ones. We can in fact learn something about it by noting that, since $m^2 = 0$, and from (2.210), we have

$$W \cdot W|k\rangle = 0, \quad P \cdot P|k\rangle = 0 \tag{2.214}$$

and from (2.209)

$$W \cdot P|k\rangle = 0,$$

so W^μ and P^μ are orthogonal and both lightlike. This means they must be proportional,

$$(W^\mu - \lambda P^\mu)|k\rangle = 0, \tag{2.215}$$

and we have the result that the state of a massless particle is characterised by *one number* λ, which is the ratio of W^μ and P^μ and so has the dimensions of angular momentum. It is called *helicity*. If parity is included, the helicity takes on two values, λ and $-\lambda$. What seems (at least to me) to be a mystery is why λ is integral or half-integral.

This reproduces what we know about the neutrino and the photon (assuming the neutrino, or one of them, is massless). A left-handed massless neutrino obeys the Weyl equation and has $\lambda = -\frac{1}{2}$. Photons come in both right- and left-circularly polarised states, with $\lambda = \pm 1$ – but not $\lambda = 0$, which would appear if the photon were massive.

2.8 Maxwell and Proca equations

We turn now to particles of spin 1. Photons have no mass and are described by Maxwell's equations, and massive spin 1 particles (for example, the intermediate bosons $W^\pm$ of weak interactions) are described by the Proca equation.

Maxwell's equations are, of course, well known. Our only concern here is to show how they are cast in a *manifestly* Lorentz covariant form. It is clear that they are, in fact, Lorentz covariant: it was Einstein's observation that Maxwell's equations were Lorentz covariant that gave birth to the theory of relativity. We seek only a notation which expresses the covariance as neatly as possible.

Maxwell's equations (in Heaviside–Lorentz rationalised units, so that $e^2/4\pi hc = \alpha = 1/137$) are

$$(a)\ \operatorname{div}\mathbf{B} = 0, \quad (b)\ \operatorname{curl}\mathbf{E} + \frac{\partial \mathbf{B}}{\partial t} = 0,$$

$$(c)\ \operatorname{div}\mathbf{E} = \rho, \quad (d)\ \operatorname{curl}\mathbf{B} - \frac{\partial \mathbf{E}}{\partial t} = \mathbf{j}. \tag{2.216}$$

(a) tells us there are no magnetic charges. (b) is Faraday's law; a changing magnetic field produces an electric field. (c) is Gauss's law; the total charge inside a closed surface may be obtained by integrating the normal component of **E** over the surface. (*d*) is Ampere's law, curl $\mathbf{B} = \mathbf{j}$, with Maxwell's additional term $\partial \mathbf{E}/\partial t$, stating that changing electric fields produce magnetic fields. The equations (*a*) and (*b*) are known as the homogeneous equations, (*c*) and (*d*) as the inhomogeneous ones.

Introducing the 4-vector potential

$$A^\mu = (\phi, \mathbf{A}) \tag{2.217}$$

with

$$\mathbf{B} = \operatorname{curl} \mathbf{A}, \quad \mathbf{E} = -\frac{\partial \mathbf{A}}{\partial t} - \nabla \phi, \tag{2.218}$$

equations (*a*) and (*b*) are *automatically satisfied*, since $\operatorname{div}\operatorname{curl} \equiv 0$ and $\operatorname{curl}\operatorname{grad} \equiv 0$. Now observe that the right-hand sides of equations (2.218) are the components of a 4-dimensional curl, defined by

$$F^{\mu\nu} = -F^{\nu\mu} = \partial^\mu A^\nu - \partial^\nu A^\mu. \tag{2.219}$$

It has components (recall that $\partial^i = -\partial_i$)

$$\begin{aligned} F^{0i} &= \partial^0 A^i - \partial^i A^0 \\ &= \left(\frac{\partial \mathbf{A}}{\partial t} + \nabla \phi\right)_i \\ &= -E^i \end{aligned} \tag{2.220}$$

and

$$\begin{aligned} F^{ij} &= \partial^i A^j - \partial^j A^i \\ &= -\varepsilon^{ijk} B^k, \end{aligned} \tag{2.221}$$

where $\varepsilon^{ijk} = \varepsilon_{ijk}$ is the totally antisymmetric Levi–Civita symbol (2.124). Equations (2.219) and (2.220) may be displayed in matrix form, with the rows and columns corresponding to the numbers 0, 1, 2, 3:

$$F^{\mu\nu} = \begin{pmatrix} 0 & -E^1 & -E^2 & -E^3 \\ E^1 & 0 & -B^3 & B^2 \\ E^2 & B^3 & 0 & -B^1 \\ E^3 & -B^2 & B^1 & 0 \end{pmatrix}. \tag{2.222}$$

$F^{\mu\nu}$ is called the *electromagnetic field tensor*. It transforms, under Lorentz transformations, like an antisymmetric second rank tensor.

$$F^{\mu\nu} \to \Lambda^\mu{}_\alpha \Lambda^\nu{}_\beta F^{\alpha\beta}.$$

To summarise so far: if we write the electric and magnetic fields in terms of the tensor $F^{\mu\nu}$, then the statement that $F^{\mu\nu}$ is a 4-dimensional curl means that the first two (homogeneous) Maxwell equations are automatically satisfied.

Now consider the inhomogeneous equations. It is straightforward to verify that they are both contained in the covariant equation

$$\partial_\mu F^{\mu\nu} = j^\nu \tag{2.223}$$

with

$$j^\nu = (\rho, \mathbf{j}). \tag{2.224}$$

For putting $\nu = 0$ gives

$$\partial_1 F^{10} + \partial_2 F^{20} + \partial_3 F^{30} = \rho,$$
$$\operatorname{div} \mathbf{E} = \rho,$$

which is (c), and putting $\nu = 1$ gives

$$\partial_0 F^{01} + \partial_2 F^{21} + \partial_3 F^{31} = j^1,$$
$$-\frac{\partial E^1}{\partial t} + \frac{\partial}{\partial x_2} B^3 - \frac{\partial}{\partial x_3} B^2 = j^1,$$

which is the '1' component of (d).

It is perhaps useful here to insert a remark about gauge transformations. Although (2.217) specifies the electric and magnetic fields in terms of $\mathbf{A}$ and ϕ, it does not do so *uniquely*, for under a *gauge transformation*

$$\mathbf{A} \rightarrow \mathbf{A} - \nabla\chi, \quad \phi \rightarrow \phi + \frac{\partial \chi}{\partial t} \tag{2.225}$$

which has the covariant form

$$A^\mu \rightarrow A^\mu + \partial^\mu \chi, \tag{2.226}$$

where χ is an arbitrary scalar function, $\mathbf{E}$ and $\mathbf{B}$ remain unchanged; equivalently $F^{\mu\nu}$ is unchanged:

$$F^{\mu\nu} \rightarrow F^{\mu\nu} + (\partial^\mu \partial^\nu - \partial^\nu \partial^\mu)\chi = F^{\mu\nu}. \tag{2.227}$$

Substituting (2.219) into (2.223) we see that A^μ satisfies

$$\Box A^\nu - \partial^\nu(\partial_\mu A^\mu) = j^\nu. \tag{2.228}$$

We may now make use of the freedom (2.226) and choose a particular χ so that the transformed A^μ satisfies the *Lorentz gauge condition*:

$$\partial_\mu A^\mu = \frac{\partial \phi}{\partial t} + \nabla \cdot \mathbf{A} = 0. \tag{2.229}$$

In this 'choice of gauge' (2.228) becomes

$$\Box A^\mu = j^\mu, \tag{2.230}$$

which of course stands for the well-known equations

$$\frac{\partial^2 \phi}{\partial t^2} - \nabla^2 \phi = \rho, \quad \frac{\partial^2 \mathbf{A}}{\partial t^2} - \nabla^2 \mathbf{A} = \mathbf{j}, \tag{2.231}$$

whose solutions give the Liénard–Wiechert potentials. *In vacuo*, equation (2.230) becomes

$$\Box A^\mu = 0, \tag{2.232}$$

which means that the electromagnetic field, when its quantum nature is fully exploited, will be seen to correspond to massless particles (which therefore travel at the speed of light; hence relativity, which was where we came in).

We have now cast Maxwell's equations into a manifestly covariant form. The homogeneous equations (*a*) and (*b*) are summarised in equation (2.219). The inhomogeneous equations (*c*) and (*d*) are summarised in (2.223). We shall now show that there is a neat way of combining equations (2.218) and (2.219), so that no explicit reference to the vector potential A^μ need be made. From (2.219) it follows that

$$\partial^\lambda F^{\mu\nu} + \partial^\mu F^{\nu\lambda} + \partial^\nu F^{\lambda\mu} = 0 \tag{2.233}$$

as may trivially be checked. Now we define the *dual tensor* $\tilde{F}^{\mu\nu}$ by

$$\tilde{F}^{\mu\nu} = \tfrac{1}{2}\varepsilon^{\mu\nu\rho\sigma} F_{\rho\sigma} \tag{2.234}$$

where $\varepsilon^{\mu\nu\rho\sigma}$ is the Levi-Civita symbol in four dimensions (with $\varepsilon^{0123} = 1$). Its elements are easily seen to be

$$\tilde{F}^{\mu\nu} = \begin{pmatrix} 0 & -B^1 & -B^2 & -B^3 \\ B^1 & 0 & E^3 & -E^2 \\ B^2 & -E^3 & 0 & E^1 \\ B^3 & E^2 & -E^1 & 0 \end{pmatrix}. \tag{2.235}$$

Because of the antisymmetry of $\varepsilon^{\mu\nu\rho\sigma}$, it follows that the equation

$$\partial_\mu \tilde{F}^{\mu\nu} = 0 \tag{2.236}$$

yields (2.233); alternatively, by looking at (2.235) the reader will soon satisfy himself that it gives Maxwell's equations (*a*) and (*b*). In conclusion, then, Maxwell's equations may be written in the compact form

■
$$\partial_\mu F^{\mu\nu} = j^\nu, \quad \partial_\mu \tilde{F}^{\mu\nu} = 0. \tag{2.237}$$

Massive spin 1 particles obey equations which generalise Maxwell's equations. They are known as the *Proca equations*:

■
$$F^{\mu\nu} = \partial^\mu A^\nu - \partial^\nu A^\mu; \quad \partial_\mu F^{\mu\nu} + m^2 A^\nu = 0. \tag{2.238}$$

Taking the divergence of this we have

$$m^2 \partial_\nu A^\nu = 0, \tag{2.239}$$

and since $m^2 \neq 0$, we find $\partial_\nu A^\nu = 0$; the Lorentz condition, as it were, always holds, and we have lost the freedom of gauge transformations which Maxwell's equations had. In fact, since $F^{\mu\nu}$ is gauge invariant, it follows directly from (2.238) that *the equations for massive spin 1 particles are not gauge invariant.* Substituting (2.239) into (2.238) gives

$$(\Box + m^2)A^\mu = 0 \tag{2.240}$$

as well as

$$\partial_\mu A^\mu = 0. \tag{2.241}$$

Equation (2.240) shows, as expected, that we have particles of mass m. Equation (2.241) is one condition imposed on the four components of A^μ, so there are only three independent components. This is indeed appropriate for a massive spin 1 particle.

Let us conclude this section with a general remark on wave equations, which, for convenience, are reproduced together.

$$\text{(Klein–Gordon)} \quad (\Box + m^2)\phi = 0, \tag{2.16}$$

$$\text{(Dirac)} \quad (\mathrm{i}\gamma^\mu\partial_\mu - m)\psi = 0, \tag{2.96}$$

$$(\Box + m^2)\psi = 0, \tag{2.98}$$

$$\text{(Maxwell)} \quad \partial_\mu F^{\mu\nu} = 0, \tag{2.223}$$

$$\Box A^\mu = 0, \tag{2.232}$$

$$\text{(Proca)} \quad \partial_\mu F^{\mu\nu} + m^2 A^\nu = 0, \tag{2.238}$$

$$(\Box + m^2)A^\mu = 0. \tag{2.240}$$

In the case of spin $\frac{1}{2}$ and spin 1 fields, there are *first order* wave equations, but there is no first order wave equation solely for scalar fields‡. On the other hand, every component of the spin $\frac{1}{2}$ and spin 1 field satisfies a Klein–Gordon equation (with $m = 0$ for the photon), which is, after all, only a requirement of relativity ($E^2 - p^2 = m^2$) and quantum theory ($E \to \mathrm{i}(\partial/\partial t)$, $\mathbf{p} \to -\mathrm{i}\nabla$). Thus the Dirac, Maxwell and Proca equations are of a different type from the Klein–Gordon equation. We saw above that the Dirac equation could be *derived* by considering the transformation of spinors under the Lorentz group, and it may be shown (see Weinberg 1964; Johnson *et al.* 1965), though we did not do so here, that the Maxwell and Proca equations can be obtained in the same way. So these equations, for fields with non-zero spin, are simply a relation between the spin components; in Weinberg's words, they are a confession that we have too many spin components. The Klein–Gordon equation is not of this nature, since we only have one component.

One final observation: in our derivation of the Dirac equation we relied crucially on the assumption that the components of the spin $\frac{1}{2}$ field form a *linear vector space*, suitable as a basis for constructing a representation of the

‡ The Duffin–Kemmer equations (Duffin (1938), Kemmer (1939)) are first order equations which describe spin 0 *and* spin 1 fields. I thank Dr David Owen for bringing this to my attention.

Lorentz group. This assumption, while it may look mathematically innocuous, is physically highly non-trivial, for it corresponds to a *principle of superposition* and therefore to wave–particle duality and the quantum theory. In other words, the fields we have found are *already* quantum fields. The statement, often found in the literature, that we must now subject these fields to 'second quantisation' is, in this light, misleading. It is better to say that we shall explore further the implication that these fields are quantum fields, by (for example) writing down the commutation relations which must hold between them; this we do in Chapter 4.

2.9 Maxwell's equations and differential geometry

Maxwell's equations (2.237) relate antisymmetric tensors and vectors, but, as indicated by the indices, do so component by component. Compared with equations such as $\nabla \cdot \mathbf{B} = 0$, this may be considered a retrograde step; $\nabla \cdot \mathbf{B}$ is a more economical notation than $\nabla^i B_i$, to which it is equivalent. So, we are led to ask, is there a way of writing Maxwell's equations in terms of the tensor F and current j, without making explicit reference to the components? Because of the development of differential geometry, there is indeed a way to do this, and Maxwell's equations take on the elegant form $\mathrm{d}F = 0$, $\mathrm{d}^*F = J$; the antisymmetry of the field tensor F is automatically included! In this section we shall explain this notation. A common reaction of physicists to this type of mathematical development is one of impatience. After all, they point out, the equation $\mathrm{d}^*F = J$ has to be translated into the form $\partial_\mu F^{\mu\nu} = j^\nu$ before it can be dealt with in a particular co-ordinate system. This may be true, but in the opinion of a growing number of physicists the development of notation actually corresponds to a deepening in our understanding. The point is that, like many developments in contemporary mathematics, this one has evolved by introducing new concepts, and making distinctions that have not in the past been made. In the case of Maxwell's equations, light is shed on their geometrical interpretation. As will be discussed in the next chapter, electromagnetism is a gauge theory, having a $U(1)$ invariance group. Gauge theories with non-Abelian invariance groups ($SU(2) \times U(1)$, $SU(3)$) play a central role in contemporary particle physics, and the geometrical interpretation of them, along the lines mentioned, may indeed play no minor part in the understanding of their ultimate significance.

To begin, consider the meaning of ordinary line and surface integrals:

$$\left.\begin{aligned} I_1 &= \int_C F_x\,\mathrm{d}x + F_y\,\mathrm{d}y + F_z\,\mathrm{d}z = \int_C \mathbf{F}\cdot\mathrm{d}\mathbf{r}, \\ I_2 &= \int_S (G_x\,\mathrm{d}y \wedge \mathrm{d}z + G_y\,\mathrm{d}z \wedge \mathrm{d}x + G_z\,\mathrm{d}x \wedge \mathrm{d}y) \\ &= \int_S \mathbf{G}\cdot\mathrm{d}\mathbf{S}. \end{aligned}\right\} \tag{2.242}$$

I_1 and I_2 are *numbers*. I_1 is the integral of something over a line C, and I_2 the integral of something else over a surface S. In some sense, then, the 'something' is *dual* to the 'line', since when they are 'combined' (by the integral) the result is a pure number. Similarly, for I_2, the 'something else' is dual to the 'surface'. We systematise this by coining new words; the line and the surface are called 'chains', and the objects integrated over the chains are called 'differential forms' or simply 'forms'. Thus forms are dual to chains.

We shall call a line a 1-chain, since it has one dimension, a surface a 2-chain, etc., and denote the generic chain C_n, with n dimensions. So we have

$$\left.\begin{array}{ll} C_0 & \text{0-chain = point,} \\ C_1 & \text{1-chain = line,} \\ C_2 & \text{2-chain = area,} \\ C_3 & \text{3-chain = volume,} \\ C_n & n\text{-chain.} \end{array}\right\} \tag{2.243}$$

Now the *boundary* of an n-chain is an $(n-1)$-chain. The boundary of an area is a line, and that of a line two points. We define a *boundary operator* ∂ which maps C_n into C_{n-1}

$$C_n \xrightarrow{\partial} C_{n-1} \quad \text{or} \quad \partial C_n = C_{n-1}. \tag{2.244}$$

Some chains have no boundaries: the surface of a sphere is a 2-chain (area) with no boundary, and a closed line like a circle is a 1-chain with no boundary. Such *closed chains* are called *cycles* and denoted Z_n. Since they have no boundary, it is clear that

$$\partial Z_n = 0. \tag{2.245}$$

(Z_n is actually the kernel of the mapping (2.244).) On the other hand, there are some chains which *themselves are boundaries* of higher dimensional chains, and these are denoted B_n:

$$B_n = \partial C_{n+1}. \tag{2.246}$$

(B_{n-1} is actually the image of the map (2.244).) For example, a closed surface B_2 is the boundary of a volume, and a closed line B_1 is the boundary of an area. It is clear that the B_ns themselves have no boundary (are closed):

$$\partial B_n = 0. \tag{2.247}$$

Combining these last two equations gives

■ $$\partial^2 = 0. \tag{2.248}$$

In words, 'the boundary of a boundary is zero', or a chain which is a boundary is also closed. An interesting consideration is whether the converse holds: is a closed chain necessarily the boundary of another chain? In Euclidean spaces the answer is yes, so that $Z_n = B_n$. In general, however, there are closed chains which are not boundaries, so $Z_n \supset B_n$. For example, on a torus, a closed curve like C_1 in Fig. 2.3 is not the boundary of any part of the surface of the torus, whereas C_2 obviously is. Similarly, on the space of a circle S^1, the circle itself is not the boundary of any part of the space; it may not be thought of as the boundary of the area enclosed, because the area, which is 2-dimensional, is not part of S^1, which is 1-dimensional. This completes what we need to say about chains.

We now turn to *forms*. As mentioned above, the integral of a form over a chain is a number. We write

$$\int_{C_n} \omega_n \equiv \int_{C_n} f_{i_1 \dots i_n} \, \mathrm{d}x_{i_1} \wedge \mathrm{d}x_{i_2} \wedge \cdots \wedge \mathrm{d}x_{i_n} = \text{number}. \tag{2.249}$$

The wedge product, $\wedge$, appears above because the *orientation* of a curve or surface, etc., is important. The existence of integrals implies a duality between forms and chains, which we shall see manifested. A 1-form ω_1 is something to be integrated over a line (1-chain), so in 3-dimensional space is of the form $A\,\mathrm{d}x + B\,\mathrm{d}y + C\,\mathrm{d}z$. Other forms follow the same pattern, so we have (in 3-dimensional space)

$$\left.\begin{array}{lll} \omega_0 & \text{0-form} & \text{function,} \\ \omega_1 & \text{1-form} & A\,\mathrm{d}x + B\,\mathrm{d}y + C\,\mathrm{d}z, \\ \omega_2 & \text{2-form} & f\,\mathrm{d}x \wedge \mathrm{d}y + g\,\mathrm{d}y \wedge \mathrm{d}z + h\,\mathrm{d}z \wedge \mathrm{d}x, \\ \omega_3 & \text{3-form} & F\,\mathrm{d}x \wedge \mathrm{d}y \wedge \mathrm{d}z, \end{array}\right\} \tag{2.250}$$

where

$$\mathrm{d}x \wedge \mathrm{d}y = -\mathrm{d}y \wedge \mathrm{d}x, \quad \mathrm{d}x \wedge \mathrm{d}x = 0, \text{etc.} \tag{2.251}$$

Because of (2.251), it is clear that in an n-dimensional space there are n-forms, but not $(n+1)$- or higher degree forms. Clearly, if we differentiate an n-form we will get something like an $(n+1)$-form: we shall get *precisely* an $(n+1)$-form if we build into the differentiation the antisymmetrisation above. Hence

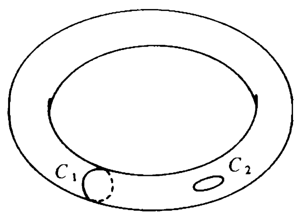

Fig. 2.3. Closed curves on a torus; C_1 is not the boundary of a surface belonging to the torus, but C_2 is.

we define a so-called *exterior derivative operator* d:

$$\mathrm{d}\omega_n = \omega_{n+1}. \tag{2.252}$$

Its action on a 1-form (in 3-space) is

$$\begin{aligned}\mathrm{d}(A\,\mathrm{d}x + B\,\mathrm{d}y + C\,\mathrm{d}z) &= \frac{\partial A}{\partial y}\mathrm{d}y \wedge \mathrm{d}x + \frac{\partial A}{\partial z}\mathrm{d}z \wedge \mathrm{d}x + \frac{\partial B}{\partial x}\mathrm{d}x \wedge \mathrm{d}y \\ &\quad + \frac{\partial B}{\partial z}\mathrm{d}z \wedge \mathrm{d}y + \frac{\partial C}{\partial x}\mathrm{d}x \wedge \mathrm{d}z + \frac{\partial C}{\partial y}\mathrm{d}y \wedge \mathrm{d}z \\ &= \left(\frac{\partial B}{\partial x} - \frac{\partial A}{\partial y}\right)\mathrm{d}x \wedge \mathrm{d}y + \left(\frac{\partial C}{\partial y} - \frac{\partial B}{\partial z}\right)\mathrm{d}y \wedge \mathrm{d}z \\ &\quad + \left(\frac{\partial A}{\partial z} - \frac{\partial C}{\partial x}\right)\mathrm{d}z \wedge \mathrm{d}x. \end{aligned} \tag{2.253}$$

By studying this example, readers should learn how d operates on any form, and so convince themselves that the 2-form

$$\omega_2 = f\,\mathrm{d}x \wedge \mathrm{d}y + g\,\mathrm{d}y \wedge \mathrm{d}z + h\,\mathrm{d}z \wedge \mathrm{d}x$$

has exterior derivative

$$\mathrm{d}\omega_2 = \left(\frac{\partial f}{\partial z} + \frac{\partial g}{\partial x} + \frac{\partial h}{\partial y}\right)\mathrm{d}x \wedge \mathrm{d}y \wedge \mathrm{d}z. \tag{2.254}$$

In the first example, (2.253), the quantities

$$\frac{\partial C}{\partial y} - \frac{\partial B}{\partial z}, \quad \frac{\partial A}{\partial z} - \frac{\partial C}{\partial x}, \quad \frac{\partial B}{\partial x} - \frac{\partial A}{\partial y}$$

are the three components of curl $\mathbf{F}$ where $\mathbf{F} = \mathbf{i}A + \mathbf{j}B + \mathbf{k}C$. In the second example, putting $(g, h, f) = \mathbf{W}$, a vector, the quantity $\partial g/\partial x + \partial h/\partial y + \partial f/\partial z$ is div $\mathbf{W}$.

Now note that, in view of (2.254), if we calculate the exterior derivative of (2.253), we get identically zero; in other words

$$\mathrm{d}[\mathrm{d}(A\,\mathrm{d}x + B\,\mathrm{d}y + C\,\mathrm{d}z)] = \mathrm{d}^2(A\,\mathrm{d}x + B\,\mathrm{d}y + C\,\mathrm{d}z) = 0$$

or, in general,

■ $$\mathrm{d}^2 = 0. \tag{2.255}$$

In component language, this reads

$$\text{div}\,\text{curl} = 0;$$

d is sometimes called the *coboundary operator*, to emphasise the fact that $\mathrm{d}^2 = 0$ is the dual of $\partial^2 = 0$, equation (2.248); $\mathrm{d}^2 = 0$ is known as the *Poincaré lemma*.

An n-form ω_n is called *closed* if $\mathrm{d}\omega_n = 0$.

An n-form ω_n is called *exact* if it is the derivative of an $(n-1)$-form,

$\omega_n = \mathrm{d}\omega_{n-1}$.

The Poincaré lemma tells us that all exact forms are closed, since $\mathrm{d}(\mathrm{d}\omega_{n-1}) = \mathrm{d}^2\omega_{n-1} = 0$, but it is not in general true that all closed forms are exact, though this is true in Euclidean spaces. Again, this is because of the duality between chains and forms: in Euclidean spaces, all closed chains are boundaries.

Well-known results follow from *Stokes' formula*, which states that if ω is a p-form, and c a $(p+1)$-chain, then

$$\int_{\partial c} \omega = \int_c \mathrm{d}\omega. \tag{2.256}$$

As an example, put $p = 2$. Let ω then be the 2-form (in 3-dimensional space),

$$\omega_2 = A_x \mathrm{d}y \wedge \mathrm{d}z + A_y \mathrm{d}z \wedge \mathrm{d}x + A_z \mathrm{d}x \wedge \mathrm{d}y,$$

and let C_3 be a domain V, with boundary ∂V. Then Stokes' formula gives

$$\int_{\partial V} A_x \mathrm{d}y \wedge \mathrm{d}z + A_y \mathrm{d}z \wedge \mathrm{d}x + A_z \mathrm{d}x \wedge \mathrm{d}y$$

$$= \int_V \left(\frac{\partial A_x}{\partial x} + \frac{\partial A_y}{\partial y} + \frac{\partial A_z}{\partial z}\right) \mathrm{d}x \wedge \mathrm{d}y \wedge \mathrm{d}z$$

where we have used (2.254); and this is

$$\oint_{\partial V} \mathbf{A} \cdot \mathrm{d}\mathbf{S} = \int_V \operatorname{div} \mathbf{A}\, \mathrm{d}V \tag{2.257}$$

which is the divergence theorem (or Gauss' theorem). As an exercise, the reader may show that with $p = 1$ we recover Stokes' theorem

$$\oint_{\partial A} \mathbf{A} \cdot \mathrm{d}\mathbf{l} = \int_A \operatorname{curl} \mathbf{A} \cdot \mathrm{d}\mathbf{S}. \tag{2.258}$$

We have now seen the relation between the exterior derivative operator d and the usual differential operators of grad, div and curl. However, as seen in the derivation of (2.257) above, if the components of $\mathbf{A}$ are the coefficients of a 2-form, $\operatorname{div} \mathbf{A}$ is the coefficient of the 3-form obtained by operating d on the 2-form. In the ordinary language of vectors, the operator $\boldsymbol{\nabla}$ changes a scalar into a vector, and a vector either into a scalar (div) or into an axial vector (curl). There is, however, an operator which does not change the vectorial character; this is the *Laplacian* ∇^2 (the d'Alembertian $\Box$ in four dimensions). $\nabla^2\phi$ is a scalar, $\nabla^2\mathbf{A}$ is a vector, etc. How is this represented in the language of forms? d changes a p-form into a $(p+1)$-form, so we need to combine it with another operator (δ) which changes a p-form into a $(p-1)$-form. We shall now show how this is done.

For definiteness, let us work in 3-dimensional space. The space of 1-forms is obviously 3-dimensional, with basis $\mathrm{d}x$, $\mathrm{d}y$ and $\mathrm{d}z$. The space of 2-forms is also 3-dimensional; in fact the bases may be written as follows:

$$\left.\begin{array}{ll} \text{Basis of } \omega_p \text{ for } n = 3 & \\ \omega_0\text{:} & 1, \\ \omega_1\text{:} & \mathrm{d}x, \mathrm{d}y, \mathrm{d}z, \\ \omega_2\text{:} & \mathrm{d}x \wedge \mathrm{d}y, \mathrm{d}y \wedge \mathrm{d}z, \mathrm{d}z \wedge \mathrm{d}x, \\ \omega_3; & \mathrm{d}x \wedge \mathrm{d}y \wedge \mathrm{d}z. \end{array}\right\} \tag{2.259}$$

Obviously, there are no 4-forms in 3-dimensional space. It is clear that the dimensionality of the space of p-forms is the same as that of the space of $(n - p)$-forms, so we may define an operator which converts one into the other. It is known as the *Hodge* (Hodge star) operator* or *duality transformation*. In a Euclidean (flat) space it is defined by

$$*(\mathrm{d}x^{i_1} \wedge \mathrm{d}x^{i_2} \wedge \cdots \wedge \mathrm{d}x^{i_p}) = \frac{1}{(n-p)!}\varepsilon_{i_1 i_2 \ldots i_p i_{p+1} \ldots i_n} \mathrm{d}x^{i_{p+1}} \wedge \mathrm{d}x^{i_{p+2}} \wedge \cdots \wedge \mathrm{d}x^{i_n}. \tag{2.260}$$

Thus in the case $n = 3$ we have the following basis for $*\omega$:

$$\left.\begin{array}{ll} \text{Bases of } *\omega_p \text{ for } n = 3 & \\ *\omega_0\text{:} & \mathrm{d}x \wedge \mathrm{d}y \wedge \mathrm{d}z, \\ *\omega_1\text{:} & \mathrm{d}y \wedge \mathrm{d}z, \mathrm{d}z \wedge \mathrm{d}x, \mathrm{d}x \wedge \mathrm{d}y, \\ *\omega_2\text{:} & \mathrm{d}z, \mathrm{d}x, \mathrm{d}y, \\ *\omega_3\text{:} & 1. \end{array}\right\} \tag{2.261}$$

Repeating the operator * on a p-form ω_p gives

$$**\omega_p = (-1)^{p(n-p)}\omega_p. \tag{2.262}$$

(To convince oneself that sign changes occur, simply consider the case of $p = 1$, $n = 2$.) It is clear that whereas $\mathrm{d}\omega_p \sim \omega_{p+1}$, $\mathrm{d}(*\omega_p) \sim *\omega_{p-1}$, so we define the operator δ:

$$\delta = (-1)^{np+n+1} * \mathrm{d}*, \tag{2.263}$$

where p is the degree of the form ω_p on which δ is applied and n is the dimension of the space; δ is called the *adjoint exterior derivative operator*, and it should be clear that $\delta\omega$ is of degree $(p - 1)$.

As an example, let us show that δ changes the 1-form $\mathbf{v} \cdot \mathrm{d}\mathbf{s}$ into a 0-form:

$$\begin{aligned}\delta(\mathbf{v}\cdot d\mathbf{s}) &= \delta(v_x\,dx + v_y\,dy + v_z\,dz)\\ &= -{*}\,d{*}(v_x\,dx + v_y\,dy + v_z\,dz)\\ &= -{*}\,d(v_x\,dy\wedge dz + v_y\,dz\wedge dx + v_z\,dx\wedge dy)\\ &= -\overset{*}{\left(\frac{\partial v_x}{\partial x} + \frac{\partial v_y}{\partial y} + \frac{\partial v_z}{\partial z}\right)}dx\wedge dy\wedge dz\\ &= -\operatorname{div}\mathbf{v}.\end{aligned} \tag{2.264}$$

It is easy to see that, like d, δ has a square of zero:

$$\begin{aligned}\delta\delta &= (-1)^{np+n+1}(-1)^{n(p-1)+n+1}{*}d{*}{*}d{*}\\ &= (-1)^{n}{*}\,d^2{*},\end{aligned}$$

and in view of (2.255) we have

$$\delta^2 = 0. \tag{2.265}$$

Finally, the *Laplacian* Δ changes p-forms into p-forms and is defined by

$$\Delta = (d + \delta)^2 = d\delta + \delta d. \tag{2.266}$$

After this long preamble, it is now an easy matter to show how Maxwell's equations may be put into a geometric (or 'intrinsic') form. The space we work in is of course 4-dimensional Minkowski space–time. Lowering the indices on the field tensor $F^{\mu\nu}$ in (2.222) gives ($E_x \equiv E_1$, etc.)

$$\left.\begin{aligned}&F_{01} = E_x, \quad F_{02} = E_y, \quad F_{03} = E_z,\\ &F_{12} = -B_z, \quad F_{31} = -B_y, \quad F_{23} = -B_x.\end{aligned}\right\} \tag{2.267}$$

We then define the *Faraday 2-form F* by

$$\begin{aligned}F &= -\tfrac{1}{2}F_{\mu\nu}\,dx^\mu\wedge dx^\nu\\ &= (E_x\,dx + E_y\,dy + E_z\,dz)\wedge dt\\ &\quad + B_z\,dx\wedge dy + B_x\,dy\wedge dz + B_y\,dz\wedge dx.\end{aligned} \tag{2.268}$$

The dual form (also a 2-form) $*F$ is (see (2.260))

$$\begin{aligned}{*}F &= -\tfrac{1}{2}F_{\mu\nu}{*}(dx^\mu\wedge dx^\nu)\\ &= -E_x\,dy\wedge dz - E_y\,dz\wedge dx - E_z\,dx\wedge dy\\ &\quad + (B_x\,dx + B_y\,dy + B_z\,dz)\wedge dt.\end{aligned} \tag{2.269}$$

It is seen that the components of $*F$ in the basis $dx^\mu\wedge dx^\nu$ are $-\tfrac{1}{2}\tilde{F}_{\mu\nu}$ as shown in (2.235)

$${*}F = -\tfrac{1}{2}\tilde{F}_{\mu\nu}\,dx^\mu\wedge dx^\nu. \tag{2.270}$$

Finally, we define the current density 3-form J:

$$J = (j_x \mathrm{d}y \wedge \mathrm{d}z + j_y \mathrm{d}z \wedge \mathrm{d}x + j_z \mathrm{d}x \wedge \mathrm{d}y) \wedge \mathrm{d}t - \rho \mathrm{d}x \wedge \mathrm{d}y \wedge \mathrm{d}z. \tag{2.271}$$

A simple exercise shows that Maxwell's equations are

■ $$\mathrm{d}F = 0, \quad \mathrm{d}^*F = J. \tag{2.272}$$

This may be seen by putting $F = -\frac{1}{2}F_{\mu\nu}\mathrm{d}x^\mu \wedge \mathrm{d}x^\nu$ and then observing that $\mathrm{d}F = 0$ implies equation (2.233), which is equivalent to the two homogeneous Maxwell's equations. Alternatively, using equation (2.268), we find explicitly

$$\begin{aligned}\mathrm{d}F &= \frac{\partial E_x}{\partial y}\mathrm{d}y \wedge \mathrm{d}x \wedge \mathrm{d}t + \frac{\partial E_x}{\partial z}\mathrm{d}z \wedge \mathrm{d}x \wedge \mathrm{d}t + \frac{\partial E_y}{\partial x}\mathrm{d}x \wedge \mathrm{d}y \wedge \mathrm{d}t + \cdots \\ &\quad + \frac{\partial B_z}{\partial z}\mathrm{d}z \wedge \mathrm{d}x \wedge \mathrm{d}y + \frac{\partial B_z}{\partial t}\mathrm{d}t \wedge \mathrm{d}x \wedge \mathrm{d}y + \cdots \\ &= \left(\frac{\partial E_y}{\partial x} - \frac{\partial E_x}{\partial y} + \frac{\partial B_z}{\partial t}\right)\mathrm{d}x \wedge \mathrm{d}y \wedge \mathrm{d}t + \cdots + (\operatorname{div}\mathbf{B})\mathrm{d}x \wedge \mathrm{d}y \wedge \mathrm{d}z.\end{aligned} \tag{2.273}$$

$\mathrm{d}F = 0$ then implies the two equations

$$\operatorname{curl}\mathbf{E} + \frac{\partial \mathbf{B}}{\partial t} = 0; \quad \operatorname{div}\mathbf{B} = 0,$$

which are the homogeneous Maxwell equations. Similar manipulations show that $\mathrm{d}^*F = J$ yields the inhomogeneous equations.

In Euclidean space (which, for our purposes, may be extended to Minkowski space), the converse of the Poincaré lemma holds: all closed forms are exact, so if $\mathrm{d}F = 0$, then there is a 1-form A such that

$$F = \mathrm{d}A. \tag{2.274}$$

The 1-form A will be, in a co-ordinate basis,

$$A = A_\mu \mathrm{d}x^\mu \tag{2.275}$$

and it follows immediately that equation (2.274) is equivalent to $F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu$, as in (2.219). The 1-form A has a further geometric significance; it is the *connection form* which is used to define covariant derivatives – this is treated in the next chapter. If, however, we start from the point of view that the presence of electromagnetism gives a covariant derivative, and therefore a connection 1-form A, then $F = \mathrm{d}A$ is known as the 'curvature' 2-form, and the identity $\mathrm{d}F = 0$ is known as the *Bianchi identity*.

Summary‡

[2]Scalar particles are described by the Klein–Gordon equation, but the difficulties of interpretation are that the probability density is not positive definite, and negative energy states exist. It is concluded that the Klein–Gordon equation is not suitable as a single-particle equation. [3]The Dirac equation describes spin $\frac{1}{2}$ particles, and may be derived from the group $SL(2, C)$ which is shown to embody Lorentz transformations, extended by parity. The Dirac equation is found to have positive probability density, but has negative, as well as positive energy states. Massless spin $\frac{1}{2}$ particles are shown to obey the Weyl equation. [4]The hypothesis that the negative energy states are filled leads to the successful prediction of antiparticles. [5]It is shown how to construct spinors which satisfy the Dirac equation, and the transformation properties of bilinear forms $\bar{\psi}O\psi$ are studied, as well as various algebraic identities involving spinors and γ-matrices. [6]The Dirac equation gives the correct gyromagnetic ratio for the electron. [7]It is observed that the 'Casimir' invariant operator for spin, $s(s+1)$, is not constructed from generators of the Lorentz group, but from those of the inhomogeneous Lorentz (= Poincaré) group, whose other Casimir invariant is $(\text{mass})^2$. Particles with $m^2 > 0$ have a 'little group' $SU(2)$, which is interpreted as spin, whereas those with $m^2 = 0$ and $m^2 < 0$ have non-compact little groups, so their spin is not described by a rotation group. [8]Maxwell's equations are exhibited in manifestly covariant form, and the Proca equations, for massive spin 1 particles, are written down. [9]Chains and differential forms are explained, and Maxwell's equations exhibited in terms of differential forms.

Guide to further reading

Good accounts of spinors and the relation between $O(3)$ and $SU(2)$ are to be found in van der Waerden (1974), Landau & Lifshitz (1977), Misner, Thorne & Wheeler (1973), Sexl & Urbantke (1976), and Normand (1980). The topological distinction between $SU(2)$ and $O(3)$ is explained in Sexl & Urbantke (1976, §7.6), D. Speiser, in Gürsey (1964), and F. Gürsey, in DeWitt & DeWitt (1964). An excellent account (in French) of the Lorentz group and the connection with $SL(2, C)$ appears in A.S. Wightman, in DeWitt & Omnès (1960). See also R. Omnès & M. Froissart in DeWitt & Jacob (1965), and Berestetskii, Lifshitz & Pitaevskii (1971). For full accounts of the Dirac equation, see, for example, Bjorken & Drell (1964), Itzykson & Zuber (1980). Wigner's analysis of the Poincaré group appears in Wigner (1939); see also International Atomic Agency (1963). The relativistic treatment of spin is considered in DeWitt & Omnès (1960) and DeWitt & Jacob (1965). An early, and rather brief, account is given by Pauli (1965). The connection between the Poincaré group and wave equations goes back to Bargmann & Wigner (1948),

‡ Superscripts refer to section numbers.

and is explored further by Weinberg (1964) and Johnson *et al.* (1965). Introductions to differential forms, suitable for the physicist, are Flanders (1963), Choquet-Bruhat, Morette-DeWitt & Bleik-Dillard (1977), Choquet-Bruhat (1968), Schutz (1980), Eguchi, Gilkey & Hanson (1980), von Westenholz (1978). Less complete, but very readable, accounts appear in Misner, Thorne & Wheeler (1973, chs. 3, 4, 8, 9) and Misner & Wheeler (1957).