2
The electromagnetic field

Classical electrodynamics serves both as a point of reference and as the point of departure for the development of covariant field theories of matter and radiation. It was the observation that Maxwell's equations predict a universal speed of light *in vacuo* which led to the special theory of relativity, and this, in turn, led to the importance of *perspective* in identifying generally applicable physical laws. It was realized that the symmetries of special relativity meant that electromagnetism could be reformulated in a compact form, using a vector notation for spacetime unified into a single parameter space. The story of covariant fields therefore begins with Maxwell's four equations for the electromagnetic field in $3+1$ dimensions.

2.1 Maxwell's equations

In their familiar form, Maxwell's equations are written (in SI units)

$$\vec{\nabla}\cdot\mathbf{E} = \frac{\rho_e}{\epsilon_0} \tag{2.1a}$$

$$\vec{\nabla}\times\mathbf{E} = -\frac{\partial\mathbf{B}}{\partial t} \tag{2.1b}$$

$$\vec{\nabla}\cdot\mathbf{B} = 0 \tag{2.1c}$$

$$c^2(\vec{\nabla}\times\mathbf{B}) = \frac{\mathbf{J}}{\epsilon_0} + \frac{\partial\mathbf{E}}{\partial t}. \tag{2.1d}$$

ρ_e is the charge density, $\mathbf{J}$ is the electric current density and $c^2 = (\epsilon_0\mu_0)^{-1}$ is the speed of light in a vacuum squared. These are valid, as they stand, in inertial frames in flat $(3+1)$ dimensional spacetimes. The study of covariant field theory begins by assuming that these equations are true, in the sense that any physical laws are 'true' – i.e. that they provide a suitably idealized description of the physics of electromagnetism. We shall not attempt to follow the path which

https://doi.org/10.1017/9781009289887.004 Published online by Cambridge University Press

led to their discovery, nor explore their limitations. Rather, we are interested in summarizing their form and substance, and in identifying symmetries which allow them to be expressed in an optimally simple form. In this way, we hope to learn something deeper about their meaning, and facilitate their application.

2.1.1 Potentials

This chapter may be viewed as a demonstration of how applied covariance leads to a maximally simple formulation of Maxwell's equations. A more complete understanding of electromagnetic covariance is only possible after dealing with the intricacies of chapter 9, which discusses the symmetry of spacetime. Here, the aim is to build an algorithmic understanding, in order to gain a familiarity with key concepts for later clarification.

In texts on electromagnetism, Maxwell's equations are solved for a number of problems by introducing the idea of the vector and scalar potentials. The potentials play an important role in modern electrodynamics, and are a convenient starting point for introducing covariance.

The electromagnetic potentials are introduced by making use of two theorems, which allow Maxwell's equations to be re-written in a simplified form. In a covariant formulation, one starts with these and adds the idea of a unified *spacetime*. Spacetime is the description of space and time which treats the apparently different parameters x and t in a symmetrical way. It does not claim that they are equivalent, but only that they may be treated together, since both describe different aspects of the *extent* of a system. The procedure allows us to discover a simplicity in electromagnetism which is not obvious in eqns. (2.1).

The first theorem states that the vanishing divergence of a vector implies that it may be written as the curl of some other vector quantity $\mathbf{A}$:

$$\vec{\nabla} \cdot \mathbf{v} = 0 \quad \Rightarrow \quad \mathbf{v} = \vec{\nabla} \times \mathbf{A}. \tag{2.2}$$

The second theorem asserts that the vanishing of the curl of a vector implies that it may be written as the gradient of some scalar ϕ:

$$\vec{\nabla} \times \mathbf{v} = 0 \quad \Rightarrow \quad \mathbf{v} = \vec{\nabla}\phi. \tag{2.3}$$

The deeper reason for both these theorems, which will manifest itself later, is that the curl has an *anti-symmetric* property. The theorems, as stated, are true in a homogeneous, isotropic, flat space, i.e. in a system which does not have irregularities, but they can be generalized to any kind of space. From these, one defines two *potentials*: a vector potential A_i and a scalar ϕ, which are auxiliary functions (fields) of space and time.

The physical electromagnetic field is the derivative of the potentials. From eqn. (2.1c), one defines

$$\mathbf{B} = \vec{\nabla} \times \mathbf{A}. \tag{2.4}$$

https://doi.org/10.1017/9781009289887.004 Published online by Cambridge University Press

This form completely solves that equation. One equation has now been automatically and completely solved by re-parametrizing the problem in terms of a new variable. Eqn. (2.1c) tells us now that

$$\vec{\nabla} \times \mathbf{E} = -\frac{\partial}{\partial t}(\vec{\nabla} \times \mathbf{A})$$
$$\vec{\nabla} \times \left(\mathbf{E} + \frac{\partial \mathbf{A}}{\partial t}\right) = 0. \tag{2.5}$$

Consequently, according to the second theorem, one can write

$$\mathbf{E} + \frac{\partial \mathbf{A}}{\partial t} = -\vec{\nabla}\phi, \tag{2.6}$$

giving

$$\mathbf{E} = -\vec{\nabla}\phi - \frac{\partial \mathbf{A}}{\partial t}. \tag{2.7}$$

The minus sign on the right hand side of eqn. (2.6) is the convention which is used to make attractive forces positive and repulsive forces negative.

Introducing potentials in this way is not a necessity: many problems in electromagnetism can be treated by solving eqns. (2.1) directly, but the introduction often leads to significant simplifications when it is easier to solve for the potentials than it is to solve for the fields.

The potentials themselves are a mixed blessing: on the one hand, the re-parametrization leads to a number of helpful insights about Maxwell's equations. In particular, it reveals *symmetries*, such as the gauge symmetry, which we shall explore in detail later. It also allows us to write the matter–radiation interaction in a *local* form which would otherwise be impossible. The price one pays for these benefits is the extra conceptual layers associated with the potential and its gauge invariance. This confuses several issues and forces us to deal with constraints, or conditions, which uniquely define the potentials.

2.1.2 Gauge invariance

Gauge invariance is a symmetry which expresses the freedom to re-define the potentials arbitrarily without changing their physical significance. In view of the theorems above, the fields **E** and **B** are invariant under the re-definitions

$$\mathbf{A} \to \mathbf{A}' = \mathbf{A} + \vec{\nabla}s$$
$$\phi \to \phi' = \phi - \frac{\partial s}{\partial t}. \tag{2.8}$$

These re-definitions are called *gauge transformations*, and $s(x)$ is an arbitrary scalar function. The transformation means that, when the potentials are used

https://doi.org/10.1017/9781009289887.004 Published online by Cambridge University Press

as variables to solve Maxwell's equations, the parametrization of physics is not unique. Another way of saying this is that there is a freedom to choose between one of many different values of the potentials, each of which leads to the same values for the physical fields **E** and **B**. One may therefore choose whichever potential makes the solution easiest. This is a curious development. Why make a definite problem arbitrary? Indeed, this freedom can cause problems if one is not cautious. However, the arbitrariness is unavoidable: it is deeply connected with the symmetries of spacetime (the Lorentz group). Occasionally gauge invariance leads to helpful, if abstract, insights into the structure of the field theory. At other times, it is desirable to eliminate the fictitious freedom it confers by introducing an auxiliary condition which pins down a single $\phi, \mathbf{A}$ pair for each value of $\mathbf{E}, \mathbf{B}$. As long as one uses a potential as a tool to solve Maxwell's equations, it is necessary to deal with gauge invariance and the multiplicity of equivalent solutions which it implies.

2.1.3 4-vectors and $(n+1)$-vectors

Putting the conceptual baggage of gauge invariance aside for a moment, one proceeds to make Maxwell's equations covariant by combining space and time in a unified vector formulation. This is easily done by looking at the equations of motion for the potentials. The equations of motion for the vector potentials are found as follows: first, substituting for the electric field in eqn. (2.1a) using eqn. (2.7), one has

$$-\nabla^2\phi - \frac{\partial}{\partial t}(\nabla \cdot \mathbf{A}) = \frac{\rho_e}{\epsilon_0}. \tag{2.9}$$

Similarly, using eqn. (2.4) in (2.1d), one obtains

$$c^2\vec{\nabla} \times (\vec{\nabla} \times \mathbf{A}) = \frac{\mathbf{J}}{\epsilon_0} + \frac{\partial}{\partial t}\left(-\vec{\nabla}\phi - \frac{\partial \mathbf{A}}{\partial t}\right). \tag{2.10}$$

Using the vector identity

$$\vec{\nabla} \times (\vec{\nabla} \times \mathbf{A}) = \vec{\nabla}(\vec{\nabla} \cdot \mathbf{A}) - \nabla^2\mathbf{A} \tag{2.11}$$

to simplify this, one obtains

$$c^2\left(\frac{1}{c^2}\frac{\partial^2}{\partial t^2} - \nabla^2\right)\mathbf{A} = \frac{\mathrm{j}}{\epsilon_0} - \vec{\nabla}\left(\frac{\partial \phi}{\partial t} + c^2(\vec{\nabla} \cdot \mathbf{A})\right). \tag{2.12}$$

It is already apparent from eqns. (2.8) that the potentials $\phi, \mathbf{A}$ are not unique. This fact can now be used to tidy up eqn. (2.12), by making a choice for ϕ and $\mathbf{A}$:

$$\vec{\nabla} \cdot \mathbf{A} + \frac{1}{c^2}\frac{\partial \phi}{\partial t} = 0. \tag{2.13}$$

https://doi.org/10.1017/9781009289887.004 Published online by Cambridge University Press

The right hand side of eqn. (2.13) is chosen to be zero, but, of course, any constant would do. Making this arbitrary (but not random) choice, is called *choosing a gauge*. It partially fixes the freedom to choose the scalar field s in eqns. (2.8). Specifically, eqn. (2.13) is called the Lorentz gauge. This common choice is primarily used to tidy up the equations of motion, but, as noted above, at some point one has to make a choice anyway so that a single pair of vector potentials (scalar, vector) corresponds to only one pair of physical fields ($\mathbf{E}$, $\mathbf{B}$).

The freedom to choose the potentials is not entirely fixed by the adoption of the Lorentz condition, however, as we may see by substituting eqn. (2.8) into eqn. (2.13). Eqn. (2.13) is not completely satisfied; instead, one obtains a new condition

$$\left(\nabla^2 - \frac{\partial^2}{\partial t^2}\right) s = 0. \tag{2.14}$$

A second condition is required in general to eliminate all of the freedom in the vector potentials.

General covariance is now within reach. The symmetry with which space and time, and also ϕ and $\mathbf{A}$, enter into these equations leads us to define spacetime vectors and derivatives:

$$\partial_\mu = \left(\frac{1}{c}\partial_t, \vec{\nabla}\right) \tag{2.15}$$

$$x^\mu = \begin{pmatrix} ct \\ \mathbf{x} \end{pmatrix}, \tag{2.16}$$

with Greek indices $\mu, \nu = 0, \ldots, n$ and $x^0 \equiv ct$. Repeated indices are summed according to the usual Einstein summation convention, and we define[1]

$$\Box = \partial^\mu \partial_\mu = -\frac{1}{c^2}\partial_t^2 + \nabla^2. \tag{2.17}$$

In n space dimensions and one time dimension ($n = 3$ normally), the $(n + 1)$ dimensional vector potential is defined by

$$A^\mu = \begin{pmatrix} \phi/c \\ \mathbf{A} \end{pmatrix}. \tag{2.18}$$

Using these $(n + 1)$ dimensional quantities, it is now possible to re-write eqn. (2.12) in an extremely beautiful and fully covariant form. First, one re-writes eqn. (2.10) as

$$-\Box\, \mathbf{A} = \frac{\mathbf{J}}{c^2\epsilon_0} - \vec{\nabla}\partial_\mu A^\mu. \tag{2.19}$$

[1] In some old texts, authors wrote $\Box^2$ for the same operator, since it is really a four-sided (four-dimensional) version of ∇^2.

https://doi.org/10.1017/9781009289887.004 Published online by Cambridge University Press

Next, one substitutes the gauge condition eqn. (2.13) into eqn. (2.9), giving

$$-\Box\,\phi = \frac{\rho_e}{\epsilon_0}. \tag{2.20}$$

Finally, the $(n+1)$ dimensional current is defined by

$$J^\mu = \begin{pmatrix} c\rho_e \\ \mathbf{J} \end{pmatrix}, \tag{2.21}$$

and we end up with the $(n+1)$ dimensional field equation

$$-\Box\,A^\mu = \mu_0 J^\mu, \tag{2.22}$$

where $c^2 = (\mu_0\epsilon_0)^{-1}$ has been used. The potential is still subject to the constraint in eqn. (2.13), which now appears as

$$\partial_\mu A^\mu = 0. \tag{2.23}$$

2.1.4 The field strength

The new attention given to the potential A_μ should not distract from the main aim of electromagnetism: namely to solve Maxwell's equations for the electric and magnetic fields. These two physical components also have a covariant formulation; they are now elegantly unified as the components of a rank 2 *tensor* which is denoted $F_{\mu\nu}$ and is defined by

$$F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu; \tag{2.24}$$

the tensor is anti-symmetric

$$F_{\mu\nu} = -F_{\nu\mu}. \tag{2.25}$$

This anti-symmetry, which was alluded to earlier, is the reason for the gauge invariance. The form of eqn. (2.24) is like a $(3+1)$ dimensional curl, expressed in index notation. The explicit components of this field tensor are the components of the electric and magnetic field components, in a Cartesian basis $\mathbf{E} = (E_1, E_2, E_3)$, etc.:

$$F_{\mu\nu} = \begin{pmatrix} 0 & -E_1/c & -E_2/c & -E_3/c \\ E_1/c & 0 & B_3 & -B_2 \\ E_2/c & -B_3 & 0 & B_1 \\ E_3/c & B_2 & -B_1 & 0 \end{pmatrix}. \tag{2.26}$$

In chapter 9, it will be possible to provide a complete understanding of how the symmetries of spacetime provide an explanation for why the electric and

https://doi.org/10.1017/9781009289887.004 Published online by Cambridge University Press

magnetic components of this field appear to be separate entities, in a fixed reference frame.

With the help of the potentials, *three* of Maxwell's equations (eqns. (2.1a,c,d)) are now expressed in covariant form. Eqn. (2.1c) is solved implicitly by the vector potential. The final equation (and also eqn. (2.1c), had one not used the vector potential) is an algebraic *identity*, called the Jacobi or Bianchi identity. Moreover, the fact that it is an identity is only clear when we write the equations in covariant form. The final equation can be written

$$\epsilon^{\mu\nu\lambda\rho}\partial_\mu F_{\nu\lambda} = 0, \tag{2.27}$$

where $\epsilon^{\mu\nu\lambda\rho}$ is the completely anti-symmetric tensor in four dimensions, defined by its components in a Cartesian basis:

$$\epsilon^{\mu\nu\lambda\rho} = \begin{cases} +1 & \mu\nu\lambda\rho = 0123 \text{ and even permutations} \\ -1 & \mu\nu\lambda\rho = 0132 \text{ and other odd permutations} \\ 0 & \text{otherwise.} \end{cases} \tag{2.28}$$

This equation is not a condition on $F_{\mu\nu}$, in spite of appearances. The anti-symmetry of both $\epsilon^{\mu\nu\lambda\rho}$ and $F_{\mu\nu}$ implies that the expansion of eqn. (2.27), in terms of components, includes many terms of the form $(\partial_\mu\partial_\nu - \partial_\nu\partial_\mu)A_\lambda$, the sum of which vanishes, provided A_λ contains no singularities. Since the vector potential is a continuous function in all physical systems,[2] the truth of the identity is not in question here.

The proof that this identity results in the two remaining Maxwell's equations applies only in $3+1$ dimensions. In other numbers of dimensions the equations must be modified. We shall not give it here, since it is easiest to derive using the index notation and we shall later re-derive our entire formalism consistently in that framework.

2.1.5 Covariant field equations using $F_{\mu\nu}$

The vector potential has been used thus far, because it was easier to identify the structure of the $(3+1)$ dimensional vectors than to guess the form of $F^{\mu\nu}$, but one can now go back and re-express the equations of motion in terms of the so-called physical fields, or field strength $F_{\mu\nu}$. The arbitrary choice of gauge in eqn. (2.22) is then eliminated.

Returning to eqn. (2.9) and adding and subtracting $\partial_0^2\phi$, one obtains

$$-\Box\,\phi - \partial_0(\partial_\nu A^\nu) = \frac{\rho_e}{\epsilon_0}. \tag{2.29}$$

[2] The field strength can never change by more than a step function, because of Gauss' law: the field is caused by charges, and a point charge (delta function) is the most singular charge that exists physically. This ensures the continuity of A_μ.

https://doi.org/10.1017/9781009289887.004 Published online by Cambridge University Press

Adding this to eqn. (2.19) (without choosing a value for $\partial_\nu A^\nu$), one has

$$-\Box A^\mu = \frac{J^\mu}{c^2\epsilon_0} - \partial^\mu(\partial_\nu A^\nu). \tag{2.30}$$

Taking the last term on the right hand side over to the left and using eqn. (2.17) yields

$$\partial_\nu(\partial^\mu A^\nu - \partial^\nu A^\mu) = \frac{J^\mu}{c^2\epsilon_0}. \tag{2.31}$$

The parenthesis on the left hand side is now readily identified as

$$\partial_\nu F^{\mu\nu} = \mu_0 J^\mu. \tag{2.32}$$

This is the covariant form of the field equations for the physical fields. It incorporates two of the four Maxwell equations as before (eqn. (2.1c) is implicit in the structure we have set up). The final eqn. (2.27) is already expressed in terms of the physical field strength, so no more attention is required.

2.1.6 Two invariants

There are two invariant, scalar quantities (no free indices) which can be written down using the physical fields in $(3+1)$ dimensions. They are

$$\mathcal{F} = F^{\mu\nu}F_{\mu\nu} \tag{2.33}$$

$$\mathcal{G} = \epsilon^{\mu\nu\lambda\rho}F_{\mu\nu}F_{\lambda\rho}. \tag{2.34}$$

The first of these evaluates to

$$\mathcal{F} = 2\left(\mathbf{B}^2 - \frac{1}{c^2}\mathbf{E}^2\right). \tag{2.35}$$

In chapter 4 this quantity is used to construct the *action* of the system: a generating function the dynamical behaviour. The latter gives

$$\mathcal{G} = \mathbf{E}\cdot\mathbf{B}. \tag{2.36}$$

In four dimensions, this last quantity vanishes for a self-consistent field: the electric and magnetic components of a field (resulting from the same source) are always perpendicular. In other numbers of dimensions the analogue of this invariant does not necessarily vanish.

The quantity $\mathcal{F}$ has a special significance. It turns out to be a Lagrangian, or generating functional, for the electromagnetic field. It is also related to the energy density of the field by a simple transformation.

https://doi.org/10.1017/9781009289887.004 Published online by Cambridge University Press

2.1.7 Gauge invariance and physical momentum

As shown, Maxwell's equations and the physical field $F_{\mu\nu}$ are invariant under gauge transformations of the form

$$A_\mu \to A_\mu + (\partial_\mu s). \tag{2.37}$$

It turns out that, when considering the interaction of the electromagnetic field with matter, the dynamical variables for matter have to change under this gauge transformation in order to uphold the invariance of the field equations.

First, consider classical particles interacting with an electromagnetic field. The force experienced by classical particles with charge q and velocity $\mathbf{v}$ is the Lorentz force

$$F_{\mathrm{EM}} = q(\mathbf{E} + \mathbf{v} \times \mathbf{B}). \tag{2.38}$$

The total force for an electron in an external potential V and an electromagnetic field is therefore

$$\frac{\mathrm{d}p_i}{\mathrm{d}t} = -e(E_i + \epsilon_{ijk} v_j B_k) - \partial_i V. \tag{2.39}$$

Expressing E and B in terms of the vector potential, we have

$$\partial_t(p_i - eA_i) = -eF_{ij}\dot{x}_j - \partial_i(V + eA_t). \tag{2.40}$$

This indicates that, apart from a gauge-invariant Biot–Savart contribution in the first term on the right hand side of this equation, the electromagnetic interaction is achieved by replacing the momentum p_i and the energy E by

$$p_\mu \to (p_\mu - eA_\mu). \tag{2.41}$$

The Biot–Savart term can also be accounted for in this way if we go over to a relativistic, Lorentz-covariant form of the equations:

$$\frac{\mathrm{d}}{\mathrm{d}\tau}(p_\mu - eA_\mu) + F_{\mu\nu} I_l^\nu = 0, \tag{2.42}$$

where $I_l^\mu = -e\mathrm{d}x^\mu/\mathrm{d}\tau \sim I\mathrm{d}\mathbf{l}$ is the current in a length of wire $\mathrm{d}x$ (with dimensions current $\times$ length) and τ is the proper time. In terms of the more familiar current density, we have

$$\frac{\mathrm{d}}{\mathrm{d}\tau}(p_\mu - eA_\mu) + \int \mathrm{d}\sigma\, F_{\mu\nu} J^\nu = 0. \tag{2.43}$$

We can now investigate what happens under a gauge transformation. Clearly, these equations of motion can only be invariant if p_μ also transforms so as to cancel the term, $\partial_\mu s$, in eqn. (2.37). We must have in addition

$$p_\mu \to p_\mu + e\partial_\mu s. \tag{2.44}$$

https://doi.org/10.1017/9781009289887.004 Published online by Cambridge University Press

Without a deeper appreciation of *symmetry*, this transformation is hard to understand. Arising here in a classical context, where symmetry is not emphasized, it seems unfamiliar. What is remarkable, however, is that the group theoretical notions of quantum theory of matter makes the transformation very clear. The reason is that the state of a quantum mechanical system is formulated very conveniently as a vector in a group theoretical vector space. Classically, positions and momenta are not given a state-space representation. In the quantum formulation, gauge invariance is a simple consequence of the invariance of the equations of motion under changes of the arbitrary complex phase of the quantum state or wavefunction.

In covariant vector language, the field equations are invariant under a redefinition of the vector potential by

$$A_\mu \to A_\mu + (\partial_\mu s), \tag{2.45}$$

where $s(x)$ is any scalar field. This symmetry is not only a mathematical curiosity; it also has a physical significance, which has to do with conservation.

2.1.8 Wave solutions to Maxwell's equations

The equation for harmonic waves $W(x)$, travelling with speed v, is given by

$$\left(\nabla^2 - \frac{1}{v^2}\frac{\partial^2}{\partial t^2}\right) W(x) = 0. \tag{2.46}$$

If the speed of the waves is $v = c$, this may be written in the compact form

$$-\Box\, W(x) = 0. \tag{2.47}$$

It should already be apparent from eqn. (2.22) that Maxwell's equations have wavelike solutions which travel at the speed of light. Writing eqn. (2.22) in terms of the field strength tensor, we have

$$-\Box\, F_{\mu\nu} = \mu_0(\partial_\mu J_\nu - \partial_\nu J_\mu). \tag{2.48}$$

In the absence of electric charge $J_\mu = 0$, the solutions are free harmonic waves. When $J_\mu \neq 0$, Maxwell's equations may be thought of as the equations of forced oscillations, but this does not necessarily imply that all the solutions of Maxwell's equations are wavelike. The Fourier theorem implies that any function may be represented by a suitable linear super-position of waves. This is understood by noting that the source in eqn. (2.48) is the spacetime 'curl' of the current, which involves an extra derivative. Eqn. (2.32) is a more useful starting point for solving the equations for many sources. The free wave solutions for the field are linear combinations of plane waves with constant coefficients:

$$A_\mu(k) = C_k \exp(\mathrm{i}k_\mu x^\mu). \tag{2.49}$$

https://doi.org/10.1017/9781009289887.004 Published online by Cambridge University Press

By substituting this form into the equation

$$-\Box A_\mu = 0, \tag{2.50}$$

one obtains a so-called dispersion relation for the field:

$$k_\mu k^\mu = k^2 = \mathbf{k}^2 - \omega^2/c^2 = 0. \tag{2.51}$$

This equation may be thought of as a constraint on the allowed values of k. The total field may be written compactly in the form

$$A_\mu(x) = \int \frac{\mathrm{d}^{n+1}k}{(2\pi)^{n+1}} \mathrm{e}^{\mathrm{i}k_\mu x^\mu}\ A_\mu(k)\ \delta(k^2), \tag{2.52}$$

where $A(k)_\mu$ represents the amplitude of the wave with wavenumber k_i, and the vector index specifies the polarization of the wave modes. From the gauge condition in eqn. (2.23), we have

$$k_\mu A(k)^\mu = 0. \tag{2.53}$$

The delta-function constraint in eqn. (2.52) ensures that the combination of waves satisfies the dispersion relation in eqn. (2.51). If we use the property of the delta function expressed in Appendix A, eqn. (A.15), then eqn. (2.52) may be written

$$A_\mu(x) = \hat{\epsilon}_\mu \int \frac{\mathrm{d}^{n+1}k}{(2\pi)^{n+1}} \mathrm{e}^{\mathrm{i}(k_i x^i - \omega t)}\ A(k)\ \frac{1}{c\,k_i}\left(\frac{\partial\omega}{\partial k^i}\right) \times \left(\delta(k_0 - \sqrt{\mathbf{k}^2}) + \delta(k_0 + \sqrt{\mathbf{k}^2})\right). \tag{2.54}$$

The delta functions ensure that the complex exponentials are waves travelling at the so-called phase velocity

$$v^i_{\mathrm{ph}} = \pm\frac{\omega}{k_i} \tag{2.55}$$

where ω and k_i satisfy the dispersion relation. The amplitude of the wave clearly changes at the rate

$$v^i_{\mathrm{gr}} = \frac{\partial\omega}{\partial k_i}, \tag{2.56}$$

known as the group velocity. By choosing the coefficient $C(k)$ for each frequency and wavelength, the super-position principle may be used to build up any complementary (steady state) solution to the Maxwell field. We shall use this approach for other fields in chapter 5.

https://doi.org/10.1017/9781009289887.004 Published online by Cambridge University Press

2.2 Conservation laws

The simple observation of 'what goes in must come out' applies to many physical phenomena, including electromagnetism, and forms a predictive framework with rich consequences. Conservation is a physical fact, which must be reflected in dynamical models. Just as economics uses money as a book-keeping parameter for transactions, so physics accounts for transactions (interactions) with energy, charge and a variety of similarly contrived labels which have proven useful in keeping track of 'stock' in the physical world.

2.2.1 Current conservation

Perhaps the central axiom of electromagnetism is the conservation of total electric charge. An algebraic formulation of this hypothesis provides an important relationship, which will be referred to many times. Consider the electric current I, defined in terms of the rate of flow of charge:

$$I = \int \mathrm{d}\sigma \cdot \mathbf{J} = \frac{\mathrm{d}Q}{\mathrm{d}t}. \tag{2.57}$$

Expressing the charge Q as the integral over the charge density, one has

$$\int \nabla \cdot \mathbf{J} \mathrm{d}\sigma = -\partial_t \int \rho_e \mathrm{d}\sigma. \tag{2.58}$$

Comparing the integrand on the left and right hand sides gives

$$\frac{\partial \rho_e}{\partial t} + \vec{\nabla} \cdot \mathbf{J} = 0, \tag{2.59}$$

or, in index notation,

$$\partial_i J^i = -\partial_t \rho_e. \tag{2.60}$$

This may now be expressed in 4-vector language (or $(n+1)$-vector language), and the result is:

$$\partial_\mu J^\mu = 0. \tag{2.61}$$

This result is known as a *continuity condition* or a *conservation law*. All conservation laws have this essential form, for some $(n+1)$ dimensional current vector J^μ. The current is then called a conserved current. In electromagnetism one can verify that the conservation of current is compatible with the equations of motion very simply in eqn. (2.32) by taking the 4-divergence of the equation:

$$\partial_\mu \partial_\nu F^{\mu\nu} = \mu_0 \, \partial_\mu J^\mu = 0. \tag{2.62}$$

The fact that the left hand size is zero follows directly from the anti-symmetrical and non-singular nature of $F_{\mu\nu}$.

https://doi.org/10.1017/9781009289887.004 Published online by Cambridge University Press

2.2.2 Poynting's vector and energy conservation

The electromagnetic field satisfies a continuity relation of its own. This relation is arrived at by considering the energy flowing though a unit volume of the field. The quantities defined below will later re-emerge in a more general form as the so-called field theoretical energy–momentum tensor.

The passage of energy through an electromagnetic system may be split up into two contributions. The first is the work done on any electric charges contained in the volume. This may be expressed in terms of the current density and the electric field as follows. The rate at which work is done on a moving charge is given by the force vector dotted with the rate of change of the displacement (i.e. the velocity), $\mathbf{F} \cdot \mathbf{v}$. The force, in turn, is given by the charge multiplied by the electric field strength $q\mathbf{E}$, which we may write in terms of the charge density ρ_e inside a spatial volume $\mathrm{d}\sigma$ as $\rho_e \mathbf{E} \mathrm{d}\sigma$. The rate at which work is done on charges may now be expressed in terms of an external source or current, by identifying the external current to be the density of charge which is flowing out of some volume of space with a velocity $\mathbf{v}$

$$\mathbf{J}_{\text{ext}} = \rho_e \mathbf{v}. \tag{2.63}$$

We have

$$\text{Rate of work} = \mathbf{E} \cdot \mathbf{J}_{\text{ext}} \mathrm{d}\sigma. \tag{2.64}$$

The second contribution to the energy loss from a unit volume is due to the flux of radiation passing through the surface (S) which bounds the infinitesimal volume (σ). This flux of radiation is presently unknown, so we shall refer to it as $\mathcal{S}$. If we call the total energy density $\mathcal{H}$, then we may write that the loss of energy from an infinitesimal volume is the sum of these two contributions:

$$-\partial_t \int_\sigma \mathcal{H} \, \mathrm{d}\sigma = \int_S \mathcal{S} \cdot \mathrm{d}\mathbf{S} + \int_\sigma \mathbf{E} \cdot \mathbf{J}_{\text{ext}} \mathrm{d}\sigma. \tag{2.65}$$

In 1884, Poynting identified $\mathcal{H}$ and $\mathbf{S}$ using Maxwell's equations. We shall now do the same. The aim is to eliminate the current $\mathbf{J}_{\text{ext}}$ from this relation in order to express $\mathcal{H}$ and $\mathbf{S}$ in terms of $\mathbf{E}$ and $\mathbf{B}$ only. We begin by using the fourth Maxwell equation (2.1d) to replace $\mathbf{J}_{\text{ext}}$ in eqn. (2.65):

$$\mathbf{E} \cdot \mathbf{J}_{\text{ext}} = \frac{\mathbf{E} \cdot (\nabla \times \mathbf{B})}{\mu_0} - \epsilon_0 \, \mathbf{E} \cdot \partial_t \mathbf{E}. \tag{2.66}$$

Using the vector identity in Appendix A, eqn. (A.71), we may write

$$\mathbf{E} \cdot (\nabla \times \mathbf{B}) = \nabla \cdot (\mathbf{B} \times \mathbf{E}) + \mathbf{B} \cdot (\nabla \times \mathbf{E}). \tag{2.67}$$

The second Maxwell eqn. (2.1b) may now be used to replace $\nabla \times \mathbf{E}$, giving

$$\mathbf{E} \cdot (\nabla \times \mathbf{B}) = \nabla \cdot (\mathbf{B} \times \mathbf{E}) - \mathbf{B} \partial_t \mathbf{B}. \tag{2.68}$$

https://doi.org/10.1017/9781009289887.004 Published online by Cambridge University Press

Finally, noting that

$$\frac{1}{2}\partial_t(\mathbf{X}\cdot\mathbf{X}) = \mathbf{X}\partial_t\mathbf{X}, \tag{2.69}$$

and using this with $\mathbf{X} = \mathbf{E}$ and $\mathbf{X} = \mathbf{B}$ in eqns. (2.66) and (2.68), we may write:

$$\mathbf{E}\cdot\mathbf{J}_{\text{ext}} = \frac{\nabla\cdot(\mathbf{B}\times\mathbf{E})}{\mu_0} - \frac{1}{2}\partial_t\left(\epsilon_0\mathbf{E}\cdot\mathbf{E} + \frac{1}{\mu_0}\mathbf{B}\cdot\mathbf{B}\right). \tag{2.70}$$

This equation has precisely the form of eqn. (2.65), and the pieces can now be identified:

$$\begin{aligned} S^0 = \mathcal{H} &= \frac{1}{2}\left(\epsilon_0\mathbf{E}\cdot\mathbf{E} + \frac{1}{\mu_0}\mathbf{B}\cdot\mathbf{B}\right) \\ &\equiv \frac{1}{2}\left(\mathbf{E}\cdot\mathbf{D} + \mathbf{B}\cdot\mathbf{H}\right) \end{aligned} \tag{2.71}$$

$$\begin{aligned} S^i = \mathbf{S} &= \frac{\mathbf{E}\times\mathbf{B}}{c\mu_0} \\ &\equiv \frac{\mathbf{E}\times\mathbf{H}}{c}. \end{aligned} \tag{2.72}$$

The new fields $\mathbf{D} = \epsilon_0\mathbf{E}$ and $\mu_0\mathbf{H} = \mathbf{B}$ have been defined. The energy density $\mathcal{H}$ is often referred to as a Hamiltonian for the free electromagnetic field, whereas $\mathbf{S}$ is referred to as the Poynting vector.

$$\partial_\mu S^\mu = \left(F_{0\mu}J^\mu_{\text{ext}}\right) \tag{2.73}$$

is the rate at which work is done by an infinitesimal volume of the field. It is clear from the appearance of an explicit zero component in the above that this argument cannot be the whole story. One expects a generally covariant expression. The expression turns out to be

$$\partial_\mu\theta^{\mu\nu}_{\text{Maxwell}} = F^{\mu\nu}J_\mu, \tag{2.74}$$

where $\theta_{\mu\nu}$ is the energy–momentum tensor. Notice how it is a surface integral which tells us about flows in and out of a volume of space. One meets this idea several times, in connection with boundary conditions and continuity.

2.3 Electromagnetism in matter

To describe the effect of matter on the electromagnetic field in a covariant way, one may use either a microscopic picture of the field interacting with matter at the molecular level, or a macroscopic, effective field theory, which hides the details of these interactions by defining equivalent fields.

https://doi.org/10.1017/9781009289887.004 Published online by Cambridge University Press

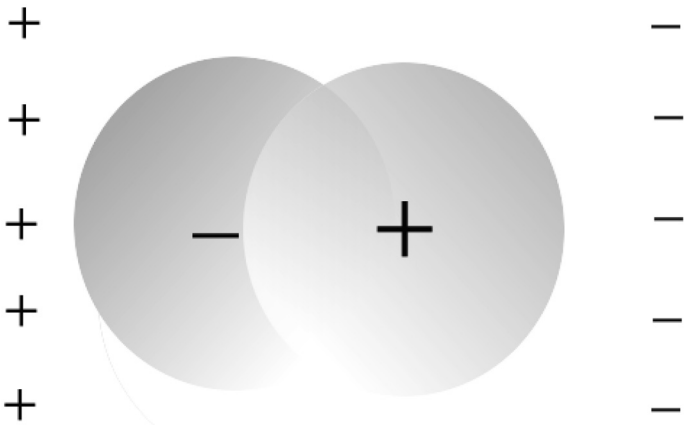

Fig. 2.1. Matter is not electrically neutral at the microscopic level.

2.3.1 Dielectrics

One tends to think of ordinary matter as being electrically neutral, but of course it is composed of atoms and molecules, which have a centre of positive charge and a centre of negative charge – and these two centres do not necessarily lie at the same place. The more symmetrical a molecule is, the more neutral it is: for instance, the noble gases have highly symmetrical electron orbits and thus have almost no polarizability on average; the water molecule, on the other hand, has an asymmetry which allows a trickle of water to be attracted by a charged comb.

When an electric field is switched on in the vicinity of a dielectric material, the centres of positive and negative charge in each molecule are forced apart slightly (see figure 2.1) in a substance-dependent way. We say that such a molecule has a certain polarizability.

For classical external fields, atoms and molecules behave like dipoles, i.e. there is a clear separation of the charge into two parts: a positive pole and a negative pole. But we would be doing a disservice to the radiation field (not to mention the quantum theory) if we did not recall that the field has a wave nature and a characteristic wavelength. Molecules behave like dipoles if the wavelength of the external field is large compared to the size of the molecule – since then there is a clear direction to the field at every point inside the molecule's charge cloud. If, on the other hand, the wavelength of the field is of the order of the size of the molecule or less, then the field can reverse direction inside the molecule itself. The charge then gets re-distributed into a more complex pattern, and so-called quadrapole moments and perhaps higher 'pole moments' must be taken into account. In this text, we shall only consider the dipole approximation.

The simplest way to model the polarization of an atom or molecule is to view it as opposite charges coupled together by a spring. This model is adequate for

https://doi.org/10.1017/9781009289887.004 Published online by Cambridge University Press

many materials, provided the external electric field is not too strong. Materials which behave like this are called *linear* media. Initially, the centres of positive and negative charge are in the same place, but, as the external field is switched on, they are pulled further and further apart, balanced by the restoring force of the spring. This separation of charge creates a new local field, which tends to cancel the external field inside the molecule, but to reinforce it at the poles. If the charge clouds have charge q and the spring constant is κ then the force equation is simply

$$F = -\kappa s = Eq, \tag{2.75}$$

where s is the displacement of the charges from one another, in the rest frame of the atoms. The separation multiplied by the charge gives the effective contribution to the field at the poles of a single molecule, denoted the *dipole moment* $\mathbf{d}$:

$$\mathbf{d} = |\mathbf{s}|q = \frac{q^2}{\kappa}\mathbf{E}. \tag{2.76}$$

The quantity q^2/κ is denoted by α and is called the *polarizability*; it denotes the effective strength of the resistance to polarization. The *polarization* field is

$$\mathbf{P} = \rho_N \mathbf{d} = \rho_N \alpha \mathbf{E} \tag{2.77}$$

where ρ_N is the total number of molecules per unit volume. It is proportional to the field of particles displacements $s^i(x)$ and it hides some non-covariant assumptions (see the next section). Normally speaking, one writes $q = -e$, where $-e$ is the charge on the electron. Then,

$$\alpha_{\text{static}} = \frac{-q^2}{\kappa}. \tag{2.78}$$

If one considers time-varying fields, or specifically waves of the form

$$\mathbf{E} = \mathbf{E}_0 \mathrm{e}^{\mathrm{i}(\mathbf{k}\cdot\mathbf{x}-\omega t)}, \tag{2.79}$$

it is found that, for a single *optically active* electron (i.e. one in an orbit which can be affected by an external field), the equation of motion is now that of a damped harmonic oscillator:

$$m(\omega_0 + \mathrm{i}\gamma\omega - \omega^2)\mathbf{s} = -e\mathbf{E}_0, \tag{2.80}$$

where $\omega_0^2 = \kappa/m$ and γ is a damping term. Using this equation to replace for $\mathbf{s}$ in eqn. (2.76), we get

$$\alpha(\omega) = \frac{q^2/m}{(\omega_0^2 + \mathrm{i}\gamma\omega - \omega^2)}. \tag{2.81}$$

https://doi.org/10.1017/9781009289887.004 Published online by Cambridge University Press

Thus the polarizability is a frequency-dependent quantity. This explains why a prism can split a mixture of white light into its component frequencies. A further definition is of interest, namely the *electric susceptibility* $\chi_{\rm e} = N\alpha(\omega)/\epsilon_0$. For ρ_N particles per unit volume, this is often expressed in terms of the plasma frequency $\omega_{\rm p}^2 = Ne^2/m$. Thus,

$$\mathbf{P} = \epsilon_0 \chi_{\rm e} \mathbf{E}. \tag{2.82}$$

This is closely related to the change in the refractive index, $n^2 = 1 + \chi_{\rm e}$, of a material due to polarization, when $\mu_{\rm r} = 1$ (which it nearly always is). In real space, we note from eqn. (2.80) that the polarization satisfies a differential equation

$$(\partial_t^2 - \gamma\,\partial_t + \omega_0^2)\mathbf{P} = \frac{q^2}{m}\rho_N \mathbf{E} \tag{2.83}$$

and thus the real space susceptibility can be thought of as a Green function for the differential operator and $\mathbf{E}$ may be thought of as a source.

$$\mathbf{P}(t) = \epsilon_0 \int {\rm d}t\; \chi(t - t')\mathbf{E}. \tag{2.84}$$

$\chi(t-t')$ is taken to satisfy retarded boundary conditions, which, in turn, implies that its real and imaginary parts in terms of ω are related. The relationship is referred to as a *Kramers–Kronig relation*, and is simply a consequence of the fact that a retarded Green function is *real*.

2.3.2 Covariance and relative motion: the Doppler effect

The frequency-dependent expressions above are true only in the rest frame of the atoms. The results are not covariant with respect to moving frames of reference. When one studies solid state systems, such as glasses and crystals, these expressions are quite adequate, because the system has a naturally preferred rest frame and the atoms in the material do not move relative to one another, on average. However, in gaseous matter or plasmas, this is not the case: the thermal motion of atoms relative to one another can be important, because of the Doppler effect. This fact can be utilized to good effect; for example, in laser cooling the motion of atoms relative to a laser radiation field can be used to bring all of the atoms into a common rest frame by the use of a resonant, frequency-dependent interaction. A Galilean-covariant expression can be written by treating the field as one co-moving mass, or as a linear super-position of co-moving masses. With only one component co-moving, the transformation of the position of an atom in the displacement field can be written

$$\mathbf{x}(t) \to \mathbf{x} + \mathbf{v}t, \tag{2.85}$$

https://doi.org/10.1017/9781009289887.004 Published online by Cambridge University Press

where $\mathbf{v}$ is the velocity of the motion relative to some reference frame (usually the container of the gaseous matter, or the laboratory frame). This means that the invariant form $(kx - \omega t)$ is transformed into

$$\mathbf{k}\cdot\mathbf{x} - \omega t \to \mathbf{k}\cdot(\mathbf{x} + \mathbf{v}t) - \omega t = \mathbf{k}\cdot\mathbf{x} - \omega_\beta t, \tag{2.86}$$

where

$$\omega_\beta = \omega(1 - \hat{\mathbf{k}}\cdot\beta) = \omega(1 - \hat{\mathbf{k}}\cdot\mathbf{v}/c). \tag{2.87}$$

Thus, the expressions above can be used, on replacing ω with a sum over all ω_β, and integrated over all the values of the velocity vector β^i of which the field is composed. The polarizability takes the form

$$\alpha(\omega) = \frac{q^2/m}{(\omega_0^2 + \mathrm{i}\gamma\omega - \omega_\beta^2)}. \tag{2.88}$$

where

$$\omega_\beta = (1 - \hat{k}^i\beta_i)\omega. \tag{2.89}$$

2.3.3 Refractive index in matter

It appears that the introduction of a medium destroys the spacetime covariance of the equations of motion. In fact this is false. What is interesting is that covariance can be restored by re-defining the $(n + 1)$ dimensional vectors so as to replace the speed of light in a vacuum with the *effective* speed of light in a medium. The speed of light in a dielectric medium is

$$v = \frac{c}{n} \tag{2.90}$$

where $n = \epsilon_\mathrm{r}\mu_\mathrm{r} > 1$ is the refractive index of the medium.

Before showing that covariance can be restored, one may consider the equation of motion for waves in a dielectric medium from two perspectives. The purpose is to relate the multifarious fields to the refractive index itself. It is also to demonstrate that the polarization terms can themselves be identified as currents which are set in motion by the electric field. In other words, we will show the equivalence of (i) $\mathbf{P} \neq 0$, but $J_\mu = 0$, and (ii) $\mathbf{P} = 0$ with J_μ given by the current resulting from charges on springs! Taking

$$J^0 = c\rho_e J^i = -\rho_N ec\frac{\mathrm{d}s^i}{\mathrm{d}t}, \tag{2.91}$$

the current is seen to be a result of the net charge density set in motion by the field. This particular derivation applies only in $3 + 1$ dimensions.

https://doi.org/10.1017/9781009289887.004 Published online by Cambridge University Press

To obtain the wave equation we begin with the Bianchi identity

$$\epsilon_{ijk}\partial_j E_k + \partial_t B_i = 0, \tag{2.92}$$

and then operate from the left with $\epsilon_{ilm}\partial_l$. Using the identity (A.38) (see Appendix A) for the product of two anti-symmetric tensors, we obtain

$$\left[\nabla^2 E_i - \partial_i(\partial^j E_j)\right] + \epsilon_{ilm}\partial_l\partial_t B_i = 0. \tag{2.93}$$

Taking ∂_t of the fourth Maxwell equation, one obtains

$$\frac{1}{\mu_0\mu_\mathrm{r}}\epsilon_{ijk}\partial_j\partial_t B_k = \partial_t J_i + \epsilon_0\epsilon_\mathrm{r}\frac{\partial^2 E_i}{\partial t^2}. \tag{2.94}$$

These two equations can be used to eliminate B_i, giving an equation purely in terms of the electric field. Choosing the charge distribution to be isotropic (uniform in space), we have $\partial_i\rho_e = 0$, and thus

$$\left[\nabla^2 - \frac{n^2}{c^2}\frac{\partial^2}{\partial t^2}\right] E_i = \mu_0\mu_\mathrm{r}\partial_t J_i. \tag{2.95}$$

In this last step, we used the definition of the refractive index in terms of ϵ_r:

$$n^2 = \epsilon_r\mu_\mathrm{r} = (1+\chi_\mathrm{e})\mu_\mathrm{r}. \tag{2.96}$$

This result is already suggestive of the fact that Maxwell's equations in a medium can be written in terms of an effective speed of light.

We may now consider the two cases: (i) $\mathbf{P} \neq 0$, but $J_\mu = 0$,

$$\left[\nabla^2 - \frac{n^2}{c^2}\frac{\partial^2}{\partial t^2}\right] E_i = 0; \tag{2.97}$$

and (ii) $\mathbf{P} = 0$ ($n = 1$), $J_\mu \neq 0$.

$$\left[\nabla^2 - \frac{1}{c^2}\frac{\partial^2}{\partial t^2}\right] E_i = \mu_0\mu_\mathrm{r}\frac{-\rho_N e^2\omega^2/m \cdot E_i}{(\omega_0^2 + \mathrm{i}\gamma\omega - \omega^2)}. \tag{2.98}$$

The differential operators on the left hand side can be replaced by $\mathbf{k}^2$ and ω^2, by using the wave solution (2.79) for the electric field to give a 'dispersion relation' for the field. This gives:

$$\begin{aligned}\frac{\mathbf{k}^2}{\omega^2} &= \frac{1}{c^2}\left(1 + \frac{\mu_\mathrm{r}}{\epsilon_0}\frac{\rho_N e^2\omega^2/m}{(\omega_0^2 + \mathrm{i}\gamma\omega - \omega^2)}\right) \\ &= \frac{\mathbf{n}^2}{c^2}.\end{aligned} \tag{2.99}$$

https://doi.org/10.1017/9781009289887.004 Published online by Cambridge University Press

So, from this slightly cumbersome expression for the motion of charges, one derives the microscopic form of the refractive index. In fact, comparing eqns. (2.99) and (2.98), one sees that

$$n^2 = 1 + \frac{\rho_N \alpha(\omega) \mu_{\rm r}}{\epsilon_0}. \tag{2.100}$$

Since $\mu_{\rm r}$ is very nearly unity in all materials that waves penetrate, it is common to ignore this and write

$$n^2 \sim 1 + \chi_{\rm e}. \tag{2.101}$$

The refractive index is a vector in general, since a material could have a different index of refraction in different directions. Such materials are said to be anisotropic. One now has both microscopic and macroscopic descriptions for the interaction of radiation with matter, and it is therefore possible to pick and choose how one wishes to represent this physical system. The advantage of the microscopic formulation is that it can easily be replaced by a quantum theory at a later stage. The advantage of the macroscopic field description is that it is clear why the form of Maxwell's equations is unaltered by the specific details of the microscopic interactions.

2.4 Aharonov–Bohm effect

The physical significance of the vector potential A_μ (as opposed to the field $F_{\mu\nu}$) was moot prior to the arrival of quantum mechanics. For many, the vector potential was merely an artifice, useful in the computation of certain boundary value problems. The formulation of quantum mechanics as a local field theory established the vector potential as the fundamental *local* field, and the subsequent attention to gauge symmetry fuelled pivotal developments in the world of particle physics. Today, it is understood that there is no fundamental difference between treating the basic electromagnetic interaction as a rank 2 anti-symmetric tensor $F_{\mu\nu}$, or as a vector with the additional requirement of gauge invariance. They are equivalent representations of the problem. In practice, however, the vector potential is the easier field to work with, since it couples locally. The price one pays lies in ensuring that gauge invariance is maintained (see chapter 9).

The view of the vector potential as a mathematical construct was shaken by the discovery of the Aharonov–Bohm effect. This was demonstrated is a classic experiment of electron interference through a double slit, in which electrons are made to pass through an area of space in which $A_\mu \neq 0$ but where $F_{\mu\nu} = 0$. The fact that a change in the electron interference pattern was produced by this configuration was seen as direct evidence for the physical reality of A_μ. Let us examine this phenomenon.

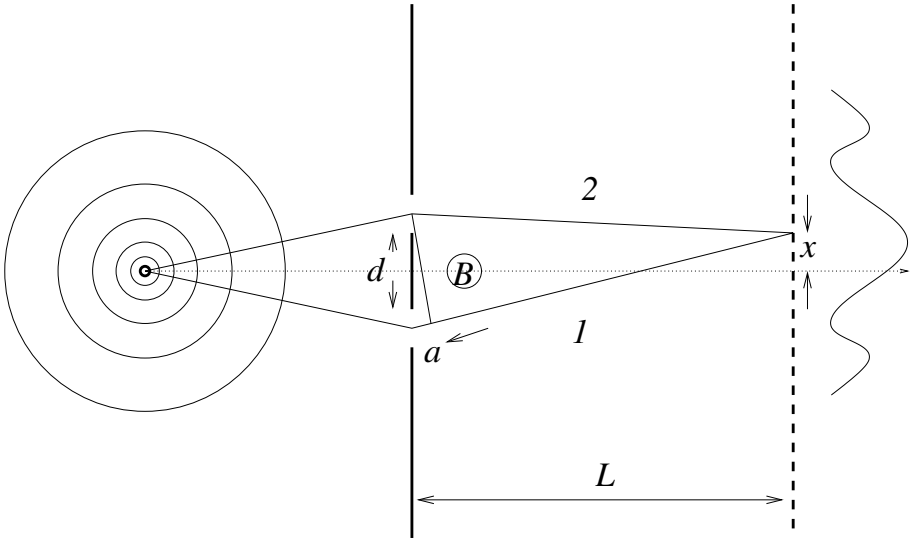

Fig. 2.2. The Aharonov–Bohm experiment.

The physical layout of the double-slit experiment is shown in figure 2.2. An electron source fires electrons at the slits, these pass through the slits and interfere in the usual way, forming an interference pattern on the screen at the end of their path. In order to observe the Aharonov–Bohm effect, one places a solenoid on the far side of the slits, whose magnetic field is constrained within a cylinder of radius R. The vector potential arising from the solenoid geometry is not confined to the inside of the solenoid however. It also extends outside of the solenoid, but in such a way as to produce no magnetic field.

What is remarkable is that, when the solenoid is switched on, the interference pattern is shifted by an amount x. This indicates that a phase shift $\Delta\theta$ is introduced between the radiation from the two slits, and is caused by the presence of the solenoid. If the distance L is much greater than x and a then we have

$$\begin{aligned} a &\sim \frac{x}{L}\, d \\ \Delta\theta &= 2\pi \left(\frac{L_1 - L_2}{\lambda} \right) = \frac{2\pi a}{\lambda} \\ x &= \left(\frac{L\lambda}{2\pi d} \right) \Delta\theta. \end{aligned} \tag{2.102}$$

The phase difference can be accounted for by the gauge transformation of the electron field by the vector potential. Although the absolute value of the vector potential is not gauge-invariant, the potential difference between the paths is.

https://doi.org/10.1017/9781009289887.004 Published online by Cambridge University Press

The vector potential inside and outside the solenoid position is

$$\begin{aligned}(r < R): \quad & A_\phi = \frac{1}{2}Br, \quad A_r = A_z = 0 \\ (r > R): \quad & A_\phi = \frac{BR^2}{2r}, \quad A_r = A_z = 0.\end{aligned} \tag{2.103}$$

The magnetic field in the regions is

$$\begin{aligned}B_z &= \nabla_r A_\phi - \nabla_\phi A_r \\ &= \nabla_r A_\phi \\ &= \frac{1}{r}\left[\frac{\partial}{\partial_r}(rA_\phi) - \frac{\partial}{\partial\phi}A_r\right] \\ &= 0 \quad (r < R) \\ &= B \quad (r > R).\end{aligned} \tag{2.104}$$

The phase difference can be determined, either from group theory, or from quantum mechanics to be

$$\exp(\mathrm{i}\theta) = \exp\left(\mathrm{i}\frac{e}{\hbar}\int_P A^i \mathrm{d}x_i\right), \tag{2.105}$$

where 'P' indicates the integral along a given path. Around the closed loop from one slit to the screen and back to the other slit, the phase difference is (using Stokes' theorem)

$$\begin{aligned}\Delta\theta &= \theta_1 - \theta_2 \\ &\sim \frac{e}{\hbar}\oint A_\phi \,\mathrm{d}r \\ &= \frac{e}{\hbar}\int (\vec{\nabla}\times\mathbf{B})\cdot\mathrm{d}\mathbf{S} \\ &= \frac{e}{\hbar}\int \mathbf{B}\cdot\mathrm{d}\mathbf{S}.\end{aligned} \tag{2.106}$$

The phase shift therefore results from the paths having to travel around the solenoid, i.e. in a loop where magnetic flux passes through a small part of the centre. Note, however, that the flux does not pass through the path of the electrons, only the vector potential is non-zero for the straight-line paths.

There are two ways of expressing this: (i) electrons must be affected by the vector potential, since the field is zero for any classical path from the slits to the screen; or (ii) electrons are stranger than we think: they seem to be affected by a region of space which is classically inaccessible to them. The viewpoints are really equivalent, since the vector potential is simply an analytic extension of the field strength, but the result is no less amazing. It implies a non-locality in

https://doi.org/10.1017/9781009289887.004 Published online by Cambridge University Press

the action of the magnetic field: action at a distance, and not only at a distance, but from within a container. If one chooses to believe in the vector potential as a fundamental field, the behaviour seems less objectionable: the interaction is then local. There is no action at a distance, and what happens inside the solenoid is of less interest.

Whether one chooses to view this as evidence for the physical reality of the vector potential or of the strangeness of quantum mechanics is a matter of viewpoint. Indeed, the reality of any field is only substantiated by the measurable effect it has on experiments. However, there are deeper reasons for choosing the interpretation based on the reality of the vector potential, which have to do with locality and topology, so at the very least this gives us a new respect for the utility of the vector potential. In view of the utility of A_μ and its direct appearance in dynamical calculations, it seems reasonable to accept it as the fundamental field in any problem which is simplified by that assumption.