Chapter I. Basics (12 Sep 2020)

A. Maxwell's Equations: macroscopic version.

The traditional development of graduate-level E&M begins with statics and leaves dynamics to the second semester. However, by now everyone has already seen Maxwell's Equations in undergrad E&M. Therefore, it makes sense to begin as Jackson does, by reviewing these equations, for the moment accepting them as-is and leaving their experimental foundations in the Coulomb and Ampère force laws, charge conservation, and Faraday's induction measurements to Ch. 2.

Here, we use these equations to develop easily derived quantities such as potentials, boundary conditions, and energy densities. This provides opportunities to review vector calculus, Gauss' and Stokes' Theorems, and the structure of conservation equations in general. Underlying mathematical details are given in Appendix 1. These extensions also provide excellent opportunities to extract the physics from the math, although in most cases full details wait for later chapters.

In the two sets of units in common use, the macroscopic Maxwell Equations are:

cgs:	SI:	
$\nabla \cdot \vec{D} = 4\pi\rho;$	$\nabla \cdot \vec{D} = \rho;$	(1.1a,b)
$\nabla \cdot \vec{B} = 0;$	$\nabla \cdot \vec{B} = 0;$	(1.1c,d)
$\nabla \times \vec{E} + \frac{1}{c}\frac{\partial \vec{B}}{\partial t} = 0;$	$\nabla \times \vec{E} + \frac{\partial \vec{B}}{\partial t} = 0;$	(1.1e,f)
$\nabla \times \vec{H} = \frac{4\pi}{c}\vec{J} + \frac{1}{c}\frac{\partial \vec{D}}{\partial t};$	$\nabla \times \vec{H} = \vec{J} + \frac{\partial \vec{D}}{\partial t};$	(1.1g,h)

where $\vec{E}$, $\vec{D}$, $\vec{B}$, and $\vec{H}$ are the electric field, displacement field, magnetic flux density, and magnetic field intensity, respectively; ρ is the charge density; and $\vec{J} = \rho\vec{v}$ is the current density, where $\vec{v}$ is the velocity of ρ in the "lab" or reference frame. The first set of equations is written in cgs (centimeter-gram-second) units, and the second in SI (System Internationale, formerly rmks or rationalized meter-kilogram-second) units.

These equations are supplemented by a set of phenomenological equations called the constitutive relations:

cgs:	SI:	
$\vec{D} = \varepsilon \vec{E} = \vec{E} + 4\pi \vec{P}$	$\vec{D} = \varepsilon_r \varepsilon_o \vec{E} = \vec{E} + \vec{P}$	(1.2a,b)
$\vec{B} = \mu \vec{H} = \vec{H} + 4\pi \vec{M}$	$\vec{B} = \mu_r \mu_o \vec{H} = \vec{H} + \vec{M}$	(1.2c,d)

where in the SI system ε_o and μ_o are the permittivity and permeability, respectively, of free space, and ε_r and μ_r are dimensionless relative values. In the SI system the use of ε_o and μ_o is mandatory, but in the cgs system ε_o and μ_o do not appear. The systems connect via $\varepsilon_{cgs} = (\varepsilon_r)_{SI}$ and $\mu_{cgs} = (\mu_r)_{SI}$. In materials, $\vec{P}$ and $\vec{M}$ are the electric and magnetic dipole densities, respectively, which may be either induced or built-in. If $\vec{J}$ is a response instead of a source term, as for example in a conductor, then an additional constitutive relation is

$$\vec{J} = \sigma \vec{E}\,; \tag{1.2e}$$

where σ is the conductivity. Clearly, ε, μ, and σ encode the properties of a material. These are phenomenological parameters in the above equations, but they are developed from atomic-scale processes in Chs. 7-9.

Why two sets of units? This question is addressed more fully in Sec. C. For now, we note that the cgs system spotlights the physics, while the SI system is essential for practical calculations. For example, a code-limited household-circuit current of 15 A in SI units is a current of 4.5×10^{10} statamperes in cgs units. The more practical alternative is obvious. Since most of you already speak at least two languages, there is really no excuse for you not to become fluent in two sets of units. In any case I list the factors that allow you to mindlessly convert any cgs E&M equation to its SI equivalent (and vice versa) in Appendix 2.

To complete the equations review, add the Lorentz force law and charge conservation. The Lorentz-force law is not fundamental. It is derived from Coulomb's Law and Faraday's Law of Induction, as we show in Ch. 2. Nevertheless, it is a useful description of the force $\vec{F}_q$ acting on a charge q moving with a velocity $\vec{v}$ in a region containing both electric and magnetic fields. In the two sets of units

cgs:	SI:	
$\vec{F}_q = q\vec{E} + \frac{q}{c}\vec{v}\times\vec{B}\,;$	$\vec{F}_q = q\vec{E} + q\vec{v}\times\vec{B}\,.$	(1.3a,b)

These equations also show that the fundamental fields interacting with materials are $\vec{E}$ and $\vec{B}$, not $\vec{D}$ and $\vec{H}$. The final equation in our collection of fundamental relations is the charge-conservation relation,

$$\nabla \cdot \vec{J} + \frac{\partial \rho}{\partial t} = 0 . \qquad (1.4)$$

This is the same in both systems of units.

In addition to summarizing electrodynamics in four equations, Maxwell added the second term on the right-hand side of Eq. (1.1g) to make it consistent with charge conservation. That this term is necessary is easily verified by taking the divergence of Eq. (1.1g) and using Eqs. (1.1a) and (1.4). It is almost certain that Ampère realized that this term is already contained in his force law, so we can only speculate why the opportunity to include it was left to Maxwell. We return to this in Ch. 2.

Regarding special relativity, Maxwell's Equations are covariant, meaning that they are the same in any inertial system (no accelerations). This is supported by all evidence to date. The surprising but incontrovertible conclusion is that *c* must therefore be a universal constant independent of the motion of either source or observer. It is worth pausing a moment to let this sink in. We can be grateful for this, because if *c* were not constant, general solutions of Maxwell's Equations would be too difficult to contemplate. The most stringent test of the invariance of *c* to date is the LIGO experiment. In addition to its primary role of detecting gravitational waves, it can also be considered an enormously more sensitive version of the Michelson-Morley experiment. The complete absence of any daily or seasonal effect now places *c* to be constant to approximately parts in 10^{19}.

B. Units.

A discussion of the relative merits of the cgs and SI systems of units in different contexts does not usually rate a separate section in an E&M text. However, this proved to be necessary when I found that incoming students, who to a person were well skilled in the SI system, could not answer the elementary entrance-exam question: what happens to Maxwell's Equations when $c \to \infty$? The answer is immediately evident in the cgs version of Eqs. (1.1): magnetic effects disappear completely, and all that is left is electrostatics. This simply recognizes that for $c \to \infty$, information about any change in a charge distribution $\rho(\vec{r})$ is transmitted instantly everywhere in the universe, so the wave equation, which requires the existence of a magnetic field, is irrelevant. In contrast, *no* change occurs in the SI equations, explaining why confusion about the role of *c* arises.

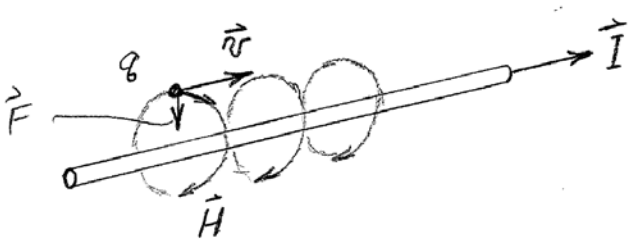
Yet for the physics to be consistent, in the $c \to \infty$ limit the SI equations must also reduce to electrostatics. This reduction occurs through the constitutive relations, not Maxwell's Equations themselves. The key to understanding the situation starts with the relation $c = 1/\sqrt{\mu_o \varepsilon_o}$, although this is initially of little help because it is not immediately obvious whether $\varepsilon_o \to 0$, $\mu_o \to 0$, or both. The answer is $\mu_o \to 0$, which follows from Eq. (1.2b). If $\varepsilon_o \to 0$ then $\vec{E}$ would be infinite if $\vec{D}$ were finite. Consequently, in the SI system μ_o must approach zero, so the effect of $c \to \infty$ enters through the constitutive relation $\vec{B} = \mu_r \mu_o \vec{H}$. With $\vec{B} = 0$, Eq. (1.1c) is satisfied trivially, and Eq. (1.1e) reduces to electrostatics.

Curiously, in the SI system $\vec{H}$ remains finite. Given that $\vec{H}_{cgs} = \sqrt{4\pi\mu_o}\vec{H}_{SI}$, a finite $\vec{H}_{SI}$, even though awkward from a physics perspective, is not inconsistent with its cgs counterpart. A finite value of $\vec{H}_{SI}$ causes no contradictions in principle, because the role of $\vec{H}_{SI}$ is to align magnetic dipoles. If magnetism is gone, magnetic dipoles are also gone, so a finite $\vec{H}_{SI}$ is more of a mathematical curiosity than a physical contradiction.

A better reason for using the cgs system is that inverse powers of *c* are a convenient way of organizing physical phenomena, and in so doing, avoiding errors due to overlooked terms. The classic example is the contribution of the scalar potential to radiation. As mentioned above, electrostatics is independent of *c*, which can be described as phenomena of order $c^0 = 1$. At the next order, magnetism, radiation, scattering, and diffraction all depend on *c* as c^{-1}. Special relativity, where frames in relevant motion no longer satisfy the Galilean requirement $t = t'$ and lengths contract in the direction of motion, involve *c* as c^{-2} (think $\gamma = 1/\sqrt{1 - v^2/c^2}$). The cgs system exhibits its dependences on *c* explicitly.

Unless special relativity is brought in specifically, classical electrodynamics treats electrodynamic phenomena to orders c^0 and c^{-1}. The associated phenomena can be described as statics and retardation physics, respectively (although terms $\sim c^{-2}$, appear, these are always combinations of two contributions $\sim c^{-1}$.) By ignoring this hierarchy, standard treatments of radiation, scattering, and diffraction (Jackson included) make the common error of failing to recognize the c^{-1} contribution of the scalar potential to these phenomena (more on this in Ch. 15). However, by limiting analyses to order c^{-1} we limit ourselves to nonrelativistic phenomena. We cannot, for example, describe the forward projection of radiation emitted from charges moving at relativistic velocities in synchrotrons, which is a consequence of time passing much more slowly in the rest frame of the charge relative to that in the lab frame. Here, a full relativistic treatment is required. This is done in Ch. 17.

Further physics-based justification of the cgs system follows from a thought experiment. Consider the force on a point charge *q* moving with a velocity $\vec{v}$ parallel to an electrically neutral wire carrying a current *I*, as shown in the diagram. By the Biot-Savart law, the current generates a magnetic field that, with the directions shown, deflects q toward the wire according to



$$\vec{F} = q\vec{E} + q\vec{v} \times \vec{B}, \tag{1.5}$$

in SI units. Because the wire is neutral in the lab frame, $\vec{E} = 0$.

Now consider the diagram from the perspective of an observer moving with *q*. In this frame $\vec{v} = 0$. However, the force on q obviously cannot depend on the motion of the observer. Thus in the moving frame of reference, the loss of $q\vec{v} \times \vec{B}$ must be compensated by the appearance of an electric field. In this example, in the moving system the uncharged wire appears to be charged! That $\vec{B}$ can turn into $\vec{E}$ (and *vice versa*) suggests that at some level $\vec{E}$ and $\vec{B}$ are different aspects of the same entity. From special relativity we know that this is the second-rank field tensor. Therefore, it is reasonable to express $\vec{E}$ and $\vec{B}$ in the same units. In the cgs system this is achieved by dividing $\vec{v}$ by the universal constant *c*. This is justified quantitatively in special relativity. Brau shows for example that all of Maxwell's Equations follow from Coulomb's Law and special relativity.

We can extract additional information from the result. Because typical laboratory speeds are much less than *c*, under most conditions we can expect magnetic forces to be significantly weaker than electric forces. That, and the fact that electrostatics is independent of *c*, are two of the reasons that courses in classical electrodynamics begin with electrostatics. In practical applications, motors make up for this weakness by moving enormous numbers of electrons simultaneously. But it is worth noting that if *c* were twice as large in our universe, for a given current, motors would generate only half as much torque.

The fundamental constants on which these units depend are an interesting topic in their own right. A more thorough discussion is given in Appendix 3, but some comments here are appropriate. Since 1983, *c* is *defined* to have the value $c = 2.99792458 \times 10^8$ m/s. Up until 20 May 2019 μ_o was *defined* to have the value $\mu_o = 4\pi \times 10^{-7}$ (weber)/(A m), leaving $\varepsilon_o = 1/\mu_o c^2$ also as a defined quantity. On 20 May 2019, the 144th anniversary of the first international meeting held to standardize measurements, additional recommendations of the Comité International des Poids et Mesures (CIPM) (English translation: International Committee for Weights and Measures) were adopted and also went into effect. These include elevating the elementary charge *e*, Planck's constant *h*, Boltzmann's constant *k*, and the Avogadro number N_A to defined status. As part of the fallout, ε_o and μ_o are now both derived quantities, uncertain to about 2 parts in 10^{10}.

The apparently bizarre situation where physical constants are defined rather than measured can be understood by considering *c*. At present (2020) we can measure time to a relative accuracy of about 3.5 parts in 10^{19}. This is many orders of magnitude better than we can measure length. Hence rather than to attempt to determine *c* using a (relatively) inaccurate standard of length, we *define c* and measure length in terms of *c* and time. As our understanding of physics and measurement technology continue to improve, the CIPM can be expected to make further changes.

Returning to the constitutive equations, in Chs. 7 and 9 we will derive ε and μ, respectively, from models that illustrate the physics leading to these quantities when fields are applied to a material. Briefly, we will find that ε, and by extension optical properties, can be obtained by applying Newton's Second Law, $\vec{F} = m\vec{a}$, to the atomic-

scale charges in a material. Thus for electrostatics, $\vec{E}$ provides the drive and $\vec{D}$ is the response. $\vec{D}$ includes the mechanical strain induced by $\vec{E}$. This is consistent with the constitutive relation, which shows that $\vec{E}$ is the fundamental field.

However, the constitutive-relation reasoning does not work with magnetism, where the same logic would lead to the incorrect conclusion that $\vec{H}$ is the fundamental field. Different physics is involved in the two cases. Because there is no magnetic charge, μ in a material is a consequence of aligning existing dipoles. The drive for doing this is provided by $\vec{H}$. The response is $\vec{B}$, which also measures the increase in free energy resulting from the reduction of disorder.

The practical consequences of the different mechanisms behind μ and ε is significant. Dielectric relaxation times are of the order of reciprocal phonon frequencies, or $\tau \sim 10^{-13}$ s. In the magnetic case, relaxation is entropy-driven, and is about 10 orders of magnitude slower. Consequently, we will not see magnetic integrated circuits any time soon. Nevertheless, we can achieve relatively rapid magnetic responses by constructing artificial structures called metamaterials that simulate magnetic behavior by providing for storage of magnetic energy through resonances. Interest has recently arisen in so-called "negative index" materials, where both μ and ε are negative. These materials must be fabricated, because equivalent materials are not found in nature. We discuss this in a later chapter as well.

A final comment, repeating the one made above: since most of you already speak at least two languages, the challenge of working with two sets of units should not prove to be insurmountable.

C. Maxwell's Equations, microscopic version.

As discussed in detail in Ch. 4, the macroscopic Maxwell Equations describe the behavior of macroscopic, or volume-averaged, fields. Points to keep in mind include the facts that physics takes place in real space, in real time, and on the atomic scale, and that information is always lost by averaging. The justification for the macroscopic equations is that it makes no more sense to follow the motion of ca. 10^{23} charges/cm^3 in a typical material than it does to follow the motion of the 2.7×10^{19} molecules/cm^3 in a gas at STP. In thermodynamics we deal with this by defining macroscopic parameters such as pressure, temperature, energy density, etc., then develop equations such as $PV = nRT$ that these macroscopic parameters satisfy. These parameters and equations are far more useful in practical applications than the microscopic ones. The macroscopic Maxwell Equations fall in the same category. They are also conveniently empirical: ε and μ can be treated phenomenologically, so for a given material they can be measured, and the measured values used to calculate everyday phenomena ranging from optical properties such as reflection, transmission, and scattering, to the behavior of ferromagnets.

However, to obtain a complete picture we must determine what information is sacrificed by averaging and how it vanishes. Accordingly, we return to the microscopic Maxwell Equations, where every charge is treated explicitly and, according to the Lorentz force law, the only fields are $\vec{E}$ and $\vec{B}$. We can then determine whether this

lost information is relevant, or of no practical importance. This also allows us to construct atomic-scale models for ε and μ, and therefore obtain the atomic-scale bases for $\vec{D}$ and $\vec{H}$, which are also done in later chapters.

The microscopic Maxwell Equations are

cgs:	SI:	
$\nabla \cdot \vec{E} = 4\pi\rho;$	$\nabla \cdot \vec{E} = \rho/\varepsilon_o;$	(1.6a,b)
$\nabla \cdot \vec{B} = 0;$	$\nabla \cdot \vec{B} = 0.$	(1.6c,d)
$\nabla \times \vec{E} + \frac{1}{c}\frac{\partial \vec{B}}{\partial t} = 0;$	$\nabla \times \vec{E} + \frac{\partial \vec{B}}{\partial t} = 0;$	(1.6e,f)
$\nabla \times \vec{B} = \frac{4\pi}{c}\vec{J} + \frac{1}{c}\frac{\partial \vec{E}}{\partial t};$	$\nabla \times \vec{B} = \mu_o \vec{J} + \mu_o \varepsilon_o \frac{\partial \vec{E}}{\partial t};$	(1.6g,h)

where for a point charge q located at $\vec{r}'$, $\rho(\vec{r}) = q\delta(\vec{r} - \vec{r}')$. With $\vec{D}$ and $\vec{H}$ gone, the "clean" appearance of the SI equations is lost: ε_o and μ_o must be written explicitly.

To summarize what we discuss in detail in Chs. 4 and 6, the macroscopic and microscopic pictures are connected as follows. Using the microscopic electric field $\vec{E}(\vec{r})$ as an example, we define the macroscopic equivalent $< \vec{E}(\vec{r}) >$ as

$$< \vec{E}(\vec{r}) > = \int_V d^3r' W(\vec{r} - \vec{r}')\vec{E}(\vec{r}'), \tag{1.7}$$

where $W(\vec{r} - \vec{r}')$ is a function that is positive definite, integrates to 1, varies slowly on an atomic scale, but varies rapidly enough on a laboratory scale to follow laboratory-scale features. The micro-macro connection also involves local fields, which are discussed in Ch. 7. Those familiar with Fourier analysis have already recognized Eq. (1.7) as a convolution in direct space, so the Fourier transform of $< \vec{E}(\vec{r}) >$ is equal to the product of the Fourier transforms of $\vec{E}(\vec{r})$ and $W(\vec{r})$. Because $\vec{E}(\vec{r})$ varies rapidly in direct space, it has Fourier coefficients that extend to high wave numbers $\vec{k}$ in reciprocal space. But because $W(\vec{r})$ is slowly varying in direct space, its Fourier transform is a narrow distribution about $\vec{k} = 0$. Thus $W(\vec{r} - \vec{r}')$ acts as a filter to eliminate the high-wave-number components of $\vec{E}(\vec{r})$, leaving a result that varies slowly on the atomic scale.

From the discussion on averaging, it may be surprising for students to learn that optical measurements can obtain nanostructural details of sizes one or two orders of magnitude less than the wavelengths of light used in these measurements. This appears to violate the Rayleigh criterion for resolution. However, your cell phone is direct evidence of this. Integrated-circuits technology is critically dependent on reflectance-based measurements in the visible-near ultraviolet part of the spectrum to monitor and control the lithographic, deposition, and etching processes needed to fabricate devices.

Feature sizes dropped below visible-near ultraviolet wavelengths with the 180 nm node (the terminology for a given class of feature sizes) in 1999. We are now (2020) in the 7 nm fabrication node, with a limited number of 5 nm components also in production. The apparent contradiction with the Rayleigh criterion is resolved by noting that reflectance-related measurements can now be made with sufficiently high accuracy so that useful information about details nominally suppressed by averaging can be extracted.

The alert reader may have noticed that Maxwell's Equations appear to impose 8 conditions on only 6 unknowns, which are the three vector components each of $\vec{E}$ and $\vec{B}$. However, overdetermination does not occur, as can be shown by taking the divergences of Eqs, (1.6e,f) and (1.6g,h). The divergence of the Faraday-Maxwell Equation (1.6e,f) is zero, and hence is consistent with Eqs. (1.6c,d). The divergence of the Ampère-Maxwell Equation (1.6g,h) is also zero, and hence consistent with Eqs. (1.6a,b) together with the charge conservation Eq. (1.4). This consistency ensures that overdetermination does not occur.

D. Potentials.

Maxwell's Equations are written in terms of fields and their interaction with materials. However, solutions of these equations are expedited by defining scalar and vector potentials ϕ and $\vec{A}$, respectively. As mathematical tools, potentials are useful for two reasons. First, they are defined so the two homogeneous Maxwell Equations are satisfied identically, so these need not be considered further. Second, when Maxwell's Equations are converted to equivalent equations for ϕ and $\vec{A}$, the Laplacian operator ∇^2 is introduced. This is significant because for ∇^2, Green functions exist. Green functions convert differential equations (for example the Maxwell Equations themselves) into integral equations. This is an enormous step forward, because integrals can be evaluated deterministically using numerical techniques, approximated to any desired level of accuracy, and may even have analytic solutions. As a result, potentials play a major role in the rest of the notes.

Some comments about potentials are necessary to establish a physics perspective. We shall see that there is considerable flexibility in their definitions, with different definitions termed gauges. The two most common are the Coulomb and Lorentz gauges, defined below. It is important to realize that, whatever the gauge, the result is a solution of Maxwell's Equations, which are written in terms of fields, not potentials. Since any gauge will get you there, from a mathematics perspective differences among gauges are completely irrelevant.

However, from the physics perspective the differences are significant. Potentials, properly defined, yield physical understanding as well. As a simple example, in the Lorentz gauge ϕ and $\vec{A}$ satisfy the wave equation, which tells us that information cannot propagate faster than the speed of light. While this constraint is also built into Eqs. (1.1), you can examine them all day and not come to this conclusion. Further, it is well known that the scalar potential measures the work required to move a charge from one region of space to another where the scalar potentials of the two regions differ.

The physical reality of the vector potential is less obvious. It was long considered by physicists to be nothing more than a convenient mathematical construct, despite its central role in driving eddy currents (engineers recognized its physical significance first). Changing the viewpoint of physicists required discovering the Aharonov-Bohm effect, where the phase of the wave function of an electron was found to be affected in traversing a region of space where $\vec{B}$, $\vec{E}$, and ϕ are identically zero but $\vec{A}$ is not. In quantum mechanics, potentials are now considered more fundamental than fields.

For simplicity and to focus on the vector math, the following development is based on the microscopic form of Maxwell's Equations in cgs units. The macroscopic treatment for scalar ε and μ follows similarly, and is left as a homework assignment. The situation for tensorial ε and μ, necessary to describe metamaterials and crystal optics, is more challenging and is left to a later chapter.

To begin, Eqs. (1.6c), (1.6d) and (1.6e), (1.6f) can be satisfied identically by defining $\vec{A}$ and ϕ as

$$\vec{B} = \nabla \times \vec{A}; \tag{1.8a}$$

$$\vec{E} = -\frac{1}{c}\frac{\partial \vec{A}}{\partial t} - \nabla\phi \quad \text{(cgs)}; \qquad \vec{E} = -\frac{\partial \vec{A}}{\partial t} - \nabla\phi \quad \text{(SI)}. \tag{1.8b,c}$$

Equation (1.8a) has the same form in both cgs and SI units. The proofs of Eqs. (1.8b,c) follow by direct substitution:

$$\nabla \times \vec{E} + \frac{1}{c}\frac{\partial \vec{B}}{\partial t} = \nabla \times \left(-\frac{1}{c}\frac{\partial \vec{A}}{\partial t} - \nabla\phi \right) + \frac{1}{c}(\nabla \times \vec{A}) = 0\,. \tag{1.9}$$

In principle only the vector-potential term in Eqs. (1.8b,c) is needed to satisfy Eq. (1.8c), but if the term $(-\nabla\phi)$ is not included, it is impossible to make Poisson's Equation consistent with Ampère's Equation. Rather than complicate the math by working only with the vector-potential part of Eqs. (1.8b,c) then fixing the damage later, we continue with Eqs. (1.8b,c) as written and provide the justification for $(-\nabla\phi)$ below.

Substituting these relations into the third Maxwell equation and invoking the vector-calculus identity $\nabla \times (\nabla\times) = \nabla(\nabla\cdot) - \nabla^2$ leads to

$$\begin{aligned} \nabla \times \nabla \times \vec{A} &= \nabla(\nabla \cdot \vec{A}) - \nabla^2 \vec{A} \\ &= \frac{4\pi}{c}\vec{J} - \frac{1}{c^2}\frac{\partial^2 \vec{A}}{\partial t^2} - \frac{1}{c}\frac{\partial}{\partial t}\nabla\phi \end{aligned} \tag{1.10}$$

in cgs units. Moving terms around and being a bit cavalier with the ordering of the gradient and time-derivative operations, which is allowed if the functions on which they operate are regular, this equation is converted into

$$(\nabla^2 - \frac{1}{c^2}\frac{\partial^2}{\partial t^2})\vec{A} = -\frac{4\pi}{c}\vec{J} + \nabla\left(\nabla \cdot \vec{A} + \frac{1}{c}\frac{\partial \phi}{\partial t} \right). \tag{1.11}$$

Substituting Eq. (1.8b) into Eq. (1.5a) yields

$$\frac{1}{c}\frac{\partial(\nabla\cdot\vec{A})}{\partial t}+\nabla^2\phi=-4\pi\rho\,. \tag{1.12}$$

These scalar- and vector-potential versions of the two inhomogeneous Maxwell equations are more complicated than necessary because ϕ and $\vec{A}$ appear in each of Eqs. (1.11) and (1.12). We can fix this interdependency in several ways. In the *Coulomb gauge*, $\nabla\cdot\vec{A}$ is simply *defined* to be zero, in which case

$$(\nabla^2-\frac{1}{c^2}\frac{\partial^2}{\partial t^2})\vec{A}=-\frac{4\pi}{c}\vec{J}+\frac{1}{c}\frac{\partial}{\partial t}(\nabla\phi); \tag{1.13a}$$

$$\nabla^2\phi=-4\pi\rho\,. \tag{1.13b}$$

Interpreting Eq. (1.13b) from a physics perspective, we would conclude that ϕ is updated instantly everywhere in the universe to any change in ρ, so as far as ϕ is concerned, c is infinite. This solution for ϕ then becomes a source term in Eq. (1.13a). It can be shown more generally that the resulting solution for $\vec{A}$ eliminates this nonphysical behavior at the field level, so causality is not violated. However, we do not prove it here. The present assumption needing proof is that $\nabla\cdot\vec{A}$ can be defined arbitrarily. We will prove this later.

Even though the equations remain consistent with relativity at the field level, it still makes more sense to formulate everything so the equations are consistent with relativity at the potential level. This can be accomplished simply by eliminating the gradient term in Eq. (1.11). This is conveniently done by setting

$$\nabla\cdot\vec{A}+\frac{1}{c}\frac{\partial\phi}{\partial t}=0\,. \tag{1.14}$$

This is the definition of the *Lorentz gauge*. With this condition, the inhomogeneous Maxwell equations reduce to

cgs:	SI:	
$(\nabla^2-\frac{1}{c^2}\frac{\partial^2}{\partial t^2})\vec{A}=-\frac{4\pi}{c}\vec{J}$;	$(\nabla^2-\mu_o\varepsilon_o\frac{\partial^2}{\partial t^2})\vec{A}=-\mu_o\vec{J}$;	(1.15a,b)
$(\nabla^2-\frac{1}{c^2}\frac{\partial^2}{\partial t^2})\phi=-4\pi\rho$;	$(\nabla^2-\mu_o\varepsilon_o\frac{\partial^2}{\partial t^2})\phi=-\rho/\varepsilon_o$.	(1.15c,d)

The potentials ϕ and $\vec{A}$ now both satisfy wave equations, so the effects of changes in ρ or $\vec{J}$ now both propagate at the speed of light. In addition, the wave operator $(\nabla^2-\frac{1}{c^2}\frac{\partial^2}{\partial t^2})$ is covariant, i.e., has the same form in any inertial system – but only if *c* is

a universal constant. In Ch. 2 we find additionally that the Lorentz gauge is the only one consistent with both the Colomb and Ampère force equations.

Capitalizing on covariance, we now introduce notation from special relativity, identifying $(\phi, \vec{A})$ as the *potential four-vector* and $(c\rho, \vec{J})$ as the *current 4-vector*. Given this, Eq. (1.14) is a *4-divergence*, where the four-dimensional gradient operator is $(\nabla, \frac{1}{c}\frac{\partial}{\partial t})$. The proof that these are indeed 4-vectors is done in Ch. 17. Our accomplishments so far can thus be summarized as

$$(\nabla^2 - \frac{1}{c^2}\frac{\partial^2}{\partial t^2})(\phi, \vec{A}) = -\frac{4\pi}{c}(c\rho, \vec{J}). \tag{1.16}$$

Equation (1.16) has advantages beyond simplified notation. The wave operator has a Green-function solution, which allows ϕ and $\vec{A}$ to be expressed in integral form. This forms the basis of our later treatments of radiation, scattering, and diffraction. Next, 4-vectors are invariant under a Lorentz transformation. Thus $(\phi^2 - \vec{A}^2)$ and $((c\rho)^2 - \vec{J}^2)$ are relativistic invariants, independent of the motion of the source or observer in any inertial system even though ϕ, $\vec{A}$, ρ, and $\vec{J}$ are not. These relations are useful in transforming charges, currents, potentials, and fields from one moving coordinate system to another. This is another aspect covered in Ch. 17.

Several minor details: the reader may have noticed that we could have added a gradient term to $\vec{A}$ and the result would still satisfy Eqs. (1.c,d), the only difference being $\vec{A} \to \vec{A}' = \vec{A} + \nabla\Lambda$. However, this only affects what we mean by $\vec{A}$. By repeating the calculations, it is easily seen that the same equations are obtained if $\vec{A}'$ replaces $\vec{A}$. Thus for simplicity we set $\Lambda = 0$.

Second, we defined $\vec{A}$ to ensure that $\nabla \cdot \vec{B} = 0$, but $\vec{A}$ also enters into the equation for $\vec{E}$. Does the Lorentz gauge introduce a consistency problem? Let's check. Substituting the Lorentz gauge into $\nabla \cdot \vec{E}$ leads to

$$\nabla \cdot \vec{E} = -\frac{1}{c}\frac{\partial}{\partial t}\left(-\frac{1}{c}\frac{\partial \phi}{\partial t}\right) - \nabla^2\phi \tag{1.17a}$$

$$= \left(\frac{1}{c^2}\frac{\partial^2}{\partial t^2} - \nabla^2\right)\phi = 4\pi\rho, \tag{1.17b}$$

which is the negative of Eq. (1.14a). Consistency is indeed realized.

Third, the above developments assume that $\nabla \cdot \vec{A}$ can be chosen to have any convenient value. When we discuss gauge transformations, we find this to be the case.

Finally, what happens if we neglect $(-\nabla\phi)$ in Eqs. (1.9a,b) and use only $\vec{E} = -(1/c)(\partial\vec{A}/\partial t)$? Ampère's and Poisson's Equations become

$$\left(\nabla^2 - \frac{1}{c^2}\frac{\partial^2}{\partial t^2}\right)\vec{A} = -\frac{4\pi}{c}\vec{J} + \nabla(\nabla\cdot\vec{A}); \tag{1.18a}$$

$$-\frac{1}{c}\frac{\partial}{\partial t}(\nabla\cdot\vec{A}) = 4\pi\rho. \tag{1.18b}$$

These expressions clearly overdetermine $\vec{A}$, so both cannot be satisfied simultaneously. The scalar potential is necessary to resolve the contradiction. However, we already knew that ϕ is essential from electrostatics.

One characteristic of the Lorentz gauge that is not generally appreciated is the role of ϕ in ensuring that $\vec{E}$ and $\vec{B}$ are perpendicular to the propagation direction $\vec{k}$ of a plane wave. For the moment we accept the result that in the far-field radiation zone and the locally planar approximation, $\vec{A}(\vec{r},t)$ is given by

$$\vec{A}(\vec{r},t) = \vec{A}_o e^{i\vec{k}\cdot\vec{r}-i\omega t}, \tag{1.19}$$

where $\hat{A}_o$ is the direction of the acceleration of the charge q that gives rise to $\vec{A}$. Since q can be accelerated in any direction, $\vec{A}$ can be oriented in any direction as well. Substituting Eq. (1.19) in Eq. (1.1c) gives $\nabla\cdot\vec{B} = i\vec{k}\cdot\vec{B} = 0$, which shows that $\vec{B}$ is already orthogonal to $\vec{k}$. To determine $\vec{E}$, we first substitute the plane-wave expression in Eq. (1.14), finding

$$\phi = \frac{c}{\omega}\vec{k}\cdot\vec{A} = \hat{k}\cdot\vec{A}, \tag{1.20}$$

taking advantage of the dispersion relation $ck/\omega = 1$ obtained from Eq. (1.18a) in source-free regions of space. Substituting Eqs. (1.18) and (1.19) in Eq. (1.9a), we find

$$\vec{E}(\vec{r},t) = \frac{i\omega}{c}(\vec{\vec{I}} - \hat{k}\hat{k})\cdot\vec{A}(\vec{r},t) = -\frac{i\omega}{c}\left(\hat{k}\times(\hat{k}\times\vec{A}(\vec{r},t)\right) \tag{1.21a,b}$$

where $\vec{\vec{I}}$ is the second-rank unit tensor. This is obviously also orthogonal to $\vec{k}$ and to $\vec{B}$. ϕ therefore removes the longitudinal contribution of $\vec{A}$. Because $\vec{A}$ is relatively easy to calculate for an oscillating charge, Eqs. (1.21) provide a simple way of determining $\vec{E}$ in the far-field limit. We use this to advantage in Chs. 15 and 16.

The above development also provides an indication of what is meant by electrostatics and magnetostatics. For electrical phenomena, setting all time derivatives equal to zero has the same effect as letting $c \to \infty$. The same terms vanish in both cases. Thus electrostatics is electrodynamics with no time delays; everything settles instantly into steady-state conditions. We will see this more explicitly when we discuss radiation, diffraction, and scattering. This conclusion does not contradict Coulomb's results, which were obtained with static charge distributions. The $c \to \infty$ limit will be assumed implicitly in Chs. 3-5.

The same assertion cannot be made for magnetostatics, because magnetic effects are already of first order in 1/*c*. Here, time derivatives must be considered explicitly, beginning with the charge-conservation equation. From Eq. (1.4) it follows that magnetostatics is defined by the condition

$$\nabla \cdot \vec{J} = 0 . \tag{1.22}$$

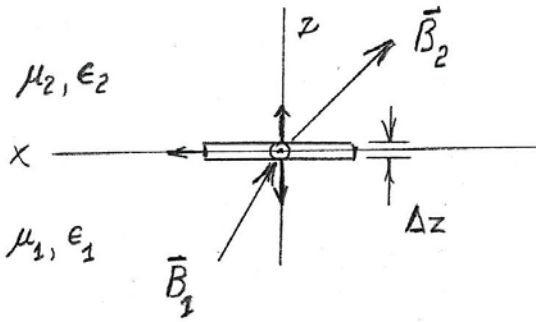
It might seem odd that a configuration containing a moving current can be described as static. "Quasistatic" or "stationary" might be more appropriate terminology.

A final comment on sources and responses. In the above we have assumed that $\vec{J}$ and ρ are source terms, that is, supplied externally and independent of potentials or fields. For conductors, where $\vec{J} = \sigma \vec{E}$ and $\vec{E}$ is given by Eqs. (1.9a,b), this is clearly not the case. In this situation the wave equation and the equivalent of the Lorentz gauge both contain an additional term, and the equation for $\vec{A}$ is homogeneous. We only call attention to this here, leaving the details to a homework assignment.

E. Boundary conditions.

The solutions of Eqs. (1.1) are complicated if ε and μ depend on $\vec{r}$, so we usually assume that ε and μ are constants. Yet some types of spatial variations, for example a discontinuous change in ε and/or μ at an abrupt interface between two materials, where gradients diverge and the mathematics are ill-defined, lead to solutions that are well defined. We deal with these boundary-condition situations by capitalizing on Gauss' or Stokes' Theorems and the divergence or curl nature of Maxwell's Equations to cast these calculations into equivalent calculations where all functions are defined in adjacent regions where they are regular. These calculations reduce to defining volumes, surfaces, and path integrals in ways that capitalize on the properties of the configuration, particularly symmetry, to prove the propositions. We shall use these procedures throughout the course. Therefore, the processes by which we obtain these results are as important as the results themselves. Note that the results apply equally well to "virtual" interfaces, where no physical boundaries are present.

We start with the easiest equation to deal with, Eq. (1.1c). Consider a locally planar interface between materials 1 and 2 containing magnetic flux densities $\vec{B}_1$ and $\vec{B}_2$, as shown in the diagram at the right. With an abrupt transition from μ_1 to μ_2, the normal gradient diverges, hence a direct approach is impossible. However, the fact that Eq. (1.1c) is a divergence relation suggests that we use Gauss' Theorem to establish a connection between $\vec{B}_1$ and $\vec{B}_2$ by constructing a volume with one surface in material 1 and another in material 2, then use Eq. (1.1c) and Gauss' Theorem to obtain the boundary condition.



Accordingly, let the local *z* axis be perpendicular to the interface, with material 1 below and material 2 above. Now, construct a surface *S* that encloses a rectilinear "Gaussian pillbox" of dimensions $\Delta x \times \Delta y \times \Delta z$ that straddles the interface, as shown in side view in the diagram. Three of the six normal vectors $\hat{n}$ are also shown. Next, apply Gauss' Theorem. The result is

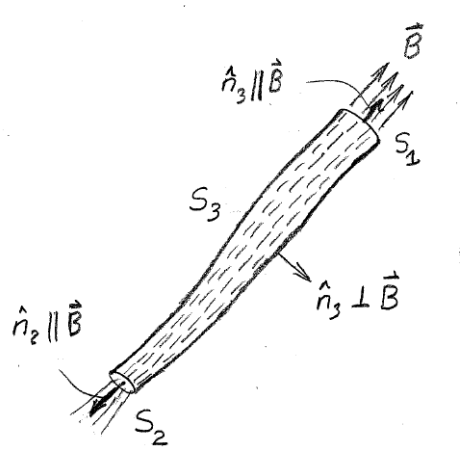
$$\begin{aligned}\int_V d^3r' \nabla \cdot \vec{B} &= \int_S d^2r' \hat{n} \cdot \vec{B} \\ &= \left(\hat{z} \cdot \vec{B}_2 + (-\hat{z}) \cdot \vec{B}_1\right) \Delta x \Delta y \\ &= 0, \end{aligned} \tag{1.23}$$

where we have used Eq. (1.1c), identified $\hat{n}_1 = (-\hat{z})$ and $\hat{n}_2 = \hat{z}$, and eliminated possible contributions from the sidewalls by taking the limit $\Delta z \to 0$. The result is

$$\hat{z} \cdot \vec{B}_1 = \hat{z} \cdot \vec{B}_2 . \tag{1.24}$$

We conclude that the normal components of $\vec{B}$ are continuous across the interface. From a physics perspective, this is simply a consequence of the absence of magnetic charge.

Taking a detour into additional physics, a similar construction can be used to prove that magnetic flux lines are continuous. The diagram at the right shows a region of space containing a magnetic flux density $\vec{B}$. Construct *S* such that the side S_3 is everywhere parallel to $\vec{B}$ (normal vectors $\hat{n}_3$ everywhere perpendicular to $\vec{B}$), and the ends S_1 and S_2 are everywhere perpendicular to $\vec{B}$ (normal vectors $\hat{n}_1$ and $\hat{n}_2$ parallel and antiparallel, respectively, to $\vec{B}$.) Applying Gauss' Theorem to the enclosed volume *V* yields



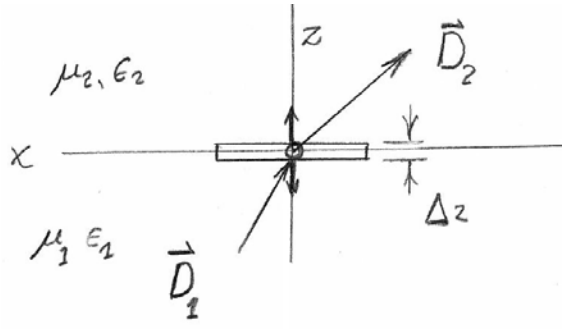
$$\begin{aligned}\int_V d^3r' \nabla \cdot \vec{B} &= \int_S d^2r' \hat{n} \cdot \vec{B} \\ &= \int_{S_1} d^2r' \hat{n}_1 \cdot \vec{B} + \int_{S_2} d^2r' \hat{n}_2 \cdot \vec{B} + 0 = \Phi_1 - \Phi_2 \\ &= 0, \end{aligned} \tag{1.25}$$

since the sidewalls contribute nothing, and by definition Φ_1 and Φ_2 are the magnetic fluxes passing through S_1 and S_2, respectively. The result is

$$\Phi_1 = \Phi_2 . \tag{1.26}$$

This equation is again a consequence of the fact that there is no magnetic charge.

The derivation places no restrictions on the length of *S*. With no magnetic charge, S must either extend to infinity, or curve back upon itself to form a loop. The first option is nonphysical, so we are left with the second. Consequently, we conclude that flux lines are continuous. In the final step, the cross section of *S* is reduced to where it contains only a single flux quantum $h/2e$, and the definition of a flux line is complete.



If there is no magnetic charge, how do flux line originate? In materials systems the answer is straightforward: from pre-existing dipoles, which can be considered as infinitesimally spaced pairs of infinitely large magnetic charges of equal and opposite signs. Magnetic fields are also produced by induction, where closed loops form without the aid of dipoles.

The electric-field equivalent, shown at the bottom of the previous page, is more interesting because Eq. (1.1a) involves ρ as well as $\vec{D}$. Applying Gauss' Theorem to Eq. (1.1a) with the coordinate system and surface shown, the relevant equations are

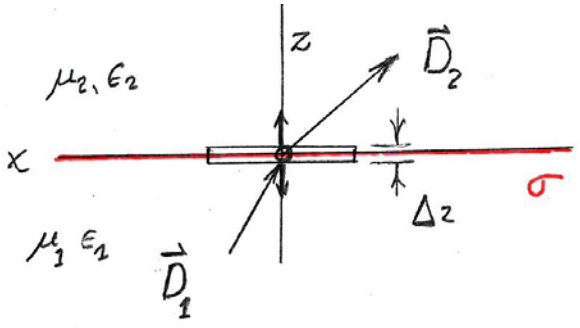
$$\int_V d^3r' \nabla \cdot \vec{D} = \int_S d^2r' \hat{n} \cdot \vec{D}$$

$$= \left(\hat{z} \cdot \vec{D}_2 + (-\hat{z}) \cdot \vec{D}_1 \right) \Delta x \Delta y = (\vec{D}_2 - \vec{D}_1) \cdot \hat{z} \Delta x \Delta y$$

$$= 4\pi \int_V d^3r' \rho(\vec{r}') \;\; = 4\pi \int_A d^2r' \left(\int_{-\Delta z}^{\Delta z_2} dz' \rho(\vec{r}') \right)$$

$$= 4\pi \left(\int_{-\Delta z_1}^{\Delta z_2} dz' \rho(z') \right) \Delta x \Delta y \,, \qquad (1.27)$$

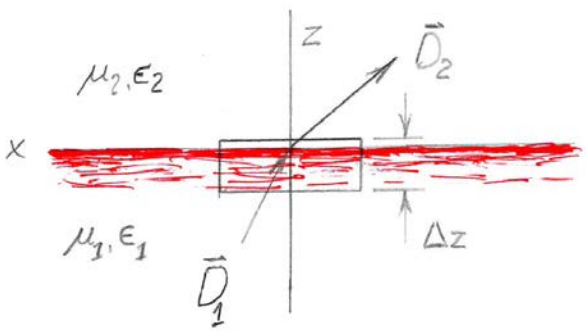
where Δz_1 and Δz_2 are the (positive-definite) *distances* that the integral over z' extends into media 1 and 2, respectively. If $\rho(\vec{r}) = 0$ throughout the region, then the result reduces to that discussed for $\vec{B}$: the normal component of $\vec{D}$ is continuous across the interface.



Next, suppose that $\rho(\vec{r}) = \rho(z) = \sigma\delta(z)$, where σ is the surface charge density. This is shown in the diagram on the right. An infinite value over zero thickness leads to another type of singularity, but one that can also be handled with Gauss' Theorem. The result is

$$\hat{z} \cdot (\vec{D}_2 - \vec{D}_1) = 4\pi\sigma \,. \qquad (1.28)$$

Finally, suppose that $\rho(\vec{r})$ is distributed in a finite but narrow region of medium 1 adjacent to medium 2, as shown in the diagram on the right. final diagram. This occurs for example in the space-charge region of a semiconductor or Thomas-Fermi screening in a metal. In that case we make Δz_1 large enough to pick up the nonzero charge, defining

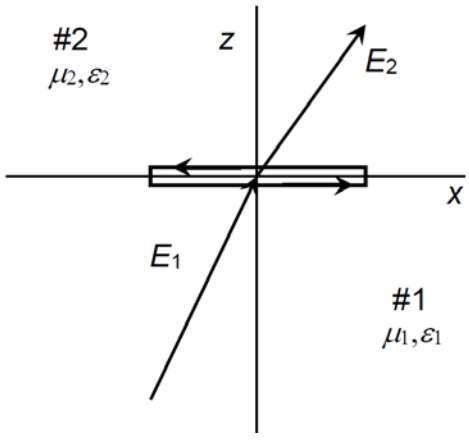


$$\sigma = \int_{-\Delta z_1}^{\Delta z_2} dz' \, \rho(\vec{r}') \cdot \tag{1.29}$$

We have therefore not only established the continuity condition on the normal components of the limiting values of $\vec{D}$ on either side of the interface, but also the definition of the surface charge density σ. In the limit that $(\Delta z_1 + \Delta z_2) \to 0$, then $\rho(\vec{r}) \to \sigma \, \delta(z)$.

The integral definition of σ is relevant because ideal interface-charge distributions $\rho(\vec{r}) = \sigma \, \delta(z)$ do not exist in practice. The closest approximation occurs for static electric fields applied to a metal. In this case the induced "screening" charge density decays approximately exponentially with a Thomas-Fermi screening length of the order of an atomic spacing. Other materials give different results. The space-charge region of a semiconductor may extend inward as much as a micron from the surface. The key point is that if σ is to be interpreted as an interface charge density, then $(\Delta z_1 + \Delta z_2)$ must be large enough to pick up *all* the interface charge. More details regarding metals and semiconductors are provided in later chapters.

We now consider tangential components. The tangential case is more complicated because there are two dimensions parallel to the interface. Tangential components suggest line integrals, so we look to Stokes' Theorem and Eqs. (1.1e) and (1.1g). The simpler expression is Eq. (1.1e), so we start with the electric-field case. Again, suppose a locally planar interface between the two materials containing electric fields $\vec{E}_1$ and $\vec{E}_2$, as shown in the diagram at the right. As usual, take $\hat{z}$ to be normal to the interface and to point toward material 2. For simplicity, define $\hat{x}$ by supposing that $\vec{E}_1$ lies in the *xz* plane, or $\vec{E}_1 = \hat{x}E_{1x} + \hat{z}E_{1z}$. Since we know nothing yet about the tangential components of $\vec{E}_2$, let $\vec{E}_2 = \hat{x}E_{2x} + \hat{y}E_{2y} + \hat{z}E_{2z}$.



To apply Stokes' Theorem, define a rectilinear path of length Δx and height Δz straddling the interface as shown, and start the path integration at the lower left corner. Then writing the length element $d\vec{l}$ as $d\vec{l} = \hat{l}(s)ds$, where s is the monotonically increasing distance along the path and $\hat{l}(s)$ is its local direction,

$$\int_S d^2r'\,\hat{n}\cdot\nabla\times\vec{E} = \int_0^{\Delta x} dx \int_0^{\Delta z} dz\,(-\hat{y})\cdot\left(-\frac{1}{c}\frac{\partial\vec{B}}{\partial t}\right)_{\Delta z\to 0} = 0$$

$$= \oint_C \vec{E}\cdot d\vec{\ell} = \int_0^{\Delta x} ds\,\hat{x}\cdot\vec{E}_1 + \int_{\Delta x}^{\Delta x+\Delta z} ds\,\hat{z}\cdot\vec{E} + \int_{\Delta x+\Delta z}^{2\Delta x+\Delta z} ds\,(-\hat{x})\cdot\vec{E}_2 + \int_{2\Delta x+\Delta z}^{2\Delta x+2\Delta z} ds\,(-\hat{z})\cdot\vec{E}$$

$$= \hat{x}\cdot(\vec{E}_1 - \vec{E}_2) = E_{1x} - E_{2x}. \tag{1.30}$$

We eliminate possible contributions from $\partial\vec{B}/\partial t$ and the sides of length Δz by taking the limit $\Delta z \to 0$. Then

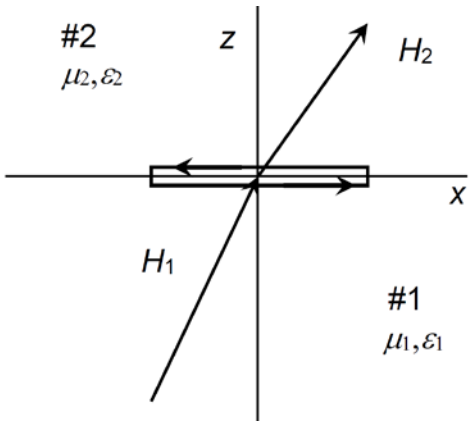
$$E_{1x} = E_{2x}. \tag{1.31}$$

Thus the tangential component E_x is continuous across the interface. Repeating the calculation for a loop in the yz plane yields the analogous result, $E_{1y} = E_{2y}$. However, $E_{1y} = 0$ by construction. Hence $E_{2y} = 0$, that is, $\vec{E}_1$ and $\vec{E}_2$ are coplanar, and their tangential components at the interface are equal.

The result can be summarized in general as

$$\hat{z}\times\vec{E}_1 = \hat{z}\times\vec{E}_2. \tag{1.32}$$

With Eq. (1.27) describing the normal component, a complete picture is obtained.



The magnetic equivalent, Eq. (1.1g), presents more challenges because of the current term. This case must be treated similarly to that of Eq. (1.1a). Define the x direction according to $\vec{H}_1 = \hat{x}H_{1x} + \hat{z}H_{1z}$ and perform the path integration as above. Then applying Stokes' Theorem and neglecting any possible contributions from $\partial\vec{D}/\partial t$ and the parts of the path of length Δz, we are left with

$$\oint_C \vec{H}\cdot d\vec{\ell} = \cdot\left(\hat{x}\cdot\vec{H}_1 + (-\hat{x})\cdot\vec{H}_2\right)\Delta x$$

$$= \int_0^{\Delta x} dx \int_{-\Delta z_1}^{\Delta z_2} dz \left((-\hat{y}) \cdot \nabla \times \vec{H} \right) = \int_0^{\Delta x} dx \int_{-\Delta z_1}^{\Delta z_2} dz \left(\frac{4\pi}{c} (-\hat{y}) \cdot \vec{J} \right). \tag{1.33}$$

As with the situation involving interface charge, this expression is useful only if Δz_1 and Δz_2 are large enough so the integration range includes all regions where $\vec{J}$ is nonzero. In this case $\vec{H}_1$ and $\vec{H}_2$ assume their limiting values, and we can write

$$H_{1x} - H_{2x} = \frac{4\pi}{c} K_y, \tag{1.34}$$

where the surface current $\vec{K}$ is defined as

$$\vec{K} = \int_{-\Delta z_1}^{\Delta z_2} dz' \left(\vec{J} + \frac{1}{4\pi} \frac{\partial \vec{D}}{\partial t} \right) \approx \int_{-\Delta z_1}^{\Delta z_2} dz' \, \vec{J} \,. \tag{1.35}$$

Repeating the same calculation for the loop in the *yz* plane with the coordinates shown yields

$$H_{1y} - H_{2y} = -\frac{4\pi}{c} K_x. \tag{1.36}$$

Equations (1.35) and (1.36) can be combined to yield

$$\hat{z} \times (\vec{H}_1 - \vec{H}_2) = \frac{4\pi}{c} \vec{K} \,. \tag{1.37}$$

If the materials in which $\vec{K}$ is generated are isotropic, Eq. (1.37) reduces to the usual condition that $\vec{H}_1$ and $\vec{H}_2$ are coplanar. If neither material is a conductor so that an interface current is not present, then the results reduce further to the condition that the tangential components of $\vec{H}$ are continuous across an interface. With the normal components connected by Eq. (1.23), a complete picture is obtained.

F. Energy.

We conclude Ch. 1 by calculating intensity and energy density directly from Maxwell's Equations. Jackson does this in Ch. 6, but these topics are worth discussing now not only because energy is fundamental to physics but also because this is an excellent example of how mathematics should work, the result contains important physics that is easily extracted, and the derivation involves only a few short steps beyond what we have already accomplished. We work with the macroscopic rather than the microscopic equations, because the macroscopic case contains more physics.

Because energy is conserved, the physics dictates that the result must have same characteristics as those of Eq. (1.4), which expresses conservation of charge. The terms will either be time derivatives, which in this case represent the accumulation or depletion of energy stored in the electric and magnetic fields, or divergences, which represent extraction of energy (intensity) being transported through the region under investigation. Finally, this will be a *differential* expression describing incremental changes rather than absolute values.

The development starts with the classical-mechanics expression for differential work done *on* a point charge *q by* a force $\vec{F}$ when *q* is displaced by a distance increment $d\vec{l}$. The differential work done is

$$dW = \vec{F} \cdot d\vec{l} \,. \tag{1.38}$$

Our goal is to express this as differential work as work done by fields, which by conservation of energy can then be used to identify intensities and energy densities of fields. Start by substituting the Lorentz force for $\vec{F}$:

$$dW = \left(q\vec{E} + \frac{q}{c}\vec{v} \times \vec{B} \right) \cdot d\vec{l} \,. \tag{1.39}$$

Now divide Eq. (1.39) by *dt* to convert it to power:

$$\frac{dW}{dt} = \left(q\vec{E} + \frac{q}{c}\vec{v} \times \vec{B} \right) \cdot \frac{d\vec{l}}{dt} = \left(q\vec{E} + \frac{q}{c}\vec{v} \times \vec{B} \right) \cdot \vec{v} \,, \tag{1.40a}$$

$$= q\vec{E} \cdot \vec{v} \,. \tag{1.40b}$$

Equation (1.40b) follows because $d\vec{l}/dt = \vec{v}$ and $\vec{v} \cdot (\vec{v} \times \vec{B}) = 0$. Thus we have already learned that a static magnetic field cannot change the energy of a moving charge *q*.

Next, convert everything to densities by dividing by a volume element ΔV. The result is

$$\frac{1}{\Delta V}\frac{dW}{dt} = \frac{dU}{dt} = \vec{E} \cdot \frac{q\vec{v}}{\Delta V} = \vec{E} \cdot (\rho\vec{v}) = \vec{E} \cdot \vec{J} \,, \tag{1.41}$$

where U is the energy density, $q = \rho \Delta V$, and as usual $\vec{J} = \rho \vec{v}$.

Now, eliminate $\vec{J}$ using the macroscopic version of Ampère's Law, Eq. (1.1g). The form of Eq. (1.1g) that we need is

$$\vec{J} = \frac{c}{4\pi} \nabla \times \vec{H} - \frac{1}{4\pi}\frac{\partial \vec{D}}{\partial t} \,. \tag{1.42}$$

Substituting Eq. (1.42) into Eq. (1.41) gives

$$\frac{dU}{dt} = \frac{c}{4\pi} \vec{E} \cdot \nabla \times \vec{H} - \frac{1}{4\pi} \vec{E} \cdot \frac{\partial \vec{D}}{\partial t} \,. \tag{1.43}$$

Being a time derivative the second term on the right has the correct form, but the first term needs work. We can fix this with the vector identity and the Maxwell-Faraday Equation:

$$\nabla \cdot (\vec{E} \times \vec{H}) = \vec{H} \cdot (\nabla \times \vec{E}) - \vec{E} \cdot (\nabla \times \vec{H}) \tag{1.44a}$$

$$= \vec{H} \cdot \left(-\frac{1}{c}\frac{\partial \vec{B}}{\partial t} \right) - \vec{E} \cdot (\nabla \times \vec{H}) \,. \tag{1.44b}$$

Substituting this for the "incorrect" term in Eq. (1.42) yields

$$\frac{dU}{dt} = -\nabla \cdot \left(\frac{c}{4\pi} \vec{E} \times \vec{H} \right) - \frac{1}{4\pi} \vec{H} \cdot \frac{\partial \vec{B}}{\partial t} - \frac{1}{4\pi} \vec{E} \cdot \frac{\partial \vec{D}}{\partial t}, \tag{1.45}$$

which clearly has the correct form. If μ and ε are constants, the second and third terms can be written in absolute form as

$$\frac{dU}{dt} = -\nabla \cdot \left(\frac{c}{4\pi} \vec{E} \times \vec{H} \right) - \frac{\partial}{\partial t} \left(\frac{1}{8\pi} \vec{H} \cdot \vec{B} \right) - \frac{\partial}{\partial t} \left(\frac{1}{8\pi} \vec{E} \cdot \vec{D} \right). \tag{1.46}$$

By comparing time derivatives on the left and right sides of Eq. (1.46), the magnetic- and electric-field expressions in large brackets are seen to be the energy densities U_M and U_E of magnetic and electric fields when μ and ε are constant.

Having done the math, we now consider the physics. First, if the fields are doing work on the object, then by energy conservation they must decrease in time. The signs of the last two terms on the right side of Eq. (1.45) are consistent with this.

Second, given that the last two terms describe stored energy, then the divergence term must describe energy flow, that is, intensity (again, see Eq. (1.4)). We recognize this as the Poynting vector. Recalling that the tangential components of $\vec{E}$ and $\vec{H}$ are continuous at an interface, we see that it is no accident that energy flow is written in terms of these two fields. If either $\vec{D}$ or $\vec{B}$ or both were involved, then energy would accumulate at interfaces, which obviously does not happen. Physics is nothing if not consistent.

Considering Eq. (1.44) further, it is also no accident that the time derivatives act on $\vec{B}$ and $\vec{D}$, and not $\vec{E}$ and $\vec{H}$. We recall that $\vec{D} = \vec{E} + 4\pi\vec{P}$, where $\vec{P}$ is the dipole density. On the atomic scale, $\vec{E}$ induces dipoles by moving charges out of their equilibrium positions, thereby storing energy as mechanical strain. This mechanical energy is taken into account automatically if the time derivative operates on $\vec{D}$, but not if it operates on $\vec{E}$. Similarly, in the magnetic case energy is stored by aligning dipoles by applying $\vec{H}$. This energy is taken into account automatically if the time derivative operates on $\vec{B}$, but not if it operates on $\vec{H}$. Again, physics is nothing if not consistent.

Next, Eq. (1.45) is a *differential* rather than an absolute relation. For linear systems energy densities start from zero and can be expressed in closed form, as noted above. However, in nonlinear situations the time integration may have to be done numerically, and the starting energy may not be zero. To see how these integrations are accomplished, multiply both sides by a time increment Δt. The result is

$$\Delta U = -\nabla \cdot \left(\frac{c}{4\pi} \vec{E} \times \vec{H} \right) \Delta t - \frac{1}{4\pi} \vec{H} \cdot \Delta \vec{B} - \frac{1}{4\pi} \vec{E} \cdot \Delta \vec{D}. \tag{1.46}$$

Equation (1.46) shows that energy density increases according to the areas under the $\vec{H}$ vs. $\vec{B}$ and $\vec{E}$ vs. $\vec{D}$ curves. Unfortunately, magnetic data are usually given as magnetization curves $\vec{B}$ vs. $\vec{H}$, not $\vec{H}$ vs. $\vec{B}$.

The magnetic case is also complicated by the fact that owing to their domain structure, considerable energy may already be stored in ferromagnets even when the material is nominally unmagnetized. With ferromagnetic materials energy calculations almost always involve numerical integration. We say more about this in Ch. 7.

Next, we note that Eq. (1.45) is valid for any real fields at any time *t*. When we discuss radiation, our interest will shift from fields to average power. In so doing we perform time averages to cast Eq. (1.45) into its time-averaged equivalent. The time-averaged version of Eq. (1.45) is usually the one that gets the most attention.

Finally, the fact that fields possess energy shows that they are more than simple mathematical conveniences. They are physical entities in their own right. Fields also contain momentum, which will be discussed in a later chapter.