Chapter 1

Introductory concepts

1.1 Notation, conventions, and symbology

Any book that covers a broad range of topics will likely harbor some problems with notation and symbology. This results from having the same symbol used in different areas to represent different quantities, and also from having too many quantities to represent. Rather than invent new symbols, we choose to stay close to the standards and warn the reader about any symbol used to represent more than one distinct quantity.

The basic nature of a physical quantity is indicated by typeface or by the use of a diacritical mark. Scalars are shown in ordinary typeface: q, Φ, for example. Vectors are shown in boldface: $\mathbf{E}, \mathbf{\Pi}$. Dyadics are shown in boldface with an overbar: $\bar{\boldsymbol{\epsilon}}, \bar{\mathbf{A}}$. Frequency dependent quantities are indicated by a tilde, whereas time dependent quantities are written without additional indication; thus we write $\tilde{\mathbf{E}}(\mathbf{r}, \omega)$ and $\mathbf{E}(\mathbf{r}, t)$. (Some quantities, such as impedance, are used in the frequency domain to interrelate Fourier spectra; although these quantities are frequency dependent they are seldom written in the time domain, and hence we do not attach tildes to their symbols.) We often combine diacritical marks: for example, $\tilde{\bar{\boldsymbol{\epsilon}}}$ denotes a frequency domain dyadic. We distinguish carefully between phasor and frequency domain quantities. The variable ω is used for the frequency variable of the Fourier spectrum, while $\check{\omega}$ is used to indicate the constant frequency of a time harmonic signal. We thus further separate the notion of a phasor field from a frequency domain field by using a *check* to indicate a phasor field: $\check{\mathbf{E}}(\mathbf{r})$. However, there is often a simple relationship between the two, such as $\check{\mathbf{E}} = \tilde{\mathbf{E}}(\check{\omega})$.

We designate the field and source point position vectors by $\mathbf{r}$ and $\mathbf{r}'$, respectively, and the corresponding relative displacement or distance vector by $\mathbf{R}$:

$$\mathbf{R} = \mathbf{r} - \mathbf{r}'.$$

A hat designates a vector as a unit vector (e.g., $\hat{\mathbf{x}}$). The sets of coordinate variables in rectangular, cylindrical, and spherical coordinates are denoted by

$$(x, y, z), \qquad (\rho, \phi, z), \qquad (r, \theta, \phi),$$

respectively. (In the spherical system ϕ is the azimuthal angle and θ is the polar angle.) We freely use the "del" operator notation ∇ for gradient, curl, divergence, Laplacian, and so on.

The SI (MKS) system of units is employed throughout the book.

© 2001 by CRC Press LLC

1.2 The field concept of electromagnetics

Introductory treatments of electromagnetics often stress the role of the field in force transmission: the individual fields **E** and **B** are defined via the mechanical force on a small test charge. This is certainly acceptable, but does not tell the whole story. We might, for example, be left with the impression that the EM field always arises from an interaction between charged objects. Often coupled with this is the notion that the field concept is meant merely as an aid to the calculation of force, a kind of notational convenience not placed on the same physical footing as force itself. In fact, fields are more than useful — they are fundamental. Before discussing electromagnetic fields in more detail, let us attempt to gain a better perspective on the field concept and its role in modern physical theory. Fields play a central role in any attempt to describe physical reality. They are as real as the physical substances we ascribe to everyday experience. In the words of Einstein [63],

> "It seems impossible to give an obvious qualitative criterion for distinguishing between matter and field or charge and field."

We must therefore put fields and particles of matter on the same footing: both carry energy and momentum, and both interact with the observable world.

1.2.1 Historical perspective

Early nineteenth century physical thought was dominated by the *action at a distance* concept, formulated by Newton more than 100 years earlier in his immensely successful theory of gravitation. In this view the influence of individual bodies extends across space, instantaneously affects other bodies, and remains completely unaffected by the presence of an intervening medium. Such an idea was revolutionary; until then *action by contact*, in which objects are thought to affect each other through physical contact or by contact with the intervening medium, seemed the obvious and only means for mechanical interaction. Priestly's experiments in 1766 and Coulomb's torsion-bar experiments in 1785 seemed to indicate that the force between two electrically charged objects behaves in strict analogy with gravitation: both forces obey inverse square laws and act along a line joining the objects. Oersted, Ampere, Biot, and Savart soon showed that the magnetic force on segments of current-carrying wires also obeys an inverse square law.

The experiments of Faraday in the 1830s placed doubt on whether action at a distance really describes electric and magnetic phenomena. When a material (such as a dielectric) is placed between two charged objects, the force of interaction decreases; thus, the intervening medium does play a role in conveying the force from one object to the other. To explain this, Faraday visualized "lines of force" extending from one charged object to another. The manner in which these lines were thought to interact with materials they intercepted along their path was crucial in understanding the forces on the objects. This also held for magnetic effects. Of particular importance was the number of lines passing through a certain area (the *flux*), which was thought to determine the amplitude of the effect observed in Faraday's experiments on electromagnetic induction.

Faraday's ideas presented a new world view: electromagnetic phenomena occur in the region surrounding charged bodies, and can be described in terms of the laws governing the "field" of his lines of force. Analogies were made to the stresses and strains in material objects, and it appeared that Faraday's force lines created equivalent electromagnetic

© 2001 by CRC Press LLC

stresses and strains in media surrounding charged objects. His law of induction was formulated not in terms of positions of bodies, but in terms of lines of magnetic force. Inspired by Faraday's ideas, Gauss restated Coulomb's law in terms of flux lines, and Maxwell extended the idea to time changing fields through his concept of displacement current.

In the 1860s Maxwell created what Einstein called "the most important invention since Newton's time"— a set of equations describing an entirely field-based theory of electromagnetism. These equations do not model the forces acting between bodies, as do Newton's law of gravitation and Coulomb's law, but rather describe only the dynamic, time-evolving structure of the electromagnetic field. Thus bodies are not seen to interact with each other, but rather with the (very real) electromagnetic field they create, an interaction described by a supplementary equation (the Lorentz force law). To better understand the interactions in terms of mechanical concepts, Maxwell also assigned properties of stress and energy to the field.

Using constructs that we now call the electric and magnetic fields and potentials, Maxwell synthesized all known electromagnetic laws and presented them as a system of differential and algebraic equations. By the end of the nineteenth century, Hertz had devised equations involving only the electric and magnetic fields, and had derived the laws of circuit theory (Ohm's law and Kirchoff's laws) from the field expressions. His experiments with high-frequency fields verified Maxwell's predictions of the existence of electromagnetic waves propagating at finite velocity, and helped solidify the link between electromagnetism and optics. But one problem remained: if the electromagnetic fields propagated by stresses and strains on a medium, how could they propagate through a vacuum? A substance called the *luminiferous aether*, long thought to support the transverse waves of light, was put to the task of carrying the vibrations of the electromagnetic field as well. However, the pivotal experiments of Michelson and Morely showed that the aether was fictitious, and the physical existence of the field was firmly established.

The essence of the field concept can be conveyed through a simple thought experiment. Consider two stationary charged particles in free space. Since the charges are stationary, we know that (1) another force is present to balance the Coulomb force between the charges, and (2) the momentum and kinetic energy of the system are zero. Now suppose one charge is quickly moved and returned to rest at its original position. Action at a distance would require the second charge to react immediately (Newton's third law), but by Hertz's experiments it does not. There appears to be no change in energy of the system: both particles are again at rest in their original positions. However, after a time (given by the distance between the charges divided by the speed of light) we find that the second charge does experience a change in electrical force and begins to move away from its state of equilibrium. But by doing so it has gained net kinetic energy and momentum, and the energy and momentum of the system seem larger than at the start. This can only be reconciled through field theory. If we regard the field as a physical entity, then the nonzero work required to initiate the motion of the first charge and return it to its initial state can be seen as increasing the energy of the field. A disturbance propagates at finite speed and, upon reaching the second charge, transfers energy into kinetic energy of the charge. Upon its acceleration this charge also sends out a wave of field disturbance, carrying energy with it, eventually reaching the first charge and creating a second reaction. At any given time, the net energy and momentum of the system, composed of both the bodies and the field, remain constant. We thus come to regard the electromagnetic field as a true physical entity: an entity capable of carrying energy and momentum.

© 2001 by CRC Press LLC

1.2.2 Formalization of field theory

Before we can invoke physical laws, we must find a way to describe the *state* of the system we intend to study. We generally begin by identifying a set of *state variables* that can depict the physical nature of the system. In a mechanical theory such as Newton's law of gravitation, the state of a system of point masses is expressed in terms of the instantaneous positions and momenta of the individual particles. Hence $6N$ state variables are needed to describe the state of a system of N particles, each particle having three position coordinates and three momentum components. The time evolution of the system state is determined by a supplementary force function (e.g., gravitational attraction), the initial state (initial conditions), and Newton's second law $\mathbf{F} = d\mathbf{P}/dt$.

Descriptions using finite sets of state variables are appropriate for action-at-a-distance interpretations of physical laws such as Newton's law of gravitation or the interaction of charged particles. If Coulomb's law were taken as the force law in a mechanical description of electromagnetics, the state of a system of particles could be described completely in terms of their positions, momenta, and charges. Of course, charged particle interaction is not this simple. An attempt to augment Coulomb's force law with Ampere's force law would not account for kinetic energy loss via radiation. Hence we abandon[1] the mechanical viewpoint in favor of the field viewpoint, selecting a different set of state variables. The essence of field theory is to regard electromagnetic phenomena as affecting all of space. We shall find that we can describe the field in terms of the four vector quantities **E**, **D**, **B**, and **H**. Because these fields exist by definition at each point in space and each time t, a finite set of state variables cannot describe the system.

Here then is an important distinction between field theories and mechanical theories: the state of a field at any instant can only be described by an infinite number of state variables. Mathematically we describe fields in terms of functions of continuous variables; however, we must be careful not to confuse all quantities described as "fields" with those fields innate to a scientific field theory. For instance, we may refer to a temperature "field" in the sense that we can describe temperature as a function of space and time. However, we do *not* mean by this that temperature obeys a set of physical laws analogous to those obeyed by the electromagnetic field.

What special character, then, can we ascribe to the electromagnetic field that has meaning beyond that given by its mathematical implications? In this book, **E**, **D**, **B**, and **H** are integral parts of a *field-theory description* of electromagnetics. In any field theory we need two types of fields: a *mediating field* generated by a source, and a field describing the source itself. In free-space electromagnetics the mediating field consists of **E** and **B**, while the source field is the distribution of charge or current. An important consideration is that the source field must be independent of the mediating field that it "sources." Additionally, fields are generally regarded as unobservable: they can only be measured indirectly through interactions with observable quantities. We need a link to mechanics to observe **E** and **B**: we might measure the change in kinetic energy of a particle as it interacts with the field through the Lorentz force. The Lorentz force becomes the force function in the mechanical interaction that uniquely determines the (observable) mechanical state of the particle.

A field is associated with a set of *field equations* and a set of *constitutive relations*. The field equations describe, through partial derivative operations, both the spatial distribution and temporal evolution of the field. The constitutive relations describe the effect

[1] Attempts have been made to formulate electromagnetic theory purely in action-at-a-distance terms, but this viewpoint has not been generally adopted [69].

© 2001 by CRC Press LLC

of the supporting medium on the fields and are dependent upon the physical state of the medium. The state may include macroscopic effects, such as mechanical stress and thermodynamic temperature, as well as the microscopic, quantum-mechanical properties of matter.

The value of the field at any position and time in a bounded region V is then determined uniquely by specifying the sources within V, the initial state of the fields within V, and the value of the field or finitely many of its derivatives on the surface bounding V. If the boundary surface also defines a surface of discontinuity between adjacent regions of differing physical characteristics, or across discontinuous sources, then *jump conditions* may be used to relate the fields on either side of the surface.

The variety of forms of field equations is restricted by many physical principles including reference-frame invariance, conservation, causality, symmetry, and simplicity. Causality prevents the field at time $t = 0$ from being influenced by events occurring at subsequent times $t > 0$. Of course, we prefer that a field equation be mathematically robust and well-posed to permit solutions that are unique and stable.

Many of these ideas are well illustrated by a consideration of electrostatics. We can describe the electrostatic field through a mediating scalar field $\Phi(x, y, z)$ known as the electrostatic potential. The spatial distribution of the field is governed by Poisson's equation

$$\frac{\partial^2 \Phi}{\partial x^2} + \frac{\partial^2 \Phi}{\partial y^2} + \frac{\partial^2 \Phi}{\partial z^2} = -\frac{\rho}{\epsilon_0}, \boldsymbol{\theta}$$

where $\rho = \rho(x, y, z)$ is the source charge density. No temporal derivatives appear, and the spatial derivatives determine the spatial behavior of the field. The function ρ represents the spatially-averaged distribution of charge that acts as the source term for the field Φ. Note that ρ incorporates no information about Φ. To uniquely specify the field at any point, we must still specify its behavior over a boundary surface. We could, for instance, specify Φ on five of the six faces of a cube and the normal derivative $\partial\Phi/\partial n$ on the remaining face. Finally, we cannot directly observe the static potential field, but we can observe its interaction with a particle. We relate the static potential field theory to the realm of mechanics via the electrostatic force $\mathbf{F} = q\mathbf{E}$ acting on a particle of charge q.

In future chapters we shall present a classical field theory for macroscopic electromagnetics. In that case the mediating field quantities are $\mathbf{E}$, $\mathbf{D}$, $\mathbf{B}$, and $\mathbf{H}$, and the source field is the current density $\mathbf{J}$.

1.3 The sources of the electromagnetic field

Electric charge is an intriguing natural entity. Human awareness of charge and its effects dates back to at least 600 BC, when the Greek philosopher Thales of Miletus observed that rubbing a piece of amber could enable the amber to attract bits of straw. Although *charging by friction* is probably still the most common and familiar manifestation of electric charge, systematic experimentation has revealed much more about the behavior of charge and its role in the physical universe. There are two kinds of charge, to which Benjamin Franklin assigned the respective names *positive* and *negative*. Franklin observed that charges of opposite kind attract and charges of the same kind repel. He also found that an increase in one kind of charge is accompanied by an increase in the

© 2001 by CRC Press LLC

other, and so first described the principle of *charge conservation*. Twentieth century physics has added dramatically to the understanding of charge:

1. Electric charge is a fundamental property of matter, as is mass or dimension.
2. Charge is *quantized*: there exists a smallest quantity (*quantum*) of charge that can be associated with matter. No smaller amount has been observed, and larger amounts always occur in integral multiples of this quantity.
3. The charge quantum is associated with the smallest subatomic particles, and these particles interact through electrical forces. In fact, matter is organized and arranged through electrical interactions; for example, our perception of physical contact is merely the macroscopic manifestation of countless charges in our fingertips pushing against charges in the things we touch.
4. Electric charge is an *invariant*: the value of charge on a particle does not depend on the speed of the particle. In contrast, the mass of a particle increases with speed.
5. Charge acts as the source of an electromagnetic field; the field is an entity that can carry energy and momentum away from the charge via propagating waves.

We begin our investigation of the properties of the electromagnetic field with a detailed examination of its source.

1.3.1 Macroscopic electromagnetics

We are interested primarily in those electromagnetic effects that can be predicted by classical techniques using continuous sources (charge and current densities). Although macroscopic electromagnetics is limited in scope, it is useful in many situations encountered by engineers. These include, for example, the determination of currents and voltages in lumped circuits, torques exerted by electrical machines, and fields radiated by antennas. Macroscopic predictions can fall short in cases where quantum effects are important: e.g., with devices such as tunnel diodes. Even so, quantum mechanics can often be coupled with classical electromagnetics to determine the macroscopic electromagnetic properties of important materials.

Electric charge is not of a continuous nature. The quantization of atomic charge — $\pm e$ for electrons and protons, $\pm e/3$ and $\pm 2e/3$ for quarks — is one of the most precisely established principles in physics (verified to 1 part in 10^{21}). The value of e itself is known to great accuracy:

$$e = 1.60217733 \times 10^{-19} \text{ Coulombs (C)}.$$

However, the discrete nature of charge is not easily incorporated into everyday engineering concerns. The strange world of the individual charge — characterized by particle spin, molecular moments, and thermal vibrations — is well described only by quantum theory. There is little hope that we can learn to describe electrical machines using such concepts. Must we therefore retreat to the macroscopic idea and ignore the discretization of charge completely? A viable alternative is to use atomic theories of matter to estimate the useful scope of macroscopic electromagnetics.

Remember, we are completely free to postulate a theory of nature whose scope may be limited. Like continuum mechanics, which treats distributions of matter as if they were continuous, macroscopic electromagnetics is regarded as valid because it is verified by experiment over a certain range of conditions. This applicability range generally corresponds to dimensions on a laboratory scale, implying a very wide range of validity for engineers.

© 2001 by CRC Press LLC

Macroscopic effects as averaged microscopic effects. Macroscopic electromagnetics can hold in a world of discrete charges because applications usually occur over physical scales that include vast numbers of charges. Common devices, generally much larger than individual particles, "average" the rapidly varying fields that exist in the spaces between charges, and this allows us to view a source as a continuous "smear" of charge. To determine the range of scales over which the macroscopic viewpoint is valid, we must compare averaged values of microscopic fields to the macroscopic fields we measure in the lab. But if the effects of the individual charges are describable only in terms of quantum notions, this task will be daunting at best. A simple compromise, which produces useful results, is to extend the macroscopic theory right down to the microscopic level and regard discrete charges as "point" entities that produce electromagnetic fields according to Maxwell's equations. Then, in terms of scales much larger than the classical radius of an electron ($\approx 10^{-14}$ m), the expected rapid fluctuations of the fields in the spaces between charges is predicted. Finally, we ask: over what spatial scale must we average the effects of the fields and the sources in order to obtain agreement with the macroscopic equations?

In the spatial averaging approach a convenient weighting function $f(\mathbf{r})$ is chosen, and is normalized so that $\int f(\mathbf{r})\, dV = 1$.

An example is the Gaussian distribution

$$f(\mathbf{r}) = (\pi a^2)^{-3/2} e^{-r^2/a^2},$$

where a is the approximate radial extent of averaging. The spatial average of a microscopic quantity $F(\mathbf{r}, t)$ is given by

$$\langle F(\mathbf{r}, t) \rangle = \int F(\mathbf{r} - \mathbf{r}', t)\, f(\mathbf{r}')\, dV'. \tag{1.1}$$

The scale of validity of the macroscopic model can be found by determining the averaging radius a that produces good agreement between the averaged microscopic fields and the macroscopic fields.

The macroscopic volume charge density. At this point we do not distinguish between the "free" charge that is unattached to a molecular structure and the charge found near the surface of a conductor. Nor do we consider the dipole nature of polarizable materials or the microscopic motion associated with molecular magnetic moment or the magnetic moment of free charge. For the consideration of free-space electromagnetics, we assume charge exhibits either three degrees of freedom (*volume charge*), two degrees of freedom (*surface charge*), or one degree of freedom (*line charge*).

In typical matter, the microscopic fields vary spatially over dimensions of 10^{-10} m or less, and temporally over periods (determined by atomic motion) of 10^{-13} s or less. At the surface of a material such as a good conductor where charge often concentrates, averaging with a radius on the order of 10^{-10} m may be required to resolve the rapid variation in the distribution of individual charged particles. However, within a solid or liquid material, or within a free-charge distribution characteristic of a dense gas or an electron beam, a radius of 10^{-8} m proves useful, containing typically 10^6 particles. A diffuse gas, on the other hand, may have a particle density so low that the averaging radius takes on laboratory dimensions, and in such a case the microscopic theory must be employed even at macroscopic dimensions.

Once the averaging radius has been determined, the value of the charge density may be found via (1.1). The volume density of charge for an assortment of point sources can

© 2001 by CRC Press LLC

be written in terms of the three-dimensional Dirac delta as

$$\rho^o(\mathbf{r}, t) = \sum_i q_i \delta(\mathbf{r} - \mathbf{r}_i(t)),$$

where $\mathbf{r}_i(t)$ is the position of the charge q_i at time t. Substitution into (1.1) gives

$$\rho(\mathbf{r}, t) = \langle \rho^o(\mathbf{r}, t) \rangle = \sum_i q_i f(\mathbf{r} - \mathbf{r}_i(t)) \tag{1.2}$$

as the averaged charge density appropriate for use in a macroscopic field theory. Because the oscillations of the atomic particles are statistically uncorrelated over the distances used in spatial averaging, the time variations of microscopic fields are not present in the macroscopic fields and temporal averaging is unnecessary. In (1.2) the time dependence of the spatially-averaged charge density is due entirely to bulk motion of the charge aggregate (macroscopic charge motion).

With the definition of macroscopic charge density given by (1.2), we can determine the total charge $Q(t)$ in any macroscopic volume region V using

$$Q(t) = \int_V \rho(\mathbf{r}, t)\, dV. \tag{1.3}$$

We have

$$Q(t) = \sum_i q_i \int_V f(\mathbf{r} - \mathbf{r}_i(t))\, dV = \sum_{\mathbf{r}_i(t) \in V} q_i.$$

Here we ignore the small discrepancy produced by charges lying within distance a of the boundary of V. It is common to employ a box B having volume ΔV:

$$\begin{cases} f(\mathbf{r}) = 1/\Delta V, & \mathbf{r} \in B, \\ 0, & \mathbf{r} \notin B. \end{cases}$$

In this case

$$\rho(\mathbf{r}, t) = \frac{1}{\Delta V} \sum_{\mathbf{r} - \mathbf{r}_i(t) \in B} q_i.$$

The size of B is chosen with the same considerations as to atomic scale as was the averaging radius a. Discontinuities at the edges of the box introduce some difficulties concerning charges that move in and out of the box because of molecular motion.

The macroscopic volume current density. Electric charge in motion is referred to as *electric current*. Charge motion can be associated with external forces and with microscopic fluctuations in position. Assuming charge q_i has velocity $\mathbf{v}_i(t) = d\mathbf{r}_i(t)/dt$, the charge aggregate has volume current density

$$\mathbf{J}^o(\mathbf{r}, t) = \sum_i q_i \mathbf{v}_i(t)\, \delta(\mathbf{r} - \mathbf{r}_i(t)).$$

Spatial averaging gives the macroscopic volume current density

$$\mathbf{J}(\mathbf{r}, t) = \langle \mathbf{J}^o(\mathbf{r}, t) \rangle = \sum_i q_i \mathbf{v}_i(t) f(\mathbf{r} - \mathbf{r}_i(t)). \tag{1.4}$$

© 2001 by CRC Press LLC

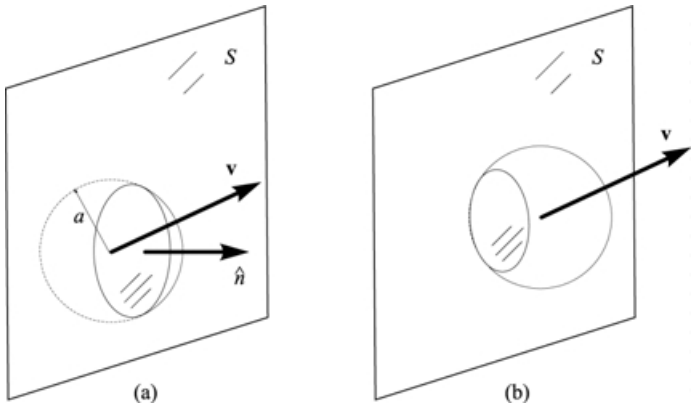

Figure 1.1: Intersection of the averaging function of a point charge with a surface S, as the charge crosses S with velocity $\mathbf{v}$: (a) at some time $t = t_1$, and (b) at $t = t_2 > t_1$. The averaging function is represented by a sphere of radius a.

Spatial averaging at time t eliminates currents associated with microscopic motions that are uncorrelated at the scale of the averaging radius (again, we do not consider the magnetic moments of particles). The assumption of a sufficiently large averaging radius leads to

$$\mathbf{J}(\mathbf{r}, t) = \rho(\mathbf{r}, t)\, \mathbf{v}(\mathbf{r}, t). \tag{1.5}$$

The total flux $I(t)$ of current through a surface S is given by

$$I(t) = \int_S \mathbf{J}(\mathbf{r}, t) \cdot \hat{\mathbf{n}}\, dS$$

where $\hat{\mathbf{n}}$ is the unit normal to S. Hence, using (4), we have

$$I(t) = \sum_i q_i \frac{d}{dt}(\mathbf{r}_i(t) \cdot \hat{\mathbf{n}}) \int_S f(\mathbf{r} - \mathbf{r}_i(t))\, dS$$

if $\hat{\mathbf{n}}$ stays approximately constant over the extent of the averaging function and S is not in motion. We see that the integral effectively intersects S with the averaging function surrounding each moving point charge (Figure 1.1). The time derivative of $\mathbf{r}_i \cdot \hat{\mathbf{n}}$ represents the velocity at which the averaging function is "carried across" the surface.

Electric current takes a variety of forms, each described by the relation $\mathbf{J} = \rho\mathbf{v}$. Isolated charged particles (positive and negative) and charged insulated bodies moving through space comprise *convection currents*. Negatively-charged electrons moving through the positive background lattice within a conductor comprise a *conduction current*. Empirical evidence suggests that conduction currents are also described by the relation $\mathbf{J} = \sigma\mathbf{E}$ known as *Ohm's law*. A third type of current, called *electrolytic current*, results from the flow of positive or negative ions through a fluid.

1.3.2 Impressed vs. secondary sources

In addition to the simple classification given above we may classify currents as *primary* or *secondary*, depending on the action that sets the charge in motion.

© 2001 by CRC Press LLC

It is helpful to separate primary or "impressed" sources, which are independent of the fields they source, from secondary sources which result from interactions between the sourced fields and the medium in which the fields exist. Most familiar is the conduction current set up in a conducting medium by an externally applied electric field. The impressed source concept is particularly important in circuit theory, where independent voltage sources are modeled as providing primary voltage excitations that are independent of applied load. In this way they differ from the secondary or "dependent" sources that react to the effect produced by the application of primary sources.

In applied electromagnetics the primary source may be so distant that return effects resulting from local interaction of its impressed fields can be ignored. Other examples of primary sources include the applied voltage at the input of an antenna, the current on a probe inserted into a waveguide, and the currents producing a power-line field in which a biological body is immersed.

1.3.3 Surface and line source densities

Because they are spatially averaged effects, macroscopic sources and the fields they source cannot have true spatial discontinuities. However, it is often convenient to work with sources in one or two dimensions. Surface and line source densities are idealizations of actual, continuous macroscopic densities.

The entity we describe as a *surface charge* is a continuous volume charge distributed in a thin layer across some surface S. If the thickness of the layer is small compared to laboratory dimensions, it is useful to assign to each point $\mathbf{r}$ on the surface a quantity describing the amount of charge contained within a cylinder oriented normal to the surface and having infinitesimal cross section dS. We call this quantity the *surface charge density* $\rho_s(\mathbf{r}, t)$, and write the volume charge density as

$$\rho(\mathbf{r}, w, t) = \rho_s(\mathbf{r}, t) f(w, \Delta),$$

where w is distance from S in the normal direction and Δ in some way parameterizes the "thickness" of the charge layer at $\mathbf{r}$. The continuous density function $f(x, \Delta)$ satisfies

$$\int_{-\infty}^{\infty} f(x, \Delta)\, dx = 1$$

and

$$\lim_{\Delta \to 0} f(x, \Delta) = \delta(x).$$

For instance, we might have

$$f(x, \Delta) = \frac{e^{-x^2/\Delta^2}}{\Delta\sqrt{\pi}}. \tag{1.6}$$

With this definition the total charge contained in a cylinder normal to the surface at $\mathbf{r}$ and having cross-sectional area dS is

$$dQ(t) = \int_{-\infty}^{\infty} [\rho_s(\mathbf{r}, t)\, dS]\, f(w, \Delta)\, dw = \rho_s(\mathbf{r}, t)\, dS,$$

and the total charge contained within any cylinder oriented normal to S is

$$Q(t) = \int_S \rho_s(\mathbf{r}, t)\, dS. \tag{1.7}$$

© 2001 by CRC Press LLC

We may describe a line charge as a thin "tube" of volume charge distributed along some contour Γ. The amount of charge contained between two planes normal to the contour and separated by a distance dl is described by the *line charge density* $\rho_l(\mathbf{r}, t)$. The volume charge density associated with the contour is then

$$\rho(\mathbf{r}, \rho, t) = \rho_l(\mathbf{r}, t) f_s(\rho, \Delta),$$

where ρ is the radial distance from the contour in the plane normal to Γ and $f_s(\rho, \Delta)$ is a density function with the properties

$$\int_0^\infty f_s(\rho, \Delta) 2\pi\rho \, d\rho = 1$$

and

$$\lim_{\Delta \to 0} f_s(\rho, \Delta) = \frac{\delta(\rho)}{2\pi\rho}.$$

For example, we might have

$$f_s(\rho, \Delta) = \frac{e^{-\rho^2/\Delta^2}}{\pi \Delta^2}. \tag{1.8}$$

Then the total charge contained between planes separated by a distance dl is

$$dQ(t) = \int_0^\infty [\rho_l(\mathbf{r}, t)\, dl]\, f_s(\rho, \Delta) 2\pi\rho \, d\rho = \rho_l(\mathbf{r}, t)\, dl$$

and the total charge contained between planes placed at the ends of a contour Γ is

$$Q(t) = \int_\Gamma \rho_l(\mathbf{r}, t)\, dl. \tag{1.9}$$

We may define surface and line currents similarly. A surface current is merely a volume current confined to the vicinity of a surface S. The volume current density may be represented using a surface current density function $\mathbf{J}_s(\mathbf{r}, t)$, defined at each point $\mathbf{r}$ on the surface so that

$$\mathbf{J}(\mathbf{r}, w, t) = \mathbf{J}_s(\mathbf{r}, t) f(w, \Delta).$$

Here $f(w, \Delta)$ is some appropriate density function such as (1.6), and the surface current vector obeys $\hat{\mathbf{n}} \cdot \mathbf{J}_s = 0$ where $\hat{\mathbf{n}}$ is normal to S. The total current flowing through a strip of width dl arranged perpendicular to S at $\mathbf{r}$ is

$$dI(t) = \int_{-\infty}^\infty [\mathbf{J}_s(\mathbf{r}, t) \cdot \hat{\mathbf{n}}_l(\mathbf{r})\, dl]\, f(w, \Delta)\, dw = \mathbf{J}_s(\mathbf{r}, t) \cdot \hat{\mathbf{n}}_l(\mathbf{r})\, dl$$

where $\hat{\mathbf{n}}_l$ is normal to the strip at $\mathbf{r}$ (and thus also tangential to S at $\mathbf{r}$). The total current passing through a strip intersecting with S along a contour Γ is thus

$$I(t) = \int_\Gamma \mathbf{J}_s(\mathbf{r}, t) \cdot \hat{\mathbf{n}}_l(\mathbf{r})\, dl.$$

We may describe a line current as a thin "tube" of volume current distributed about some contour Γ and flowing parallel to it. The amount of current passing through a plane normal to the contour is described by the *line current density* $J_l(\mathbf{r}, t)$. The volume current density associated with the contour may be written as

$$\mathbf{J}(\mathbf{r}, \rho, t) = \hat{\mathbf{u}}(\mathbf{r}) J_l(\mathbf{r}, t) f_s(\rho, \Delta),$$

© 2001 by CRC Press LLC

where $\hat{\mathbf{u}}$ is a unit vector along Γ, ρ is the radial distance from the contour in the plane normal to Γ, and $f_s(\rho, \Delta)$ is a density function such as (1.8). The total current passing through any plane normal to Γ at $\mathbf{r}$ is

$$I(t) = \int_0^\infty [J_l(\mathbf{r}, t)\hat{\mathbf{u}}(\mathbf{r}) \cdot \hat{\mathbf{u}}(\mathbf{r})]\, f_s(\rho, \Delta) 2\pi\rho\, d\rho = J_l(\mathbf{r}, t).$$

It is often convenient to employ *singular* models for continuous source densities. For instance, it is mathematically simpler to regard a surface charge as residing only in the surface S than to regard it as being distributed about the surface. Of course, the source is then discontinuous since it is zero everywhere outside the surface. We may obtain a representation of such a charge distribution by letting the thickness parameter Δ in the density functions recede to zero, thus concentrating the source into a plane or a line. We describe the limit of the density function in terms of the δ-function. For instance, the volume charge distribution for a surface charge located about the xy-plane is

$$\rho(x, y, z, t) = \rho_s(x, y, t) f(z, \Delta).$$

As $\Delta \to 0$ we have

$$\rho(x, y, z, t) = \rho_s(x, y, t) \lim_{\Delta \to 0} f(z, \Delta) = \rho_s(x, y, t)\delta(z).$$

It is a simple matter to represent singular source densities in this way as long as the surface or line is easily parameterized in terms of constant values of coordinate variables. However, care must be taken to represent the δ-function properly. For instance, the density of charge on the surface of a cone at $\theta = \theta_0$ may be described using the distance normal to this surface, which is given by $r\theta - r\theta_0$:

$$\rho(r, \theta, \phi, t) = \rho_s(r, \phi, t)\delta\left(r[\theta - \theta_0]\right).$$

Using the property $\delta(ax) = \delta(x)/a$, we can also write this as

$$\rho(r, \theta, \phi, t) = \rho_s(r, \phi, t)\frac{\delta(\theta - \theta_0)}{r}.$$

1.3.4 Charge conservation

There are four fundamental conservation laws in physics: conservation of energy, momentum, angular momentum, and charge. These laws are said to be *absolute*; they have never been observed to fail. In that sense they are true empirical laws of physics.

However, in modern physics the fundamental conservation laws have come to represent more than just observed facts. Each law is now associated with a fundamental symmetry of the universe; conversely, each known symmetry is associated with a conservation principle. For example, energy conservation can be shown to arise from the observation that the universe is symmetric with respect to time; the laws of physics do not depend on choice of time origin $t = 0$. Similarly, momentum conservation arises from the observation that the laws of physics are invariant under translation, while angular momentum conservation arises from invariance under rotation.

The law of conservation of charge also arises from a symmetry principle. But instead of being spatial or temporal in character, it is related to the invariance of electrostatic potential. Experiments show that there is no absolute potential, only potential difference. The laws of nature are invariant with respect to what we choose as the "reference"

© 2001 by CRC Press LLC

potential. This in turn is related to the invariance of Maxwell's equations under gauge transforms; the values of the electric and magnetic fields do not depend on which gauge transformation we use to relate the scalar potential Φ to the vector potential $\mathbf{A}$.

We may state the conservation of charge as follows:

The net charge in any closed system remains constant with time.

This does not mean that individual charges cannot be created or destroyed, only that the total charge in any isolated system must remain constant. Thus it is possible for a positron with charge e to annihilate an electron with charge $-e$ without changing the net charge of the system. Only if a system is not closed can its net charge be altered; since moving charge constitutes current, we can say that the total charge within a system depends on the current passing through the surface enclosing the system. This is the essence of the continuity equation. To derive this important result we consider a closed system within which the charge remains constant, and apply the Reynolds transport theorem (see § A.2).

The continuity equation. Consider a region of space occupied by a distribution of charge whose velocity is given by the vector field $\mathbf{v}$. We surround a portion of charge by a surface S and let S deform as necessary to "follow" the charge as it moves. Since S always contains precisely the same charged particles, we have an isolated system for which the time rate of change of total charge must vanish. An expression for the time rate of change is given by the Reynolds transport theorem (A.66); we have[2]

$$\frac{DQ}{Dt} = \frac{D}{Dt}\int_{V(t)} \rho \, dV = \int_{V(t)} \frac{\partial \rho}{\partial t} \, dV + \oint_{S(t)} \rho \mathbf{v} \cdot \mathbf{dS} = 0.$$

The "D/Dt" notation indicates that the volume region $V(t)$ moves with its enclosed particles. Since $\rho\mathbf{v}$ represents current density, we can write

$$\int_{V(t)} \frac{\partial \rho(\mathbf{r}, t)}{\partial t} \, dV + \oint_{S(t)} \mathbf{J}(\mathbf{r}, t) \cdot \mathbf{dS} = 0. \tag{1.10}$$

In this *large-scale form* of the continuity equation, the partial derivative term describes the time rate of change of the charge density for a fixed spatial position $\mathbf{r}$. At any time t, the time rate of change of charge density integrated over a volume is exactly compensated by the total current exiting through the surrounding surface.

We can obtain the continuity equation in *point form* by applying the divergence theorem to the second term of (1.10) to get

$$\int_{V(t)} \left[\frac{\partial \rho(\mathbf{r}, t)}{\partial t} + \nabla \cdot \mathbf{J}(\mathbf{r}, t) \right] dV = 0.$$

Since $V(t)$ is arbitrary we can set the integrand to zero to obtain

$$\frac{\partial \rho(\mathbf{r}, t)}{\partial t} + \nabla \cdot \mathbf{J}(\mathbf{r}, t) = 0. \tag{1.11}$$

[2]Note that in Appendix A we use the symbol $\mathbf{u}$ to represent the velocity of a material and $\mathbf{v}$ to represent the velocity of an artificial surface.

© 2001 by CRC Press LLC

This expression involves the time derivative of ρ with $\mathbf{r}$ fixed. We can also find an expression in terms of the material derivative by using the transport equation (A.67). Enforcing conservation of charge by setting that expression to zero, we have

$$\frac{D\rho(\mathbf{r},t)}{Dt} + \rho(\mathbf{r},t)\,\nabla\cdot\mathbf{v}(\mathbf{r},t) = 0. \tag{1.12}$$

Here $D\rho/Dt$ is the time rate of change of the charge density experienced by an observer moving with the current.

We can state the large-scale form of the continuity equation in terms of a stationary volume. Integrating (1.11) over a stationary volume region V and using the divergence theorem, we find that

$$\int_V \frac{\partial\rho(\mathbf{r},t)}{\partial t}\,dV = -\oint_S \mathbf{J}(\mathbf{r},t)\cdot\mathbf{dS}.$$

Since V is not changing with time we have

$$\frac{dQ(t)}{dt} = \frac{d}{dt}\int_V \rho(\mathbf{r},t)\,dV = -\oint_S \mathbf{J}(\mathbf{r},t)\cdot\mathbf{dS}. \tag{1.13}$$

Hence any increase of total charge within V must be produced by current entering V through S.

Use of the continuity equation. As an example, suppose that in a bounded region of space we have

$$\rho(\mathbf{r},t) = \rho_0 r e^{-\beta t}.$$

We wish to find $\mathbf{J}$ and $\mathbf{v}$, and to verify both versions of the continuity equation in point form. The spherical symmetry of ρ requires that $\mathbf{J} = \hat{\mathbf{r}}J_r$. Application of (1.13) over a sphere of radius a gives

$$4\pi\frac{d}{dt}\int_0^a \rho_0 r e^{-\beta t} r^2\,dr = -4\pi J_r(a)a^2.$$

Hence

$$\mathbf{J} = \hat{\mathbf{r}}\beta\rho_0\frac{r^2}{4}e^{-\beta t}$$

and therefore

$$\nabla\cdot\mathbf{J} = \frac{1}{r^2}\frac{\partial}{\partial r}(r^2 J_r) = \beta\rho_0 r e^{-\beta t}.$$

The velocity is

$$\mathbf{v} = \frac{\mathbf{J}}{\rho} = \hat{\mathbf{r}}\beta\frac{r}{4},$$

and we have $\nabla\cdot\mathbf{v} = 3\beta/4$. To verify the continuity equations, we compute the time derivatives

$$\frac{\partial\rho}{\partial t} = -\beta\rho_0 r e^{-\beta t},$$

$$\begin{aligned}\frac{D\rho}{Dt} &= \frac{\partial\rho}{\partial t} + \mathbf{v}\cdot\nabla\rho \\ &= -\beta\rho_0 r e^{-\beta t} + \left(\hat{\mathbf{r}}\beta\frac{r}{4}\right)\cdot\left(\hat{\mathbf{r}}\rho_0 e^{-\beta t}\right) \\ &= -\frac{3}{4}\beta\rho_0 r e^{-\beta t}.\end{aligned}$$

© 2001 by CRC Press LLC

Note that the charge density decreases with time less rapidly for a moving observer than for a stationary one (3/4 as fast): the moving observer is following the charge outward, and $\rho \propto r$. Now we can check the continuity equations. First we see

$$\frac{D\rho}{Dt} + \rho \nabla \cdot \mathbf{v} = -\frac{3}{4}\beta\rho_0 r e^{-\beta t} + (\rho_0 r e^{-\beta t})\left(\frac{3}{4}\beta\right) = 0,$$

as required for a moving observer; second we see

$$\frac{\partial \rho}{\partial t} + \nabla \cdot \mathbf{J} = -\beta\rho_0 r e^{-\beta t} + \beta\rho_0 e^{-\beta t} = 0,$$

as required for a stationary observer.

The continuity equation in fewer dimensions. The continuity equation can also be used to relate current and charge on a surface or along a line. By conservation of charge we can write

$$\frac{d}{dt}\int_S \rho_s(\mathbf{r}, t)\, dS = -\oint_\Gamma \mathbf{J}_s(\mathbf{r}, t) \cdot \hat{\mathbf{m}}\, dl \tag{1.14}$$

where $\hat{\mathbf{m}}$ is the vector normal to the curve Γ and tangential to the surface S. By the surface divergence theorem (B.20), the corresponding point form is

$$\frac{\partial \rho_s(\mathbf{r}, t)}{\partial t} + \nabla_s \cdot \mathbf{J}_s(\mathbf{r}, t) = 0. \tag{1.15}$$

Here $\nabla_s \cdot \mathbf{J}_s$ is the surface divergence of the vector field $\mathbf{J}_s$. For instance, in rectangular coordinates in the $z = 0$ plane we have

$$\nabla_s \cdot \mathbf{J}_s = \frac{\partial J_{sx}}{\partial x} + \frac{\partial J_{sy}}{\partial y}.$$

In cylindrical coordinates on the cylinder $\rho = a$, we would have

$$\nabla_s \cdot \mathbf{J}_s = \frac{1}{a}\frac{\partial J_{s\phi}}{\partial \phi} + \frac{\partial J_{sz}}{\partial z}.$$

A detailed description of vector operations on a surface may be found in Tai [190], while many identities may be found in Van Bladel [202].

The equation of continuity for a line is easily established by reference to Figure 1.2. Here the net charge exiting the surface during time Δt is given by

$$\Delta t[I(u_2, t) - I(u_1, t)].$$

Thus, the rate of net increase of charge within the system is

$$\frac{dQ(t)}{dt} = \frac{d}{dt}\int \rho_l(\mathbf{r}, t)\, dl = -[I(u_2, t) - I(u_1, t)]. \tag{1.16}$$

The corresponding point form is found by letting the length of the curve approach zero:

$$\frac{\partial I(l, t)}{\partial l} + \frac{\partial \rho_l(l, t)}{\partial t} = 0, \tag{1.17}$$

where l is arc length along the curve. As an example, suppose the line current on a circular loop antenna is approximately

$$I(\phi, t) = I_0 \cos\left(\frac{\omega a}{c}\phi\right)\cos \omega t,$$

© 2001 by CRC Press LLC

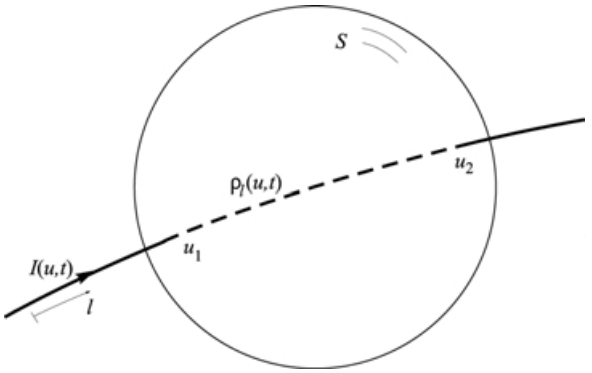

Figure 1.2: Linear form of the continuityequation.

where a is the radius of the loop, ω is the frequency of operation, and c is the speed of light. We wish to find the line charge density on the loop. Since $l = a\phi$, we can write

$$I(l,t) = I_0 \cos\left(\frac{\omega l}{c}\right)\cos\omega t.$$

Thus

$$\frac{\partial I(l,t)}{\partial l} = -I_0\frac{\omega}{c}\sin\left(\frac{\omega l}{c}\right)\cos\omega t = -\frac{\partial \rho_l(l,t)}{\partial t}.$$

Integrating with respect to time and ignoring any constant (static) charge, we have

$$\rho(l,t) = \frac{I_0}{c}\sin\left(\frac{\omega l}{c}\right)\sin\omega t$$

or

$$\rho(\phi,t) = \frac{I_0}{c}\sin\left(\frac{\omega a}{c}\phi\right)\sin\omega t.$$

Note that we could have used the chain rule

$$\frac{\partial I(\phi,t)}{\partial l} = \frac{\partial I(\phi,t)}{\partial \phi}\frac{\partial \phi}{\partial l} \quad \text{and} \quad \frac{\partial \phi}{\partial l} = \left[\frac{\partial l}{\partial \phi}\right]^{-1} = \frac{1}{a}$$

to calculate the spatial derivative.

We can apply the volume density continuity equation (1.11) directly to surface and line distributions written in singular notation. For the loop of the previous example, we write the volume current density corresponding to the line current as

$$\mathbf{J}(\mathbf{r},t) = \hat{\phi}\,\delta(\rho - a)\delta(z)I(\phi,t).$$

Substitution into (1.11) then gives

$$\nabla\cdot[\hat{\phi}\delta(\rho-a)\delta(z)I(\phi,t)] = -\frac{\partial\rho(\mathbf{r},t)}{\partial t}.$$

The divergence formula for cylindrical coordinates gives

$$\delta(\rho-a)\delta(z)\frac{\partial I(\phi,t)}{\rho\,\partial\phi} = -\frac{\partial\rho(\mathbf{r},t)}{\partial t}.$$

© 2001 by CRC Press LLC

Next we substitute for $I(\phi, t)$ to get

$$-\frac{I_0}{\rho}\frac{\omega a}{c}\sin\left(\frac{\omega a}{c}\phi\right)\delta(\rho - a)\delta(z)\cos\omega t = -\frac{\partial\rho(\mathbf{r}, t)}{\partial t}.$$

Finally, integrating with respect to time and ignoring any constant term, we have

$$\rho(\mathbf{r}, t) = \frac{I_0}{c}\delta(\rho - a)\delta(z)\sin\left(\frac{\omega a}{c}\phi\right)\sin\omega t,$$

where we have set $\rho = a$ because of the presence of the factor $\delta(\rho - a)$.

1.3.5 Magnetic charge

We take for granted that electric fields are produced by electric charges, whether stationary or in motion. The smallest element of electric charge is the electric *monopole*: a single discretely charged particle from which the electric field diverges. In contrast, experiments show that magnetic fields are created only by currents or by time changing electric fields; hence, magnetic fields have moving electric charge as their source. The elemental source of magnetic field is the magnetic *dipole*, representing a tiny loop of electric current (or a spinning electric particle). The observation made in 1269 by Pierre De Maricourt, that even the smallest magnet has two poles, still holds today.

In a world filled with symmetry at the fundamental level, we find it hard to understand why there should not be a source from which the magnetic field diverges. We would call such a source *magnetic charge*, and the most fundamental quantity of magnetic charge would be exhibited by a *magnetic monopole*. In 1931 Paul Dirac invigorated the search for magnetic monopoles by making the first strong theoretical argument for their existence. Dirac showed that the existence of magnetic monopoles would imply the quantization of electric charge, and would thus provide an explanation for one of the great puzzles of science. Since that time magnetic monopoles have become important players in the "Grand Unified Theories" of modern physics, and in cosmological theories of the origin of the universe.

If magnetic monopoles are ever found to exist, there will be both positive and negatively charged particles whose motions will constitute currents. We can define a macroscopic magnetic charge density ρ_m and current density $\mathbf{J}_m$ exactly as we did with electric charge, and use conservation of magnetic charge to provide a continuity equation:

$$\nabla\cdot\mathbf{J}_m(\mathbf{r}, t) + \frac{\partial\rho_m(\mathbf{r}, t)}{\partial t} = 0. \tag{1.18}$$

With these new sources Maxwell's equations become appealingly symmetric. Despite uncertainties about the existence and physical nature of magnetic monopoles, magnetic charge and current have become an integral part of electromagnetic theory. We often use the concept of *fictitious* magnetic sources to make Maxwell's equations symmetric, and then derive various equivalence theorems for use in the solution of important problems. Thus we can put the idea of magnetic sources to use regardless of whether these sources actually exist.

© 2001 by CRC Press LLC

1.4 Problems

1.1 Write the volume charge density for a singular surface charge located on the sphere $r = r_0$, entirely in terms of spherical coordinates. Find the total charge on the sphere.

1.2 Repeat Problem 1.1 for a charged half plane $\phi = \phi_0$.

1.3 Write the volume charge density for a singular surface charge located on the cylinder $\rho = \rho_0$, entirely in terms of cylindrical coordinates. Find the total charge on the cylinder.

1.4 Repeat Problem 1.3 for a charged half plane $\phi = \phi_0$.

© 2001 by CRC Press LLC