame axes by

$$\theta - \theta'' = d\theta = \frac{dv_0}{v}\left(1 - \frac{1}{\gamma}\right). \tag{5.46}$$

In equation (5.46), dv_0 is a velocity *in* the instantaneous rest frame, whereas v is a velocity *of* that frame relative to the centre of the circle. It is more convenient to express the result in terms of quantities all in the latter frame. The change in velocity as observed in S is $dv = dv_0/\gamma$ (by using the velocity addition equations, (3.20)). Hence

$$d\theta = \frac{dv}{v}\left(\gamma - 1\right). \tag{5.47}$$

The motion completes one circuit of the circle when $\int dv = 2\pi v$, at which point the net rotation angle of the axes is $2\pi(\gamma - 1)$, in agreement with (5.40).

We conclude that the axes in which the particle is momentarily at rest, when chosen such that each set is parallel to the previous one for an observer on the particle, are found to rotate in the reference frame of the centre of the circle (and therefore in any inertial reference frame) at the rate

$$\frac{d\theta}{dt} = \frac{a}{v}\left(\gamma - 1\right), \tag{5.48}$$

and it is easy to check that the directions are such that this can be written in vector notation

$$\boldsymbol{\omega}_T = \frac{\mathbf{a} \wedge \mathbf{v}}{c^2}\frac{\gamma^2}{1+\gamma}, \tag{5.49}$$

where we made use of equation (3.10). In fact the derivation did not need to assume the motion was circular, and we can always choose to align the axes with the local velocity, so we have proved the vector result (5.49) for any motion where the acceleration is perpendicular to the velocity. By analysing a product of two Lorentz boosts, it can be shown that the result is valid in general.

Application

In order to apply (5.49) to dynamical problems, one uses a standard kinematic result for rotating frames (whether classical or relativistic), namely that if some vector $\mathbf{s}$ has a rate of change $(d\mathbf{s}/dt)_{\rm rot}$ in a frame rotating with angular velocity $\boldsymbol{\omega}_T$, then its rate of change in a non-rotating frame is

$$\left(\frac{d\mathbf{s}}{dt}\right)_{\rm nonrot} = \left(\frac{d\mathbf{s}}{dt}\right)_{\rm rot} + \boldsymbol{\omega}_T \wedge \mathbf{s}. \tag{5.50}$$

For example, suppose by 'the rest frame' of an accelerating particle we mean one of the sequence of instantaneous inertial rest frames of the particle. The dynamical equations applying in the rest frame will dictate the proper rate of change $d\mathbf{s}/d\tau$ of any given vector $\mathbf{s}$ describing some property of the particle. For a particle describing circular motion the sequence of instantaneous rest frames can be regarded as a single rotating frame, to which (5.50) applies, with the substitution

$$\left(\frac{d\mathbf{s}}{dt}\right)_{\rm rot} = \frac{1}{\gamma}\left(\frac{d\mathbf{s}}{d\tau}\right)_{\rm rest\ frame}. \tag{5.51}$$

For motion which is curved but not circular, the equation applies to each short segment of the trajectory.

For electrons in atoms, there is a centripetal acceleration given by the Coulomb attraction to the nucleus, $\mathbf{a} = -e\mathbf{E}/m$ where $\mathbf{E}$ is the electric field at the electron, calculated in the rest frame of the nucleus, and $-e$ is the charge on an electron. For atoms such as hydrogen, the velocity $v \ll c$ so we can use $\gamma \simeq 1$ and we obtain to good approximation,

$$\boldsymbol{\omega}_T = \frac{e\mathbf{v} \wedge \mathbf{E}}{2mc^2}. \tag{5.52}$$

The spin-orbit interaction calculated in the instantaneous rest frame of the *electron* gives a Larmor precesssion frequency

$$\boldsymbol{\omega}_L = \frac{-g_{\rm s}\mu_B}{\hbar}\frac{\mathbf{v} \wedge \mathbf{E}}{c^2}, \tag{5.53}$$

where $g_{\rm s}$ is the gyromagnetic ratio of the spin of the electron and the Bohr magneton is $\mu_B = e\hbar/2m$. To find what is observed in an inertial frame, such as the rest frame of the nucleus, we must add the Thomas precession to the Larmor precession,

$$\boldsymbol{\omega} = \boldsymbol{\omega}_L + \boldsymbol{\omega}_T \;=\; \left(\frac{-g_{\rm s}\mu_B}{\hbar} + \frac{e}{2m}\right)\frac{\mathbf{v} \wedge \mathbf{E}}{c^2} \;=\; -\frac{e}{2m}(g_{\rm s} - 1)\frac{\mathbf{v} \wedge \mathbf{E}}{c^2} \tag{5.54}$$

If we now substitute the approximate value $g_s = 2$, we find that the Thomas precession frequency for this case has the opposite sign and half the magnitude of the rest frame Larmor frequency. This means that the precession frequency observed in the rest frame of the nucleus will be half that in the electron rest frame. More precisely, the impact is to replace g_s by $g_s - 1$ (not $g_s/2$): it is an additive, not a multiplicative correction (see exercise **??**).

The above argument treated the motion as if classical rather than quantum mechanics was adequate. This is wrong. However, upon reexamining the argument starting from Schrödinger's equation, one finds that the spin-orbit interaction gives a contribution to the potential energy of the system, and the precession of the spin of the electron may still be observed. For example when the electron is in a non-stationary state (a superposition of states of different orientation), the spin direction precesses at ω_L in the rest frame of the electron, and at $\omega_L + \omega_T$ in the rest frame of the nucleus. This precession must be related to the gap between energy levels by the universal factor $\hbar$, so it follows that the Thomas precession factor (a kinematic result) must influence the observed energy level splittings (a dynamic result).

5.8 The Lorentz group*

A product of two rotations is a rotation, but a product of two Lorentz boosts is not always a Lorentz boost (c.f. eq. (5.43)). This invites one to look into the question, to what general class of transformations does the Lorentz transformation belong?

We define the Lorentz transformation as that general type of transformation of coordinates that preserves the interval $(ct)^2 - x^2 - y^2 - z^2$ unchanged. Using eq. (3.41) this definition is conveniently written

$$L \equiv \{\Lambda : \Lambda^T g \Lambda = g\}. \tag{5.55}$$

where L denotes the set of all Lorentz transformations, and in this section we will use the symbol Λ instead of $\mathcal{L}$ to denote individual Lorentz transformations. g is the Minkowski metric defined in (3.39).

We will now prove that the set L is in fact a *group*, and furthermore it can be divided into 4 distinct parts, one of which is a sub-group. Here a mathematical *group* is a set of entities that can be combined in pairs, such that the combination rule is associative (i.e. $(ab)c = a(bc)$), the set is closed under the combination rule, there is an identity element and every element has an inverse. *Closure* here means that for every pair of elements in the set, their combination is also in the set. We can prove all these properties for the Lorentz group by using matrices that satisfy (5.55). The operation or 'combination rule' of the group will be matrix multiplication. The matrices are said to be a *representation* of the group.

1. *Associativity.* This follows from the fact that matrix multiplication is associative.

2. *Closure.* The net effect of two successive Lorentz transformations $\mathsf{X} \to \mathsf{X}' \to \mathsf{X}''$ can be written $\mathsf{X}'' = \Lambda_2\Lambda_1\mathsf{X}$. The combination $\Lambda_2\Lambda_1$ is a Lorentz transformation, since it satisfies (5.55):

$$(\Lambda_2\Lambda_1)^T g\Lambda_2\Lambda_1 = \Lambda_1^T\Lambda_2^T g\Lambda_2\Lambda_1 = \Lambda_1^T g\Lambda_1 = g.$$

3. *Inverses.* We have to show that the inverse matrix Λ^{-1} exists and is itself a Lorentz transformation. To prove its existence, take determinants of both sides of $\Lambda^T g\Lambda = g$ to obtain

$$|\Lambda|^2|g| = |g|$$

but $|g| = -1$ so

$$|\Lambda|^2 = 1, \qquad |\Lambda| = \pm 1. \tag{5.56}$$

Since $|\Lambda| \neq 0$ we deduce that the matrix Λ does have an inverse. To show that Λ^{-1} satisfies (5.55) we need a related formula. First consider

$$(\Lambda g)(\Lambda^T g\Lambda)(g\Lambda^T) = \Lambda g^3\Lambda^T = \Lambda g\Lambda^T$$

Now pre-multiply by $(\Lambda g\Lambda^T g)^{-1}$:

$$\Lambda g\Lambda^T = g^{-1} = g \tag{5.57}$$

where we used $(AB)^{-1} = B^{-1}A^{-1}$ for any pair of matrices A, B, and we can be sure that $(\Lambda g\Lambda^T g)^{-1}$ exists because $|\Lambda g\Lambda^T g| = |\Lambda|^2|g|^2 = 1$. Now to show that Λ^{-1} is a Lorentz transformation, take the inverse of both sides of (5.57):

$$\begin{aligned} (\Lambda g\Lambda^T)^{-1} &= g, \qquad \Rightarrow (\Lambda^T)^{-1}g^{-1}\Lambda^{-1} = g \\ \Rightarrow (\Lambda^{-1})^T g\Lambda^{-1} &= g \end{aligned} \tag{5.58}$$

which shows Λ^{-1} satisfies the condition (5.55).

4. *Identity element.* The identity matrix satisfies (5.55) and so can serve as the identity element of the Lorentz group.

Since the complete set of 4×4 real matrices can themselves be considered as a representation of a 16-dimensional real space, we can think of the Lorentz group as a subset of 16-dimensional real space. The defining condition (5.55) might appear to set 16 separate conditions, which would reduce the space to a single point, but there is some repetition since g is symmetric, so there is a continuous 'space' of solutions. There are 10 linearly independent conditions (a symmetric 4×4 matrix has 10 independent elements); it follows that L is a six-dimensional subset of R^{16}. That is, a general member of the set can be specified by 6 real parameters; you can think of these as 3 to specify a rotation and 3 to specify a velocity.

We can move among some members of the Lorentz group by continuous changes, such as by a change in relative velocity between reference frames or a change in rotation angle. However we can show that not all parts of the group are continuously connected in this way. The condition (5.56) is interesting because it is not possible to change the determinant of a matrix discontinuously by a continuous change in its elements. This means that we can identify two subsets:

$$\begin{aligned} L_\uparrow &\equiv \{\Lambda \in L : |\Lambda| = +1\} \\ L_\downarrow &\equiv \{\Lambda \in L : |\Lambda| = -1\} \end{aligned} \tag{5.59}$$

and one cannot move between $L_\uparrow$ and $L_\downarrow$ by a continuous change of matrix elements. The subsets are said to be *disconnected*. One can see that the subset $L_\downarrow$ is not a group because it is not closed (the product of any two of its members lies in $L_\uparrow$), but it is not hard to prove that $L_\uparrow$ is a group, and therefore a sub-group of L. An important member of $L_\downarrow$ is the spatial inversion through the origin, also called the *parity* operator:

$$P \equiv (t \to t,\ \mathbf{r} \to -\mathbf{r}).$$

Its matrix representation is

$$P = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix}. \tag{5.60}$$

What is interesting is that if $\Lambda \in L_\uparrow$ then $P\Lambda \in L_\downarrow$. Thus to understand the whole group it suffices to understand the sub-group $L_\uparrow$ and the effect of P. The action of P is to reverse the direction of vector quantities such as the position vector or momentum vector; the subscript arrow notation $L_\uparrow, L_\downarrow$ is a reminder of this. Members of $L_\uparrow$ are said to be *proper* and members of $L_\downarrow$ *improper*. Rotations are in $L_\uparrow$, reflections are in $L_\downarrow$.

We can divide the Lorentz group a second time by further use of (5.55). We adopt the notation Λ^μ_ν for the (μ, ν) component of Λ. Examine the $(0, 0)$ component of (5.55). If we had the matrix product $\Lambda^T\Lambda$ this would be $\sum_\mu (\Lambda^\mu_0)^2$, but the g matrix in the middle introduces a sign change, so we obtain

$$\begin{aligned} -(\Lambda^0_0)^2 + \sum_{i=1}^{3} (\Lambda^i_0)^2 &= g_{00} = -1 \\ \Rightarrow \qquad \Lambda^0_0 &= \pm \left(1 + \sum_{i=1}^{3} (\Lambda^i_0)^2 \right)^{1/2}. \end{aligned} \tag{5.61}$$

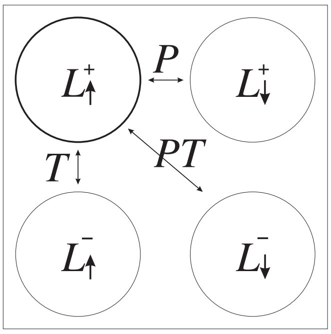


Figure 5.9: The structure of the Lorentz group. The proper orthochronous set $L_\uparrow^+$ is a subgroup. It is continuous and 6-dimensional. The other subsets can be obtained from it.

The sum inside the square root is always positive since we are dealing with real matrices, and we deduce that

$$\text{either } \Lambda_0^0 \geq 1 \quad \text{or } \Lambda_0^0 \leq -1.$$

That is, the time-time component of a Lorentz transformation can either be greater than or equal to 1, or less than or equal to -1, but there is a region in the middle, from -1 to 1, that is forbidden. It follows that the transformations with $\Lambda_0^0 \geq 1$ form a set disconnected from those with $\Lambda_0^0 \leq 1$. We define

$$L^+ \equiv \{\Lambda \in L : \Lambda_0^0 \geq 1\} \tag{5.62}$$
$$L^- \equiv \{\Lambda \in L : \Lambda_0^0 \leq 1\}. \tag{5.63}$$

An important member of L^- is the time-reversal operator

$$T \equiv (t \to -t, \; \mathbf{r} \to \mathbf{r})$$

whose matrix representation is[4]

$$T = \begin{pmatrix} -1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}. \tag{5.64}$$

It is now straightforward to define the sub-sets $L_\uparrow^+$ $L_\downarrow^+$ $L_\uparrow^-$ $L_\downarrow^-$ as intersections of the above. It is easy to show furthermore (left as an exercise for the reader) that $L_\uparrow^+$ is a group and the

[4]The time-reversal operator is not the same as the Minkowski metric, although they may look the same in a particular coordinate system such as rectangular coordinates. Their difference is obvious as soon as one adopts another coordinate system such as polar coordinates.

operators P, T and PT allow one-to-one mappings between $L_\uparrow^+$ and the other distinct sets, as shown in figure 5.9.

A member of $L_\uparrow^+$ is called a 'proper orthochronous' Lorentz transformation. It can be shown (see chapter 17) that a general member of this group can be written[5]

$$\Lambda = e^{-\boldsymbol{\rho}\cdot\mathbf{K}-\boldsymbol{\theta}\cdot\mathbf{S}} \tag{5.65}$$

where $\boldsymbol{\rho}$ is a rapidity vector, $\boldsymbol{\theta}$ is a rotation angle (the direction of the vector specifying the axis of rotation), and $\mathbf{K}$ and $\mathbf{S}$ are the following sets of matrices:

$$S_x = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 \\ 0 & 0 & 1 & 0 \end{pmatrix}, \quad K_x = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}, \tag{5.66}$$

$$S_y = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \end{pmatrix}, \quad K_y = \begin{pmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}, \tag{5.67}$$

$$S_z = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}, \quad K_x = \begin{pmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{pmatrix}. \tag{5.68}$$

For example, a boost in the x direction would be given by $\boldsymbol{\theta} = 0$ and $\boldsymbol{\rho} = (\rho, 0, 0)$ with $\tanh\rho = v/c$. The S matrices are said to 'generate' rotations, and the K matrices to 'generate' boosts.

All known fundamental physics is invariant under proper orthochronous Lorentz transformations, but examples of both parity violation and time reversal violation are known in weak radioactive processes. Thus one cannot always ask for Lorentz invariance under the whole Lorentz group, but as far as we know it is legitimate to require invariance under transformations in $L_\uparrow^+$. This group is also called the 'restricted' Lorentz group.

5.8.1 Further group terminology

[*Section omitted in lecture-note version.*]

[5]The exponential of a matrix M is defined $\exp(M) \equiv 1 + M + M^2/2! + M^3/3! + \cdots$. It can be calculated from $\exp(M) = U \exp(M_D) U^\dagger$ where M_D is a diagonalized form of M, i.e. $M_D = U^\dagger M U$ where U is the (unitary) matrix whose columns are the normalized eigenvectors of M.

5.9 Exercises

[*Section omitted in lecture-note version.*]