Contents

CHAPTER 1

Introduction: Electromagnetic Theory without Myths

The full development of the theory of electromagnetism in the nineteenth century stands as one of the greatest achievements in the history of physics. The theory of electromagnetism as formulated by Maxwell is a mathematically consistent theory that provides an excellent description of an extremely wide range of physical phenomena. Of course, Maxwell's electromagnetism is a classical theory that cannot properly describe phenomena in which the quantum properties of the electromagnetic field play an important role, but the quantum field theory of the electromagnetic field is built upon the foundation of the classical theory.

Maxwell's equations relate the electric and magnetic fields, $\boldsymbol{E}$ and $\boldsymbol{B}$, to each other and to the charge density, ρ, and the current density, $\boldsymbol{J}$. That is, $\rho(\boldsymbol{x})$ is the electric charge per unit volume at $\boldsymbol{x}$, and for any unit vector $\hat{\boldsymbol{n}}$ at $\boldsymbol{x}$, $\boldsymbol{J}(\boldsymbol{x}) \cdot \hat{\boldsymbol{n}}$ gives the flux of charge per unit area through an area element perpendicular to $\hat{\boldsymbol{n}}$. Maxwell's equations in SI units[1] are as follows:

$$\nabla \cdot \boldsymbol{E} = \frac{\rho}{\epsilon_0}, \tag{1.1}$$

$$\nabla \times \boldsymbol{B} - \frac{1}{c^2}\frac{\partial \boldsymbol{E}}{\partial t} = \mu_0 \boldsymbol{J}, \tag{1.2}$$

$$\nabla \cdot \boldsymbol{B} = 0, \tag{1.3}$$

$$\nabla \times \boldsymbol{E} + \frac{\partial \boldsymbol{B}}{\partial t} = 0. \tag{1.4}$$

Here ρ and $\boldsymbol{J}$ must satisfy the charge-current conservation equation

$$\frac{\partial \rho}{\partial t} + \nabla \cdot \boldsymbol{J} = 0, \tag{1.5}$$

[1] As discussed in the preface, SI units have the unfortunate feature that the three constants $\epsilon_0 \approx 8.85 \times 10^{-12}$ F/m (the vacuum permittivity), $\mu_0 \approx 1.26 \times 10^{-6}$ H/m (the vacuum permeability), and $c \approx 3.00 \times 10^8$ m/s (the speed of light) appearing in the equations of electromagnetism are not independent but satisfy $\epsilon_0 \mu_0 c^2 = 1$. Consequently, the appearance of formulas in SI units can be changed in nontrivial-looking ways by using this identity.

since otherwise, no solutions to eqs. (1.1) and (1.2) exist. Apart from this restriction, $\rho(t, \boldsymbol{x})$ and $\boldsymbol{J}(t, \boldsymbol{x})$ can be specified arbitrarily.

Maxwell's equations have survived without modification for more than one and a half centuries (i.e., the equations I have written above are equivalent to those given by Maxwell). However, our understanding of electromagnetism at a fundamental level has progressed greatly since the time of Maxwell. Despite this fact, many outdated ways of thinking about electromagnetism remain prevalent. This is strongly reinforced by the quasi-historical way in which electromagnetism is usually taught, even at the graduate level: One normally starts with Coulomb's law in electrostatics, with point charges taken as "fundamental." This motivates the introduction of an electric field $\boldsymbol{E}$ satisfying eq. (1.1) as well as $\nabla \times \boldsymbol{E} = 0$ (i.e., eq. (1.4) with $\partial \boldsymbol{B}/\partial t = 0$). Energy is assigned to the electrostatic interaction via an analysis of the mechanical work done when moving point charges quasi-statically. Similarly, in magnetostatics, one normally starts with the Biot-Savart law for the force between current elements. This motivates the introduction of a magnetic field $\boldsymbol{B}$ satisfying eq. (1.2) with $\partial \boldsymbol{E}/\partial t = 0$ as well as eq. (1.3). The dynamical terms in $\boldsymbol{E}$ and $\boldsymbol{B}$ are then introduced to get the full Maxwell equations in the form given above. A scalar potential, ϕ, and vector potential, $\boldsymbol{A}$, satisfying

$$\boldsymbol{E} = -\nabla \phi - \frac{\partial \boldsymbol{A}}{\partial t}, \tag{1.6}$$

$$\boldsymbol{B} = \nabla \times \boldsymbol{A}, \tag{1.7}$$

also are introduced at some stage as a convenient way of solving the Maxwell equations (1.3) and (1.4).

This manner in which the theory of electromagnetism is presented encourages a number of unhealthy ways of thinking about the theory, which I have referred to as "myths" in the title of this chapter. The most pernicious of these myths are the following: (i) The field strengths, $\boldsymbol{E}$ and $\boldsymbol{B}$, are taken to be fundamental, whereas the potentials, ϕ and $\boldsymbol{A}$, are viewed as quantities that are introduced merely as a convenience. (ii) The energy, momentum, and stress properties of the electromagnetic field are considered to be properties derived or guessed from interactions with charged matter and conservation laws rather than properties of the electromagnetic field having a fundamental status comparable to that of Maxwell's equations themselves. For example, in this regard, it is often stated that the momentum density of the electromagnetic field is ambiguous up to the addition of the curl of a vector field, since it is not uniquely determined by energy conservation. (iii) Electromagnetic fields are considered to be *produced* by charged matter (as opposed to the fact that electromagnetic fields *interact* with charged matter). (iv) Point charges are taken to be a fundamental description of charged matter, despite blatant mathematical inconsistencies associated with them, such as infinite self-energy.

In the following sections, I do my best to debunk these myths. There is, of course, a serious pedagogical problem with my doing this, since to fully follow all of the discussion in this chapter, readers will need to have a considerable knowledge of electromagnetic theory. While it would be reasonable to hope that readers will have a considerable knowledge of electromagnetic theory by the time they have gotten to the end of this book, it is not reasonable to assume such knowledge at the beginning. Indeed, many of the points discussed here will be properly explained in detail only in the last two chapters of this book. It is not necessary that the reader follow all details of the discussion in this chapter—since everything said in this chapter will be elucidated in the remainder of the book—but it is important that the reader gain a sense of the viewpoint on classical electromagnetism that I take. I feel that it is highly preferable to begin this book in this way rather than to get started on the wrong foot by taking the usual quasi-historical path. In the succeeding chapters, I develop the subject in a largely conventional way—starting with electrostatics and magnetostatics before moving to full electrodynamics—but the viewpoint taken will always be fully compatible with the discussion in this chapter.

Before discussing the above myths, I wish to make some comments about the relationship of classical electrodynamics to special relativity. Maxwell's equations are not compatible with the spacetime structure of pre-relativity physics unless one has a "preferred rest frame." This, by itself, was not troubling in the nineteenth century, since it was believed that there was a mechanical medium—the "luminiferous aether"—through which electromagnetic fields propagated. Such an aether would naturally provide a preferred rest frame. However, the lack of evidence in the Michelson-Morley experiment for a preferred rest frame as well as other problems with the theory of the aether resulted in severe difficulties that were ultimately resolved by the theory of special relativity. In the theory of special relativity, the Newtonian time function t (defining an "absolute notion" of simultaneity) and the metric of space are replaced by a single quantity: the metric of spacetime. Classical electrodynamics is fully compatible with the spacetime structure of special relativity, without the need for an aether.

The structure of classical electrodynamics is considerably simpler when formulated within the framework of special relativity. I wait until chapter 8 to give a proper discussion of the formulation of electromagnetism within special relativity, but I wish to make a few remarks here, so that the reader can get some flavor of what this formulation looks like without waiting until near the end of this book. In special relativity, the scalar potential, ϕ, and vector potential, $\boldsymbol{A}$, are seen to be the time and space components of a single "4-(dual-)vector potential"

$$A_\mu = (-\phi/c, \boldsymbol{A}). \tag{1.8}$$

The electric and magnetic fields are seen to arise from a single field strength tensor

$$F_{\mu\nu} = \frac{\partial A_\nu}{\partial x^\mu} - \frac{\partial A_\mu}{\partial x^\nu}, \tag{1.9}$$

with $x^\mu = (x^0 = ct, x^1, x^2, x^3)$. Since $F_{\mu\nu} = -F_{\nu\mu}$, it has 6 independent components. For an observer at rest in these coordinates, the electric field corresponds to the 3 time-space components of $F_{\mu\nu}$

$$E_i = cF_{i0}, \qquad i = 1, 2, 3, \tag{1.10}$$

whereas the magnetic field corresponds to the 3 independent space-space components of $F_{\mu\nu}$,

$$B_i = F_{jk}, \qquad i = 1, 2, 3, \tag{1.11}$$

where (i, j, k) is a cyclic permutation of $(1, 2, 3)$. In particular, since observers who move relative to each other define different "time directions" in spacetime, what would be claimed by one observer to be a "pure electric field" will be seen by another observer to be a combination of electric and magnetic fields. The "invariant description" of the field strengths is given by $F_{\mu\nu}$. Maxwell's equations can then be written in terms of $F_{\mu\nu}$, the spacetime metric, and the charge-current 4-vector:

$$J^{\mu} = (c\rho, \boldsymbol{J}). \tag{1.12}$$

Although the special relativistic formulation of classical electrodynamics has the major advantage of simplicity, it has the major disadvantage of unfamiliarity. Most readers are unlikely to be familiar with the distinction between, for example, vectors and dual vectors, and the role played by the spacetime metric in the equations of physics. While these concepts are not inordinately difficult to explain—and I explain them in chapter 8—it would be too much of a distraction to do so before presenting the theory of electromagnetism. Therefore, I defer the discussion of special relativity until chapter 8 and, with the exception of a few side comments, I do not use special relativistic notation for classical electrodynamics until that point. However, it is important that the reader be aware of the fact that classical electrodynamics is compatible with the spacetime structure of special relativity even if we use a notation that does not make it manifestly so.

1.1 The Fundamental Electromagnetic Variables Are the Potentials, Not the Field Strengths

The electromagnetic field is a fundamental constituent of nature. Its existence does not need to be justified or explained any more (or less) than the existence of, say, electrons needs to be justified or explained. The electromagnetic field is a "gauge field," the same basic type of field that also describes the W and Z bosons and gluons. Indeed, the electromagnetic field together with the W and Z fields comprise a unified "electroweak gauge field" that describes both the electromagnetic and weak interactions. However, for the ("low-energy") phenomena of interest in this book, the electromagnetic field effectively decouples from its electroweak partners and can be considered on its own.

I defer giving a mathematically complete discussion of electromagnetism as a gauge field until chapter 9. What is necessary for the reader to be aware of now is that the fundamental description of the electromagnetic field is given in terms of the *potentials* ϕ and $\boldsymbol{A}$, not the *field strengths* $\boldsymbol{E}$ and $\boldsymbol{B}$. As explained below, there are situations where the potentials contain more information than can be obtained from the field strengths. However, ϕ and $\boldsymbol{A}$ do not uniquely describe the electromagnetic field: the potentials $\phi', \boldsymbol{A}'$ and $\phi, \boldsymbol{A}$ are considered to be physically equivalent (i.e., they represent the same electromagnetic field) if they differ by a *gauge transformation*, that is, if for some function $\chi(t, \boldsymbol{x})$, we have[2]

$$\phi' = \phi - \frac{\partial \chi}{\partial t}, \qquad \boldsymbol{A}' = \boldsymbol{A} + \nabla \chi. \tag{1.13}$$

In other words, an electromagnetic field is an equivalence class of potentials $\phi, \boldsymbol{A}$ under the transformation eq. (1.13).

It is easily verified that the field strengths, $\boldsymbol{E}$ and $\boldsymbol{B}$, defined by eqs. (1.6) and (1.7), are gauge invariant. Furthermore, it is not difficult to show that in any simply connected[3]

[2] In special relativistic notation, a gauge transformation can be expressed more simply as $A_\mu \to A_\mu + \partial\chi/\partial x^\mu$.

[3] A simply connected region is one in which every closed loop can be continuously deformed to a point.

spacetime region, if $\phi_1, \boldsymbol{A}_1$ and $\phi_2, \boldsymbol{A}_2$ give rise to the same field strengths $\boldsymbol{E}$ and $\boldsymbol{B}$, then $\phi_1, \boldsymbol{A}_1$ and $\phi_2, \boldsymbol{A}_2$ can differ at most by a gauge transformation. Thus, in any simply connected region, $\boldsymbol{E}$ and $\boldsymbol{B}$ contain all of the information contained in ϕ and $\boldsymbol{A}$. Since all physically measurable quantities must be gauge invariant, it is very convenient in many circumstances to work with $\boldsymbol{E}$ and $\boldsymbol{B}$ rather than ϕ and $\boldsymbol{A}$. In many contexts, electromagnetic phenomena can be fully described in terms of $\boldsymbol{E}$ and $\boldsymbol{B}$.

However, as we shall see in chapter 9, the coupling of the electromagnetic field to fundamental charged matter (namely, charged fields) can be described only in terms of the potentials, not the field strengths. Furthermore, there are physically relevant situations where $\boldsymbol{E}$ and $\boldsymbol{B}$ do not contain all of the information about the electromagnetic field. As an example, consider the region outside an infinite solenoid. Suppose that inside the solenoid, there is a nonvanishing, uniform magnetic field, but outside the solenoid, we have $\boldsymbol{E} = \boldsymbol{B} = 0$. Since the region outside the solenoid is not simply connected, the fact that $\boldsymbol{E}$ and $\boldsymbol{B}$ vanish in that region does not imply that the potentials are gauge equivalent to zero there. Indeed, eq. (1.7) implies, via Stokes's theorem, that when $\boldsymbol{B} \neq 0$ *inside* the solenoid, we have $\oint \boldsymbol{A} \cdot \boldsymbol{dl} \neq 0$ for any loop *outside* the solenoid that encloses it. (Note that $\oint \boldsymbol{A} \cdot \boldsymbol{dl}$ is gauge invariant, i.e., its value does not change under eq. (1.13).) A quantum mechanical charged particle that stays entirely outside the solenoid will be affected by this vector potential, as it will produce a relative phase shift in the parts of the wave function that go around the solenoid in different directions, producing a physically measurable shift in the resulting interference pattern. This phenomenon, known as the Aharonov-Bohm effect, is sometimes attributed to the weirdness of quantum mechanics. However, the effect has nothing to do with quantum mechanics—the same effect would occur for a classical charged field. And there is nothing weird about the effect, once one recognizes that the electromagnetic field is represented, at a fundamental level, by the potentials $\phi, \boldsymbol{A}$ (modulo gauge), not the field strengths $\boldsymbol{E}, \boldsymbol{B}$.

Thus, while for many purposes, it is convenient to introduce and work with the field strengths $\boldsymbol{E}$ and $\boldsymbol{B}$, it is important to recognize that the fundamental description of the electromagnetic field is given by the potentials ϕ and $\boldsymbol{A}$. The Maxwell equations (1.3) and (1.4) should be viewed as consequences of the definitions of $\boldsymbol{E}$ and $\boldsymbol{B}$ given by eqs. (1.6) and (1.7).

1.2 Electromagnetic Energy, Momentum, and Stress Are an Integral Part of the Theory

The electromagnetic field, like all other forms of matter, has energy, momentum, and stress properties. These properties, like Maxwell's equations (1.1)–(1.4), are an integral part of the theory.

As discussed much more fully in chapter 9, classical electrodynamics can be viewed as arising from the Lagrangian density

$$\mathcal{L} = \frac{1}{2}\left(\epsilon_0 |\boldsymbol{E}|^2 - \frac{1}{\mu_0}|\boldsymbol{B}|^2\right) - \phi\rho + \boldsymbol{A} \cdot \boldsymbol{J}. \tag{1.14}$$

Here, as discussed in section 1.1, the dynamical variables are ϕ and $\boldsymbol{A}$, and the Euler-Lagrange equations are obtained by varying $\mathcal{L}$ with respect to these variables; $\boldsymbol{E}$ and $\boldsymbol{B}$ are viewed as the functions of ϕ and $\boldsymbol{A}$ defined by eqs.(1.6) and (1.7). In eq. (1.14), the charge density ρ and current density $\boldsymbol{J}$ are treated as externally prescribed, nondynamical quantities.[4] The Euler-Lagrange equations arising from the variation of eq. (1.14) with respect to ϕ and $\boldsymbol{A}$ are precisely Maxwell's equations (1.1)–(1.2). The additional Maxwell equations (1.3)–(1.4) follow from the definitions (1.6) and (1.7) of $\boldsymbol{E}$ and $\boldsymbol{B}$, respectively. The fact that the Lagrangian must be viewed as a function of ϕ and $\boldsymbol{A}$—and the terms representing the coupling of the electromagnetic field to charged matter cannot even be written down in terms of $\boldsymbol{E}$ and $\boldsymbol{B}$—is further manifestation of the fact that the fundamental dynamical variables in electomagnetism are ϕ and $\boldsymbol{A}$.

The energy, momentum, and stress properties of the electromagnetic field are determined by its coupling to gravity. The coupling to gravity is obtained by generalizing the Lagrangian (1.14) for the spacetime of special relativity to curved spacetime. This can be done in a very simple and natural way, which is unique if one does not allow derivatives of the metric to appear in the Maxwell Lagrangian. The stress-energy-momentum tensor of the electromagnetic field is then obtained by functional differentiation of the Lagrangian with respect to the spacetime metric, since this is what appears as a source term for gravity in Einstein's equation of general relativity. I briefly indicate how this works in section 9.1. The only point I wish to make here is that, just as the Lagrangian (1.14) gives rise to Maxwell's equations, its natural generalization to curved spacetime gives rise to the following formulas for the energy density $\mathcal{E}$, momentum density $\boldsymbol{\mathcal{P}}$, and stress tensor Θ_{ij} of the electromagnetic field:

$$\mathcal{E} = \frac{1}{2}\left(\epsilon_0|\boldsymbol{E}|^2 + \frac{1}{\mu_0}|\boldsymbol{B}|^2\right), \tag{1.15}$$

$$\boldsymbol{\mathcal{P}} = \epsilon_0 \boldsymbol{E} \times \boldsymbol{B}, \tag{1.16}$$

$$\Theta_{ij} = \epsilon_0 E_i E_j + \frac{1}{\mu_0} B_i B_j - \frac{1}{2}\delta_{ij}\left(\epsilon_0|\boldsymbol{E}|^2 + \frac{1}{\mu_0}|\boldsymbol{B}|^2\right). \tag{1.17}$$

These formulas should be viewed as having fundamental status in the theory of electromagnetism, comparable to that of Maxwell's equations.

In principle, the validity of eqs. (1.15)–(1.17) could be tested by observing the gravitational effects of electromagnetic fields. Electromagnetic fields make nontrivial contributions to the mass-energy of ordinary matter—certainly large enough to produce observable gravitational effects for macroscopic bodies. However, there is no way to observe these effects separately from the gravitational effects of the nonelectromagnetic constituents of matter. Thus, it would be necessary to observe the gravitational effects of free electromagnetic fields if one wishes to test eqs. (1.15)–(1.17). The gravitational effects of free electromagnetic fields are far too small to be measured in laboratory experiments. However, in the early universe, the thermally distributed electromagnetic radiation that presently constitutes the cosmic microwave background made a dominant contribution to the energy density and pressure in the universe, both of which affect the expansion of the universe. The expansion history of the universe is observed to be in accord with the electromagnetic energy density and pressure of thermal radiation obtained from the above formulas.

---

[4] Of course, the charged matter should really have its own dynamical degrees of freedom, and there should be additional terms in the Lagrangian involving the fields representing the charged matter. The coupling terms between the charged matter and electromagnetic field should then be represented in terms of $\phi, \boldsymbol{A}$, and the dynamical fields describing the charged matter. This will be seen explicitly in chapter 9.

There are important conservation laws associated with eqs. (1.15)–(1.17). In special relativity, the "flow of mass" (momentum) and "flow of energy" represent the same quantity apart from a factor of c^2, so

$$\boldsymbol{\mathcal{S}} \equiv c^2 \boldsymbol{\mathcal{P}} = c^2 \epsilon_0 \boldsymbol{E} \times \boldsymbol{B} = \frac{1}{\mu_0} \boldsymbol{E} \times \boldsymbol{B} \tag{1.18}$$

represents the flux of energy per unit volume of the electromagnetic field. A computation using Maxwell's equations yields (see section 5.1 for details)

$$\frac{\partial \mathcal{E}}{\partial t} + \nabla \cdot \boldsymbol{\mathcal{S}} = -\boldsymbol{J} \cdot \boldsymbol{E}, \tag{1.19}$$

$$\frac{\partial \mathcal{P}_i}{\partial t} - \sum_{j=1}^{3} \partial_j \Theta_{ij} = -\left[\rho E_i + (\boldsymbol{J} \times \boldsymbol{B})_i\right]. \tag{1.20}$$

In the absence of charges and currents (i.e., when $\rho = \boldsymbol{J} = 0$), the right sides of eqs. (1.19) and (1.20) vanish. In this case, eqs. (1.19) and (1.20) have the interpretation of expressing local conservation of energy and momentum of the electromagnetic field. To see this more explicitly, note that in a small volume δV about $\boldsymbol{x}$, the quantity $\delta V \nabla \cdot \boldsymbol{\mathcal{S}}$ represents the net flux of energy out of δV. By eq. (1.19), this is equal to $-\delta V \partial \mathcal{E}/\partial t$ when $\rho = \boldsymbol{J} = 0$, thus expressing local conservation of energy. Global energy conservation for the electromagnetic field is obtained by integrating eq. (1.19) over all of space, assuming that $\boldsymbol{E}$ and $\boldsymbol{B}$ vanish sufficiently rapidly near infinity. In that case, the integral over all of space of $\nabla \cdot \boldsymbol{\mathcal{S}}$ vanishes by Gauss's theorem (see chapter 2), and we obtain

$$\frac{d}{dt} \int \mathcal{E} \, d^3x = 0, \tag{1.21}$$

provided that $\rho = \boldsymbol{J} = 0$. Similarly, when $\rho = \boldsymbol{J} = 0$, eq. (1.20) expresses local conservation of momentum, and integration of eq. (1.20) over all of space yields the global momentum conservation law

$$\frac{d}{dt} \int \boldsymbol{\mathcal{P}} \, d^3x = 0. \tag{1.22}$$

When ρ and $\boldsymbol{J}$ are nonvanishing, the right sides of eqs. (1.19) and (1.20) are, in general, nonvanishing, and electromagnetic energy and momentum are not conserved by themselves. This is because electromagnetic energy and momentum can be exchanged with the energy and momentum of the charged matter. For the total (electromagnetic and matter) energy to be locally conserved, the electromagnetic field must be transferring energy to the matter at the rate

$$\frac{\partial \mathcal{E}_{\text{matter}}}{\partial t} = \boldsymbol{J} \cdot \boldsymbol{E}. \tag{1.23}$$

Similarly, for total momentum to be conserved, the electromagnetic field must be transferring momentum to the matter at the rate given by minus the right side of eq. (1.20); that is, it must be exerting a force per unit volume, $\boldsymbol{f}$, on the matter, given by

$$\boldsymbol{f} = \rho \boldsymbol{E} + \boldsymbol{J} \times \boldsymbol{B}. \tag{1.24}$$

In standard treatments of electromagnetism, the order of the arguments presented above is reversed, giving rise to some serious difficulties. Instead of starting with the eqs. (1.15)–(1.17) for energy density, momentum density, and stress, and then deriving the Lorentz force, eq. (1.24), standard treatments start with the Lorentz force—or rather, the Coulomb's law version of this expression for static point charges in electrostatics. The "work done" in quasi-statically bringing charges together from infinity is then calculated and is associated with the energy contained in the electromagnetic field. This argument eventually leads to the correct formula $\frac{\epsilon_0}{2}\int |\boldsymbol{E}|^2 d^3x$ for the energy of the electromagnetic field in electrostatics. However, this argument works in electrostatics because it is possible to move a charged body in an electric field in such a way that its rest mass (i.e., internal energy) does not change. Although this may seem obvious, the corresponding result does *not* hold in magnetostatics, because energy is required to maintain the currents in a body. This is true for permanent magnets as well as current loops. The rest mass of a magnetic dipole will change as it moves in a nonuniform magnetic field, as we shall see explicitly in section 4.3 and again in section 10.3.2. As I show in section 4.3, the electromagnetic interaction energy of a magnetic dipole $\boldsymbol{\mu}$ in an external magnetic field $\boldsymbol{B}^{\text{ext}}$ can be derived straightforwardly from eq. (1.15) and yields the value $+\boldsymbol{\mu}\cdot\boldsymbol{B}^{\text{ext}}$. However, many references give the incorrect formula $-\boldsymbol{\mu}\cdot\boldsymbol{B}^{\text{ext}}$ based on arguments using "work done," failing to take into account the change in rest mass.

The formulas (1.15) for the electromagnetic energy density and (1.16) for the electromagnetic momentum density $\mathcal{P}$ are justified in many standard treatments by taking eq. (1.23) as a starting point. It is then natural to interpret eq. (1.19) (which is derived directly from Maxwell's equations) as representing local energy conservation. One thereby can identify $\mathcal{E}$ and $\boldsymbol{\mathcal{S}} \equiv c^2\mathcal{P}$ with electromagnetic energy density and energy flux, respectively. However, this argument has the serious drawback that $\mathcal{P}$ appears in eq. (1.19) only in the form $\nabla\cdot\mathcal{P}$. This leads many authors to suggest that $\mathcal{P}$ is undefined up to the addition of the curl of a vector field. This is not correct; formulas for $\mathcal{P}$ that differ by a curl of a vector field will have different gravitational consequences, so if one has two formulas for $\mathcal{P}$ that differ by a curl, at most one of them can be valid.

In summary, rather than attempt to derive eqs. (1.15)–(1.17) from Maxwell's equations by assuming that eq. (1.23) and eq. (1.24) hold, it is much healthier to view formulas (1.15)–(1.17) as an integral part of the specification of the theory, with eq. (1.23) and eq. (1.24) then following as consequences. The conservation laws (1.19) and (1.20) provide important consistency relations between Maxwell's equations and eqs. (1.15)–(1.17), but they do not enable one to derive eqs. (1.15)–(1.17) from Maxwell's equations. Equations (1.15)–(1.17) should be viewed as fundamental aspects of electromagnetic theory, with a status similar to that of Maxwell's equations.

1.3 Electromagnetic Fields Should Not Be Viewed as Being Produced by Charged Matter

Maxwell's equations (1.1)–(1.4) together with eqs. (1.23) and (1.24) describe the *interaction* of the electromagnetic field with matter. The electromagnetic field does not, in any sense, play a subordinate role in this interaction. The electromagnetic field has its own independent dynamical degrees of freedom, and these should be thought of as being on an equal footing with the dynamical degrees of freedom of the charged matter. The electromagnetic field should not be thought of as being *produced* by charges and currents—despite the fact that ρ and $\boldsymbol{J}$ are commonly referred to as "source terms" in Maxwell's equations (and I use this terminology in this book).

The independent dynamical degrees of freedom of the electromagnetic field are characterized by the initial value formulation of Maxwell's equations, which is discussed in section 5.4. The theorem at the end of that section states the following: Specify $\rho(t, \boldsymbol{x})$ and $\boldsymbol{J}(t, \boldsymbol{x})$ on spacetime, subject to the conservation equation (1.5). Let $\boldsymbol{E}_0(\boldsymbol{x})$ and $\boldsymbol{B}_0(\boldsymbol{x})$ be arbitrary vector fields on space such that $\nabla \cdot \boldsymbol{E}_0 = \rho(t=0, \boldsymbol{x})/\epsilon_0$, and $\nabla \cdot \boldsymbol{B}_0 = 0$. Then there exists a unique solution $(\boldsymbol{E}(t, \boldsymbol{x}), \boldsymbol{B}(t, \boldsymbol{x}))$ to Maxwell's equations (1.1)–(1.4) such that $\boldsymbol{E}(t=0, \boldsymbol{x}) = \boldsymbol{E}_0(\boldsymbol{x})$, and $\boldsymbol{B}(t=0, \boldsymbol{x}) = \boldsymbol{B}_0(\boldsymbol{x})$. Thus, there are as many solutions to Maxwell's equations with a specified ρ and $\boldsymbol{J}$ as there are vector fields $(\boldsymbol{E}_0(\boldsymbol{x}), \boldsymbol{B}_0(\boldsymbol{x}))$ satisfying the above conditions on their divergence. The fact that this initial data for the electromagnetic field can be freely specified shows that the electromagnetic field has its own independent dynamical degrees of freedom. Solutions to Maxwell's equations are *not* determined by ρ and $\boldsymbol{J}$.

The dynamical degrees of freedom of the electromagnetic field are not visible in electrostatics and magnetostatics, since no time-independent solutions of Maxwell's equations with $\rho = \boldsymbol{J} = 0$ go to zero at infinity. Thus, if one specifies time-independent sources, $\partial \rho/\partial t = \partial \boldsymbol{J}/\partial t = 0$, then solutions to Maxwell's equations for $\boldsymbol{E}$ and $\boldsymbol{B}$ with $\partial \boldsymbol{E}/\partial t = \partial \boldsymbol{B}/\partial t = 0$ and with $\boldsymbol{E}$ and $\boldsymbol{B}$ going to zero at infinity are uniquely determined by ρ and $\boldsymbol{J}$. Consequently, one can uniquely associate a stationary electric field $\boldsymbol{E}$ with a stationary charge distribution ρ, and one can uniquely associate a stationary magnetic field $\boldsymbol{B}$ with a stationary current distribution $\boldsymbol{J}$. Therefore, it is possible to view the electric field in electrostatics as being "produced" by charges, and it is possible to view the magnetic field in magnetostatics as being "produced" by currents. In electrostatics, one can even get away with saying—as is frequently done—that charges exert forces on one another. This, of course, is not the case: The electromagnetic force (1.24) on a charged body is exerted by the electromagnetic field that is present at the location of the body, not by other, distant charges.

In electrodynamics, one is frequently interested in considering situations in which there is "no incoming electromagnetic radiation." As discussed in depth in section 5.2, solutions with no incoming radiation are given by the retarded Green's function applied to ρ and $\boldsymbol{J}$, and these solutions are uniquely determined by ρ and $\boldsymbol{J}$. Again, this makes it possible to take the view that, in the absence of incoming radiation, the electromagnetic fields are "produced" by the charges and currents. However, while the "no incoming radiation" condition is a useful idealization applicable to many problems, it should not be taken seriously as an initial condition for our universe. Although we certainly do not know the precise initial conditions at the "big bang," we do know that matter in the very early universe was an extremely hot and dense plasma. In such a hot and dense plasma, the electromagnetic field "produces" charges (e.g., electron-positron pairs) to much the same degree as charges "produce" electromagnetic fields. It certainly does not make any sense to think of the charges as coming first and then producing the electromagnetic fields.

Thus, although there are circumstances where one could take the view that electromagnetic fields are produced by charges, it is far healthier to think of the electromagnetic field and charged matter as independent entities that interact via Maxwell's equations and eqs. (1.23) and (1.24). Indeed, the view that electromagnetic fields are produced by charges is particularly untenable in quantum field theory, since it is essential for the understanding of such phenomena as the vacuum fluctuations of the electromagnetic field that the electromagnetic field have its own dynamical degrees of freedom, independently of the existence of charged matter.

1.4 At a Fundamental Level, Classical Charged Matter Must Be Viewed as Continuous Rather Than Point-Like

Maxwell's equations (1.1)–(1.4) were formulated above using a continuum notion of charge density ρ and current density $\boldsymbol{J}$; that is, ρ and $\boldsymbol{J}$ were taken to be smooth functions of $(t, \boldsymbol{x})$. These equations have a mathematically well-posed initial value formulation, as already mentioned in section 1.3 and as discussed in depth in section 5.4. However, in a complete theory, one must also specify the form of the charged matter and its equations of motion. As discussed further in chapter 9, at a fundamental level, charged matter is believed to consist of charged (quantum) fields. However, one can also consider "phenomenological models" of charged matter, such as a charged fluid. In any case, the equations of motion of the charged matter together with Maxwell's equations comprise a coupled system that must be solved simultaneously—since the motion of the charged matter depends on the electromagnetic field, but the dynamical evolution of the electromagnetic field depends on the charges and currents of the matter. It is essential that the coupled Maxwell–charged-matter system have a well-posed initial value formulation, so that there is no difficulty, in principle, in obtaining solutions to the coupled system for given initial conditions.

However, at least 90% of what is normally treated in electromagnetism courses does not consider the full, coupled Maxwell–matter system but instead considers the following two idealized problems:

- Type I. For a given externally specified ρ and $\boldsymbol{J}$, find the corresponding electromagnetic fields (i.e., the unique stationary solution in electrostatics and magnetostatics and/or the retarded solution in electrodynamics).
- Type II. Find the motion of a charged body for given externally specified fields $\boldsymbol{E}$ and $\boldsymbol{B}$ (i.e., neglecting the self-fields associated with the presence of the charged body).

For these idealized problems, it is very useful to introduce the notion of a point charge.

By a "point charge of charge q" moving on the worldline $\boldsymbol{X}(t)$ (with $|d\boldsymbol{X}/dt| < c$ for all t) is meant the charge-current

$$\rho(t, \boldsymbol{x}) = q\delta(\boldsymbol{x} - \boldsymbol{X}(t)), \qquad \boldsymbol{J}(t, \boldsymbol{x}) = q\frac{d\boldsymbol{X}}{dt}(t)\delta(\boldsymbol{x} - \boldsymbol{X}(t)), \tag{1.25}$$

where δ denotes the 3-dimensional Dirac delta function. This may be thought of as a limit of a charge distribution that at each t becomes more and more concentrated at the point $\boldsymbol{X}(t)$. This limit does not define a function, but it has a well-defined meaning as a distribution.[5] The charge-current eq. (1.25) satisfies eq. (1.5) in a well-defined, distributional sense.

[5] A distribution is a linear map from "test functions" (i.e., smooth functions that are nonvanishing only in a bounded region) into numbers that depends continuously on the test function in an appropriate sense. The Dirac delta function is simply the evaluation map on test functions; that is, $\delta(\boldsymbol{x} - \boldsymbol{X})$ maps the test function f into the number $f(\boldsymbol{X})$.

It can be seen from Gauss's law that for solutions to eq. (1.1) with ρ given by eq. (1.25), the electric field $\boldsymbol{E}$ must diverge near the charge as $1/|\boldsymbol{x}-\boldsymbol{X}|^2$. Consequently, by eq. (1.15), the electromagnetic energy density diverges as $1/|\boldsymbol{x}-\boldsymbol{X}|^4$, which is not integrable. Thus, the total electromagnetic energy of a point charge is infinite, and so point charges cannot be considered to be physical objects in classical electrodynamics. Nevertheless, they can be introduced in the context of problems of type I or type II above.

In problems of type I, since Maxwell's equations are linear in $\boldsymbol{E}$ and $\boldsymbol{B}$, these equations make perfectly good mathematical sense when ρ and $\boldsymbol{J}$ (and hence, $\boldsymbol{E}$ and $\boldsymbol{B}$) are distributions rather than functions. It is extremely useful to consider solutions to Maxwell's equations with a point charge-current, eq. (1.25). Such solutions are of direct interest for describing situations where the charge-current is highly localized, and more general solutions can be obtained by "superposition" (again using the linearity of Maxwell's equations).

For problems of type II, it would be quite complicated to analyze the motion of an extended charged body described by a causal dynamics compatible with special relativity, since the different electromagnetic forces on the different parts of the body would induce internal oscillations. One might therefore be tempted to take a limit of vanishing size of the body, wherein the complications due to internal dynamics should become negligible. However, such a limit at fixed q leads one back to the problem of infinite self-energy and also would require infinite mechanical stresses to keep the body from flying apart. Nevertheless, we will see in chapter 10 that it is possible to take a limit in which the size, charge, and mass of the body all scale to zero in a suitable manner. To leading order, the motion $\boldsymbol{x}(t)$ of the body becomes independent of its internal structure and is given by the Lorentz force equation[6]

$$\frac{d\boldsymbol{p}}{dt} = \boldsymbol{F} = q\left(\boldsymbol{E}^{\text{ext}} + \boldsymbol{v} \times \boldsymbol{B}^{\text{ext}}\right), \tag{1.26}$$

where $\boldsymbol{v} = d\boldsymbol{x}/dt$, and $\boldsymbol{p} = \gamma m \boldsymbol{v}$ with $\gamma = (1 - v^2/c^2)^{-1/2}$. Here $\boldsymbol{E}^{\text{ext}}$ and $\boldsymbol{B}^{\text{ext}}$ are the "externally prescribed" fields, with the influence of the charge-current of eq. (1.25) ignored. Note that the force $\boldsymbol{F}$ appearing in eq. (1.26) corresponds to eq. (1.24), with ρ and $\boldsymbol{J}$ given by eq. (1.25). The equation of motion (1.26) makes good mathematical sense and, for specified external fields, has a unique solution for a given initial position and velocity of the point charge.

One might attempt to go beyond the context of problems of type I or type II to consider the full, coupled Maxwell-matter system with point charges. In other words, one could attempt to solve Maxwell's equations with source eq. (1.25) simultaneously with eq. (1.26), where now $\boldsymbol{E}$ and $\boldsymbol{B}$ represent the full electromagnetic field, including the effects of the point charge. However, this system of equations does not make mathematical sense, since Maxwell's equations imply that $\boldsymbol{E}$ must be singular at the location of the charge, in which case, eq. (1.26) is ill defined. This is a reflection of the fact that the coupled Maxwell-matter system is nonlinear, and distributional solutions make sense for a nonlinear system only in very limited circumstances—which do not apply here.

[6] I also show in chapter 10 how to obtain leading-order corrections to Lorentz force motion taking into account the self-field of the body.

This difficulty is resolved by simply recognizing that, at a fundamental level in classical electrodynamics,[7] ρ and $\boldsymbol{J}$ must be taken to be quantities smoothly distributed in spacetime. No difficulties of the sort mentioned in the previous paragraph arise when one considers continuum charged matter. In particular, as we shall see in chapter 9, the self-consistent coupled system of Maxwell's equations and the equation of motion of a charged scalar field is well posed. The notion of a point charge is convenient to introduce in the circumstances described above, but it cannot be viewed as a fundamental description of charged matter.

[7]The same is true in quantum electrodynamics in the sense that for any physically acceptable state of charged matter, $\langle\rho\rangle$ and $\langle\boldsymbol{J}\rangle$ must be smoothly distributed in spacetime.

Index