Chapter 6 Relativistic Mechanics

6-1 Introduction

In Chapter 1 we pointed out the necessity of a frame of reference for describing dynamical processes, and noted that Newton's law of motion assumes the simple form $\mathbf{p} = \mathbf{F}$ in an inertial frame of reference. We saw further that an arbitrary number of such inertial frames exist if one does, and that the laws of motion are identical in form in all such frames. The invariance of form of the laws of motion under change of inertial reference frame constitutes *Galileo's principle of relativity*. However, for this principle to be strictly valid it is necessary that interactions between the constituents of a dynamical system be *instantaneous*, that is to say any change of the state of motion in one of the constituents must immediately be sensed by all the other constituents in interaction with it. Now, experiment shows that instantaneous interactions do not in fact exist in nature. Instead, there is a finite time lag between the "emitting" and "sensing" of the interaction between two electric charges, for example, and thus a finite velocity of propagation of interaction. We call this velocity of propagation c and identify its physical significance in a moment.

Given that all interactions propagate with a universal velocity c it is easy to see that Galileo's principle of relativity breaks down. For according to the Galilean transformation of velocities given by (1.4) of Chapter 1, the velocity c will clearly depend on the frame of reference in which the interaction is viewed. Thus, the corresponding laws of motion will become inertial frame-dependent, in violation of Galileo's relativity principle.

We must now consider the significance of the velocity of propagation c. It is known experimentally that electromagnetic waves (light) propagate in vacuo with a velocity that is *independent* of the frame of reference in which this velocity is measured (the famous Michelson-Morley experiment). Therefore, the velocity of light is certainly a candidate for c as the *universal* velocity. The thoughtful reader may wonder about other candidates for c, like the velocity of neutrinos or gravitons, for example. However, the important point here is that there can be only *one* universal velocity if all the tenets of the special theory of relativity are to be satisfied. In what follows we will identify c with the velocity of light in vacuo, $c \simeq 3 \times 10^{10}$ cm/sec. However, its real significance for relativity theory as a universal velocity must always be borne in mind.

6-2 *Einstein's Principle of Relativity*

After the violation of Galileo's principle of relativity for light waves became apparent from Michelson's experiments in the 1880's, there were numerous attempts to modify the (then) current "aether" theory of light to account for this apparent violation. None of these attempts were fully successful, although some were extremely ingenious. There simply seemed to be no way of reconciling Galileo's principle with Michelson's measurements.

The way out of this dilemma was resolved by Einstein in 1905. Einstein's special theory of relativity, or the relativity theory of inertial frames, rests on two postulates: (a) The universal velocity c is the same in all directions for all observers moving uniformly with respect to each other, and (b) all laws of nature are identical in all inertial frames of reference. Postulate (b) sounds like Galileo's relativity principle again. However, we will see that postulate (a) compels us to rethink our (somewhat intuitive) concept of the absolute nature of time. We introduce an inertial frame Σ and a second one Σ' moving with a uniform velocity $\mathbf{V}$ relative to Σ (see Fig. 6.1).

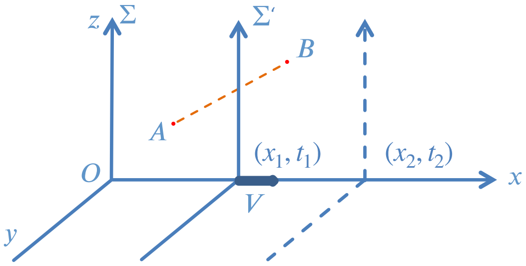

Figure 6.1: The inertial frames Σ and Σ'.

We attach observers rigidly to Σ and Σ' and supply them with identical devices for measuring distances and quantity of elapsed time in the respective frames. Both observers record the emission and absorption of a light ray travelling from A to B. The observer in Σ assigns coordinates (x_a, y_a, z_a) and a time t_a to the event of emission at A, and coordinates (x_b, y_b, z_b) and a time t_b to the event of absorption at B. Knowing the velocity of light to be c, he concludes that

$$c^2(t_b - t_a)^2 = (x_b - x_a)^2 + (y_b - y_a)^2 + (z_b - z_c)^2, \tag{6.1}$$

since both sides of this equation express the distance AB. The observer in Σ' observes the same two events, assigns coordinates (x'_a, y'_a, z'_a, t'_a) and (x'_b, y'_b, z'_b, t'_b) to them, and concludes that

$$c^2(t'_b - t'_a)^2 = (x'_b - x'_a)^2 + (y'_b - y'_a)^2 + (z'_b - z'_a)^2, \qquad (6.2)$$

since by postulate (a) the velocity c is the *same* in Σ and Σ'.

The appearance in these relations of time intervals $t_b - t_a$ and $t'_b - t'_a$ *that depend on the reference frame to which the time measuring device* (a "clock") *measuring them, is attached*, is something completely alien to the Newtonian idea of absolute time. Instead, we are forced to assert that there is no such thing as absolute time. The quantity of elapsed time depends on which observer makes the measurement.

We thus see that it becomes a natural idea to associate the *four* coordinates (x, y, z, t) with a "happening" like the emission or absorption of a light ray. The coordinates (x, y, z, t) are said to label an *event* in a given reference frame. An event can therefore be represented as a point, called a *world* point in a fictitious four-dimensional space, or *space-time*, with axes labelled by x, y, z and t. The motion of a particle is then represented by the curve (called a *world* line) traced out by a succession of world points in this four-dimensional space.

So far (6.1) and (6.2) refer to the motion of light rays. We now generalize the points A and B to refer to *any* two events and define the interval between these two events as

$$s_{ab}^2 = c^2(t_b - t_a)^2 - (x_b - x_a)^2 - (y_b - y_a)^2 - (z_b - z_a)^2. \qquad (6.3)$$

If A and B are infinitesimally separated events at (x, y, z, t) and $(x + dx, y + dy, z + dz, t + dt)$, then $s_{ab} = ds$, where

$$ds^2 = c^2 dt^2 - dx^2 - dy^2 - dz^2. \qquad (6.4)$$

We now assert that the infinitesimal interval ds between any two events is the same in all inertial frames. There are several ways of demonstrating this assertion. We follow the elegant argument of Landau and Lifshitz[66]. Consider again the two frames Σ and Σ'. According to (6.1) and (6.2) $ds = 0$ in Σ and ds' in Σ' also vanishes, $ds' = 0$, for light ray propagation from A to B, the latter now being infinitesimally separated from A. Hence ds and ds' must be infinitesimals of the same order, or $ds' = \eta ds$, where η is a factor of proportionality. A little reflection shows that η can at most depend on the magnitude of the relative velocity $\mathbf{V}$ of Σ' relative to Σ, i.e. $\eta = \eta(|\mathbf{V}|)$. Any spatial or temporal dependence in η would violate the homogeneity of space and time, while a directional dependence on $\mathbf{V}$ would violate the isotropy of space. But if $\eta = \eta(|\mathbf{V}|)$, we may equally well view Σ from Σ' and conclude that $ds = \eta ds'$, or that $\eta^2 = 1$. However, the result must hold for any velocity, so that we cannot choose $\eta = +1$ for some velocities and $\eta = -1$ for others. Also

[66] L. Landau and E. Lifshitz, *The Classical Theory of Fields*, Addison-Wesley Inc., Cambridge, Mass., 1951), p.6.

$ds = \eta ds'$ must include the identity $ds = ds'$ when $\mathbf{V} = 0$. Hence, $\eta = 1$ and $ds = ds'$ has the same value in all inertial frames. Since $ds = ds'$ it follows directly that any finite interval has the same value, $s = s'$ in all inertial frames.

6-3 *Proper time*

We have seen that the time lapse between two events depends upon the frame of reference from which the events are viewed. So let us consider the special event of a clock *fixed rigidly* to Σ' and therefore moving with velocity $\mathbf{V}$. The moving clock is at rest in Σ', so that $dx' = dy' = dz' = 0$ and it registers a time lapse dt' (say) in the same period that it is observed to move from (xyz) to $(x + dx, y + dy, z + dz)$ in Σ in a time dt as measured by a clock attached to Σ. The invariance of intervals lets us write

$$ds^2 = c^2 dt'^2 = c^2 dt^2 - dx^2 - dy^2 - dz^2, \tag{6.5}$$

or

$$dt' = dt\sqrt{1 - \frac{v^2}{c^2}} = \frac{ds}{c}, \tag{6.6}$$

since $[(dx/dt)^2 + (dy/dt)^2 + (dz/dt)^2]^{1/2}$ is just the speed $v = |\mathbf{V}|$ of the clock attached to Σ' as seen from Σ. The interval dt' is called the *proper time*. It is the time recorded by a clock moving with a material particle. The proper time lapse for events recorded between t_1 and t_2 in Σ is given by integration as

$$t_2' - t_1' = \int_{t_1}^{t_2} dt\,\sqrt{1 - \frac{v^2}{c^2}}. \tag{6.7}$$

We comment that the speed v in (6.6) and (6.7) need *not* be uniform, i.e. v may vary with time. The point is simply that a clock performing an arbitrary motion can be considered as moving uniformly with whatever speed it has at each moment of time. We may therefore introduce a different inertial frame at each moment of time, and the argument leading to (6.6) goes through as before.

The relation (6.7) shows that the proper time interval $t_2' - t_1'$ is always *less* than the time interval $t_1 - t_2$ as measured in Σ. The moving clock runs slower than the clock at rest relative to the observer. Results of this nature lead to several apparent "paradoxes" in the theory of relativity. However, since we intend to present only as much background of the special theory as is required for discussing relativistic mechanics in a coherent way, we resist the temptation of too much disgression into those fascinating aspects of the special theory and refer the reader to the literature instead[67].

[67] A good reference for further reading is C. Moller, *The Theory of Relativity*, Oxford University Press, 1952.

6-4 The Lorentz Transformation

We now establish how the two sets of coordinates (x, y, z, t) and (x', y', z', t') of the same event in Σ and Σ' are related. To do so we introduce an imaginary time coordinate $x_4 = ict$ in place of t, and locate an event by $(x, y, z, t) \rightarrow (x_1, x_2, x_3, x_4)$ in Σ. The link between these coordinates and the corresponding coordinates (x_1', x_2', x_3', x_4') of the same event in Σ' then follows from the invariance property $s = s'$. Written out in terms of the coordinates of two events A and B, the invariance of s reads

$$-s^2 = (x_{1a} - x_{1b})^2 + (x_{2a} - x_{2b})^2 + (x_{3a} - x_{3b})^2 + (x_{4a} - x_{4b})^2 = \text{invariant} \tag{6.8}$$

if the imaginary-time coordinate x_4 is used. Equation (6.8) is just a statement of the invariance of "distance" between any two world points in the fictitious four-dimensional space, or *Euclidean space*, characterized by orthogonal axes O_{x_1,x_2,x_3,x_4}. (Alternatively, we could have introduced the notation $(x_0 = ct, x_1, x_2, x_3)$, which forms the basis of *Minkowski space*, but we will not use that here). Consequently, the transformation $(x_1, x_2, x_3, x_4) \rightarrow (x_1', x_2', x_3', x_4')$ must leave all distances in *this* four-dimensional space unchanged. If we disregard parallel displacement of axes, which merely relabels the coordinate origin, then the coordinates of the same event seen in Σ and Σ' must be connected by a rotation of coordinates in four dimensions,

$$\mathbf{x}' = A\mathbf{x}. \tag{6.9}$$

Here, A is formally the same as the rotation matrix in (3.13), $A^T A = AA^T = I$ and $\text{Det}A = +1$, while $\mathbf{x}'$ and $\mathbf{x}$ are column vectors of the coordinates of the same event in Σ and Σ'.

We now determine the elements of A under the assumption that the axes defining Σ and Σ' are parallel and that the origin of Σ' proceeds along the positive $x-$axis of Σ with velocity $\mathbf{V}$. Then the coordinates $x_2 = y$ and $x_3 = z$ remain unchanged, while $x_1 = x$ and $x_4 = ict$ must transform like

$$\begin{pmatrix} x' \\ x_4' \end{pmatrix} \begin{pmatrix} \cos\phi & \sin\phi \\ -\sin\phi & \cos\phi \end{pmatrix} \begin{pmatrix} x \\ x_4 \end{pmatrix} \tag{6.10}$$

in accordance with (**??**). The analogy here with a rotation in ordinary space is purely formal, however. The angle ϕ is merely a *parameter* in (6.10) (it will turn out to be pure imaginary) that has nothing to do with rotations in real space. However, the representation of the operator A by the 2×2 matrix on the right-hand side of (6.10) is very convenient, since the conditions on A given below (6.9) are now automatically satisfied. To determine A, we measure time from the moment that the reference frames Σ and Σ' are coincident. We follow the motion of the origin of Σ'

from Σ and deduce from the inverse transformation to (6.10) that

$$\begin{pmatrix} x \\ x_4 \end{pmatrix} \begin{pmatrix} \cos\phi & -\sin\phi \\ \sin\phi & \cos\phi \end{pmatrix} \begin{pmatrix} 0 \\ x_4' \end{pmatrix}, \tag{6.11}$$

since now $x' = 0$. This gives the ratio $x/x_4 = -\tan\phi$. However, $x/t = v$ is the speed of Σ' relative to Σ. Hence, $\tan\phi = iv/c$, so that

$$\cos\phi = \frac{1}{\sqrt{1-\frac{v^2}{c^2}}}; \quad \sin\phi = i\frac{v/c}{\sqrt{1-\frac{v^2}{c^2}}}. \tag{6.12}$$

Inserting these expressions into (6.10), one obtains the *Lorentz transformation*

$$x' = \frac{x - vt}{\sqrt{1-\frac{v^2}{c^2}}}, \quad y' = y, \quad z' = z, \quad t' = \frac{t - vx/c^2}{\sqrt{1-\frac{v^2}{c^2}}}, \tag{6.13}$$

after reverting to real time coordinates again. The inverse transformation follows upon reversing the sign of $\mathbf{V}$. (Σ has the velocity $-\mathbf{V}$ relative to Σ'):

$$x = \frac{x' + vt}{\sqrt{1-\frac{v^2}{c^2}}}, \quad y = y', \quad z = z', \quad t = \frac{t' + vx'/c}{\sqrt{1-\frac{v^2}{c^2}}}. \tag{6.14}$$

One notices that the transformation becomes singular for a reference frame moving with velocity c, and also that x and t both become imaginary for frames moving with $v > c$, corresponding to the impossibility of such motion for material particles[68].

The particular feature of the transformation (6.13) or its inverse (6.14) is that space and time coordinates get mixed up with each other. A knowledge of *both* x and t is necessary to compute x' and t' from (6.13).

We will see in the following examples of the use of the Lorentz transformation how essential this feature is if observers in Σ and Σ' are to arrive at consistent conclusions about the outcome of experiments performed in their respective frames.

(i) *Time-dilation:* Let us re-examine the phenomenon of time-dilation given by (6.6) or (6.7) from the point of view of the Lorentz transformation. A clock attached to the origin of Σ' records a time difference $t_2' - t_1'$ in the time $t_2 - t_1$ (as seen by an observer in Σ) that the origin of Σ' moves from x_1 to x_2 as measured in Σ, see Fig. 6.1. Since the clock in Σ' is at the origin, the same two events (x_1, t_1) and (x_2, t_2) in Σ have coordinates $(0, t_1')$ and $(0, t_2')$ in Σ'. Hence, from (6.14) we conclude that

$$t_2 - t_1 = \frac{t_2' - t_1'}{\sqrt{1-\beta^2}} > t_2' - t_1', \quad \beta = \frac{V}{c}, \tag{6.15}$$

since the time-dilation factor $(1-\beta^2)^{-1/2}$ is always greater than unity. Equations (6.15) and (6.7) agree of course for a constant time-dilation factor in the latter equation.

[68] This is the standard interpretation in classical physics. There are some dissenters, however. Following up ideas put forward by O.M.P. Bilaniuk, V.K. Deshpande and E.C.G. Sudershan, Am. J. Phys. **39**, 718 (1962), G. Feinberg, Phys. Rev. **159**, 1089 (1967) has developed a tentative theory of faster-than-light particles ("tachyons"). No experimental evidence exists to date for the existence of such objects.

Thus, moving clocks run slow in comparison with the clocks of the stationary observer. A similar analysis shows that the observer in Σ' (who views Σ as moving) claims that Σ's clock is running slow. Each observer thinks that the other's clock is slow!

(ii) *Simultaneity:* Time-dilation also plays havoc with the "common sense" idea of simultaneity, which now becomes a relative concept too: Events which occur simultaneously with respect to a stationary observer, are not simultaneous when viewed by a moving observer. For, suppose two events A and B a distance d apart in their rest frame Σ' occur simultaneously, i.e. have the same time coordinate (zero say) in *this* frame. Now let Σ' move with velocity $\mathbf{V}$ along the x-axis of a second frame Σ as in Fig. 6.1. The events A and B have space-time coordinates (x_a, t_a) and (x_b, t_b) in Σ. Their time-separation in Σ is thus

$$\Delta t = t_b - t_a = \frac{vd}{c^2}\frac{1}{\sqrt{1-\beta^2}}, \quad \beta = \frac{v}{c}, \tag{6.16}$$

by (6.14). But according to (6.15), the clock at A would register an interval

$$\Delta t' = \Delta t\sqrt{1-\beta^2} = \frac{vd}{c^2} \tag{6.17}$$

in the interval Δt, i.e. would be found to be running *ahead* of the clock at B by $\frac{vd}{c^2}$. The moving observer would claim that the A and B events are not simultaneous.

Lest the reader feel that such a state of affairs is pure Lewis Carroll, we examine these conclusions by way of experiment. One of the best known examples of time-dilation shows up in the decay of μ-mesons that are produced in the upper atmosphere in cosmic ray phenomena. The μ-meson (muon) decays into an electron and two neutrinos with a lifetime of about 2 microseconds when the muon is *at rest*. Even moving with a velocity near that of light, this would imply a penetration depth into the earth's atmosphere now exceeding $d \simeq 600$ meters. Yet a copious flux of muons is observed at sea level ($d \simeq 2000$ meters). How does the muon get where it got? Time-dilation tells us that if the muon has a lifetime of $\tau_0 \simeq 2\mu$ sec in its rest frame, we observe it to decay at the slower rate of $\tau_0(1-\beta^2)^{-1/2}$, and thus travel a distance $d = \beta c \tau_0 (1-\beta^2)^{-1/2}$ into our atmosphere. Experimentally, a factor $(1-\beta^2)^{-1/2} \simeq 10$ is required to explain the observed muon flux at sea level. An independent measurement[69] of the muon energy (to find β) confirms this number.

(iii) *Lorentz contraction:* But how does this state of affairs appear to the moving muon? In its rest frame it still has a lifetime τ_0 which is too short to allow it time to penetrate down to sea level. The answer to this apparent paradox is provided by an effect similar to time dilation, *viz.* a moving measuring rod is shortened relative to its length at rest. This is called the Lorentz contraction. To prove this statement, we provide the observer in Σ' with a measuring rod which is laid along the x' axis.

[69] D.H. Frisch and J.H. Smith, Am. Journ. of Physics, **31**, 342 (1962).

The ends of the rod are observed by Σ' to have coordinates x'_1 and x'_2, so its length is $l_0 = x'_2 - x'_1 > 0$ in the rest system. The *times* at which these measurements are made are immaterial. The length l_0 is called the *proper length* of the rod. What length does the observer in Σ measure? He assigns space-time coordinates (x_1, t_1) and (x_2, t_2) to the ends of the rod. Thus the length he gets will depend on the *times* that he makes his observation. Clearly what we mean by length of a rod is its length "now", i.e. he must measure the coordinates of both ends *simultaneously*. (Operationally this means placing observers along the x-axis in Σ with clocks that have previously been synchronized by light signals, say). These measurements will not be simultaneous in the rest system of the rod, but that is of no consequence. The recorded length, obtained in this way in Σ is noted as $l = x_2 - x_1$. But from (6.13)

$$l_0 = x'_2 - x'_1 = \frac{x_2 - x_1}{\sqrt{1-\beta^2}} = \frac{l}{\sqrt{1-\beta^2}}, \tag{6.18}$$

or

$$l = l_0\sqrt{1-\beta^2} < l_0. \tag{6.19}$$

The length l_0 of the moving rod is contracted by the factor $\sqrt{1-\beta^2}$. Just as moving clocks ran slow, we now discover that moving rods get shorter. Observe the complementary nature of these two statements. From the point of view of the Lorentz transformation one cannot have one without also having the other, and it is this feature which removes the muon's dilemma. For, from the point of view of the muon, it is the atmosphere that is moving. Consequently the muon only has to go *this* distance in its own lifetime τ_0, i.e. $d\sqrt{1-\beta^2} = \beta c \tau_0$, which is exactly what our earthbound observer calculates. We record in passing that for a long time it was tacitly assumed that a Lorentz contraction would be visible as a flattening of a three dimensional object in the direction of its motion. This assumption is *false*, as has been shown by Terrell *J.L. Terrell, Phys. Rev.* ***116****, 1041 (1959). See also V.F. Weisskopf, Physics Today* ***13****, 24 (1960).* The point is that what one means by "see" is the ocular image created in one's brain by light rays entering the eye at a given instant in time. These rays necessarily leave an extended body at different times, and can be shown to give the viewer the impression of a rotation, rather than a compression of the object.

(iii) *Doppler Shift*

The discussion so far has carefully avoided the question of what happens to clocks and meter sticks when viewed from non-inertial frames of reference. Such questions properly belong to the province of Einstein's General Theory of Relativity. However, it is possible to illustrate some effects of non-inertial frames within the framework of the special theory by looking at the Doppler shift. The Doppler shift, which has nothing to do with non-inertial frames *per se*, refers to the phenomena, shared by all wave-like disturbances, whereby the frequency of the emitted vibration appears to be higher (lower) to an observer moving towards (away from) the source. The analysis for a light emitter runs as follows. If the natural frequency of the source is ν_0 (i.e. the frequency recorded by an observer at rest relative to the source), then it has a frequency $\nu = \nu_0\sqrt{1-\beta^2}$ by time-dilation in the reference frame of an observer moving relative to the light source with speed $v = \beta c$. The time between successive light pulses in the observer's frame is thus v^{-1}. However, in this time the pulse travels a distance c/ν, the source a distance $\beta c/\nu$, so that the observer records pulses spaced at a distance $(1-\beta)c/\nu$ if he is approaching the source. The number of these pulses (which travel with velocity c) per second is thus recorded by the observer as

$$\nu_D = \frac{\nu}{1-\beta} = \nu_0\sqrt{\frac{1+\beta}{1-\beta}} \tag{6.20}$$

an equation which expresses the relativistic Doppler shift: The natural frequency of an approaching source is shifted to a higher value ("blued"), that of a receeding source to a lower value ("reddened").

(v) *Falling Photons turn Blue*

By means of the Doppler shift we can study the effect of a uniform gravitational field on the frequency of a light source placed in that field. The argument is based on the experimental equivalence of gravitational and inertial mass, plus the validity of Newton's laws of motion. Given these facts, one demonstrates that the motion of a mass in falling under gravity with acceleration g is indistinguishable from the motion of the same mass moving in gravity-free space, but viewed from a *non-inertial* reference frame accelerating upwards with acceleration g. This statement is an example of Einstein's *equivalence principle*. Here it follows as a consequence of known experimental facts in the special case of the constant gravitational acceleration of a mass point. Einstein elevated it to a *general* principle, applicable to arbitrary gravitational accelerations, and embracing the motion of light rays as well as particles[70]. We can use this principle to study the behavior of a light source in a uniform gravitational field. Consider a stationary source S of natural frequency ν_0 placed at a vertical height h above an observer on the surface of the earth. If h is small, the gravitational field in which source (and observer) find themselves has uniform value $g(\sim 980\ \mathrm{cm/s^2})$ pointing from source to observer. The equivalence principle maintains that any effect of this gravitational field on the source S must be the same as what would be observed if source *plus* observer were moving *upwards* with acceleration in gravity-free space. Starting from rest in the latter situation, the observer reaches a velocity $v = gh/c$ in the time h/c that a light-pulse from S takes to reach him if we assume he does not change position appreciably during this time. But since the observer is moving

[70] A concise discussion may be found in L. Landau and E. Lifschitz, *loc. cit.*, Chapter 10.

towards the source, he records a Doppler shift in the source frequency of

$$\frac{\nu_D - \nu_0}{\nu_0} \simeq \beta = \frac{gh}{c^2} \tag{6.21}$$

by (6.20) with $\beta = gh/c^2 \ll 1$. This equation holds equally for a *stationary* source in a uniform gravitational field, by the equivalence principle. It says that a source at a higher gravitational potential $\phi_S = gh$ appears "blue-shifted" to an observer at a lower potential $\phi_0 = 0$, by an amount

$$\Delta\nu = \frac{\nu_0}{c^2}(\phi_S - \phi_0), \tag{6.22}$$

where $\phi_S - \phi_0 = gh$. The validity of (6.22), which coincides with the more general result for weak but not necessarily uniform fields, $\phi_S - \phi_o \neq gh$, that one obtains in the General Theory[71], has been demonstrated in an earthbound laboratory. In a remarkable experiment, Pund and Rebka[72] have shown that photons emitted by ^{57}Fe were "blued" by just the predicted amount after "falling" a measured distance under gravity. Previously, the prediction (6.22) could only be tested in terms of the "red-shift" expected for light emitted by massive, distant stars (see Problems).

Two aspects of the result (6.22) should be emphasized. Firstly if observer and source interchange positions, so that the observer is at a higher potential relative to the source, the light reaching him will be "red-shifted". Stated more specifically in terms of atomic periods of the atoms constituting the light source, this means that all observers *agree* that the "clock" at a lower gravitational potential runs slower than an identical clock at a higher potential. This contrasts for example with the opinions of observers in uniform relative motion, who always maintain that the other's clock is running slow. The second point to make in regard to (6.22) is that the General Theory teaches us that the speeding up of an atomic clock in a gravitational field, rather than the loss of "gravitational potential energy" by the emitted photons is the correct physical interpretation of this relation.

An appreciation of the different behavior of clocks in inertial *vs* non-inertial reference frames is also the clue to understanding the oft-discussed "twin paradox". There now seems to be consensus that the travelling twin *does* age slower, and furthermore that *both* twins agree that this is so (if the effect of accelerations on the observations of the travelling twin are not considered the twins don't agree: That is the paradox). A careful discussion of the twin paradox is only really possible using the full machinery of the General Theory of Relativity. We refer the reader to Moller's book quoted at the beginning of this section for such discussions.

[71] L. Landau and E. Lifschitz, *ibid.*, Chapter 10.

[72] R.V. Pound and G.A. Rebka, Phys. Rev. Lett., **4**, 337 (1960).

6-5 *Four-vectors and Lorentz-invariance*

We saw in Sec. 6-4 that the coordinates of an event in space-time are related by $x' = Ax$, or

$$x'_\mu = \sum_{\nu=1}^{4} A_{\mu\nu} x_\nu, \mu = 1,2,3,4 \tag{6.23}$$

when viewed from two different inertial frames Σ or Σ'. Here, $A_{\mu\nu}$ form the elements of the orthogonal transformation matrix A, for which $Det A = +1$ must hold. For the special case considered previously, where Σ' moves along the positive x-axis of Σ with speed V and coincides with Σ at $t = 0$, (6.23) has the explicit form

$$\begin{pmatrix} x' \\ y' \\ z' \\ ict' \end{pmatrix} = \begin{pmatrix} \gamma & 0 & 0 & i\beta\gamma \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ -i\beta\gamma & 0 & 0 & \gamma \end{pmatrix} \begin{pmatrix} x \\ y \\ z \\ ict \end{pmatrix}, \tag{6.24}$$

where $\beta = v/c$ and $\gamma = (1-\beta^2)^{-1/2}$. Of course this is just a more succinct way of writing down the result (6.13). The inverse transformation is

$$\begin{pmatrix} x \\ y \\ z \\ ict \end{pmatrix} = \begin{pmatrix} \gamma & 0 & 0 & -i\beta\gamma \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ i\beta\gamma & 0 & 0 & \gamma \end{pmatrix} \begin{pmatrix} x' \\ y' \\ z' \\ ict' \end{pmatrix}. \tag{6.25}$$

Any four-component entity which transforms like x_μ is called a *four-vector*. Apart from the position four-vector $x_\mu = (x, y, z, ict) = (\mathbf{r}, ict)$ of a particle, the important four-vectors for describing relativistic motion are four-vectors associated with the velocity, momentum and acceleration of a particle. It is clear that the generalization of the concepts in Special Relativity has to be done with one eye on the transformation property (6.23). Otherwise we will run into trouble trying to satisfy the second part of Einstein's Principle of Relativity, viz. that all laws of motion be identical in all inertial frames. Consider the concept of particle velocity. An observer in Σ watches a particle change its position $\mathbf{r}(t)$ by $d\mathbf{r}$ in a time dt as measured by his clock and assigns it a velocity $\mathbf{v} = d\mathbf{r}/dt$. The change in four-vector position in this time is

$$dx_\mu = [d\mathbf{r}, icdt], \tag{6.26}$$

which is also a four-vector. However, the quotient dx_μ/dt is *not*, since both dx_μ and dt change when viewed from different reference frames. Therefore, how the time-dependence of x_μ is to be characterized requires some care. We have seen that the proper time interval ds/c, i.e. the time interval as measured in a reference frame attached to the particle and moving with it, is the same in all inertial frames. We distinguish the time interval by the symbol $d\tau$ from now on, $d\tau = ds/c$. Then, not dx_μ/dt, but rather

$$v_\mu = \frac{dx_\mu}{d\tau} = \frac{dx_\mu}{dt}\frac{dt}{d\tau} = [\frac{\mathbf{v}}{\sqrt{1-\beta^2}}, \frac{ic}{\sqrt{1-\beta^2}}] \tag{6.27}$$

transforms like a four-vector if we regard the position four-vector as a function of the proper time of the particle, $x_\mu = x_\mu(\tau)$. The construct v_μ is called the *four-velocity*. The last step in (6.27) follows from the time-dilation between observer-time dt and proper-time $d\tau = dt\sqrt{1-\beta^2}$, where $v = \beta c$ is the particle velocity. A second differentiation with respect to τ produces another four-vector,

$$f_\mu = \frac{dv_\mu}{d\tau} = [\frac{1}{\sqrt{1-\beta^2}}\frac{d}{dt}(\frac{\mathbf{v}}{\sqrt{1-\beta^2}}), \frac{ic}{\sqrt{1-\beta^2}}\frac{d}{dt}(\frac{1}{\sqrt{1-\beta^2}})], \tag{6.28}$$

called the *four-acceleration*.

Rotations in real three-dimensional space leave the length, or norm, of a vector unchanged. The same is true for the norm of a four-vector, which is unchanged by the transformation (6.23). This is called the *Lorentz-invariance* of the norm. Similarly, the scalar product $\sum_\mu A_\mu B_\mu$ of two four-vectors A_μ and B_μ is a Lorentz invariant. The property of Lorentz invariance is a very useful one, since it allows us to calculate norms and scalar products of four-vectors in any convenient frame. For example, the norm of the four-velocity v_μ in the frame attached to the particle, or *rest frame* (we denote all such quantities by a prime) is simply $\sum v'^2_\mu = -c^2$. Hence, by Lorentz invariance

$$\sum_\mu v_\mu^2 = -c^2 \tag{6.29}$$

holds for v_μ in any inertial frame. Likewise, the scalar product $\sum_\mu f_\mu v_\mu$ vanishes in any inertial frame, since

$$\sum_\mu f_\mu v_\mu = \frac{1}{2}\frac{d}{d\tau}(\sum_\mu v_\mu^2) = \frac{1}{2}\frac{d}{d\tau}(\sum_\mu v'^2_\mu) = 0 \tag{6.30}$$

and $\sum_\mu v'^2_\mu$ is constant in the rest frame of the particle.

The construction of v_μ and f_μ, as well as their properties, only depended on the *form* of the transformation (6.23) and the relation between proper time and observer time. The actual values of the $A_{\mu\nu}$ were not relevant. The developments in the following paragraphs will also be of this nature. Examination of the behavior of entities under transformations that respect the fundamental invariance of the infinitesimal interval $ds = cd\tau$ will become the guiding element when we discuss particle kinematics and particle dynamics. Such a point of departure has a much more dependable mathematical "feel" than the repeated introduction of

"Gedanken" experiments involving clocks and measuring rods. However, we have to use the Lorentz transformation explicitly whenever the values of a four-vector are required in two different frames. The question of how to combine velocities in relativity is a case in point. An observer in a frame $\sum'$, moving with velocity V along the x-axis of a second frame Σ (see Fig. 6.1 again), records the velocity of a particle as v'. What velocity does this particle have with respect to Σ? From (6.14), the space and time increments as seen from Σ and Σ' are related by

$$dx = \frac{dx' + V dt'}{\sqrt{1 - V^2/c^2}}, \quad dy = dy', \quad dz = dz', \quad dt = \frac{dt' + \frac{V}{c^2} dx'}{\sqrt{1 - V^2/c^2}}, \tag{6.31}$$

so that, since $\mathbf{v} = d\mathbf{r}/dt$ is the velocity recorded in Σ, $\mathbf{v}' = d\mathbf{r}'/dt$ that recorded in Σ',

$$v_x = \frac{dx' + V dt'}{dt' + \frac{V}{c^2} dx'} = \frac{v'_x + V}{1 + V v'_x/c^2} \tag{6.32}$$

$$v_y = \frac{dy'}{dt' + \frac{V}{c^2} dx'} \sqrt{1 - \frac{V^2}{c^2}} = \frac{v'_y \sqrt{1 - \frac{V^2}{c^2}}}{1 + V v'_x/c^2} \tag{6.33}$$

$$v_z = \frac{dz'}{dt' + \frac{V}{c^2} dx'} \sqrt{1 - \frac{V^2}{c^2}} = \frac{v'_z \sqrt{1 - \frac{V^2}{c^2}}}{1 + V v'_x/c^2}. \tag{6.34}$$

The curious structure of these equations runs contrary to intuition as with so many results in relativistic physics. However, the common denominator $1 + V v'_x/c^2$ is essential if the velocity c is not to be exceeded. For consider a light ray moving along the x-axis. Its velocity is recorded as c in Σ'. Now let Σ' be moving with velocity c relative to Σ. The velocity Σ records is $v_x = c$ according to (6.32). However, the *direction* that a light ray travels relative to observers in Σ and Σ' *is* affected by their relative motion (abberation of light, see Problems).

6-6 Momentum and Energy

We now wish to study the motion of a free particle of mass m moving with velocity v, by means of the Principle of Least Action that was discussed in Chapter 1. We do so by constructing an integral for the action function S that satisfies the following two conditions. (i) The integral must be Lorentz-invariant in order to satisfy the Principle of Relativity and (ii) its integrand must be a differential of the first order. The only Lorentz-invariant differential of first order that refers to a single particle is the invariant interval ds introduced in (6.4). In addition the integrand for S must have the dimensions of energy × time. The only combination of this nature that is available for a free particle is

$mc^2 d\tau = mcds$. Hence,

$$S_{\text{free}} = -mc \int_{s_1}^{s_2} ds \tag{6.35}$$

is an expression for S with the required attributes. The integral is taken along a world line of the particle that connects the two events s_1 and s_2 in its history. The minus sign ensures that S has a *minimum* value if the world line connecting s_1 and s_2 is a straight line, corresponding to the actual motion of the particle along a straight world line in free space (A little reflection will show that the distance $\int_{s_1}^{s_2} ds$ is a maximum when the integration is along a straight world line. This contrasts with the case in real space where integration along a straight line gives the shortest distance in free space. The difference arises because of the imaginary time coordinate in ds).

A useful expression for S is obtained by introducing the observer's time interval dt via $ds = c\sqrt{1-\beta^2}dt$ for ds in (6.35). Then

$$S_{\text{free}} = -mc^2 \int_{t_1}^{t_2} \sqrt{1 - \frac{v^2}{c^2}}\, dt, \tag{6.36}$$

which in turn identifies a possible Lagrange function for a free particle.

$$L_{\text{free}} = -mc^2 \sqrt{1 - \frac{v^2}{c^2}}. \tag{6.37}$$

For small velocities, $L_{\text{free}} \simeq -mc^2 + \frac{1}{2}mv^2$, which differs from the nonrelativistic Lagrangian $\frac{1}{2}mv^2$ by the constant $-mc^2$. However, we have seen that L is not unique. An equally suitable function for a free particle would be

$$L'_{\text{free}} = L_{\text{free}} + mc^2 = -mc^2(\sqrt{1 - \frac{v^2}{c^2}} - 1). \tag{6.38}$$

The momentum $\mathbf{p}$ is given by

$$\mathbf{p} = \frac{\partial L_{\text{free}}}{\partial \mathbf{v}} = \frac{m\mathbf{v}}{\sqrt{1 - \frac{v^2}{c^2}}}, \tag{6.39}$$

irrespective of which form for L_{free} is used. However, the first form, (6.37) is preferable. This preference has to do with the expression for the energy of a free particle. From (1.78) of Chapter 1, the energy associated with the system described by L_{free} is

$$\mathcal{E} = \boldsymbol{p} \cdot \boldsymbol{v} - L_{\text{free}} = \frac{mc^2}{\sqrt{1 - \frac{v^2}{c^2}}}. \tag{6.40}$$

Thus, $\mathcal{E}$ reduces, not to zero, but to the so-called *rest mass* energy mc^2 when $\mathbf{v} = 0$. By contrast, L'_{free} describes a system with energy $E_k =$

$\mathcal{E} - mc^2$ measured relative to the rest mass energy. The quantity

$$E_k = \mathcal{E} - mc^2 = mc^2[\frac{1}{\sqrt{1-\frac{v^2}{c^2}}} - 1] \tag{6.41}$$

is called the *kinetic energy*.

The fact that the energy $\mathcal{E}$ is more fundamental than E_k emerges from its transformation properties under Lorentz transformations. The fact that $\mathbf{p}$, as given by (6.39) and $\mathcal{E}$ together constitute the components of the following four-vector,

$$p_\mu = [\mathbf{p}, \frac{i}{c}\mathcal{E}] \tag{6.42}$$

called the *four-momentum*, or momentum-energy four-vector, means the $\mathbf{p}$ and $\mathcal{E}$ are connected with the corresponding values $\mathbf{p}'$ and $\mathcal{E}'$ in a moving frame Σ' via (6.25), i.e.

$$\begin{pmatrix} p_x \\ p_y \\ p_z \\ i\mathcal{E}/c \end{pmatrix} = \begin{pmatrix} \gamma & 0 & 0 & -i\beta\gamma \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ i\beta\gamma & 0 & 0 & \gamma \end{pmatrix} \begin{pmatrix} p'_x \\ p'_y \\ p'_z \\ i\mathcal{E}'/c \end{pmatrix}. \tag{6.43}$$

Since $\mathbf{p}' = 0$ and $\mathcal{E} = mc^2$ in the rest frame of the particle, the Lorentz-invariance of $\sum_\mu p_\mu^2$ provides a particularly convenient way of determining the energy-momentum relation for a free particle. The rest frame value is $\sum_\mu p_\mu'^2 = -(mc)^2$. Therefore by Lorentz invariance $\sum_\mu p_\mu^2 = -(mc)^2$, or

$$p^2 - \frac{\mathcal{E}^2}{c^2} = -(mc)^2, \tag{6.44}$$

where $\mathbf{p}$ and $\mathcal{E}$ refer to the observer's frame. Elimination of $(1-\frac{v^2}{c^2})^{\frac{1}{2}}$ between (6.39) and (6.40) provides us with the particle velocity

$$\mathbf{v} = \mathbf{p}\frac{c^2}{\mathcal{E}} = \frac{\partial \mathcal{E}}{\partial \mathbf{p}}. \tag{6.45}$$

The energy-momentum relation (6.44) for a relativistic particle leads to a hyperbolic, rather than a parabolic relation between energy and momentum. We illustrate this in Fig. 6.2 after rewriting (6.44) as

$$(\frac{p}{mc})^2 + (\frac{\mathcal{E}}{mc^2})^2 = -1 \tag{6.46}$$

which is the equation for a hyperbola with semi-axes mc and mc^2.

Since the gradient $\partial\mathcal{E}/\partial\mathbf{p}$ determines the particle velocity at each point according to (6.45), the asymptotes $\mathcal{E} = \pm cp$ in Fig. 6.2 describe the energy-momentum relation of "particles" moving with the speed of light. Such "particles" must necessarily have zero rest mass. Otherwise the energy of such particles would always be infinite according to (6.40).

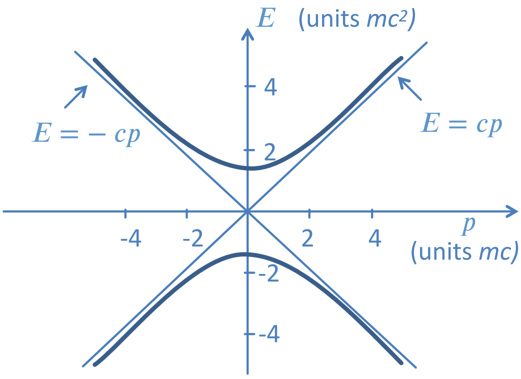

Figure 6.2: The energy-momentum relation for a relativistic free particle.

Photons, or "particles" or light are an example of entities satisfying the energy-momentum relation $\mathcal{E} = cp$. In fact, the four-vector character of $[\mathbf{p}, i\mathcal{E}/c]$, together with the photon picture of light $\mathcal{E} = cp$, where $\mathcal{E} = h\nu$, h = Planck's constant = 6.63×10^{-34} joule sec, is the energy carried by the photons constituting a light wave of frequency ν, leads to the formulae describing the Doppler shift and aberration of light in a natural way (see Problems.)

Returning to the relation (6.44) for material particles, we note that Fig. 6.2 provides visual assurances that the particle speed v cannot exceed c, since the gradient $\partial\mathcal{E}/\partial\mathbf{p}$ of the hyperbola is always *less* than the gradient of the relevant asymptote. The other aspect to notice about Fig. 6.2 is that the energy-momentum relation is two-valued, there being two possible values $\pm\mathcal{E}$ of the energy for each value of $\mathbf{p}$. These negative energy values, also shown in Fig. 6.2 are properly dismissed as unphysical in classical physics. However, all this changes in quantum physics, where it can be shown that the negative energy solutions do have a physical interpretation[73] in terms of a "sea" of occupied "antiparticle" states.

The inclusion of the rest mass energy mc^2 in $\mathcal{E}$ is an essential feature of (6.44). This has the consequence that the separate conservation of mass and energy of classical physics is amalgamated into a single mass-energy conservation law expressed by (6.44). Mass has to be considered a form of energy that is not distinct from other forms of energy and is not conserved separately. Consider an atomic nucleus by way of a concrete example. Its mass is M say, taken as a whole, and so the nucleus has energy Mc^2 at rest. However, the constituent nucleons, A in number, possess energy by virtue of their kinetic motion and mutual interaction, in addition to their rest energies $m_i c^2$. Consequently the sum $\sum_{i=1}^{A} m_i c^2 \neq Mc^2$. A part of the rest energy Mc^2 of the nucleus resides in the nucleon interations. In fact, the rest mass energy Mc^2 of a *stable* nucleus is always smaller than the sum of the masses of its constitutent

[73] See for example, J. D. Bjorken and S. D. Drell, *Relativistic Quantum Mechanics,* (McGraw-Hill Book Company, New York and London, 1964), Chapter 5.

nucleons by an amount

$$-B = (Mc^2 - \sum_{i=1}^{A} m_i c^2), \tag{6.47}$$

where $B > 0$ is called the binding energy. An energy B has to be supplied to break up M into its constituents m_i. Conversely, if $B < 0$, the system is unstable and will disintegrate with emission of energy, usually in the form of kinetic energy of the decay products. The advent of nuclear power sources is made possible by the conversion of mass-energy into some other (usable) form of energy. In the fission of uranium nuclei by neutron bombardment for example, the mass energies of the uranium and its fission products differ by about 200 MeV. This amounts to an energy release of nearly 10^{14} joules/kg of fissile material!

The conversion of a part of the rest mass energy into other forms of energy during fission is certainly one graphic illustration of the validity of (6.40) for describing the energy of a free particle. The other aspect of this relation, i.e. how the energy of a relativistic free particle depends on its speed and the existence of a limiting speed, has been beautifully illustrated by examining the speed and energy gain of electrons after passing down the accelerating tube of a linear accelerator (W. Bertozzi, Am. J. Phys. **32**, 551 (1964)). The experiment measures the speed v and kinetic energy E_k *independently* that the electrons attain in a given accelerating potential. A plot of the square of the speed attained for a given kinetic energy is shown in Fig. 6.3.

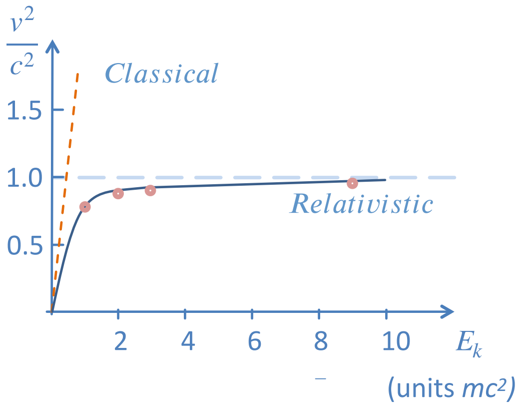

Figure 6.3: Experimental confirmation of the relativistic relation between the speed and the kinetic energy of relativistic electrons. The open circles denote the experimental points.

The deviation of the experimental points from the non-relativistic relation $v^2/c^2 = 2E_k/mc^2$ in favour of the relativistic prediction

$$\frac{v^2}{c^2} = 1 - \frac{1}{(1 + \frac{E_k}{mc^2})^2}, \tag{6.48}$$

which follows from (6.41), gives a convincing illustration of the correctness of the latter over the former. The figure also suggests that no appreciable increase in speed is brought about by increasing the energy supplied to the electron, once this energy exceeds the rest mass energy $mc^2 = 0.511$ MeV of the electron.

6-7 *Particles and Fields*

So far, we have avoided the question of how to introduce the idea of interactions into Special Relativity. We now consider this question, again from the point of view of constructing a suitable action function S for the interacting system. A prototype of the sort of problem one faces is the relativistic motion of a charge e in an electromagnetic field. This problem is relatively simple because of the known fact that the vector and scalar potentials $\mathbf{A}$ and ϕ describing the electromagnetic field in free space form the components of a four-vector

$$A_\mu = [\mathbf{A}, i\frac{\phi}{c}], \tag{6.49}$$

called the *four-potential* of the field. This result is a consequence of the *automatic* invariance of Maxwell's equations under Lorentz transformations. It is thus a simple matter to construct a Lorentz-invariant expression representing the particle-field interaction term in S. We write

$$S = S_{\text{free}} + S_{\text{int}}, \tag{6.50}$$

where S_{free} refers to a free particle. As for S_{int}, describing the particle-field interaction, we know that this term must (i) be a Lorentz-invariant, and (ii) contain quantities that refer to both the particle as well as field variables in order to describe a particle-field interaction. The simplest expression meeting these requirements is

$$S_{\text{int}} = e\int_{x_1}^{x_2} \sum_\mu A_\mu dx_\mu, \tag{6.51}$$

where $dx_\mu = v_\mu d\tau$ is the four-displacement in the proper time interval $d\tau$ and A_μ is evaluated at the particle-position $x_\mu = x_\mu(\tau)$. The electric charge e, which is a Lorentz-invariant property of the particle, serves as the "coupling constant" between particle and field. Multiplying S_{int} by additional constant amounts to setting the units of charge and electromagnetic field. The choice (6.51) is equivalent to using MKS units throughout. Writing $dx = (v_\mu/c)\, ds$ and adding the resulting expression for S_{int} to S_{free}, one has

$$S = \int_{s_1}^{s_2} [-mc + \frac{e}{c}\sum_\mu v_\mu A_\mu]\, ds. \tag{6.52}$$

Introducing the observer's time interval dt in place of ds again, this becomes

$$S = \int_{t_1}^{t_2} [-mc^2\sqrt{1-\frac{v^2}{c^2}} - e\phi + e(\mathbf{v}\cdot\mathbf{A})]\, dt, \tag{6.53}$$

since $\sum_\mu A_\mu \, ds = (-\phi + \mathbf{v} \cdot \mathbf{A}) \, dt$. The Lagrange function for a charged particle in an electromagnetic field can now be read off as

$$L = -mc^2 \sqrt{1 - \frac{v^2}{c^2}} - e\phi + e(\mathbf{v} \cdot \mathbf{A}). \tag{6.54}$$

This expression only differs from its non-relativistic analogue (2.73) of Chapter 2, in the term describing the free particle. Since the electromagnetic interaction piece is already Lorentz-invariant (although we had no way of ascertaining this in Chapter 2), it comes through unchanged in form. We can now "turn the crank" once more to obtain the momentum and energy,

$$\mathbf{p} = \frac{\partial L}{\partial \mathbf{v}} = \frac{m\mathbf{v}}{\sqrt{1 - \frac{v^2}{c^2}}} + e\mathbf{A} \tag{6.55}$$

and

$$\mathcal{E} = \frac{mc^2}{\sqrt{1 - \frac{v^2}{c^2}}} + e\phi, \tag{6.56}$$

for a charged particle in an electromagnetic field.

Since the contribution of the interaction terms with the electromagnetic field are the same in these results as in the non-relativistic case of Chapter 2, the equation of motion of the charge is still given by (2.77), but with the mechanical momentum modified by the factor $(1 - v^2/c^2)^{1/2}$. Hence,

$$\frac{d}{dt}\left(\frac{m\mathbf{v}}{\sqrt{1 - \frac{v^2}{c^2}}}\right) = e\mathbf{E} + e(\mathbf{v} \times \mathbf{B}) \tag{6.57}$$

replaces the classical equation of motion for a charge in an electromagnetic field $(\mathbf{E} \cdot \mathbf{B})$.

To find the energy-momentum relation in the presence of an electromagnetic field, we observe that $[\mathbf{p} - e\mathbf{A}, \frac{i}{c}(\mathcal{E} - e\phi)]$ is a four-vector. Lorentz invariance of the norm of this four-vector yields

$$(\mathbf{p} - e\mathbf{A})^2 - \frac{1}{c^2}(\mathcal{E} - e\phi)^2 = -mc^2. \tag{6.58}$$

6-8 A Lorentz-Invariant Lagrangian

Although the action integrals in (6.51) and (6.52) are Lorentz-invariant, the Lagrange functions they lead to are not. Other formulations of the relativistic particle problem exist which use Lagrange functions that are manifestly Lorentz-invariant. By analogy with the arguments of the non-relativistic case, consider an action function given by the integral

$$S = \int_{\tau_1}^{\tau_2} \mathcal{L}[x_\mu(\tau), v_\mu(\tau)] \, d\tau \tag{6.59}$$

over the proper time interval $\tau_2 - \tau_1$ associated with the particle. The function $\mathcal{L}$ depends on the position and velocity four-vectors $x_\mu(\tau)$ and

$v_\mu(\tau)$. Note that the four-velocity components v_μ are not independent, since $\sum_\mu v_\mu^2 = -c^2$. Introducing a suitable Lagrange-multiplier to take care of this, one finds that the variation problem $\delta S = 0$ with $\delta x_\mu(\tau)$ vanishing at the end points is satisfied, provided that

$$\frac{d}{d\tau}\left(\frac{\partial \mathcal{L}}{\partial v_\mu}\right) = \frac{\partial \mathcal{L}}{\partial x_\mu}, \quad \mu = 1,2,3,4, \tag{6.60}$$

a by now familiar equation. A suitable Lagrange function of this type for a particle in an electromagnetic field is

$$\mathcal{L} = \frac{1}{2}\sum_\mu m v_\mu^2 + e\sum_\mu v_\mu A_\mu. \tag{6.61}$$

Notice that the "kinetic" term $\sum m v_\mu^2 = -mc^2$ is actually a constant. However, as stressed in Chapter 1, it is the *functional form* of $\mathcal{L}$ that matters. The required manipulations on the function $\mathcal{L}$ are simple. One has

$$\frac{d}{d\tau}(m v_\mu + e A_\mu) = e\sum_\nu \frac{\partial A_\nu}{\partial x_\mu} v_\nu, \tag{6.62}$$

a result which may be written as

$$\frac{d}{d\tau}(m v_\mu) = e\sum_\nu F_{\mu\nu} v_\nu \tag{6.63}$$

if we transpose $dA_\mu/d\tau = \sum_\nu (\partial A_\mu/\partial x_\nu) v_\nu$ to the right hand side and introduce the abbreviation

$$F_{\mu\nu} = \frac{\partial A_\nu}{\partial x_\mu} - \frac{\partial A_\mu}{\partial x_\nu} = -F_{\nu\mu}. \tag{6.64}$$

The entity $F_{\mu\nu}$ is by construction an antisymmetric tensor of rank two and therefore has six independent entries. For this reason $F_{\mu\nu}$ often goes by the name *electromagnetic six-vector*, which terminology becomes clear on writing out its components in terms of the components of the electromagnetic fields **E** and **B**,

$$\mathbf{E} = -\operatorname{grad}\phi - \frac{\partial \mathbf{A}}{\partial t}, \quad \mathbf{B} = \operatorname{curl}\mathbf{A}. \tag{6.65}$$

Then,

$$(F_{\mu\nu}) = \begin{pmatrix} 0 & B_z & -B_y & -\frac{i}{c}E_x \\ -B_z & 0 & B_x & -\frac{i}{c}E_y \\ B_y & -B_x & 0 & -\frac{i}{c}E_z \\ \frac{i}{c}E_x & \frac{i}{c}E_y & \frac{i}{c}E_z & 0 \end{pmatrix}. \tag{6.66}$$

The equation of motion given by (6.63) is manifestly invariant in form (or as we say, covariant) under Lorentz transformations. This circumstance is brought about because it is written explicitly in terms of four-vectors and/or derivatives thereof with respect to proper time that is a Lorentz-invariant concept. In terms of ordinary three-vectors and derivatives with respect to the observer's time, the equation of motion splits into two parts. The space part of (6.63) just duplicates (6.57). For the fourth, or "time" component, one finds

$$\frac{d}{dt}(\frac{mc^2}{\sqrt{1-\frac{v^2}{c^2}}}) = e\mathbf{E}\cdot\mathbf{v}, \tag{6.67}$$

expressing the fact that the increase in energy in time dt is supplied by the work done by the electric field on the charge in this interval. The covariant form of the equation of motion for the charge thus combines Newton's law of motion plus the rate of working of the applied forces (a consequence of Newton's law of motion) into a single, compact equation.

The results given in (6.57) and (6.67) suggest a way of introducing the idea of "force" into Special Relativity. When written in terms of the proper time of the particle, these equations become

$$\frac{d\mathbf{p}}{d\tau} = \frac{1}{\sqrt{1-\frac{v^2}{c^2}}}\mathbf{F}^{(l)} \tag{6.68}$$

and

$$\frac{d}{d\tau}(\frac{mc^2}{\sqrt{1-\frac{v^2}{c^2}}}) = \frac{\mathbf{v}\cdot\mathbf{F}^{(l)}}{\sqrt{1-\frac{v^2}{c^2}}} \tag{6.69}$$

in terms of the Lorentz force $\mathbf{F}^{(l)} = e\mathbf{E} + e(\mathbf{v}\times\mathbf{B})$ acting on the charge. Multiplying the lower equation by i/c shows that the set (6.68) and (6.69) is equivalent to the single equation

$$\frac{d\mathbf{p}}{d\tau} = f_\mu, \quad \mu = 1,2,3,4, \tag{6.70}$$

where the *four-force* f_μ is given by

$$f_\mu = [\frac{\mathbf{F}}{\sqrt{1-\frac{v^2}{c^2}}}, \frac{i}{c}\frac{\mathbf{v}\cdot\mathbf{F}}{\sqrt{1-\frac{v^2}{c^2}}}]. \tag{6.71}$$

Here F (now not necessarily $F^{(l)}$) is the force on the particle as it would appear in the non-relativistic equation of motion. Alternatively, one simply can *define* f_μ by (6.70). In either case, the four-vector nature of f_μ guarantees the invariance in form, covariance of (6.70) under Lorentz transformations. Note that $\sum_\mu v_\mu f_\mu = 0$, i.e. the four-force is always perpendicular to the four-velocity, in view of (6.30).

The form and transformation properties of (6.70) suggest that it is a suitable generalization of Newton's second law of motion for relativistic particles. The definition of "force" embodied in this equation also synthesises the two fundamental conservation laws of classical physics, those of energy and momentum, into a single law of the conservation of four-momentum when the four-force is zero. That is, if $f_\mu = 0$, then

$$p_\mu = [\mathbf{p}, \frac{i}{c}\mathcal{E}] = \text{constant}, \tag{6.72}$$

which can only hold if both the space and time components $\mathbf{p}$ and $\mathcal{E}$ of p_μ are constant.

The special theory of relativity, providing as it does a fertile field for new insights and gross misconceptions, remains a somewhat aloof, if essential element of classical physics. For this reason, we have attempted to keep the presentation on a reasonably elementary level and close to the basic principles of relativity. The most natural field of application for most of the material in this chapter concerns the motion of charged particles in given electromagnetic fields, the generation of such fields by charges in arbitrary motion, and finally the "feedback" of the field generated by a charge in motion, on that motion. Such problems require a simultaneous study, which is not attempted here, of the equations of motion governing both particle motion and field motion in the presence of particle-field coupling terms[74]. Neither have we made any mention (beyond the name!) of the principles underlying the General Theory of Relativity[75].

Problems

6-1. Some stars, called white dwarfs, have nearly the mass of the sun, but the size of the earth. Assume that (6.22) describes the frequency shift experienced by light emitted from a white dwarf and calculate a value for $\Delta\nu/\nu_0$. How do Doppler shifts, as per (6.20) enter into the picture? Can you suggest a procedure for untangling the two effects?

6-2. Doppler shift. A light source of frequency ν_0 moves towards an observer with speed $V = \beta c$. Show that the formula (6.20) for the frequency ν_D that is observed follows from the transformation properties of the associated energy $\mathcal{E}' = h\nu_0$ and momentum $p' = \mathcal{E}'/c$ of the photons emitted by the source.

6-3. The photons referred to in problem 6-2 have to be emitted by a light source, e.g. an excited atom or nucleus. Examine, via the conservation of four-momentum how the recoil of the emitting source influences the previous result for ν_D.

6-4. Aberration of light. A light source, moving with velocity $V = \beta c$ along the x-axis of a stationary reference frame, emits photons in a direction θ' with respect to its direction of travel, as observed in the rest frame of the source. Show that, according to an observer at rest, these photons are emitted in a direction θ with respect to the direction of travel, where

$$\cos\theta = \frac{\cos\theta' + \beta}{1 + \beta\cos\theta'}. \tag{6.73}$$

---

[74] See for example J.D. Jackson, *Classical Electrodynamics* (John Wiley and Sons, Inc., New York, 1962), Chapter 17.

[75] See for example, R. Adler, M. Bazin and M. Schiffer, *Introduction to General Relativity* (McGraw-Hill Book Company, New York and London, 1965).

This phenomenon is known as the *aberration of light*.

6-5. Consider reference frames Σ, Σ' and Σ'' in relative motion along a common x-axis, such that Σ'' has a velocity $V'' = \beta'' c$ relative to Σ', which in turn moves with velocity $V' = \beta' c$ relative to Σ. Prove that the velocity of Σ'' as seen by Σ is $V = \beta c$, where

$$\beta = \frac{\beta' + \beta''}{1 + \beta' \beta''}, \tag{6.74}$$

by considering the successive Lorentz transformation $\Sigma \to \Sigma' + \Sigma''$ as constituting the single transformation $\Sigma \to \Sigma''$. Hint: the transformations $\Sigma \to \Sigma'$ and $\Sigma' \to \Sigma''$ are "rotations" through angles ϕ' and ϕ'', where $\tan \phi' = i\beta'$ etc., that are equivalent to the single "rotation" $\phi = \phi' + \phi''$. Set $\tan \phi = i\beta$ and use the addition theorem for the tangent function.

6-6. Photons of sufficiently high energy can lose energy in matter by the process of pair production, where an electron-positron pair is created. Show that this process cannot take place without the presence of a third particle (an atomic nucleus) to take care of momentum and energy conservation.

6-7. There has been some speculation on the literature (see e.g. O.M. Bilaniuk und E.C.G. Sudershan, Physics Today **22**, 43 (1966) for an elementary exposition) regarding the possible existence and properties of particles ("tachyons") that always move with a velocity greater than that of light. Accepting this hypothesis, show that the every-velocity relation for tachyons implies that the rest mass parameter m in (6.40) must be replaced by an imaginary parameter im. Find the corresponding energy-momentum and energy-momentum-velocity relations for tachyons. From these show that the faster a tachyon goes, the *lower* is its energy.

6-8. The fact that interactions propagate with a finite speed implies that the concept of a "rigid" body has to be examined with care from the point of view of Relativity. A paradox that illustrates the sort of problems one runs into is the following (E.M. Dewan, Am. J. Phys. **31**, 342 (1963)): A polevaulter runs into a barn with his pole held horizontally. The rest length of the pole is such that it will not fit into the barn, but according to an observer in the barn, its Lorentz contracted length will fit in. Once the polvaulter is inside the barn, the barn door is slammed shut. Can the polevaulter "explain" the fact that the door can be shut behind him, since according to him, it is the barn that has contracted so that the pole cannot possibly fit into it?