1.2 It is a remarkable fact (to be commented on further in Chapter 9) that the observed hadronic states are extremely limited as to possible quantum numbers. For example, no doubly charged meson (e.g. 'π^{++}') has ever been discovered. Such a state can *not* be made out of $q\bar{q}$ combinations of known quarks—and hence its absence confirms the simple quark model picture. Suggest

(a) another mesonic state that can *not* be made out of known $q\bar{q}$ combinations, and

(b) a possible baryonic state that can *not* be made out of known qqq combinations;

and for each state in (a), (b), give a reaction in which your proposed state could, if it existed, be produced while respecting standard strong interaction conservation laws (e.g. $\pi^+ p \rightarrow$ 'π^{++}' + n).

Examine the evidence for the existence of baryons with $S = +1$.

2

ELECTROMAGNETISM AS A GAUGE THEORY

2.1 Introduction

The previous chapter reviewed briefly some of the main reasons for thinking that the simplest constituents of matter are quarks and leptons. As we have seen, both quarks and leptons appear to be pointlike down to the smallest distance scales currently attainable by the highest-energy particle accelerators. We must now introduce the main concern of this book—namely, the nature of the forces between these 'elementary particles'.

One of the relevant forces—electromagnetism—has been well understood in its classical guise for many years. Over a century ago, Faraday, Maxwell and others developed the theory of the electromagnetic interaction, culminating in Maxwell's paper of 1864 (Maxwell 1864). Today Maxwell's theory still stands—unlike Newton's 'classical mechanics' which was shown by Einstein to require modification at relativistic velocities—speeds approaching the velocity of light. Moreover, Maxwell's electromagnetism, when suitably married to quantum mechanics, gives us, in '*quantum electrodynamics*', or QED, what we call in Part II 'the best theory we have'. As we shall see in Chapter 6, this theory is in truly remarkable agreement with experiment. As we have already indicated, the theories of the weak and strong forces included in the standard model are generalisations of QED, and promise to be as successful as that theory. QED has therefore become the paradigmatic theory.

From today's perspective, the crucial thing about electromagnetism is that it is a theory in which the *dynamics* (i.e. the behaviour of the forces) is intimately related to a *symmetry* principle. In the everyday world, a symmetry operation is something that can be done to an object that leaves the object looking the same after the operation as before. By extension, we may consider mathematical operations—or 'transformations'—applied to the objects in our theory such that the physical laws look the same after the operations as they did before. Such transformations are usually called *invariances* of the laws. Familiar examples are, for instance, the translation and rotation invariance of all fundamental laws: Newton's laws of motion remain valid whether or not

we translate or rotate a system of interacting particles. But of course—precisely because they do apply to all laws, classical or quantum—these two invariances have no special connection with any particular force law. Instead, they constrain the form of the allowed laws to a considerable extent, but by no means uniquely determine them. Nevertheless, this line of argument leads one to speculate whether it might in fact be possible to impose further types of symmetry constraints so that the forms of the force laws *are* essentially determined. This would then be one possible answer to the question: why are the force laws the way they are? (Ultimately of course this only replaces one question by another!)

In this chapter we shall discuss electromagnetism from this point of view. This is not the historical route to the theory, but it is the one which generalises to the other two interactions. This is why we believe it important to present the central ideas of this approach in the familiar context of electromagnetism at this early stage.

A distinction that is vital to the understanding of all these interactions is that between a *global* invariance and a *local* invariance. In a global invariance the same transformation is carried out at all space–time points: it has an 'everywhere simultaneously' character. In a local invariance different transformations are carried out at different individual space–time points. In general, as we shall see, a theory that is globally invariant will not be invariant under locally varying transformations. However, by introducing new force fields that interact with the original particles in the theory in a specific way, and which also transform in a particular way under the local transformations, a sort of local invariance can be restored. We will see all these things more clearly when we go into more detail, but the important conceptual point to be grasped is this: one may view these special force fields and their interactions as existing in order to permit certain local invariances to be true. The particular local invariance relevant to electromagnetism is the well known *gauge invariance* of Maxwell's equations: in the quantum form of the theory this property is directly related to an invariance under local *phase transformations* of the quantum fields. A generalised form of this phase invariance also underlies the theories of the weak and strong interactions. For this reason they are all known as 'gauge theories'.

2.2 The Maxwell equations: current conservation

We begin by considering the basic laws of classical electromagnetism, the Maxwell equations. We use a system of units (Heaviside–Lorentz) which is convenient in particle physics (see Appendix C). Before Maxwell's work these laws were (in free space)

$$\nabla\cdot\boldsymbol{E} = \rho_{\text{em}} \quad \text{(Gauss's law)} \tag{2.1}$$

$$\nabla \times \boldsymbol{E} = -\partial\boldsymbol{B}/\partial t \quad \text{(Faraday–Lenz laws)} \tag{2.2}$$

$$\nabla\cdot\boldsymbol{B} = 0 \quad \text{(no magnetic charges)} \tag{2.3}$$

and, for steady currents,

$$\nabla \times \boldsymbol{B} = \boldsymbol{j}_{\text{em}} \quad \text{(Ampère's law).} \tag{2.4}$$

Maxwell noticed that taking the divergence of this last equation leads to conflict with the continuity equation for electric charge

$$\partial\rho_{\text{em}}/\partial t + \nabla\cdot\boldsymbol{j}_{\text{em}} = 0. \tag{2.5}$$

Since

$$\nabla\cdot(\nabla \times \boldsymbol{B}) = 0 \tag{2.6}$$

from (2.4) there follows the result

$$\nabla\cdot\boldsymbol{j}_{\text{em}} = 0. \tag{2.7}$$

This can only be true in situations where the charge density is constant in time. For the general case, Maxwell modified Ampère's law to read

$$\nabla \times \boldsymbol{B} = \boldsymbol{j}_{\text{em}} + \partial\boldsymbol{E}/\partial t \tag{2.8}$$

which is now consistent with (2.5). Equations (2.1)–(2.3), together with (2.8), constitute Maxwell's equations in free space.

The vitally important continuity equation (2.5) states that the rate of decrease of charge in any arbitrary volume Ω is due precisely and only to the flux of current out of its surface; that is, no net charge can be created or destroyed in Ω. Since Ω can be made as small as we please, this means that *electric charge must be locally conserved*: a process in which charge is created at one point and destroyed at a distant one is not allowed, despite the fact that it conserves charge overall, or 'globally'. The ultimate reason for this is that the global form of charge conservation would necessitate the instantaneous propagation of signals (such as 'now, create a positron over there'), and this conflicts with special relativity—a theory which, historically, flowered from the soil of electrodynamics. The extra term introduced by Maxwell—the 'electric displacement current'—owes its place in the dynamical equations to a local conservation requirement.

We remark at this point that we have just introduced another local/global distinction, similar to that discussed above in connection with invariances. In this case the distinction applies to a conservation law—since invariances are related to conservation laws in both classical and quantum mechanics, we should perhaps not be too surprised by this. However, as with invariances, conservation laws—such as charge conservation in electromagnetism—play a central role in gauge theories

in that they are closely related to the dynamics. The point is simply illustrated by asking how we could measure the charge of a newly created subatomic particle X. There are two conceptually different ways:

(i) We could arrange for X to be created in a reaction such as

$$A + B \rightarrow C + D + X$$

where the charges of A, B, C and D are already known. In this case we can use *charge conservation* to determine the charge of X.

(ii) We could see how particle X responded to known electromagnetic fields. This uses *dynamics* to determine the charge of X.

Either way gives the same answer: it is the conserved charge which determines the particle's response to the field. By contrast, there are several other conservation laws that seem to hold in particle physics, such as lepton number and baryon number, that apparently have no dynamical counterpart (cf the remarks at the end of §1.4). To determine the baryon number of a newly produced particle, we have to use B conservation and tot up the total baryon number on either side of the reaction. As far as we know there is no baryonic force field.

Thus gauge theories are characterised by a close interrelation between *three* conceptual elements: symmetries, conservation laws and dynamics. In fact, it is now widely believed that the *only* exact quantum number conservation laws are those which have an associated gauge theory force field—see comment (i) in §2.6 below. Thus one might suspect that baryon number is not absolutely conserved—as is indeed the case in proposed unified gauge theories of the strong, weak and electromagnetic interactions. In the above discussion we have briefly touched on the connection between two pairs of these three elements: symmetries ↔ dynamics, and conservation laws ↔ dynamics. The precise way in which the remaining link is made—between the symmetry of electromagnetic gauge invariance and the conservation law of charge—is more technical. We will discuss this connection with the help of simple ideas from quantum field theory in Chapter 4. For the present we continue with our study of the Maxwell equations and, in particular, of the gauge invariance they exhibit.

2.3 The Maxwell equations: gauge invariance

In classical electromagnetism, and especially in quantum mechanics, it is convenient to introduce the vector potential $A_\mu(x)$ in place of the fields $\boldsymbol{E}$ and $\boldsymbol{B}$. We write

$$\boldsymbol{B} = \boldsymbol{\nabla} \times \boldsymbol{A} \tag{2.9}$$

$$\boldsymbol{E} = -\boldsymbol{\nabla} V - \partial \boldsymbol{A}/\partial t \tag{2.10}$$

which defines the 3-vector potential $\boldsymbol{A}$ and the scalar potential V. With these definitions, equations (2.2) and (2.3) are then automatically satisfied.

The origin of gauge invariance in classical electromagnetism lies in the fact that the potentials $\boldsymbol{A}$ and V are not unique for given physical fields $\boldsymbol{E}$ and $\boldsymbol{B}$. The transformations which $\boldsymbol{A}$ and V may undergo while preserving $\boldsymbol{E}$ and $\boldsymbol{B}$ (and hence the Maxwell equations) unchanged are called gauge transformations, and the associated invariance of the Maxwell equations is called gauge invariance.

What are these transformations? Clearly $\boldsymbol{A}$ can be changed by

$$\boldsymbol{A} \rightarrow \boldsymbol{A}' = \boldsymbol{A} + \boldsymbol{\nabla}\chi \tag{2.11}$$

where χ is an arbitrary function, with no change in $\boldsymbol{B}$ since curl grad $\equiv 0$. To preserve $\boldsymbol{E}$, V must then change simultaneously by

$$V \rightarrow V' = V - \partial\chi/\partial t. \tag{2.12}$$

These transformations can be combined into a single compact equation by introducing the 4-vector potential†

$$A^\mu \equiv (V, \boldsymbol{A}) \tag{2.13}$$

and noting (from problem 2.1) that the differential operators $(\partial/\partial t, -\boldsymbol{\nabla})$ form the components of a 4-vector ∂^μ.
A gauge transformation is then specified by

$$\boxed{A^\mu \rightarrow A'^\mu = A^\mu - \partial^\mu\chi.} \tag{2.14}$$

The Maxwell equations can also be written in a manifestly covariant form† using the 4-current $j^\mu_{\text{em}}(x)$ given by

$$j^\mu_{\text{em}} = (\rho_{\text{em}}, \boldsymbol{j}_{\text{em}}) \tag{2.15}$$

in terms of which the continuity equation takes the form (problem 2.1)

$$\partial_\mu j^\mu_{\text{em}} = 0. \tag{2.16}$$

The Maxwell equations (2.1) and (2.8) then become (problem 2.2)

$$\partial_\mu F^{\mu\nu} = j^\nu_{\text{em}} \tag{2.17}$$

where we have defined the field strength tensor

$$F^{\mu\nu} \equiv \partial^\mu A^\nu - \partial^\nu A^\mu. \tag{2.18}$$

Since $F^{\mu\nu}$ is obviously invariant under the gauge transformation

$$A^\mu \rightarrow A'^\mu = A^\mu - \partial^\mu\chi \tag{2.14}$$

the Maxwell equations in this form are manifestly gauge invariant. The

†See Appendix A.2 for relativistic notation.

'covariant field equations' satisfied by A^μ then follow from equations (2.17) and (2.18):

$$\Box A^\nu - \partial^\nu(\partial_\mu A^\mu) = j^\nu_{\text{em}}. \tag{2.19}$$

Since gauge transformations turn out to be of central importance in the quantum theory of electromagnetism, it would be nice to have some insight into why Maxwell's equations are gauge invariant. The all-important 'fourth' equation (2.8) was inferred by Maxwell from local charge conservation, as expressed by the continuity equation

$$\partial_\mu j^\mu_{\text{em}} = 0. \tag{2.16}$$

The field equation

$$\partial_\mu F^{\mu\nu} = j^\nu_{\text{em}} \tag{2.17}$$

then of course automatically embodies (2.16). The mathematical reason it does so is that $F^{\mu\nu}$ is a four-dimensional kind of 'curl'

$$F^{\mu\nu} = \partial^\mu A^\nu - \partial^\nu A^\mu \tag{2.18}$$

which is obviously unchanged by a gauge transformation

$$A^\mu \rightarrow A'^\mu = A^\mu - \partial^\mu \chi. \tag{2.14}$$

Hence there is the suggestion that the gauge invariance is related in some way to charge conservation. However, the connection is not so simple. Wigner (1949) has given a simple argument to show that the principle that no physical quantity can depend on the absolute value of the electrostatic potential, when combined with energy conservation, implies the conservation of charge. Wigner's argument relates charge (and energy) conservation to an invariance under transformation of the electrostatic potential by a constant: charge conservation alone does not seem to require the more general space–time-dependent transformation of gauge invariance.

Changing the value of the electrostatic potential by a constant amount is an example of what we have called a *global* transformation (since the change in the potential is the same everywhere). Invariance under this global transformation is related to a conservation law: that of charge. But this global invariance is not sufficient to generate the full Maxwellian dynamics. However, as remarked by 't Hooft (1980), one can regard equations (2.11) and (2.12) as expressing the fact that a *local* change in the electrostatic potential V (the $\partial\chi/\partial t$ term in (2.12)) can be compensated—in the sense of leaving the Maxwell equations unchanged—by a corresponding local change in the magnetic vector potential A. Thus by including magnetic effects, the global invariance under a change of V by a constant can be extended to a local invariance (which is a much more restrictive condition to satisfy). Hence there is

the beginning of a suggestion that one might almost 'derive' the complete Maxwell equations, which unify electricity and magnetism, from the requirement that the theory be expressed in terms of potentials in such a way as to be invariant under local (gauge) transformations on those potentials. Certainly special relativity must play a role too: this also links electricity and magnetism, via the magnetic effects of charges as seen by an observer moving relative to them. If a 4-vector potential A^μ is postulated, and it is then demanded that the theory involve it only in a way which is insensitive to local changes of the form (2.14), one is led naturally to the idea that the physical fields enter only via the quantity $F^{\mu\nu}$, which is invariant under (2.14). From this, one might conjecture the field equation on grounds of Lorentz covariance.

It goes without saying that this is certainly not a 'proof' or 'derivation' of the Maxwell equations. Nevertheless, the idea that *dynamics* (in this case, the complete interconnection of electric and magnetic effects) may be intimately related to a *local invariance requirement* (in this case, electromagnetic gauge invariance) turns out to be a fruitful one. As indicated in §2.1 above, it is generally the case that, when a certain global invariance is generalised to a local one, the existence of a new 'compensating' field is entailed, interacting in a specified way. The first example of a dynamical theory 'derived' from a local invariance requirement seems to be the theory of Yang and Mills (1954) (see also Shaw 1955), from whose paper a crucial quotation appears at the start of Part III of this book. Their work was extended by Utiyama (1956), who developed a general formalism for such compensating fields. As we have said, these types of dynamical theories, based on local invariance principles, are called gauge theories.

It is a remarkable fact that all the interactions currently regarded as fundamental are of precisely this type. We have briefly discussed the Maxwell equations in this light, and will continue with (quantum) electrodynamics in the following two sections. Another example, but one which we shall not pursue in this book, is that of general relativity (the theory of the gravitational interaction). Utiyama (1956) showed that this theory could be arrived at by generalising the global (space–time-independent) coordinate transformations of special relativity to local ones; as with electromagnetism, the more restrictive local invariance requirements entailed the existence of a new field—the gravitational one—with an (almost) prescribed form of interaction. The two other fundamental interactions—the strong interaction between quarks, and the weak interaction between quarks and leptons—also seem to be described by gauge theories (of essentially the Yang–Mills type), as we shall see in detail in Parts III and V of this book.

In order to proceed further, we must now discuss how such ideas are incorporated into quantum mechanics.

2.4 Gauge invariance in quantum mechanics

The Lorentz force law for a non-relativistic particle of charge q moving with velocity $\boldsymbol{v}$ under the influence of both electric and magnetic fields is

$$\boldsymbol{F} = q\boldsymbol{E} + q\boldsymbol{v} \times \boldsymbol{B}. \tag{2.20}$$

It may be derived, via Hamilton's equations, from the classical Hamiltonian†

$$H = (1/2m)(\boldsymbol{p} - q\boldsymbol{A})^2 + qV. \tag{2.21}$$

The Schrödinger equation for such a particle in an electromagnetic field is

$$\left(\frac{1}{2m}(-\mathrm{i}\nabla - q\boldsymbol{A})^2 + qV\right)\psi(\boldsymbol{x}, t) = \mathrm{i}\frac{\partial\psi(\boldsymbol{x}, t)}{\partial t} \tag{2.22}$$

which is obtained from the classical Hamiltonian by the usual prescription, $p \to -\mathrm{i}\nabla$, for Schrödinger's wave mechanics ($\hbar = 1$). Notice the appearance of the operator combinations

$$\boxed{\begin{aligned} \boldsymbol{D} &\equiv \nabla - \mathrm{i}q\boldsymbol{A} \\ D^0 &\equiv \partial/\partial t + \mathrm{i}qV \end{aligned}} \tag{2.23}$$

in place of ∇ and $\partial/\partial t$, in going from the free particle Schrödinger equation to the electromagnetic field case.

The solution of $\psi(\boldsymbol{x}, t)$ of the Schrödinger equation (2.22) describes completely the state of the particle moving under the influence of the potentials V, $\boldsymbol{A}$. However, these potentials are not unique, as we have already seen: they can be changed by a gauge transformation

$$\boldsymbol{A} \to \boldsymbol{A}' = \boldsymbol{A} + \nabla\chi \tag{2.11}$$

$$V \to V' = V - \partial\chi/\partial t \tag{2.12}$$

and the Maxwell equations for the fields $\boldsymbol{E}$ and $\boldsymbol{B}$ will remain the same. This immediately raises a serious question: if we carry out such a change of potentials in equation (2.22), will the solution $\psi'(\boldsymbol{x}, t)$ of the resulting equation

$$\left(\frac{1}{2m}(-\mathrm{i}\nabla - q\boldsymbol{A}')^2 + qV'\right)\psi'(\boldsymbol{x}, t) = \mathrm{i}\frac{\partial\psi'(\boldsymbol{x}, t)}{\partial t} \tag{2.22a}$$

describe the same physics as the solution $\psi(\boldsymbol{x}, t)$ of equation (2.22)? If it does, we shall be able to assume the validity of Maxwell's theory for the

†We set $\hbar = c = 1$ throughout (see Appendix B).

quantum world; if not, some modification will be necessary, since the gauge symmetry possessed by the Maxwell equations will be violated in the quantum theory.

Since we know the relations (2.11) and (2.12) between $\boldsymbol{A}, V$ and $\boldsymbol{A}', V'$, we can actually find out what $\psi'(\boldsymbol{x}, t)$ must be in order that equation (2.22a) be consistent with (2.22). We shall state the answer and then verify it; then we shall discuss the physical interpretation. The required $\psi'(\boldsymbol{x}, t)$ is

$$\psi'(\boldsymbol{x}, t) = \exp[\mathrm{i}q\chi(\boldsymbol{x}, t)]\psi(\boldsymbol{x}, t) \tag{2.24}$$

where χ is the same space–time-dependent function as appears in equations (2.11) and (2.12). To verify this we consider

$$\begin{aligned} (-\mathrm{i}\nabla - q\boldsymbol{A}')\,\psi' &= [-\mathrm{i}\nabla - q\boldsymbol{A} - q(\nabla\chi)]\,[\exp(\mathrm{i}q\chi)\psi] \\ &= q(\nabla\chi)\exp(\mathrm{i}q\chi)\psi + \exp(\mathrm{i}q\chi)\cdot(-\mathrm{i}\nabla\psi) \\ &\quad + \exp(\mathrm{i}q\chi)\cdot(-q\boldsymbol{A}\psi) - q(\nabla\chi)\exp(\mathrm{i}q\chi)\psi. \end{aligned} \tag{2.25}$$

The first and last terms cancel leaving the result

$$(-\mathrm{i}\nabla - q\boldsymbol{A}')\psi' = \exp(\mathrm{i}q\chi)\cdot(-\mathrm{i}\nabla - q\boldsymbol{A})\psi \tag{2.26}$$

which may be written using equation (2.23) as

$$(-\mathrm{i}\boldsymbol{D}'\psi') = \exp(\mathrm{i}q\chi)\cdot(-\mathrm{i}\boldsymbol{D}\psi). \tag{2.27a}$$

Thus, although the space–time-dependent phase factor feels the action of the gradient operator ∇, it 'passes through' the combined operator $\boldsymbol{D}'$ and converts it into $\boldsymbol{D}$: in fact, comparing equations (2.24) and (2.27a), we see that $\boldsymbol{D}'\psi'$ bears to $\boldsymbol{D}\psi$ exactly the same relation as ψ' bears to ψ. In just the same way we find (cf equation 2.23))

$$(\mathrm{i}D'^0\psi') = \exp(\mathrm{i}q\chi)\cdot(\mathrm{i}D^0\psi) \tag{2.27b}$$

where we have used equation (2.12) for V'. Once again, $D'^0\psi'$ is simply related to $D^0\psi$. Repeating the operation which led to equation (2.27a), we find

$$\begin{aligned} (1/2m)\,(-\mathrm{i}\boldsymbol{D}')^2\psi' &= \exp(\mathrm{i}q\chi)\cdot(1/2m)\,(-\mathrm{i}\boldsymbol{D})^2\psi \\ &= \exp(\mathrm{i}q\chi)\cdot\mathrm{i}D^0\psi \quad \text{(using equation (2.22))} \\ &= \mathrm{i}D'^0\psi' \quad \text{(using equation (2.27b)).} \end{aligned} \tag{2.28}$$

Equation (2.28) is just (2.22a) written in the D notation of equation (2.23), so we have verified that (2.24) is the correct relationship between ψ' and ψ to ensure consistency between equations (2.22) and (2.22a).

Do ψ and ψ' describe the same physics? The answer is yes, but it is not quite trivial. It is certainly obvious that the probability densities $|\psi|^2$

and $|\psi'|^2$ are equal, since in fact ψ and ψ' in equation (2.24) are related by a *phase* transformation. However, we can be interested in other observables involving the derivative operators ∇ or $\partial/\partial t$—for example, the current, which is essentially $\psi^*(\nabla\psi) - (\nabla\psi)^*\psi$. It is easy to check that this is *not* invariant under (2.24), because the phase $\chi(\boldsymbol{x}, t)$ is $\boldsymbol{x}$-dependent. But equations (2.27a) and (2.27b) show us what we must do to construct *gauge invariant currents*: namely, we must replace ∇ by $\boldsymbol{D}$ (and in general also $\partial/\partial t$ by D^0) since then

$$\psi^{*\prime}(\boldsymbol{D}'\psi') = \psi^* \exp(-\mathrm{i}q\chi)\cdot\exp(\mathrm{i}q\chi)\cdot(\boldsymbol{D}\psi) = \psi^*\boldsymbol{D}\psi \qquad (2.29)$$

for example. Thus the identity of the physics described by ψ and ψ' is indeed ensured.

We summarise these important considerations by the statement that the gauge invariance of Maxwell's equations remains an invariance in quantum mechanics provided we make the combined transformation

$$\boxed{\begin{aligned} \boldsymbol{A} &\to \boldsymbol{A}' = \boldsymbol{A} + \nabla\chi \\ V &\to V' = V - \partial\chi/\partial t \\ \psi &\to \psi' = \exp(\mathrm{i}q\chi)\psi \end{aligned}} \qquad (2.30)$$

on the potentials and on the wavefunction.

The Schrödinger equation is non-relativistic, but the Maxwell equations are of course fully relativistic. One might therefore suspect that the prescriptions discovered here are actually true relativistically as well, and this is indeed the case. We shall introduce the spin-0 and spin-$\frac{1}{2}$ relativistic wave equations in Chapter 3. For the present we note that (2.23) can be written in manifestly covariant form as

$$\boxed{D^\mu \equiv \partial^\mu + \mathrm{i}qA^\mu} \qquad (2.31)$$

in terms of which (2.27a) and (2.27b) become

$$-\mathrm{i}D'^\mu\psi' = \exp(\mathrm{i}q\chi)\cdot(-\mathrm{i}D^\mu\psi). \qquad (2.32)$$

It follows that any wave equation involving the operator ∂^μ can be made gauge invariant under the combined transformation

$$A^\mu \to A'^\mu = A^\mu - \partial^\mu\chi$$

$$\psi \to \psi' = \exp(\mathrm{i}q\chi)\psi$$

if ∂^μ is replaced by D^μ. In fact, we seem to have a very simple prescription for obtaining the wave equation for a particle in the presence of an electromagnetic field from the corresponding *free particle* wave equation: make the replacement

$$\boxed{\partial^\mu \to D^\mu \equiv \partial^\mu + \mathrm{i}qA^\mu.} \qquad (2.33)$$

In the following section this will be seen to be the basis of the so-called 'gauge principle' whereby, in accordance with the idea advanced in the previous sections, the form of *interaction* is determined by the insistence on (local) gauge invariance.

One final remark: this new kind of derivative

$$D^\mu \equiv \partial^\mu + \mathrm{i}qA^\mu \qquad (2.31)$$

turns out to be of fundamental importance—it will be the operator which generalises from the (Abelian) phase symmetry of QED (see comment (iii) of §2.6) to the (non-Abelian) phase symmetries of our weak and strong interaction theories. It is called the *'covariant derivative'*.

2.5 The argument reversed: the gauge principle

In the preceding section, we took it as *known* that the Schrödinger equation, for example, for a charged particle in an electromagnetic field, has the form

$$[(1/2m)\,(-\mathrm{i}\nabla - qA)^2 + qV]\psi = \mathrm{i}\partial\psi/\partial t. \qquad (2.22)$$

We then checked its gauge invariance under the combined transformation

$$\begin{aligned} \boldsymbol{A} &\to \boldsymbol{A}' = \boldsymbol{A} + \nabla\chi \\ V &\to V' = V - \partial\chi/\partial t \\ \psi &\to \psi' = \exp(\mathrm{i}q\chi)\psi. \end{aligned} \qquad (2.30)$$

We now want to reverse the argument: we shall start by demanding that our theory is invariant under the *space–time-dependent phase transformation*

$$\psi(\boldsymbol{x}, t) \to \psi'(\boldsymbol{x}, t) = \exp[\mathrm{i}q\chi(\boldsymbol{x}, t)]\psi(\boldsymbol{x}, t). \qquad (2.34)$$

We shall demonstrate that such a phase invariance is not possible for a free theory, but rather requires an *interacting* theory, involving a (4-vector) field whose interactions with the charged particle are precisely determined, and which undergoes the transformation

$$\boldsymbol{A} \to \boldsymbol{A}' = \boldsymbol{A} + \nabla\chi \qquad (2.11)$$

$$V \to V' = V - \partial\chi/\partial t \qquad (2.12)$$

when $\psi \rightarrow \psi'$. The demand of this type of phase invariance will then have dictated the form of the interaction—this is the basis of the *gauge principle*.

We therefore focus attention on the phase of the wavefunction. The absolute phase of a wavefunction in quantum mechanics cannot be measured; only relative phases are measurable, via some sort of interference experiment. A simple example is provided by the diffraction of particles by a two-slit system. Downstream from the slits, the wavefunction is a coherent superposition of two components, one originating from each slit: symbolically,

$$\psi = \psi_1 + \psi_2. \tag{2.35}$$

The probability distribution $|\psi|^2$ will then involve, in addition to the separate intensities $|\psi_1|^2$ and $|\psi_2|^2$, the *interference* term

$$2\mathrm{Re}(\psi_1^* \psi_2) = 2|\psi_1||\psi_2|\cos\delta$$

where δ $(= \delta_1 - \delta_2)$ is the *phase difference* between components ψ_1 and ψ_2. The familiar pattern of alternating intensity maxima and minima is then attributed to variation in the phase difference δ. Where the components are in phase, the interference is constructive and $|\psi|^2$ has a maximum; where they are out of phase, it is destructive and $|\psi|^2$ has a minimum. It is clear that if the individual phases δ_1 and δ_2 are each shifted by the same amount, there will be no observable consequences, since only the phase difference δ enters.

The situation in which a wavefunction can be changed in a certain way without leading to any observable effects is precisely what is entailed by a symmetry or invariance principle in quantum mechanics. In the case under discussion, the invariance is that of a constant overall change in phase. In performing calculations it is necessary to make some definite choice of phase; that is, to adopt a 'phase convention'. The actual value chosen is irrelevant, as is guaranteed by the invariance principle, but some choice has to be made.

Invariance under a constant change in phase is an example of a *global* invariance, according to the terminology introduced in the previous section. We make this point quite explicit by writing out the transformation as

$$\boxed{\begin{array}{r}\psi \rightarrow \psi' = \mathrm{e}^{\mathrm{i}\alpha}\psi \\ \alpha = \text{constant}\end{array}} \quad \begin{array}{l}\text{global phase} \\ \text{invariance.}\end{array} \tag{2.36}$$

That α in (2.36) is a constant, the same for all space–time points, expresses the fact that once a phase convention (choice of α) has been made at one space–time point, the same convention must be adopted at all other points. Thus in the two-slit experiment we are not free to make a *local* change of phase: for example, as discussed by 't Hooft (1980), inserting a half-wave plate behind just one of the slits will certainly have observable consequences.

There is a sense in which this may seem an unnatural state of affairs (cf the quotation from Yang and Mills cited at the start of Part III). Once a phase convention has been adopted at one space–time point, the same convention must be adopted at all other ones: the half-wave plate must extend instantaneously across all of space, or not at all. Following this line of thought, one might then be led to 'explore the possibility' of requiring invariance under *local* phase transformations; that is, independent choices of phase convention at each space–time point. By itself, the foregoing is not a compelling motivation for such a step. However, as we pointed out in §2.3, such a move from a global to a local invariance is apparently of crucial significance in classical electromagnetism and general relativity, and seems now to provide the key to an understanding of elementary particle interactions. Let us see, then, where the demand of '*local* phase invariance'

$$\boxed{\psi(x,t) \rightarrow \psi'(x,t) = \exp[\mathrm{i}\alpha(x,t)]\psi(x,t)} \quad \text{local phase invariance} \tag{2.37}$$

leads us.

There is immediately a problem: this is *not* an invariance of the free particle Schrödinger equation or of any relativistic wave equation! For example, if the original wavefunction $\psi(\boldsymbol{x}, t)$ satisfied the free particle Schrödinger equation

$$(1/2m)(-\mathrm{i}\boldsymbol{\nabla})^2\psi(\boldsymbol{x}, t) = \mathrm{i}\partial\psi(\boldsymbol{x}, t)/\partial t$$

then the wavefunction ψ', given by the local phase transformation above, will not, since both $\boldsymbol{\nabla}$ and $\partial/\partial t$ now act on $\alpha(\boldsymbol{x}, t)$ in the phase factor. Thus local phase invariance is not an invariance of the free particle wave equation. If we wish to satisfy the demands of local phase invariance, we are obliged to modify the free particle Schrödinger equation into something for which there is a local phase invariance. But this modified equation will no longer describe a free particle: in other words, the freedom to alter the phase of a charged particle's wavefunction locally is only possible if some kind of force field is introduced in which the particle moves. In more physical terms, the invariance will now be manifested in the inability to distinguish observationally between the effect of making a local change in phase convention and the effect of some new field in which the particle moves.

What kind of field will this be? In fact, we know immediately what the answer is, since the local phase transformation

$$\psi \to \psi' = \exp[i\alpha(x, t)]\psi \tag{2.37}$$

with $\alpha = q\chi$ is just the phase transformation associated with electromagnetic gauge invariance! Thus we must modify the Schrödinger equation

$$(1/2m)(-i\nabla)^2\psi = i\partial\psi/\partial t \tag{2.38}$$

to

$$(1/2m)(-i\nabla - qA)^2\psi = (i\partial/\partial t - qV)\psi$$

and satisfy the local phase invariance

$$\psi \to \psi' = \exp[i\alpha(x, t)]\psi$$

by demanding that A and V transform by

$$\begin{aligned} A &\to A' = A + q^{-1}\nabla\alpha \\ V &\to V' = V - q^{-1}\partial\alpha/\partial t \end{aligned} \tag{2.39}$$

when $\psi \to \psi'$. But the modified wave equation is of course precisely the Schrödinger equation describing the interaction of the charged particle with the electromagnetic field described by A and V.

In a covariant treatment, A and V will be regarded as parts of a 4-vector A^μ, just as $-\nabla$ and $\partial/\partial t$ are parts of ∂^μ. Thus the presence of the vector field A^μ, interacting in a 'universal' prescribed way with any particle of charge q, is dictated by local phase invariance. A vector field such as A^μ, introduced in order to guarantee local phase invariance, is called a 'gauge field'. The principle that the interaction should be so dictated by the phase (or gauge) invariance is called the *gauge principle*: it allows us to write down the wave equation for the interaction directly from the free particle equation†. As before, the method clearly generalises to the four-dimensional case.

2.6 Comments on the gauge principle in electromagnetism

(i) A properly sceptical reader may have detected an important sleight of hand in the discussion above. Where exactly did the electromagnetic charge appear from? The trouble with our argument as so far presented is that we could have defined fields A and V so that they coupled equally to all particles—instead we smuggled in a factor q.

Actually we can do a bit better than this. We can use the fact that the electromagnetic charge is absolutely conserved, to claim that there can

†Actually, the electromagnetic interaction is uniquely specified by this procedure only for particles of spin 0 and spin $\frac{1}{2}$; see §8.4 for the case of spin 1.

be no quantum mechanical interference between states of different charge q. Hence different phase changes are allowed within each 'sector' of definite q:

$$\psi' = \exp(iq\chi)\psi \tag{2.40}$$

let us say. When this becomes a local transformation, $\chi \to \chi(x, t)$, we shall need to cancel a term $q\nabla\chi$, which will imply the presence of a '$-qA$' term, as required. Note that such an argument is only possible for an *absolutely* conserved quantum number q—otherwise we cannot split up the states of the system into non-communicating sectors specified by different values of q. Reversing this line of reasoning, a conservation law such as baryon number conservation, with no related gauge field, would therefore now be suspected of not being absolutely conserved.

We still have not tied down why q is the electromagnetic charge and not some other absolutely conserved quantum number. A proper discussion of the reasons for identifying A^μ with the electromagnetic potential and q with the particle's charge will be given in Chapter 4, with the help of quantum field theory.

(ii) Accepting these identifications, we note that the form of the interaction contains but one parameter, the electromagnetic charge q of the particle in question. It is the *same* whatever the type of particle with charge q, whether it be lepton, hadron, nucleus, ion, atom, etc. Precisely this type of 'universality' is present in the weak couplings of quarks and leptons, as we shall see in Chapter 11. This strongly suggests that some form of gauge principle must be at work in generating weak interactions as well. The associated symmetry or conservation law is, however, of a very subtle kind, as we shall discuss in Chapter 13. Incidentally, although all particles of a given charge q interact electromagnetically in a universal way, there is nothing at all in the preceding argument to indicate why, in nature, the charges of observed particles are all integer multiples of one basic charge. We shall return to this problem of charge quantisation in §15.2.

(iii) Returning to point (i), we may wish that we had not had to introduce the absolute conservation of charge as a separate axiom. As remarked earlier, at the end of §2.2, we should like to relate that conservation law to the symmetry involved, namely invariance under (2.37). It is worth looking at the nature of this symmetry in a little more detail. It is not a symmetry which—as in the case of translation and rotation invariances for instance—involved changes in the space–time coordinates x and t. Instead, it operates on the *real and imaginary parts of the wavefunction*. Let us write

$$\psi = \psi_R + i\psi_I. \tag{2.41}$$

Then

$$\psi' = e^{i\alpha}\psi = \psi_R' + i\psi_I' \tag{2.42}$$

can be written as

$$\begin{aligned} \psi_R' &= (\cos\alpha)\psi_R - (\sin\alpha)\psi_I \\ \psi_I' &= (\sin\alpha)\psi_R + (\cos\alpha)\psi_I \end{aligned} \tag{2.43}$$

from which we can see that it is indeed a kind of 'rotation', but in the ψ_R–ψ_I plane, whose 'coordinates' are the real and imaginary parts of the wavefunction. We call this plane an *internal* space, and the associated symmetry an *internal symmetry*. Thus our phase invariance can be looked upon as a kind of internal space rotational invariance.

We can imagine doing two successive such transformations

$$\psi \rightarrow \psi' \rightarrow \psi'' \tag{2.44}$$

where

$$\psi'' = e^{i\beta}\psi' \tag{2.45}$$

and so

$$\psi'' = e^{i(\alpha+\beta)}\psi = e^{i\delta}\psi \tag{2.46}$$

with $\delta = \alpha + \beta$. This is a transformation of the same form as the original one. The set of all such transformations forms what mathematicians call a *group*, in this case U(1), meaning the group of all unitary one-dimensional matrices. A unitary matrix $\mathbf{U}$ is one such that

$$\mathbf{U}\mathbf{U}^\dagger = \mathbf{U}^\dagger\mathbf{U} = \mathbf{I} \tag{2.47}$$

where $\mathbf{I}$ is the identity and † denotes the Hermitian conjugate. A one–dimensional matrix is of course a single number—in this case, a complex number. Condition (2.47) limits this to being a simple phase: the set of phase factors of the form $e^{i\alpha}$, where α is any real number, form the elements of a U(1) group. These are just the factors that enter into our gauge (or phase) transformations for wavefunctions. Thus we say that the electromagnetic gauge group is U(1).

The transformations of the U(1) group have the simple property that it does not matter in what order they are performed: referring to (2.44)–(2.46), we would have got the same final answer if we had done the β 'rotation' first and then the α one, instead of the other way around; this is because, of course,

$$\exp(i\alpha)\cdot\exp(i\beta) = \exp[i(\alpha+\beta)] = \exp(i\beta)\cdot\exp(i\alpha).$$

Mathematicians call U(1) an *Abelian* group: different transformations commute. We shall see later (in Chapters 9 and 14) that the 'internal' symmetry spaces relevant to the strong and weak gauge invariances are not so simple. The 'rotations' in these cases are more like full three-dimensional rotations of real space, rather than the two-dimensional rotation of (2.43). We know that, in general, such real space rotations do *not* commute, and the same will be true of the strong and weak rotations. Their gauge groups are called *non-Abelian*.

Once again, we shall have to wait until Chapter 4 before understanding how the symmetry represented by (2.42) is really related to the conservation law of charge.

(iv) The attentive reader may have picked up one further loose end. The vector potential $\boldsymbol{A}$ is related to the magnetic field $\boldsymbol{B}$ by

$$\boldsymbol{B} = \nabla \times \boldsymbol{A}. \tag{2.9}$$

Thus if $\boldsymbol{A}$ has the special form

$$\boldsymbol{A} = \nabla f \tag{2.48}$$

$\boldsymbol{B}$ will vanish. The question we must answer, therefore, is: how do we know that the $\boldsymbol{A}$ field introduced by our gauge principle is not of the form (2.48), leading to a trivial theory ($\boldsymbol{B} = \boldsymbol{0}$)? The answer to this question will lead us on a very worthwhile detour.

The Schrödinger equation with ∇f as the vector potential is

$$(1/2m)(-i\nabla - q\nabla f)^2\psi = E\psi. \tag{2.49}$$

We can write the formal solution to this equation as

$$\psi = \exp\left(iq\int_{-\infty}^{x}\nabla f\cdot d\boldsymbol{l}\right)\cdot\psi(f=0) \tag{2.50}$$

which may be checked by using the fact that

$$\frac{\partial}{\partial a}\int^{a} f(t)dt = f(a). \tag{2.51}$$

The notation $\psi(f=0)$ means just the free particle solution with $f = 0$; the line integral is taken along an arbitrary path ending in the point $\boldsymbol{x}$. But we have

$$df = \frac{\partial f}{\partial x}dx + \frac{\partial f}{\partial y}dy + \frac{\partial f}{\partial z}dz \equiv \nabla f\cdot d\boldsymbol{l}. \tag{2.52}$$

Hence the line integral can be done trivially and the solution becomes

$$\psi = \exp[iq(f(\boldsymbol{x}) - f(-\infty))]\cdot\psi(f=0). \tag{2.53}$$

We say that the phase factor introduced by the (in reality, field-free) vector potential $\boldsymbol{A} = \nabla f$ is *integrable*: the effect of this particular $\boldsymbol{A}$ is merely to multiply the free particle solution by an $\boldsymbol{x}$-dependent phase (apart from a trivial constant phase). Since this $\boldsymbol{A}$ should give no real electromagnetic effect, we must hope that such a change in the wavefunction is also somehow harmless. Indeed, Dirac showed (Dirac

1981, pp 92–3) that such a phase factor corresponds merely to a redefinition of the momentum operator $\hat{p}$. The essential point is that (in one dimension, say) $\hat{p}$ is defined ultimately by the commutator ($\hbar = 1$)

$$[\hat{x}, \hat{p}] = \mathrm{i}. \tag{2.54}$$

Certainly the familiar choice

$$\hat{p} = -\mathrm{i}\frac{\partial}{\partial x} \tag{2.55}$$

satisfies this commutation relation. But we can also add any function of x to $\hat{p}$, and this modified $\hat{p}$ will still be satisfactory since x commutes with any function of x. More detailed considerations by Dirac showed that this arbitrary function must actually have the form $\partial F/\partial x$, where F is arbitrary. Thus

$$\hat{p}' = -\mathrm{i}\frac{\partial}{\partial x} + \frac{\partial F}{\partial x} \tag{2.56}$$

is an acceptable momentum operator. Consider then the quantum mechanics defined by the wavefunctions $\psi(f = 0)$ and the momentum operator $\hat{p} = -\mathrm{i}\partial/\partial x$. Under the unitary transformation

$$\psi(f = 0) \rightarrow \mathrm{e}^{\mathrm{i}qf(x)}\psi(f = 0) \tag{2.57}$$

$\hat{p}$ will be transformed to

$$\hat{p} \rightarrow \mathrm{e}^{\mathrm{i}qf(x)}\hat{p}\ \mathrm{e}^{-\mathrm{i}qf(x)}. \tag{2.58}$$

But the right-hand side of this equation is just $\hat{p} - q\partial f/\partial x$ (problem 2.3), which is an equally acceptable momentum operator, identifying qf with the F of Dirac.

What of the physically interesting case in which $\boldsymbol{A}$ is *not* of the form $\boldsymbol{\nabla} f$? The equation is now

$$(1/2m)(-\mathrm{i}\boldsymbol{\nabla} - q\boldsymbol{A})^2\psi = E\psi \tag{2.59}$$

to which the solution is

$$\psi = \exp\left(\mathrm{i}q\int_{-\infty}^{x}\boldsymbol{A}\cdot\mathrm{d}\boldsymbol{l}\right)\cdot\psi(\boldsymbol{A} = 0). \tag{2.60}$$

The line integral can now not be done so trivially: one says that the $\boldsymbol{A}$-field has produced a *non-integrable phase factor*. There is more to this terminology than the mere question of whether the integral is easy to do. The crucial point is that the integral now depends on the *path followed* in reaching the point $\boldsymbol{x}$, whereas the integrable phase factor in (2.50) depends only on the end-points of the integral, not on the path joining them.

Consider two paths C_1 and C_2 (figure 2.1) from $-\infty$ to the point $\boldsymbol{x}$. The difference in the two line integrals is the integral over a *closed* curve C, which can be evaluated by Stokes' theorem:

$$\int_{C_1}^{x}\boldsymbol{A}\cdot\mathrm{d}\boldsymbol{l} - \int_{C_2}^{x}\boldsymbol{A}\cdot\mathrm{d}\boldsymbol{l} = \oint_{C}\boldsymbol{A}\cdot\mathrm{d}\boldsymbol{l} = \int\int_{S}\boldsymbol{\nabla}\times\boldsymbol{A}\cdot\mathrm{d}\boldsymbol{S} = \int\int_{S}\boldsymbol{B}\cdot\mathrm{d}\boldsymbol{S} \tag{2.61}$$

where S is any surface spanning the curve C. In this form we see that if $\boldsymbol{A} = \boldsymbol{\nabla} f$, then indeed the line integrals over C_1 and C_2 are equal since curl grad $\equiv 0$, but if $\boldsymbol{B} = \boldsymbol{\nabla}\times\boldsymbol{A}$ is not zero, the difference between the integrals is determined by the enclosed flux of $\boldsymbol{B}$.

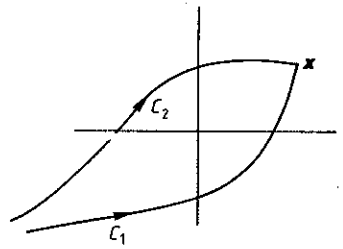

Figure 2.1 Two paths C_1 and C_2 (in two dimensions for simplicity) from $-\infty$ to the point $\boldsymbol{x}$.

The above analysis turns out to imply the existence of a remarkable phenomenon—the Aharonov–Bohm effect, named after its discoverers (Aharonov and Bohm 1959). Suppose we go back to our two-slit experiment of §2.5, only this time we imagine that a long thin solenoid is inserted between the slits, so that the components ψ_1 and ψ_2 of the split beam pass one on each side of the solenoid (figure 2.2). After passing round the solenoid, the beams are recombined, and the resulting interference pattern is observed downstream. At any point $\boldsymbol{x}$ of the pattern, the phase of the ψ_1 and ψ_2 components will be modified—relative to the $\boldsymbol{B} = \boldsymbol{0}$ case—by factors of the form (2.60). These factors depend on the respective paths, which are different for the two components ψ_1 and ψ_2. Thus the phase difference between these components, which determines the interference pattern, will involve the $\boldsymbol{B}$-dependent factor (2.61). Thus, even though the field $\boldsymbol{B}$ is essentially totally contained within the solenoid, and the beams themselves have passed through $\boldsymbol{B} = \boldsymbol{0}$ regions only, there is nevertheless an observable effect on the pattern provided $\boldsymbol{B} \neq \boldsymbol{0}$! This effect—a shift in the pattern as $\boldsymbol{B}$ varies—was first confirmed experimentally by Chambers (1960), soon after its prediction by Aharonov and Bohm. It was anticipated in work by Ehrenburg and Siday (1949); further references and discussion are contained in Berry (1984).

(v) In conclusion, we must emphasise that there is ultimately no compelling logic for the vital leap to a local phase invariance from a

global one. The latter is, by itself, both necessary and sufficient in quantum field theory to guarantee local charge conservation. Nevertheless, the gauge principle—deriving interactions from the requirement of local phase invariance—is so simple, beautiful and powerful (and apparently successful) that it has taken command of elementary particle physics. In later parts of the book we shall consider generalisations of the electromagnetic gauge principle. It will be important always to bear in mind that any attempt to base theories of non-electromagnetic interactions on some kind of gauge principle can only make sense if there is an exact symmetry involved. The reason for this will only become clear when we consider the *renormalisability* of our theories (Chapter 6).

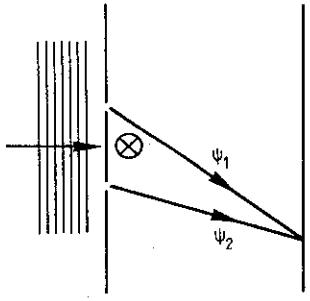

Figure 2.2 The Aharonov–Bohm effect.

Summary

Maxwell's equations and the local conservation of charge.

Electromagnetic potentials A^μ and gauge transformations.

Field strength tensor $F^{\mu\nu}$.

Lorentz covariant and gauge invariant forms of Maxwell's equations, using A^μ.

Gauge invariance in quantum mechanics requires wavefunction to change by space–time-dependent phase factor when A^μ changes by gauge transformation.

The gauge principle for generating interactions (electromagnetic case): local and global invariances.

U(1) transformations.

Non-integrable phase factor and the Aharonov–Bohm effect.

Problems

2.1 (*a*) A Lorentz transformation in the x^1 direction is given by

$$t' = \gamma(t - vx^1),\ x'^1 = \gamma(-vt + x^1)$$

$$x'^2 = x^2,\ x'^3 = x^3$$

where $\gamma = (1 - v^2)^{-1/2}$ and $c = 1$. Write down the inverse of this transformation (i.e. express (t, x^1) in terms of (t', x'^1)), and use the 'chain rule' of partial differentiation to show that, under the Lorentz transformation, the two quantities $(\partial/\partial t, -\partial/\partial x^1)$ transform in the same way as (t, x^1).

[The general result is that the four-component quantity $(\partial/\partial t, -\partial/\partial x^1, -\partial/\partial x^2, -\partial/\partial x^3) \equiv (\partial/\partial t, -\nabla)$ transforms in the same way as (t, x^1, x^2, x^3). Four-component quantities transforming this way are said to be 'contravariant four vectors', and are written with an upper four-vector index; thus $(\partial/\partial t, -\nabla) \equiv \partial^\mu$. Upper indices can be lowered by using the metric tensor $g_{\mu\nu}$, see Appendix A.2, which reverses the signs of the spatial components. Thus $\partial^\mu = (\partial/\partial t, \partial/\partial x_1, \partial/\partial x_2, \partial/\partial x_3)$. Similarly the four quantities $(\partial/\partial t, \nabla) = (\partial/\partial t, \partial/\partial x^1, \partial/\partial x^2, \partial/\partial x^3)$ transform as $(t, -x^1, -x^2, -x^3)$ and are a 'covariant four-vector', denoted, by ∂_μ.]

(*b*) Check that equation (2.5) can be written as (2.16).

2.2 How many independent components does the field strength $F^{\mu\nu}$ have? Express each component in terms of electric and magnetic field components. Hence verify that equation (2.17) correctly reproduces both equations (2.1) and (2.8).

2.3 Verify the result

$$e^{iqf(x)}\hat{p}\ e^{-iqf(x)} = \hat{p} - q\frac{\partial f}{\partial x}.$$