Ph106bc: Electrodynamics

Sunil Golwala

Winter/Spring 2024

Contents

Contents

Contents

Magnetostatics

Magnetostatics in Matter

Electrodynamics

Contents

Contents

Lectures

Ph106b

Lecture 1	2024/01/04
Lecture 2	2024/01/09
Lecture 3	2024/01/11
Lecture 4	2024/01/16
Lecture 5	2024/01/18
Lecture 6	2024/01/23
Lecture 7	2024/01/25
Lecture 8	2024/01/30
Lecture 9	2024/02/01
Lecture 10	2024/02/06
Lecture 11	2024/02/08
Lecture 12	2024/02/13
Lecture 13	2024/02/15
Lecture 14	2024/02/22
Lecture 15	2024/02/25
Lecture 16	2024/02/27
Lecture 17	2024/02/29
Lecture 18	2024/03/05
Lecture 19	2024/03/07

Ph106c

Lecture 20	2024/04/02
Lecture 21	2024/04/04
Lecture 22	2024/04/09
Lecture 23	2024/04/11
Lecture 24	2024/04/16
Lecture 25	2024/04/18
Lecture 26	2024/04/23
Lecture 27	2024/04/25
Lecture 28	2024/04/30
Lecture 29	2024/05/02
Lecture 30	2024/05/07
Lecture 31	2024/05/09
Lecture 32	2024/05/14
Lecture 33	2024/05/16
Lecture 34	2024/05/21
Lecture 35	2024/05/23
Lecture 36	2024/05/28
Lecture 37	2024/05/30
Lecture 38	N/A
Lecture 39	N/A

Lecture 1:

Introduction to Course

Basics of Electrostatics I:
Electric Force and Field
Dirac Delta Function
Integral Form of Gauss's Law
Differential Form of Gauss's Law

Date Revised: 2024/01/09 05:45
Revised lecture break,
remove misstatement about units of electric field
Date Given: 2024/01/04

Section 1
Introduction to Course

Course Material

Overview

This is a course on electrodynamics. It will review the basic material you learned in Ph1bc but will go beyond in both content as well as in mathematical sophistication.

The intended learning outcome of both Ph106b and Ph106c is for students to acquire the ability to calculate electric and magnetic potentials, fields, energies, and forces in a variety of basic physical configurations combined with an understanding of the underlying physical principles and calculation techniques. This outcome requires both an understanding of principles as well as the ability to apply them to do calculations!

The course will primarily use and follow *Introduction to Electrodynamics* by Griffiths (fourth edition). Supplementary material is drawn from Jackson and from Heald & Marion, both on electronic and physical reserve from the library. The material presented here will be self-contained, but past students have found it useful to obtain a copy of Jackson. It is certainly a book you will want if you continue in physics or a related field.

Prerequisites

Physics:

- ▶ Electricity and Magnetism: While Ph1bc is a formal prerequisite for the course, we will develop the material from scratch. However, review material will be covered quickly and a basic familiarity with the concepts will be assumed.
- ▶ Classical mechanics: Generally, mechanics at the level of Ph1a is sufficient for this course, though some optional material at the end of Ph106c will make use of Ph106a material.

Mathematics:

- ▶ Chapter 1 of Griffiths except for Sections 1.1.5 ("How Vectors Transform") and 1.5 ("The Dirac Delta Function"). We will review some prerequisite material as needed.

- ▶ Solutions to second-order linear ordinary differential equations with constant coefficients (*i.e.*, simple harmonic oscillator).

- ▶ Orthonormal functions/bases.

- ▶ Over the course, we will develop the following more sophisticated concepts:
  - ▶ Dirac Delta function.
  - ▶ Separation of variables to reduce second-order linear partial differential equations to ordinary differential equations.
  - ▶ Various specific types of orthonormal functions, specifically sinusoids, Legendre polynomials, and spherical harmonics.
  - ▶ Tensor formalism for relativity.

- ▶ Key point: Mathematics is the language of physics. You must be competent in the above basic mathematics to understand and use the material in this course. Intuition is crucial, but it must be formalized mathematically.

  However, mathematics is not just symbolic manipulation or brute force calculation. Make sure you understand the **meaning** *of every mathematical expression — i.e., carry along the intuition with the symbols! Only do algebra and explicit differentiation and integration as a last resort! We will demonstrate this approach regularly.*

Topics to be Covered

New topics for Ph106b not covered in Ph1bc

New topics for Ph106c not covered in Ph1bc

- ▶ Review of basic electrostatics — Coulomb's Law; Gauss's Law; electric field, potential, and potential energy; conductors, capacitors, and capacitance matrix.

- ▶ Advanced electrostatics — boundary value problems (BVP) for determining potentials and fields; Green Functions for BVP; multipole expansion of potential.

- ▶ Electrostatics in Matter — polarization, susceptibility, permittivity of matter; BVP with polarizable materials, energy and forces in matter.

- ▶ Magnetostatics — Lorentz force; Biot-Savart Law; Ampère's Law; vector potential; boundary conditions; multipole expansion of potential.

- ▶ Magnetostatics in Matter — magnetization, susceptibility, and permeability of matter; boundary conditions; ferromagnetism; BVP with magnetizable materials.

- ▶ Electrodynamics — electromotive force and electromagnetic induction; inductance and energy in magnetic fields; Maxwell's equations in vacuum and in matter; boundary conditions for Maxwell's equations.

- ▶ Conservation Laws — Continuity equation; Poynting's Theorem; electrodynamic momentum and energy.

- ▶ Electromagnetic Waves — in vacuum, in polarizable/magnetizable matter, in conductors, in transmission lines and waveguides.

► Potentials and Radiation — potential formulation; fields and potentials of moving point charges; radiated electromagnetic waves; antennas.

► Relativity and Electrodynamics — review of special relativity including relativistic kinematics and collisions, relativistic tensor notation, transformation of fields, transformation of field tensor, relativistic potentials, relativistic formulation of Maxwell's Equations, relativistic dynamics with EM fields, relativistic conservation theorems.

Notation; including Deviations from Griffiths

▶ We will use standard black text for material that is covered in lecture, while magenta text will be used for material that is skipped during lecture for which you remain responsible. We will skip material generally when it consists of computation or calculation that is tedious to do on the chalkboard, summarizing the results as necessary. You will need to be able to apply the skipped material as well as the techniques developed in this skipped material.

▶ Green text will be used to indicate supplementary material for which you will not be responsible.

▶ Griffiths uses boldface notation to indicate vectors and a script $\vec{r}$ to indicate the difference vector $\vec{r} - \vec{r}'$. In order to better match what can be written by hand, we use $\vec{}$ rather than boldface for vectors and we use $\vec{R}$ for the difference vector.

▶ Griffiths uses $\vec{r}$ to refer to the position of the test charge Q and $\vec{r}'$ to refer to the position of the source charge q. This seems unnecessarily confusing. We instead use q and $\vec{r}$ for the test charge and q' and $\vec{r}'$ for the source charge.

▶ Griffiths uses $\delta^3(\vec{r})$ to refer to the delta function in three spatial dimension. We use $\delta(\vec{r})$ for this for reasons that are explained after Equation 2.9.

Section 2
Review of Basics of Electrostatics

Study Guidelines

You have seen all the material in this section before in Ph1b. However, the derivations done there were not as rigorous as they could be because you were simultaneously learning vector calculus. Our goal in this section is to do more rigorous derivations to give you some practice in using the mathematical tools. We won't do any examples in lecture or the notes because they duplicate Ph1b. But you should make sure you are comfortable with the examples in Griffiths Chapter 2.

The Assumed Conditions for Electrostatics

Electrostatics is the study of electric fields, potentials, and forces under two assumptions:

▶ All electric charges sourcing the electric field are stationary and have been so for a sufficiently long time that all fields are static and thus the electric field can be written in terms of the source charges' current positions.

▶ The source charges are held fixed and cannot react to the fields from any test charges that may be stationary or moving relative to the source charges.

We will see later that, when charges are moving, it takes time for the information about the position to propagate and thus the fields at a given point depend on the configuration of the charges at earlier times.

Coulomb's Law and the Electric Field

Coulomb's Law, Electrostatic Forces, and Superposition

We begin with two empirical facts:

▶ **Coulomb's Law**: the empirical fact that the force on a test charge q at position $\vec{r}$ due to a source charge q' at $\vec{r}'$ is given by Coulomb's Law:

$$\vec{F} = \frac{1}{4\pi\epsilon_o} \frac{q'\, q}{R^2}\, \widehat{R} \qquad \text{with} \quad \vec{R} \equiv \vec{r} - \vec{r}' \qquad (2.1)$$

where $\epsilon_o = 8.85 \times 10^{-12}$ C^2 N^{-1} m^{-2} is the *permittivity of free space*. The force points along the line from q' to q as indicated by the sign of the definition of $\vec{R}$. The electric charge is in the units of *Coulombs* (C), which is a fundamental unit that cannot be written in terms of other fundamental units.

Recall that: we use $\vec{}$ rather than boldface to indicate vectors; R where Griffiths uses a script r; and a different convention from Griffiths for the symbols for the two charges and their position vectors.

▶ **Superposition**: the empirical fact that Coulomb's Law obeys the principle of superposition: the force on a test charge q at $\vec{r}$ due to N charges $\{q_i'\}$ at positions $\{\vec{r}_i'\}$ is obtained by summing the individual vector forces:

$$\vec{F} = \sum_{i=1}^{N} \vec{F}_i = \sum_{i=1}^{N} \frac{1}{4\pi\epsilon_o} \frac{q_i'\, q}{R_i^2}\, \widehat{R}_i \qquad \text{with} \quad \vec{R}_i \equiv \vec{r} - \vec{r}_i' \qquad (2.2)$$

The Electric Field

Given that any test charge q placed at the position $\vec{r}$ feels the same force per unit charge, we are motivated to abstract away the test charge and define what we call the **electric field** at that position $\vec{r}$:

$$\vec{E}(\vec{r}) = \frac{\vec{F}}{q} = \begin{cases} \frac{1}{4\pi\epsilon_o} \frac{q'}{R^2} \widehat{R} & \text{for a single source charge } q' \text{ at } \vec{r}\,' \\ \sum_{i=1}^{N} \frac{1}{4\pi\epsilon_o} \frac{q'_i}{R_i^2} \widehat{R}_i & \text{for } N \text{ source charges } \{q'_i\} \text{ at positions } \{\vec{r}\,'_i\} \end{cases}$$

$$(2.3)$$

The electric field has units of N/C.

Coulomb's Law for Continuous Charge Distributions

If a charge distribution is continuous, then the natural extension of Coulomb's Law is to *integrate* the electric field or force over the contributions from the infinitesimal charge elements dq at $\vec{r}\,'$:

$$\vec{E}(\vec{r}) = \frac{1}{4\,\pi\,\epsilon_o} \int \frac{1}{[R(\vec{r}, \vec{r}\,')]^2}\, \widehat{R}(\vec{r}, \vec{r}\,')\, dq(\vec{r}\,') \tag{2.4}$$

where $\vec{R}(\vec{r}, \vec{r}\,') = \vec{r} - \vec{r}\,'$ varies with the location $\vec{r}\,'$ of dq as the integral is performed. dq is admittedly undefined here. However, before worrying about that, let us note that the integrand is a *vector* and so this integral requires some care: we must break up $\widehat{R}$ into its components and individually integrate each component. For example, if we use Cartesian coordinates, then $\widehat{R} = \widehat{x}\left(\widehat{R} \cdot \widehat{x}\right) + \widehat{y}\left(\widehat{R} \cdot \widehat{y}\right) + \widehat{z}\left(\widehat{R} \cdot \widehat{z}\right)$, and, since the Cartesian unit vectors do not depend on the location of the infinitesimal charge $dq(\vec{r}\,')$, we may write the integral out as follows (eliding the dependence of $\vec{R}$ on $\vec{r}$ and $\vec{r}\,'$ for brevity):

$$\vec{E}(\vec{r}) = \tag{2.5}$$
$$\frac{1}{4\,\pi\,\epsilon_o}\left[\widehat{x}\int\frac{1}{R^2}\left(\widehat{R}\cdot\widehat{x}\right)dq(\vec{r}\,') + \widehat{y}\int\frac{1}{R^2}\left(\widehat{R}\cdot\widehat{y}\right)dq(\vec{r}\,') + \widehat{z}\int\frac{1}{R^2}\left(\widehat{R}\cdot\widehat{z}\right)dq(\vec{r}\,')\right]$$

which is sum of three integrals with *scalar* integrands.

Now, consider some specific charge distributions:

▶ **volume charge distribution**:

$$\vec{E}(\vec{r}) = \frac{1}{4\,\pi\,\epsilon_o} \int_{\mathcal{V}} \frac{d\tau' \rho(\vec{r}')}{R^2}\,\widehat{R}$$
with $\rho(\vec{r}')$ having units of C m^{-3}, $\vec{r}'$ running over all points in the volume distribution $\mathcal{V}$, and $d\tau'$ being the differential volume element at $\vec{r}'$ for $\mathcal{V}$
(2.6)

▶ **surface charge distribution**:

$$\vec{E}(\vec{r}) = \frac{1}{4\,\pi\,\epsilon_o} \int_{\mathcal{S}} \frac{da' \sigma(\vec{r}')}{R^2}\,\widehat{R}$$
with $\sigma(\vec{r}')$ having units of C m^{-2}, $\vec{r}'$ running over all points in the surface distribution $\mathcal{S}$, and da' being the differential area element at $\vec{r}'$ for $\mathcal{S}$
(2.7)

▶ **line charge distribution**:

$$\vec{E}(\vec{r}) = \frac{1}{4\pi\epsilon_o} \int_{\mathcal{C}} \frac{d\ell'\,\lambda(\vec{r}\,')}{R^2}\,\widehat{R}$$

with $\lambda(\vec{r}\,')$ having units of C m^{-1}, $\vec{r}\,'$ running over all points in the line distribution $\mathcal{C}$, and $d\ell'$ being the differential length element at $\vec{r}\,'$ for $\mathcal{C}$ (2.8)

Using the Dirac delta function we will define below, one can write the first two as special cases of the latter by using delta functions in the dimensions in which the charge distribution has no extent.

Aside: The Dirac Delta Function

Relating Equation 2.6 to Equation 2.3 offers us both the opportunity to rigorously connect them as well as a chance to introduce the *Dirac delta function*. (Mathematically, it is a *distribution*, not a function, but we will use the standard nomenclature.) The Dirac delta function at $\vec{r}_o$, $\delta(\vec{r} - \vec{r}_o)$, is defined by what it does when it is multiplied against an arbitrary function $f(\vec{r})$ and integrated: For any function $f(\vec{r})$ and any volume $\mathcal{V}$ containing the point $\vec{r}_o$, it holds that

$$\int f(\vec{r}\,') \, \delta(\vec{r}\,' - \vec{r}_o) \, d\tau' = \left\{ \begin{array}{ll} f(\vec{r}_o) & \vec{r}_o \in \mathcal{V} \\ 0 & \vec{r}_o \notin \mathcal{V} \end{array} \right. \qquad (2.9)$$

In particular, if $f(\vec{r})$ is unity, then the right side of the above integral is unity for $\vec{r}_o \in \mathcal{V}$: the integral of a delta function over the volume containing its $\vec{r}_o$ is 1, and, conversely, the integral of a delta function over any volume not containing its $\vec{r}_o$ vanishes.

Two notes on dimensions and notation:

▶ In order for the units in the above equation to work out, the delta function above must have units of m^{-3}. The general rule is that the delta function's units are the inverse of those of the differential that its argument says it should be integrated with. In this case, the argument is a vector in 3D space and the differential is the differential volume element $d\tau$, and so the delta function has units of m^{-3}. The units can be subtle, though. If one considers a delta function that picks out a 2D surface in 3D space (*e.g.*, for collapsing an integral of a volume charge density to one of a surface charge density), its argument will be a 3D vector, but it should have units of m^{-1} since it eliminates only one of the three dimensions. (This would be obvious if the surface were a sphere of radius a centered at $\vec{r}_o$ because then one could instead use $\delta(|\vec{r} - \vec{r}_o| - a)$, implying units of m^{-1}. Other surfaces may not be so easily defined.)

▶ Griffiths refers to the above delta function as $\delta^3(\vec{r} - \vec{r}_o)$. He does this because one can think of this delta function in terms of 1D delta functions

$$\delta^3(\vec{r} - \vec{r}_o) = \delta(x - x_o)\delta(y - y_o)\delta(z - z_o) \quad \text{where} \quad \begin{array}{l} \vec{r} = x\,\hat{x} + y\,\hat{y} + z\,\hat{z} \\ \vec{r}_o = x_o\,\hat{x} + y_o\,\hat{y} + z_o\,\hat{z} \end{array} \quad (2.10)$$

We drop the ^{3} because it is unnecessary: the dimension of the delta function is implied by its argument, the fact that it picks a single point out of 3D space. Moreover, the ^{3} notation is misleading and confusing because it suggests that δ^3 is the cube of something that has $\vec{r} - \vec{r}_o$ as its argument. It is not!

With the above, if we define the charge distribution for a set of point charges $\{q_i'\}$ at positions $\{\vec{r}_i'\}$ to be

$$\rho(\vec{r}) = \sum_{i=1}^{N} q_i' \, \delta(\vec{r} - \vec{r}_i') \tag{2.11}$$

then, when we do the integral in Equation 2.6 over any volume $\mathcal{V}$ containing all N charges, we recover the discrete version of the expression for the electric field, Equation 2.3:

$$\begin{aligned}
\vec{E}(\vec{r}) &= \frac{1}{4\pi\epsilon_o} \int_{\mathcal{V}} \sum_{i=1}^{N} \frac{d\tau' q_i' \, \delta(\vec{r}' - \vec{r}_i')}{|\vec{r} - \vec{r}'|^2} \frac{\vec{r} - \vec{r}'}{|\vec{r} - \vec{r}'|} \\
&= \frac{1}{4\pi\epsilon_o} \sum_{i=1}^{N} \int_{\mathcal{V}} d\tau' q_i' \, \delta(\vec{r}' - \vec{r}_i') \frac{\vec{r} - \vec{r}'}{|\vec{r} - \vec{r}'|^3} \\
&= \frac{1}{4\pi\epsilon_o} \sum_{i=1}^{N} q_i' \frac{\vec{r} - \vec{r}_i'}{|\vec{r} - \vec{r}_i'|^3} = \frac{1}{4\pi\epsilon_o} \sum_{i=1}^{N} \frac{q_i'}{R_i^2} \widehat{R}_i
\end{aligned} \tag{2.12}$$

Gauss's Law

Statement of Gauss's Law

The *flux* of the electric field through a surface is the integral of the component of the electric field normal to the surface over the surface:

$$\mathcal{F}_S = \int_S \vec{E} \cdot \widehat{n}(\vec{r}) \, da \qquad (2.13)$$

where $\vec{r}$ lies on the surface S and $\widehat{n}(\vec{r})$ is the surface normal at that point $\vec{r}$. Note that the flux has a sign based on the choice of the orientation of $\widehat{n}$.

Gauss's Law relates the flux of the electric field through any *closed* surface to the total charge enclosed by that surface:

$$\boxed{\mathcal{F}_S = \oint_S \vec{E} \cdot \widehat{n}(\vec{r}) \, da = \frac{1}{\epsilon_o} \int_{\mathcal{V}(S)} d\tau \, \rho(\vec{r})} \qquad (2.14)$$

where $\mathcal{V}(S)$ is the surface enclosed by S and $\oint$ indicates the integral over a closed surface. Our derivation below will take the surface normal direction to be outward from the closed volume.

Utility of Gauss's Law

Gauss's Law has three uses:

▶ For charge distributions having some amount of geometrical symmetry, it provides a way to calculate the electric field that is much easier than brute-force integration of Coulomb's Law.

▶ We will see that it will enable us to relate the electric field's *boundary conditions* at an interface between two volumes (the conditions relating the electric field components on the two sides of the interface) to the amount of charge at that interface.

▶ We can obtain a differential version of it, relating spatial derivatives of the electric field to the charge density locally. Doing so directly from Coulomb's Law is difficult (though not impossible, given what we will prove about the divergence of Coulomb's Law!).

Proof of Gauss's Law

The proof offered in Griffiths' is unnecessarily unrigorous; we follow Jackson §1.3.
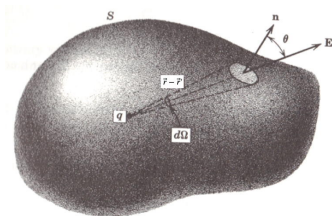
© 1999 Jackson, *Classical Electrodynamics*

First, consider a charge distribution $\rho(\vec{r})$ that lies completely inside an arbitrarily shaped closed surface $\mathcal{S}$. What is the infinitesimal flux through an infinitesimal portion da of $\mathcal{S}$ at a point $\vec{r}$ due to the infinitesimal amount of charge in the infinitesimal volume $d\tau'$ at the location $\vec{r}'$? It is

$$d^2\mathcal{F}_{\mathcal{S}}(\vec{r}, \vec{r}') = \frac{1}{4\pi\epsilon_o} \frac{d\tau'\rho(\vec{r}')}{|\vec{r} - \vec{r}'|^3} (\vec{r} - \vec{r}') \cdot \hat{n}(\vec{r})\, da \qquad (2.15)$$

The left side is a double differential because the right side is. If one considers the geometry (see diagram above), one sees that the quantity $(\vec{r} - \vec{r}') \cdot \hat{n}(\vec{r})\, da / |\vec{r} - \vec{r}'|$ is the *projected area* of the area element da normal to the unit vector $(\vec{r} - \vec{r}') / |\vec{r} - \vec{r}'|$ from $\vec{r}'$ to $\vec{r}$. Since $|\vec{r} - \vec{r}'|^2$ is the square of the distance from $\vec{r}'$ to $\vec{r}$, then the quantity $(\vec{r} - \vec{r}') \cdot \hat{n}(\vec{r})\, da / |\vec{r} - \vec{r}'|^3$ is the *solid angle* $d\Omega(\vec{r}, \vec{r}')$ subtended by da at $\vec{r}$ as viewed from $\vec{r}'$.

The corresponding mathematical formula is

$$d^2 \mathcal{F}_S(\vec{r}, \vec{r}\,') = \frac{1}{4\pi\,\epsilon_o}\, d\tau'\rho(\vec{r}\,')\, d\Omega(\vec{r}, \vec{r}\,') \tag{2.16}$$

We know that if we integrate the solid angle over the entire closed surface S surrounding our source charge point $\vec{r}\,'$, we recover 4π, so:

$$d\mathcal{F}_S(\vec{r}\,') = \frac{1}{4\pi\,\epsilon_o} \oint_S d\tau'\rho(\vec{r}\,')\, d\Omega(\vec{r}, \vec{r}\,') = \frac{1}{\epsilon_o}\, d\tau'\rho(\vec{r}\,') \tag{2.17}$$

That is, for any infinitesimal volume element $d\tau'$ at $\vec{r}\,'$, Coulomb's Law implies that the flux of the electric field due to that element through any surface S enclosing it is equal to the charge in that infinitesimal volume divided by ϵ_o.

We expect that, due to superposition, if the above is true for the flux due to an infinitesimal volume of charge, then it holds for the whole distribution of charge enclosed by $\mathcal{S}$. We can prove this by calculating the flux through $\mathcal{S}$ due to the entire charge distribution, using the fact that the distribution is fully contained inside $\mathcal{S}$ (one of our starting assumptions):

$$\mathcal{F}_\mathcal{S} = \oint_\mathcal{S} \vec{E}_\mathcal{S}(\vec{r}) \cdot \widehat{n}(\vec{r})\, da = \frac{1}{4\,\pi\,\epsilon_o} \oint_\mathcal{S} \int_{\mathcal{V}(\mathcal{S})} \frac{d\tau'\rho(\vec{r}\,')}{|\vec{r}-\vec{r}\,'|^3}\,(\vec{r}-\vec{r}\,') \cdot \widehat{n}(\vec{r})\, da$$

$$= \frac{1}{4\,\pi\,\epsilon_o} \oint_\mathcal{S} \int_{\mathcal{V}(\mathcal{S})} d\tau'\rho(\vec{r}\,')\,d\Omega(\vec{r},\vec{r}\,') = \frac{1}{4\,\pi\,\epsilon_o} \oint_\mathcal{S} \int_{\mathcal{V}(\mathcal{S})} d^2\mathcal{F}_\mathcal{S}(\vec{r},\vec{r}\,') \qquad (2.18)$$

where $\vec{E}_\mathcal{S}(\vec{r})$ is the electric field at $\vec{r}$ due to all the charge contained by $\mathcal{S}$. Note that we implicitly used superposition in the above via the formula relating $\vec{E}_\mathcal{S}(\vec{r})$ to the charge distribution. Exchanging the order of integration,

$$\mathcal{F}_\mathcal{S} = \frac{1}{4\,\pi\,\epsilon_o} \int_{\mathcal{V}(\mathcal{S})} \oint_\mathcal{S} d^2\mathcal{F}_\mathcal{S}(\vec{r},\vec{r}\,') = \frac{1}{\epsilon_o} \int_{\mathcal{V}(\mathcal{S})} d\mathcal{F}_\mathcal{S}(\vec{r}\,') = \frac{1}{\epsilon_o} \int_{\mathcal{V}(\mathcal{S})} d\tau'\rho(\vec{r}\,') \quad (2.19)$$

which is Gauss's Law.

Note how the proof depended on the $1/r^2$ dependence and superposition property of Coulomb's Law. The proof could be done in the opposite direction: Gauss's Law implies Coulomb's Law. In general, for any force, there is a simple Gauss's Law if and only if the force has a $1/r^2$ dependence. Another example is gravity, as you learned in Ph1a and Ph106a.

But we are not quite done yet, as we assumed at the start that the charge distribution vanishes outside of $\mathcal{S}$. Does the result generalize to the case where there is some charge outside of $\mathcal{S}$ so that $\vec{E}_\mathcal{S}$ receives contributions from that charge? Yes, it does.

Returning to $d^2\mathcal{F}_\mathcal{S}(\vec{r}, \vec{r}\,')$ (Equation 2.16), suppose we consider a source charge at a point $\vec{r}\,'$ that lies outside of $\mathcal{S}$. (See diagram below.) Then, for a given point $\vec{r}$ on $\mathcal{S}$ and the solid angle it subtends $d\Omega(\vec{r}, \vec{r}\,')$ as viewed from the source charge point $\vec{r}\,'$, there will be second point on $\mathcal{S}$ that has the same unit vector to the source charge point $\vec{r}\,'$ and subtends the same solid angle. But, because the direction of $\hat{n}(\vec{r})$ enters the expression for $d^2\mathcal{F}_\mathcal{S}(\vec{r}, \vec{r}\,')$, and the two points subtending the same solid angle will have opposite signs of $\hat{n}$, their two contributions cancel. Thus, the integral over $\mathcal{S}$ that yields $d\mathcal{F}_\mathcal{S}(\vec{r}\,')$ vanishes for $\vec{r}\,'$ outside of $\mathcal{S}$, and, therefore, the charge distribution at points outside of $\mathcal{S}$ do not contribute to the flux through $\mathcal{S}$, and so our derivation remains valid.
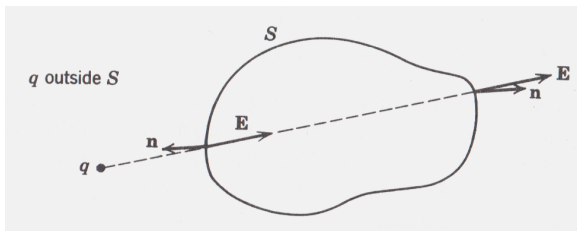
© 1999 Jackson, *Classical Electrodynamics*

The Divergence of $\vec{E}$ and the Differential Version of Gauss's Law

You learned about the divergence theorem (Gauss's theorem) in Ma1abc. Applied to $\vec{E}$, the divergence theorem says

$$\int_{\mathcal{V}(\mathcal{S})} \vec{\nabla} \cdot \vec{E}(\vec{r})\, d\tau = \oint_{\mathcal{S}} \vec{E}(\vec{r}) \cdot \hat{n}(\vec{r})\, da \qquad (2.20)$$

Gauss's Law tells us

$$\frac{1}{\epsilon_o} \int_{\mathcal{V}(\mathcal{S})} d\tau\, \rho(\vec{r}) = \oint_{\mathcal{S}} \vec{E}(\vec{r}) \cdot \hat{n}(\vec{r})\, da \qquad (2.21)$$

Combining the two, we have

$$\int_{\mathcal{V}(\mathcal{S})} \vec{\nabla} \cdot \vec{E}(\vec{r})\, d\tau = \frac{1}{\epsilon_o} \int_{\mathcal{V}(\mathcal{S})} d\tau\, \rho(\vec{r}) \qquad (2.22)$$

Since the above holds for any volume $\mathcal{V}$, the integrands must be equal, giving us the differential version of Gauss's Law:

$$\boxed{\vec{\nabla} \cdot \vec{E}(\vec{r}) = \frac{1}{\epsilon_o} \rho(\vec{r})} \qquad (2.23)$$

The differential version states that a particular combination of the spatial derivatives of the electric field at a point is related to the charge density at that point. *We will frequently employ this technique of using an equality between two integrals over an arbitrary volume or surface to conclude their integrands are equal.*

Direct Proof of Differential Version of Gauss's Law

We can prove the above differential version by simply calculating the divergence of $\vec{E}$ using Coulomb's Law, also. This is of course not independent of Gauss's Law because Gauss's Law is proven using Coulomb's Law, but it provides some exercise in vector calculus and leads us to the Dirac delta function. We take the divergence of Coulomb's Law for $\vec{E}$, noting explicitly that the divergence is a set of derivatives in $\vec{r}$:

$$\vec{\nabla}_{\vec{r}} \cdot \vec{E}(\vec{r}) = \vec{\nabla}_{\vec{r}} \cdot \int_{\mathcal{V}'} \frac{1}{4\,\pi\,\epsilon_o} \frac{d\tau'\rho(\vec{r}')}{|\vec{r}-\vec{r}'|^3} \left(\vec{r}-\vec{r}'\right) \tag{2.24}$$

The integral is over $\vec{r}'$ over the volume $\mathcal{V}'$, but the divergence is calculated in the $\vec{r}$ coordinate, so we can bring the divergence inside the integral. Note that it does not act on ρ because ρ is a function of $\vec{r}'$, not $\vec{r}$. Thus, we have

$$\vec{\nabla}_{\vec{r}} \cdot \vec{E}(\vec{r}) = \frac{1}{4\,\pi\,\epsilon_o} \int_{\mathcal{V}'} d\tau'\rho(\vec{r}')\,\vec{\nabla}_{\vec{r}} \cdot \frac{\vec{r}-\vec{r}'}{|\vec{r}-\vec{r}'|^3} \tag{2.25}$$

One could calculate the above divergence explicitly in any particular coordinate system. But it is both more rigorous and more instructive to do it using the divergence theorem, which requires us to apply a nonintuitive technique: we integrate both sides of the above equation in order to evaluate it.

We can calculate the integral of the above divergence over some arbitrary volume $\mathcal{V}$ (with surface $\mathcal{S}$, with neither $\mathcal{V}$ nor $\mathcal{S}$ necessarily related to $\mathcal{V}'$ and $\mathcal{S}'$), as we need to do for Gauss's Law, by exchanging the order of integration (no prohibition on doing so because we don't move $\vec{\nabla}_{\vec{r}}$ around) and converting the volume integral over $\vec{r}$ to an easier-to-do surface integral using the divergence theorem:

$$
\begin{aligned}
\int_{\mathcal{V}} d\tau \, \vec{\nabla}_{\vec{r}} \cdot \vec{E}(\vec{r}) &= \int_{\mathcal{V}} d\tau \, \frac{1}{4\pi\epsilon_o} \int_{\mathcal{V}'} d\tau' \rho(\vec{r}') \, \vec{\nabla}_{\vec{r}} \cdot \frac{\vec{r} - \vec{r}'}{|\vec{r} - \vec{r}'|^3} \\
&= \frac{1}{4\pi\epsilon_o} \int_{\mathcal{V}'} d\tau' \rho(\vec{r}') \int_{\mathcal{V}} d\tau \, \vec{\nabla}_{\vec{r}} \cdot \frac{\vec{r} - \vec{r}'}{|\vec{r} - \vec{r}'|^3} \\
&= \frac{1}{4\pi\epsilon_o} \int_{\mathcal{V}'} d\tau' \rho(\vec{r}') \oint_{\mathcal{S}(\mathcal{V})} da \, \widehat{n}(\vec{r}) \cdot \frac{\vec{r} - \vec{r}'}{|\vec{r} - \vec{r}'|^3}
\end{aligned} \tag{2.26}
$$

We can apply to the surface integral the same argument about solid angles that we used in proving Gauss's Law. The integrand above is just the solid angle subtended by the area element da at $\vec{r}$ as viewed from $\vec{r}'$. As before, if $\vec{r}'$ is inside $\mathcal{V}$, then the above integral yields the total solid angle, 4π. If $\vec{r}'$ is not inside of $\mathcal{V}$, then, for every area element da at $\vec{r}$, there is an area element with an equal and opposite contribution, making the integral vanish. That is, as in the proof of Gauss's Law,

$$
\oint_{\mathcal{S}(\mathcal{V})} da \, \widehat{n}(\vec{r}) \cdot \frac{\vec{r} - \vec{r}'}{|\vec{r} - \vec{r}'|^3} = \left\{ \begin{array}{ll} 4\pi & \text{if } \vec{r}' \text{ is inside } \mathcal{V} \\ 0 & \text{if } \vec{r}' \text{ is outside } \mathcal{V} \end{array} \right. \tag{2.27}
$$

The above statement says that the integral over $\mathcal{V}$ vanishes if $\vec{r}'$ is not inside $\mathcal{V}$ and yields 4π if it is inside $\mathcal{V}$. This turns the double integral over $\mathcal{V}$ and $\mathcal{V}'$ into a single integral over $\mathcal{V} \cap \mathcal{V}'$:

$$\int_{\mathcal{V}} d\tau \, \vec{\nabla}_{\vec{r}} \cdot \vec{E}(\vec{r}) = \frac{1}{4\pi\,\epsilon_o} \int_{\mathcal{V}\cap\mathcal{V}'} d\tau' \, 4\pi\,\rho(\vec{r}') = \frac{1}{\epsilon_o} \int_{\mathcal{V}\cap\mathcal{V}'} d\tau' \rho(\vec{r}') \qquad (2.28)$$

Now, consider points in $\mathcal{V}$ but outside $\mathcal{V} \cap \mathcal{V}'$. Because $\mathcal{V}'$ is the entire volume containing charge (by Coulomb's Law), the charge density vanishes in $\mathcal{V} - \mathcal{V} \cap \mathcal{V}'$. We can thus add the volume $\mathcal{V} - \mathcal{V} \cap \mathcal{V}'$ without changing the integral of the charge density because the contribution from the added volume vanishes. This changes the volume of integration from $\mathcal{V} \cap \mathcal{V}'$ to $\mathcal{V}$. Therefore,

$$\int_{\mathcal{V}} d\tau \, \vec{\nabla}_{\vec{r}} \cdot \vec{E}(\vec{r}) = \frac{1}{\epsilon_o} \int_{\mathcal{V}} d\tau' \rho(\vec{r}') \qquad (2.29)$$

The volume $\mathcal{V}$ is arbitrary, so the integrands must be equal:

$$\vec{\nabla}_{\vec{r}} \cdot \vec{E}(\vec{r}) = \frac{1}{\epsilon_o} \rho(\vec{r}) \qquad (2.30)$$

which is again the differential version of Gauss's Law.

Note the use of two nonintuitive techniques: Using the equality of integrals over an arbitrary volume to show their integrands are equal, and integrating an expression to determine what it is equal to.

Lecture 2:

Basics of Electrostatics II:
Dirac Delta Function Redux
$$\vec{\nabla} \times \vec{E} = 0$$
Electric Potential
Poisson's and Laplace's Equations

Date Revised: 2024/01/09 05:45
Date Given: 2024/01/09

Aside: Relation of the Dirac Delta Function to a Divergence, Invariance under Inversion of its Argument

Let's return to a geometric theorem we proved during the above manipulations to prove a property of the Dirac delta function. We showed above that:

$$\int_{\mathcal{V}} d\tau \, \vec{\nabla}_{\vec{r}} \cdot \frac{\vec{r} - \vec{r}'}{|\vec{r} - \vec{r}'|^3} = \oint_{\mathcal{S}(\mathcal{V})} da \, \hat{n}(r) \cdot \frac{\vec{r} - \vec{r}'}{|\vec{r} - \vec{r}'|^3} = \begin{cases} 4\pi & \text{if } \vec{r}' \text{ is inside } \mathcal{V} \\ 0 & \text{if } \vec{r}' \text{ is outside } \mathcal{V} \end{cases} \quad (2.31)$$

The far right side is proportional to the integral of the Dirac delta function, yielding

$$\int_{\mathcal{V}} d\tau \, \vec{\nabla}_{\vec{r}} \cdot \frac{\vec{r} - \vec{r}'}{|\vec{r} - \vec{r}'|^3} = 4\pi \int_{\mathcal{V}} d\tau \, \delta(\vec{r} - \vec{r}') \quad (2.32)$$

(Note the ordering of $\vec{r}$ and $\vec{r}'$ in the argument of the delta function! $\vec{r}'$ is the equivalent of $\vec{r}_o$ in Equation 2.9.) Since these integrals are equal for an arbitrary volume $\mathcal{V}$, the integrands are equal:

$$\boxed{\vec{\nabla}_{\vec{r}} \cdot \frac{\vec{r} - \vec{r}'}{|\vec{r} - \vec{r}'|^3} = 4\pi \, \delta(\vec{r} - \vec{r}')} \quad (2.33)$$

The delta function is the divergence of the $1/r^2$ law. We will find this very useful!

Since the delta function picks out the point where its argument vanishes, it doesn't matter what the sign of the argument is. One can prove this explicitly using change of variables: when the sign of the argument changes, the sign of the differential and of the limits of integration change also. Those two sign flips cancel each other. Thus

$$\boxed{\delta(\vec{r} - \vec{r}') = \delta(\vec{r}' - \vec{r})} \quad (2.34)$$

In order to introduce a new technique we will use soon, let's show this more explicitly:

$$4\pi\,\delta(\vec{r}-\vec{r}\,') = \vec{\nabla}_{\vec{r}} \cdot \frac{\vec{r}-\vec{r}\,'}{|\vec{r}-\vec{r}\,'|^3} = \vec{\nabla}_{\vec{r}} \cdot -\frac{\vec{r}\,'-\vec{r}}{|\vec{r}\,'-\vec{r}|^3} = -\vec{\nabla}_{\vec{r}} \cdot \frac{\vec{r}\,'-\vec{r}}{|\vec{r}\,'-\vec{r}|^3} \stackrel{?}{=} 4\pi\,\delta(\vec{r}\,'-\vec{r}) \qquad (2.35)$$

The last equality is not obviously true because it is not just a symbolic replacement of Equation 2.33. We can prove it by making a few more steps, though:

$$\vec{\nabla}_{\vec{r}} \cdot \frac{\vec{r}-\vec{r}\,'}{|\vec{r}-\vec{r}\,'|^3} = \vec{\nabla}_{\vec{r}-\vec{r}\,'} \cdot \frac{\vec{r}-\vec{r}\,'}{|\vec{r}-\vec{r}\,'|^3} = \left(-\vec{\nabla}_{\vec{r}\,'-\vec{r}}\right) \cdot \left(-\frac{\vec{r}\,'-\vec{r}}{|\vec{r}\,'-\vec{r}|^3}\right) = \vec{\nabla}_{\vec{r}\,'} \cdot \frac{\vec{r}\,'-\vec{r}}{|\vec{r}\,'-\vec{r}|^3} \qquad (2.36)$$

We used the following: 1) the divergence with respect to $\vec{r}-\vec{r}\,'$ is the same as the divergence with respect to $\vec{r}$ because $\vec{r}-\vec{r}\,'$ just offsets $\vec{r}$; 2) when we flip the sign on $\vec{r}-\vec{r}\,'$, we can get the argument of the divergence to match this sign flip by flipping the sign on the divergence overall (the 1D analogue would be to flip the sign on d/dx when mirroring through the origin, $x \to -x$); and, 3) $\vec{r}$ acts like an offset for $\vec{r}\,'$, and so the divergence with respect to $\vec{r}\,'-\vec{r}$ is the same as with respect to $\vec{r}\,'$. Therefore:

$$4\pi\,\delta(\vec{r}-\vec{r}\,') = \vec{\nabla}_{\vec{r}} \cdot \frac{\vec{r}-\vec{r}\,'}{|\vec{r}-\vec{r}\,'|^3} = \vec{\nabla}_{\vec{r}\,'} \cdot \frac{\vec{r}\,'-\vec{r}}{|\vec{r}\,'-\vec{r}|^3} = 4\pi\,\delta(\vec{r}\,'-\vec{r}) \qquad (2.37)$$

Note this technique of applying an offset; we will use it again.

The above derivation is a situation in which it is very important to remember what coordinate $\vec{\nabla}$ is acting on/with respect to. If we had not subscripted $\vec{\nabla}$ in the manipulation above, we would have written:

$$4\pi\,\delta(\vec{r}-\vec{r}') = \vec{\nabla}\cdot\frac{\vec{r}-\vec{r}'}{|\vec{r}-\vec{r}'|^3} = \vec{\nabla}\cdot-\frac{\vec{r}'-\vec{r}}{|\vec{r}'-\vec{r}|^3} = -\vec{\nabla}\cdot\frac{\vec{r}'-\vec{r}}{|\vec{r}'-\vec{r}|^3} \overset{?}{=} -4\pi\,\delta(\vec{r}'-\vec{r}) \qquad (2.38)$$

Rather than just yielding a last equality that was not obviously true, we have obtained an equality that seems to contradict the symmetry property! From the more detailed derivation on the previous page, we see that the seeming contradiction vanishes when we take care with the meaning of symbols by explicitly subscripting $\vec{\nabla}$. Errors of the above type are easy to make and not self-evident! Mathematics in physics is not just symbol manipulation: there is meaning that must be understood in order to be sure those manipulations are justified.

The Electric Field has Vanishing Curl

Calculating the Curl of the Electric Field

The curl of $\vec{E}$ can be shown to vanish simply by calculating it for an arbitrary charge distribution:

$$
\begin{aligned}
\vec{\nabla} \times \vec{E}(\vec{r}) &= \frac{1}{4\,\pi\,\epsilon_o}\,\vec{\nabla} \times \int_{\mathcal{V}} d\tau'\,\rho(\vec{r}')\,\frac{\vec{r} - \vec{r}'}{|\vec{r} - \vec{r}'|^3} \\
&= \frac{1}{4\,\pi\,\epsilon_o}\,\int_{\mathcal{V}} d\tau'\,\rho(\vec{r}')\,\vec{\nabla} \times \frac{\vec{r} - \vec{r}'}{|\vec{r} - \vec{r}'|^3}
\end{aligned}
\tag{2.39}
$$

We could brute-force calculate the curl in the integrand in Cartesian or spherical coordinates, but that would be painful because the function on which the curl is acting has no symmetry in the $\vec{r}$ coordinate system.

Let's take a simpler, more geometric, and more intuitive approach. As we saw above, $\vec{r}'$ is just an offset to $\vec{r}$, thus

$$\vec{\nabla}_{\vec{r}} \times \frac{\vec{r} - \vec{r}'}{|\vec{r} - \vec{r}'|^3} = \vec{\nabla}_{\vec{r} - \vec{r}'} \times \frac{\vec{r} - \vec{r}'}{|\vec{r} - \vec{r}'|^3} \qquad (2.40)$$

Note that, in doing this offset, the curl will be expressed in terms of the components of $\vec{r} - \vec{r}'$. This does not change the bounds of integration, but it may make the expression look complicated because the variable of integration is still $\vec{r}'$. Since we will show this expression, the integrand, vanishes, this bookkeeping complication is not important. If we define $\vec{s} = \vec{r} - \vec{r}'$, then we have

$$\vec{\nabla}_{\vec{r}} \times \frac{\vec{r} - \vec{r}'}{|\vec{r} - \vec{r}'|^3} = \vec{\nabla}_{\vec{s}} \times \frac{\vec{s}}{s^3} \qquad (2.41)$$

Now, the function on which the curl is acting has symmetry in the coordinate system in which the curl is acting, and hence the calculation will be simplified. You can probably see intuitively that the above curl vanishes, but let's prove it. (Note also that the change of variables would require a change to the limits of integration, but, again, because we will prove the integrand will vanish, this bookkeeping complication will not be important.)

With the above form, we can trivially apply the formula for the curl in spherical coordinates, which is listed in Griffiths. For the sake of being explicit, that formula is (with r replaced by s as the radial coordinate to avoid confusion)

$$\vec{\nabla} \times \vec{v} = \frac{1}{s}\frac{1}{\sin\theta}\left[\frac{\partial}{\partial\theta}\left(v_\phi \sin\theta\right) - \frac{\partial v_\theta}{\partial\phi}\right]\hat{s} + \frac{1}{s}\left[\frac{1}{\sin\theta}\frac{\partial v_s}{\partial\phi} - \frac{\partial}{\partial s}\left(s\, v_\phi\right)\right]\hat{\theta}$$

$$+ \frac{1}{s}\left[\frac{\partial}{\partial s}\left(s\, v_\theta\right) - \frac{\partial v_s}{\partial\theta}\right]\hat{\phi} \tag{2.42}$$

In the case considered here, s is the radial variable and the radial component of $\vec{s}/s^3$ is $1/s^2$. Thus, $\vec{v}$ has only a radial component and that radial component depends only on the radial distance from the origin. All the derivatives involving the θ and ϕ components of $\vec{v}$ vanish because the components themselves vanish, and the derivatives involving the radial component vanish because those derivatives are with respect to θ and ϕ. (Don't be confused: $\vec{v}$ itself depends on θ and ϕ because the direction of $\vec{v}$ depends on them; but the curl formula takes care of that dependence.)

Thus, we have $\vec{\nabla}_{\vec{s}} \times (\vec{s}/s^3) = 0$ and the integrand in Equation 2.39 vanishes. So:

$$\boxed{\vec{\nabla} \times \vec{E}(\vec{r}) = 0} \tag{2.43}$$

Note again that we did not brute-force differentiate, but rather we thought about how to simplify the calculational aspect (via origin offset) and then saw that made the result both geometrically/intuitively obvious and easier to demonstrate via calculation.

The Line Integral of the Electric Field

Stokes' Theorem (a mathematical theorem we will not prove here but that you saw in Ma1abc) then tells us that, for any surface $\mathcal{S}$ with boundary $\mathcal{C}(\mathcal{S})$,

$$\oint_{\mathcal{C}(\mathcal{S})} d\vec{\ell} \cdot \vec{E}(\vec{r}) = \int_{\mathcal{S}} da\,\hat{n}(\vec{r}) \cdot \left[\vec{\nabla} \times \vec{E}(\vec{r})\right] = 0 \tag{2.44}$$

The Electric Potential

Electric Potential Definition using Line Integral
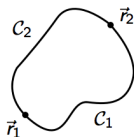
We used above the fact that the line integral of the electric field around any closed loop $\mathcal{C}$ vanishes. If we consider two points along the loop $\vec{r}_1$ and $\vec{r}_2$, $\mathcal{C}$ defines two paths along the loop from $\vec{r}_1$ to $\vec{r}_2$, $\mathcal{C}_1$ and $\mathcal{C}_2$. Let's difference the line integrals along these two paths, using the vanishing of the line integral around the loop to see that the difference vanishes:

$$\int_{\mathcal{C}_1, \vec{r}_1}^{\vec{r}_2} d\vec{\ell} \cdot \vec{E}(\vec{r}) - \int_{\mathcal{C}_2, \vec{r}_1}^{\vec{r}_2} d\vec{\ell} \cdot \vec{E}(\vec{r}) = \int_{\mathcal{C}_1, \vec{r}_1}^{\vec{r}_2} d\vec{\ell} \cdot \vec{E}(\vec{r}) + \int_{\mathcal{C}_2, \vec{r}_2}^{\vec{r}_1} d\vec{\ell} \cdot \vec{E}(\vec{r})$$

$$= \oint_{\mathcal{C}} d\vec{\ell} \cdot \vec{E}(\vec{r}) = 0 \qquad (2.45)$$

(Be careful about the endpoint ordering and signs of the two terms! The differential $d\vec{\ell}$ in a line integral has no intrinsic polarity; the polarity is set by the ordering of the endpoints. The sign of $d\vec{\ell}$ and of the endpoint ordering do not multiply; they are redundant.) Therefore,

$$\int_{\mathcal{C}_1, \vec{r}_1}^{\vec{r}_2} d\vec{\ell} \cdot \vec{E}(\vec{r}) = \int_{\mathcal{C}_2, \vec{r}_1}^{\vec{r}_2} d\vec{\ell} \cdot \vec{E}(\vec{r}) \qquad (2.46)$$

The above relation tells us that the value of the above line integral depends only on the location of its endpoints, not on the path taken. Thus, we can construct a function, *the electric potential*, $V(\vec{r})$, defining it via its differences between points:

$$V(\vec{r}_2) - V(\vec{r}_1) \equiv -\int_{\vec{r}_1}^{\vec{r}_2} d\vec{\ell} \cdot \vec{E}(\vec{r}) \qquad (2.47)$$

The fundamental theorem of calculus for line integrals in multiple dimensions implies

$$V(\vec{r}_2) - V(\vec{r}_1) = \int_{\vec{r}_1}^{\vec{r}_2} d\vec{\ell} \cdot \vec{\nabla} V(\vec{r}) \qquad (2.48)$$

where $\vec{\nabla} V(\vec{r})$ is the *gradient* of the electric potential. The above two formulae hold regardless of choice of endpoints and path, so the integrands are equal and we have

$$\vec{E}(\vec{r}) = -\vec{\nabla} V(\vec{r}) \qquad (2.49)$$

which can be viewed as an alternate definition of the potential. The offset of $V(\vec{r})$ is not defined because it has no influence on $\vec{E}(\vec{r})$, which is the quantity we began with from Coulomb's Law.

The electric potential has units of $(\text{N m}/\text{C}) = (\text{J}/\text{C})$, which we call the *volt*, V. (The appearance of J will be important when we discuss electric potential energy.) The electric field is frequently written in units of V/m instead of N/C.

It will be useful in homework to remember the mathematical definition of a gradient (another case of keeping present in your mind what the mathematical symbols mean):

$$\lim_{|\delta\vec{r}|\to 0} V(\vec{r}+\delta\vec{r}) - V(\vec{r}) = \lim_{|\delta\vec{r}|\to 0} \vec{\nabla}V(\vec{r}) \cdot \delta\vec{r} = -\lim_{|\delta\vec{r}|\to 0} \vec{E}(\vec{r}) \cdot \delta\vec{r} \qquad (2.50)$$

The above expression also serves to remind you of how to do a Taylor expansion in multiple dimensions: the dot product of the gradient and the vector differential replaces the product of the one-dimensional derivative and the scalar differential.

Relation of the Electric Potential to the Charge Distribution

We know two things now:

$$\vec{E}(\vec{r}) = \frac{1}{4\pi\epsilon_o} \int_{\mathcal{V}} d\tau' \rho(\vec{r}') \frac{\vec{r} - \vec{r}'}{|\vec{r} - \vec{r}'|^3} \qquad \text{and} \qquad V(\vec{r}_2) - V(\vec{r}_1) \equiv -\int_{\vec{r}_1}^{\vec{r}_2} d\vec{\ell} \cdot \vec{E}(\vec{r})$$

We can use these to obtain an explicit expression for the potential in terms of the charge distribution. In practice, trying to do the line integral explicitly using the definition of $\vec{E}$ is tedious and not illuminating.

Instead, let us use our understanding of the *meaning* of the mathematical expression $\vec{E}(\vec{r}) = -\vec{\nabla} V(\vec{r})$ to make an Ansatz. If we have a point charge at the origin, then the electric field points radially outward and falls off as $1/r^2$. What function's derivative gives that dependence? $V(\vec{r}) = 1/r$. This suggests to us

$$\boxed{V(\vec{r}) = \frac{1}{4\pi\epsilon_o} \int_{\mathcal{V}} d\tau' \rho(\vec{r}') \frac{1}{|\vec{r} - \vec{r}'|}} \qquad (2.51)$$

We may then prove explicitly this form is correct by taking the gradient.

First, pass $\vec{\nabla}$ inside the integral because it is $\vec{\nabla}_{\vec{r}}$ while the variable of integration is $\vec{r}'$:

$$-\vec{\nabla}_{\vec{r}} V(\vec{r}) = -\frac{1}{4\pi\epsilon_o} \int_{\mathcal{V}} d\tau' \rho(\vec{r}') \, \vec{\nabla}_{\vec{r}} \left(\frac{1}{|\vec{r} - \vec{r}'|} \right) \tag{2.52}$$

As we did earlier when calculating $\vec{\nabla} \times \vec{E}$, we change variables to $\vec{s} = \vec{r} - \vec{r}'$ to evaluate the gradient:

$$\vec{\nabla}_{\vec{r}} \left(\frac{1}{|\vec{r} - \vec{r}'|} \right) = \vec{\nabla}_{\vec{r} - \vec{r}'} \left(\frac{1}{|\vec{r} - \vec{r}'|} \right) = \vec{\nabla}_{\vec{s}} \frac{1}{s} = -\frac{\hat{s}}{s^2} = -\frac{\vec{r} - \vec{r}'}{|\vec{r} - \vec{r}'|^3} \tag{2.53}$$

where we used the formula for the gradient in spherical coordinates from Griffiths (again, with r replaced by s as the radial coordinate to avoid confusion):

$$\vec{\nabla}_{\vec{s}} T(\vec{s}) = \frac{\partial T}{\partial s} \, \hat{s} + \frac{1}{s} \frac{\partial T}{\partial \theta} \, \hat{\theta} + \frac{1}{s} \frac{1}{\sin\theta} \frac{\partial T}{\partial \phi} \, \hat{\phi} \tag{2.54}$$

We see that our form for $V(\vec{r})$ yields the correct electric field (dropping the $\vec{r}$ subscript on $\vec{\nabla}$ because, on the left-hand side, it is obvious what coordinate $\vec{\nabla}$ acts on):

$$-\vec{\nabla} V(\vec{r}) = \frac{1}{4\pi\epsilon_o} \int_{\mathcal{V}} d\tau' \rho(\vec{r}') \frac{\vec{r} - \vec{r}'}{|\vec{r} - \vec{r}'|^3} = \vec{E}(\vec{r}) \tag{2.55}$$

Comments on the Electric Potential

► **The electric potential obeys superposition**
This is trivial consequence of superposition for the electric field: because the electric potential is a linear function of the electric field, and integration is a linear operation, superposition for the electric field transfers to superposition for the electric potential. One can also see it from Equation 2.51, where the potential is a linear function of the charge density.

► **Definition of potential offset**
There are two typical choices. When the charge distribution is confined to a finite volume, the electric field vanishes at infinity, which suggests one should define the electric potential to vanish at infinity too. When the charge distribution is not confined (*e.g.*, a uniform electric field over all of space), it is typical to choose the origin to be the point at which the potential vanishes. Any other point would work, too, but will generally make the explicit functional form of $V(\vec{r})$ unnecessarily complicated if one is interested in using the above integral expression. There will be situations, however, where such a choice is the most convenient.

► **Utility of the electric potential**
The electric potential is scalar, not a vector, function, and thus applying superposition to calculate the potential due to a charge distribution, followed by taking the gradient to find the electric field, is usually much simpler than explicitly calculating the electric field.

Aside on Techniques

It is important to recognize how we almost uniformly avoided brute-force calculations of divergences, curls, and gradients so far. The only times we did those calculations explicitly were when we had rendered the calculations trivial. A key part of doing E&M successfully and with minimal pain is avoiding algebra and calculus whenever possible and instead making use of clever arguments of the type we used above. Only do algebra and calculus as a last resort! There are two reasons for this.

First, the kinds of arguments we used are more physical and help you develop intuition. For example, in proving the differential version of Gauss's Law, at no point did we explicitly take derivatives of $\vec{E}$! Incredible, right? Instead, we proved that the divergence of the $1/r^2$ law is the delta function (again, not explicitly, but by referring to the geometric proof we made for the integral version of Gauss's Law) and used that fact. We could have done the brute-force calculation in Cartesian coordinates, and it would have given the same result. But you would have derived no intuition from it.

Second, brute-force calculations are prone to oversights — like the one about the sign flip on $\vec{\nabla}$ in the delta-function symmetry derivation — as well as bookkeeping mistakes — algebraic sign flips, misapplications of the product and chain rules, etc. Doing brute-force calculations does not help you understand physics, or even mathematics. Of course, sometimes brute-force calculations are needed, but try to avoid them, and keep your wits and intuition about you as you do them!

It takes time to learn how to work this way, but we do derivations (rather than just quote results) so you can learn these techniques.

Boundary Conditions on the Electric Field and Potential

While Gauss's Law makes it possible to determine the electric field for charge distributions with sufficient symmetry, the more important application of Gauss's Law and the vanishing of $\vec{\nabla} \times \vec{E}$ is to obtain generic information on the behavior of the electric field and potential across an interface between two regions.

Boundary Condition on the Normal Component of the Electric Field
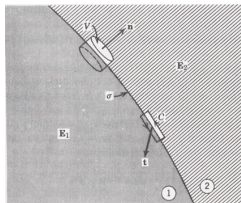
© 1999 Jackson, *Classical Electrodynamics*

Construct a Gaussian cylinder of infinitesimal height dz whose axis is normal to the interface under question at the point of interest. Let $\widehat{n}$ be the surface normal at $\vec{r}$, with orientation from region 1 to region 2. Let's calculate the flux through the cylinder's (non-infinitesimal) faces $\mathcal{S}_1$ and $\mathcal{S}_2$:

$$\mathcal{F} = \int_{\mathcal{S}_1} da\,(-\widehat{n}(\vec{r})) \cdot \vec{E}_1(\vec{r})$$

$$+ \int_{\mathcal{S}_2} da\,\widehat{n}(\vec{r}) \cdot \vec{E}_2(\vec{r}) \qquad (2.56)$$

where $\vec{E}$ is evaluated over the two faces. We neglect the flux through the cylindrical wall because we will let dz vanish in the end and so its area will vanish and it will contribute no flux. We momentarily make the assumption that there is no charge density that is singular in the direction parallel to the interface — *i.e.*, point charges or a line charge density — so that we don't have to worry about possible singularities in the electric field that might complicate the flux calculation. We allow only a surface charge density, which is delta-function singular in the z dimension but not in the dimensions parallel to the interface.

For Gauss's Law, the volume integral of the charge density enclosed has two contributions: from the non-delta-function-like *volume* charge density in the half-cylinders and from any delta-function-like *surface* charge density on the surface. The contribution of the former will vanish as we let $dz \to 0$. The latter converts the volume integral to a surface integral:

$$\mathcal{F} = \frac{1}{\epsilon_o} \int_{\mathcal{V}} d\tau \, \delta(\vec{r} - \mathcal{S}) \, \sigma(\vec{r}) = \frac{1}{\epsilon_o} \int_{\mathcal{S}} da \, \sigma(\vec{r}) \tag{2.57}$$

where $\mathcal{S}$ is the area at the interface intersected by the cylinder. (Note that this is a case where the delta function's argument requires some interpretation to understand the delta function's units. It is the $\mathcal{S}$ in the argument that implies the function has units of m^{-1} rather than m^{-3}: it is picking out a surface rather than a point and thus changing the units by one power of distance, not three.) Equating the two expressions for $\mathcal{F}$, letting $dz \to 0$, and seeing that $\mathcal{S}_1, \mathcal{S}_2 \to \mathcal{S}$ as $dz \to 0$ in the flux integral yields

$$\int_{\mathcal{S}} da \, \widehat{n}(\vec{r}) \cdot \left[\vec{E}_2(\vec{r}) - \vec{E}_1(\vec{r}) \right] = \frac{1}{\epsilon_o} \int_{\mathcal{S}} da \, \sigma(\vec{r}) \tag{2.58}$$

This holds for any choice of cylinder and thus any $\mathcal{S}$, so the integrands must be equal:

$$\boxed{\widehat{n}(\vec{r}) \cdot \left[\vec{E}_2(\vec{r}) - \vec{E}_1(\vec{r}) \right] = \frac{1}{\epsilon_o} \, \sigma(\vec{r})} \tag{2.59}$$

That is, the change in the normal component of the electric field across the interface is proportional to the surface charge density at the interface. If there is no surface charge at the interface, this component of the electric field must be continuous.

Let's now reconsider the condition we placed at the start of the derivation, that there be no charge density at the intersection of the cylinder and interface that is singular in the dimension parallel to the interface, which could consist of a set of point charges and/or a line charge density.

If that charge density is not at the contour $\mathcal{C}$ consisting of the intersection of the cylinder's wall and the interface, then the flux of its field remains entirely calculable. It may cause $\sigma(\vec{r})$ to have a delta-function singularity in one or two dimensions parallel to the interface, but no part of the derivation fails. We simply allow that type of $\sigma(\vec{r})$ in Equation 2.59.

If the parallel-dimension-singular charge density is on $\mathcal{C}$, then things are bit more complicated. If we consider the flux through the cylindrical wall anywhere but on the charge density, that flux vanishes because the field of the charge density is always parallel to the cylindrical wall as $dz \to 0$. What about *on* the charge density? Answering this question in a mathematically *explicit* fashion — *i.e.*, by calculation — is difficult, as the field not only becomes singular at this point but the direction of the singular field depends on the direction from which one approaches the charge. One can, however, conclude from this indeterminancy that there cannot be a contribution to the flux, as it would imply that the field direction is *not* indeterminate so that $\hat{n} \cdot \vec{E}$ can be nonzero. This is a mathematically valid proof by contradiction. Thus, such charge distributions do not affect the derivation and Equation 2.59 continues to hold.

We will see the above expectation confirmed in practice when we compare the potential for the point charge near the grounded sphere derived by method of images (which does not rely on Equation 2.59 in the case of such a singular charge distribution) and by separation of variables (which does).

Boundary Condition on the Tangential Component of the Electric Field
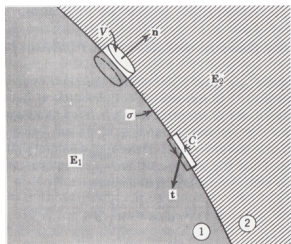
© 1999 Jackson, *Classical Electrodynamics*

Construct a rectangular loop $\mathcal{C}$ with two legs normal to the interface of interest (*i.e.*, along $\widehat{n}(\vec{r})$ at positions $\vec{r}_a$ and $\vec{r}_b$) having infinitesimal length dz and two (non-infinitesimal) legs parallel to the interface $\mathcal{C}_1$ and $\mathcal{C}_2$. Let $\widehat{t}(\vec{r})$ denote the normal to the loop area (so $\widehat{n}(\vec{r}) \cdot \widehat{t}(\vec{r}) = 0$). $\widehat{t}$ will set the orientation of the line integral we will do around the loop following the right-hand rule. The loop legs $\mathcal{C}_1$ and $\mathcal{C}_2$ parallel to the interface are parallel to the vector $\widehat{s}(\vec{r}) = \widehat{t}(\vec{r}) \times \widehat{n}(\vec{r})$. Let's calculate the line integral of $\vec{E}$ along this loop (referencing the diagram: $\vec{r}_a$ at the lower right, $\vec{r}_b$ at the upper left):

$$\oint_{\mathcal{C}} d\vec{\ell} \cdot \vec{E}(\vec{r}) = \int_{\mathcal{C}_1, \vec{r}_b - \widehat{n}(\vec{r})\frac{dz}{2}}^{\vec{r}_a - \widehat{n}(\vec{r})\frac{dz}{2}} \vec{E}_1(\vec{r}) \cdot d\vec{\ell} + \int_{\mathcal{C}_2, \vec{r}_a + \widehat{n}(\vec{r})\frac{dz}{2}}^{\vec{r}_b + \widehat{n}(\vec{r})\frac{dz}{2}} \vec{E}_2(\vec{r}) \cdot d\vec{\ell} \qquad (2.60)$$

where we neglect the contributions from the infinitesimal legs because they will vanish as $dz \to 0$. (We may apply arguments similar those just used in the derivation of the normal field boundary condition to show that these legs contribute nothing even in the case of a charge density at the interface with a delta-function singularity in the dimension parallel to the interface.)

Be careful about the signs of the integrals: $d\vec{\ell}$ for an open contour acquires its orientation from the the ordering of the endpoints; it has no intrinsic orientation until this ordering is specified. Therefore, the sign of $d\vec{\ell}$ and of the endpoint ordering do not multiply; they are redundant. Specifically, in this case, the endpoints imply that $d\vec{\ell}$ points along $+\hat{s}$ for the second term and $-\hat{s}$ for the first term and thus that the integrands have opposite sign. Do not then think that the opposite polarity of the endpoint ordering of the two terms implies another relative sign between the two integrals, with the two relative signs canceling!

The vanishing of the curl of the electric field implies the left side of the equation is zero.

We can combine the two terms on the right side by changing the endpoint ordering on the first term and recognizing that $\mathcal{C}_1 \to \mathcal{C}_2$ as $dz \to 0$ (remember: $\mathcal{C}_1$ and $\mathcal{C}_2$ themselves have no orientation: the orientation of the line integrals is set by the ordering of the endpoints). Thus, we have

$$0 = -\int_{\mathcal{C}_1, \vec{r}_a - \hat{n}(r)\frac{dz}{2}}^{\vec{r}_b - \hat{n}(r)\frac{dz}{2}} \vec{E}_1(\vec{r}) \cdot d\vec{\ell} + \int_{\mathcal{C}_2, \vec{r}_a + \hat{n}(r)\frac{dz}{2}}^{\vec{r}_b + \hat{n}(r)\frac{dz}{2}} \vec{E}_2(\vec{r}) \cdot d\vec{\ell} \xrightarrow{dz \to 0} \int_{\mathcal{C}_2, \vec{r}_a}^{\vec{r}_b} \left[\vec{E}_2(\vec{r}) - \vec{E}_1(\vec{r}) \right] \cdot d\vec{\ell}$$

With this ordering of the endpoints, we may identify $d\vec{\ell} = \hat{s}(\vec{r})\, ds$. Since the contour $\mathcal{C}_2$ is arbitrary, the integrand must vanish, yielding

$$\boxed{\hat{s}(\vec{r}) \cdot \left[\vec{E}_2(\vec{r}) - \vec{E}_1(\vec{r}) \right] = 0} \tag{2.61}$$

This expression holds for any $\hat{t}$ and thus $\hat{s}$ parallel to the surface, so it tells us that the tangential component of the electric field is continuous across any boundary (regardless of whether there is surface charge present).

Boundary Conditions on the Electric Potential

From our definition of the electric potential as the line integral of the electric field, and the corollary $\vec{E} = \vec{\nabla} V$, we can derive boundary conditions on the electric potential:

▶ **Continuity of the electric potential**
The electric potential is the line integral of the electric field. If we think about calculating the discontinuity in V by integrating $\vec{E} \cdot \hat{n} \, d\ell$ across the boundary, we recognize that, as the length of the path goes to zero, the only way to prevent the integral from vanishing is if $\vec{E} \cdot \hat{n}$ is not only nonzero but delta-function singular. The only place that can conceivably happen is at a point where a charge density becomes singular in at least one dimension (point charge or linear or surface charge density). In the same way as we argued in the derivation of the normal field boundary condition, Equation 2.59, we may also argue here that this quantity $\vec{E} \cdot \hat{n}$ still vanishes and thus V is always continuous.

We do note that, while V could become infinite near these charge densities, it must approach infinity from both sides of the boundary in the same way, and thus it remains continuous. We will see this in the example of the point charge near the grounded sphere when we do separation of variables in spherical coordinates.

▶ **Change in the normal gradient**
  This is just a direct rewriting of the boundary condition on the normal
  component of the field, Equation 2.59:

$$\frac{1}{\epsilon_o}\,\sigma(\vec{r}) = \hat{n}(\vec{r}) \cdot \left[\vec{E}_2(\vec{r}) - \vec{E}_1(\vec{r})\right] = \hat{n}(\vec{r}) \cdot \left[-\vec{\nabla}V_2(\vec{r}) + \vec{\nabla}V_1(\vec{r})\right]$$

$$\implies \boxed{\hat{n}(\vec{r}) \cdot \left[\vec{\nabla}V_2(\vec{r}) - \vec{\nabla}V_1(\vec{r})\right] = -\frac{1}{\epsilon_o}\,\sigma(\vec{r})} \qquad (2.62)$$

Note the sign!

▶ **Continuity of the tangential gradient**
  Again, this follows directly from the continuity of the tangential component of
  the electric field, Equation 2.61:

$$0 = \hat{s}(\vec{r}) \cdot \left[\vec{E}_2(\vec{r}) - \vec{E}_1(\vec{r})\right] = \hat{s}(\vec{r}) \cdot \left[-\vec{\nabla}V_2(\vec{r}) + \vec{\nabla}V_1(\vec{r})\right]$$

$$\implies \boxed{\hat{s}(\vec{r}) \cdot \left[\vec{\nabla}V_2(\vec{r}) - \vec{\nabla}V_1(\vec{r})\right] = 0} \qquad (2.63)$$

Poisson's and Laplace's Equations

It is natural to rewrite Gauss's Law in terms of the electric potential:

$$\frac{1}{\epsilon_o} \rho(\vec{r}) = \vec{\nabla} \cdot \vec{E}(\vec{r}) = -\nabla^2 V(\vec{r}) \tag{2.64}$$

Rewritten more cleanly:

$$\boxed{\nabla^2 V(\vec{r}) = -\frac{1}{\epsilon_o} \rho(\vec{r})} \tag{2.65}$$

This is known as *Poisson's Equation*.

Poisson's Equation is a partial differential equation. You know from basic calculus that a differential equation alone is not sufficient to obtain a full solution $V(\vec{r})$: constants of integration are required. For partial differential equations in multiple dimensions, the constants of integration are given by specifying *boundary conditions*, conditions for how the solution or its derivatives must behave on the boundary of the volume in which we are specifying $\rho(\vec{r})$ and would like to determine $V(\vec{r})$.

Our expression for the potential in terms of the charge distribution, Equation 2.51, is the explicit solution to this equation *for a particular boundary condition*, $V(\vec{r}) \to 0$ as $r \to \infty$. Section 3.11 will develop the concept of a Green Function, which is the generic tool for solving Poisson's Equation for arbitrary boundary conditions.

When there is no charge and the right side vanishes, Equation 2.65 is known as *Laplace's Equation*. The importance of this equation is that it implies that, in a region where there is no charge, the second derivative vanishes everywhere, which implies there can be no local maxima or minima (they would require a positive or negative second derivative). We will prove this explicitly in Section 3.1.

For completeness, let's also rewrite the curl-freeness of the electric field in terms of the electric potential. There is a mathematical theorem that the curl of a gradient always vanishes:

$$\vec{\nabla} \times (-\vec{\nabla} V) = 0 \qquad (2.66)$$

This is not surprising, as the vanishing of the curl of $\vec{E}$ is the mathematical property of $\vec{E}$ that allowed us to define the potential as a line integral, which then allowed us to write $\vec{E}$ as the gradient of the potential. The above must be true for self-consistency.

Lecture 3:

Basics of Electrostatics III:
Electric Potential Energy
Conductors

Date Revised: 2024/01/16 06:00
Revised lecture break
Date Given: 2024/01/11

Electrostatic Energy

Electric Potential Energy of a Point Charge in an Electric Field

Consider moving a point charge from $\vec{r_1}$ to $\vec{r_2}$ along a contour $\mathcal{C}$. The work done *on* the charge by the *mechanical* force pushing it in this direction is given by doing the line integral of the *negative* of the electric force along the path because that is the mechanical force that has to be exerted to move the charge against the electric force $\vec{F}_e$:

$$W_{12} = -\int_{\mathcal{C},\vec{r_1}}^{\vec{r_2}} d\vec{\ell} \cdot \vec{F}_e(\vec{r}) \tag{2.67}$$

The force is related to the electric field, and so we have

$$W_{12} = -q \int_{\mathcal{C},\vec{r_1}}^{\vec{r_2}} d\vec{\ell} \cdot \vec{E}(\vec{r}) = q \left[V(\vec{r_2}) - V(\vec{r_1}) \right] \tag{2.68}$$

That is, the work done on the charge by the mechanical force in going from $\vec{r_1}$ to $\vec{r_2}$ is given by the charge times the change in electric potential between the two positions. Note the sign: if the potential is higher at the end point, then the *mechanical* work done was positive.

Of course, this lets us to define the *electric potential energy* by

$$U(\vec{r_2}) - U(\vec{r_1}) = q \left[V(\vec{r_2}) - V(\vec{r_1}) \right] \qquad (2.69)$$

That is, the electric potential energy of the charge and the electric potential of the field are simply related. Since it was defined in terms of work done against a force, electric potential energy obviously has units of Joules (J). That is explicit in the above form, which is C (N m/C) = (N m) = J.

Note that the electric field can also do work on the charge. In this case, the sign in the above line integral for the work is flipped and work is done as the charge loses potential energy. In this case, the work done by the electric field on a charge is what gives it the kinetic energy it has at the end: the electric potential energy is converted to mechanical kinetic energy.

We are not going to write an expression for the potential energy of a continuous distribution of charge quite yet because it is difficult to distinguish two pieces: the change in potential energy that arises from moving the charge distribution in an external potential, and the change in potential energy due simply to a reconfiguration of the charge distribution in its own potential, and we will discuss the latter next.

Electric Potential Energy of a Charge Distribution

How much work must be done to assemble a distribution of charge? This energy is most easily understood by first considering the assembly of a set of point charges one-by-one by bringing them in from infinity. When the ith charge is brought in, work must be done against the electric field of the first $i - 1$ charges. Put another way, the ith charge starts with zero potential energy and ends with potential energy

$$U_i = \sum_{j=1}^{i-1} q_i \frac{1}{4\pi\epsilon_o} \frac{q_j}{|\vec{r}_i - \vec{r}_j|} \tag{2.70}$$

Thus, the total potential energy is

$$U = \frac{1}{4\pi\epsilon_o} \sum_{i=1}^{N} \sum_{j=1}^{i-1} \frac{q_i q_j}{|\vec{r}_i - \vec{r}_j|} = \frac{1}{8\pi\epsilon_o} \sum_{i,j=1, i \neq j}^{N} \frac{q_i q_j}{|\vec{r}_i - \vec{r}_j|} \tag{2.71}$$

where the factor of $1/2$ was introduced to allow i and j to both run from 1 to N. Generalizing this to a continuous charge distribution, we have

$$U = \frac{1}{8\pi\epsilon_o} \int_{\mathcal{V}} d\tau \int_{\mathcal{V}} d\tau' \frac{\rho(\vec{r})\, \rho(\vec{r}')}{|\vec{r} - \vec{r}'|} \tag{2.72}$$

In principle, the above expression is valid for any configuration of charge and potential. But sometimes we do not specify the charge distribution sourcing part of the potential — e.g., a uniform electric field over all of space — and, instead, we just state what the potential is. We can then generalize the above expression to:

$$U = \frac{1}{8\,\pi\,\epsilon_o} \int_{\mathcal{V}} d\tau \int_{\mathcal{V}} d\tau' \, \frac{\rho(\vec{r})\,\rho(\vec{r}\,')}{|\vec{r} - \vec{r}\,'|} + \int_{\mathcal{V}} d\tau\,\rho(\vec{r})\,V_{ext}(\vec{r}) \qquad (2.73)$$

where $V_{ext}(\vec{r})$ is the potential sourced by charges not explicitly indicated.

One has to be a bit careful with such situations, though, for two reasons. First, the potential energy provided is not the total electric potential energy of the system because it does not include the energy needed to put $V_{ext}(\vec{r})$ in place, only the energy needed to put $\rho(\vec{r})$ together and to put it in $V_{ext}(\vec{r})$. More importantly, because the charges sourcing $V_{ext}(\vec{r})$ are at infinity, it is not possible to start $\rho(\vec{r})$ out at infinity with zero potential energy there and then bring it into the charge configuration. We define the zero point for $V_{ext}(\vec{r})$ for convenience, e.g., at the origin, but only changes in the second term above are meaningful; the actual value of the second term is not.

Electric Potential Energy in Terms of the Electric Field

We can use the relations between potential, field, and charge density (Equations 2.6, 2.51, and 2.65) and the divergence theorem (Equation 2.20) to obtain an alternate expression for the electric potential energy in terms of the electric field as follows (now disallowing any "external potentials" $V_{ext}(\vec{r})$ of the type mentioned above, requiring us to have knowledge of all charge distributions sourcing potentials):

$$U = \frac{1}{8\pi\epsilon_o} \int_{\mathcal{V}} d\tau \int_{\mathcal{V}} d\tau' \frac{\rho(\vec{r})\,\rho(\vec{r}')}{|\vec{r}-\vec{r}'|} = \frac{1}{2}\int_{\mathcal{V}} d\tau \rho(\vec{r})\,V(\vec{r}) = -\frac{\epsilon_o}{2}\int_{\mathcal{V}} d\tau\left[\nabla^2 V(\vec{r})\right]V(\vec{r})$$

$$\overset{ibp}{=} -\frac{\epsilon_o}{2}\int_{\mathcal{V}} d\tau\,\vec{\nabla}\cdot\left[V(\vec{r})\,\vec{\nabla}V(\vec{r})\right] + \frac{\epsilon_o}{2}\int_{\mathcal{V}}|\vec{\nabla}V(\vec{r})|^2 \quad \text{with } \overset{ibp}{=} \equiv \text{integration by parts}$$

$$\overset{\substack{divergence \\ theorem}}{=} \frac{\epsilon_o}{2}\int_{\mathcal{S}(\mathcal{V})} da\,\hat{n}\cdot\left[V(\vec{r})\,\vec{E}(\vec{r})\right] + \frac{\epsilon_o}{2}\int_{\mathcal{V}}|\vec{\nabla}V(\vec{r})|^2 \tag{2.74}$$

In the last line, the first term is an integral of the product of the potential and the field at the surface of the volume. In order to get the full energy of the charge distribution, $\mathcal{V}$ must include all the charge. If we assume the charge distribution is restricted to some finite volume, then $\mathcal{V}$ is naturally the volume containing the charge distribution. But we can add volume that does not contain charge because it contributes nothing to the initial expression for the electric potential energy. (This requirement of restriction of $\rho(\vec{r})$ to finite volume is exactly the requirement of no external potentials.)

Therefore, we replace $\mathcal{V}$ with all of space and let $\mathcal{S}$ go to infinity:

$$U = \frac{\epsilon_o}{2} \int_{r \to \infty} da\, \hat{n} \cdot \left[V(\vec{r})\, \vec{E}(\vec{r}) \right] + \frac{\epsilon_o}{2} \int_{all\ space} |\vec{\nabla} V(\vec{r})|^2 \tag{2.75}$$

Because the charge distribution is restricted to the finite volume $\mathcal{V}$ and thus looks like a point charge as $r \to \infty$, the field and potential fall off like $1/r^2$ and $1/r$. The surface area of $\mathcal{S}$ only grows as r^2, so the integral goes like $1/r$ and thus vanishes as $r \to \infty$. (If the charge distribution is not restricted to a finite volume, the surface term may not vanish, requiring one to either keep the surface term or use the initial expression.)

It may seem strange that we can make this choice of $\mathcal{S}$, as changing $\mathcal{V}$ and $\mathcal{S}$ affects both integrals in the last expression. The explanation is that the choice of $\mathcal{S}$ changes the two integrals but leaves their sum constant, and taking $\mathcal{S}$ to infinity simply zeros out the first integral, leaving only the contribution of the second integral.

We thus find

$$\boxed{U = \frac{\epsilon_o}{2} \int |\vec{E}(\vec{r})|^2} \tag{2.76}$$

where the integral is over all of space. Correspondingly, the quantity $u = \frac{\epsilon_o}{2} |\vec{E}|^2$ is an energy density. We interpret this form as indicating that the potential energy created by assembling the charge distribution is stored in the field: less charge implies a smaller field and therefore less potential energy.

Superposition and Electric Potential Energy

Because the electric potential energy is a quadratic function of the charge distribution or the electric field,

> electric potential energy does not obey superposition

The energy of a sum of fields is more than just the sum of the energies of the individual fields because there are cross terms due to the potential energy of the charges in one another's fields.

Self-energy and Point Charges vs. Continuous Charge Distributions

We were slightly cavalier in going from Equation 2.71 to Equation 2.72 in that the "self-energy" term $i = j$ that was not included in the former did get included in the latter. In the point-charge version, this term is infinite because the denominator vanishes. In the continuous distribution version, $\rho(\vec{r})\,\rho(\vec{r}')\,d\tau \to 0$ as $|\vec{r} - \vec{r}'| \to 0$ as long as ρ remains finite over all space, and thus there is no infinite contribution. (If ρ included a delta function, as would be necessary to represent a point charge, then it would produce an infinite contribution because the integral would yield $\delta(\vec{0})/0$.) Thus, we must be careful and choose the appropriate formula depending on the situation.

The infinite self-energy of a point charge reflects the fact that we do not know how to assemble a point charge. In fundamental particle physics, the existence of point charges such as the electron is an assumption, not a consequence, of the theory. In fact, there is scheme, called "renormalization," by which the infinite self-energy one calculates for such a charge from Equation 2.76 is "subtracted off" in a self-consistent fashion across all situations. While this practice is accepted and applied carefully, it is not understood. String theory, which postulates that all particles are actually vibrating string-like objects with finite extent, may offer a solution, but string theory currently is not complete — it does not offer a way to calculate the Standard Model — and there is no explicit proof it is correct.

Electric Conductors

Definition and Behavior of a Conductor

We now talk about *electric conductors*, both because they are interesting and because they provide a first opportunity to use boundary conditions to determine properties of the charge distribution, field, and potential. Notice that we derive these properties without explicit calculations!

An *electric conductor* is defined to be a material in which charge is able to flow completely freely in response to an external electric field. It is assumed, *a priori*, to contain equal and opposite amounts of positive and negative electric charge that perfectly cancel everywhere in the absence of an electric field ($\rho = 0$) but that can separate in response to an electric field. One can add charge to a conductor explicitly.

Without any calculation, we know what the response of the conductor will be to an externally applied electric field: If there is any field present in the conductor, positive and negative charge densities will separate in response to the field. That separation results in an additional field whose direction is opposite the applied field because of the directions the two polarities of charge move in response to the applied field. This movement occurs until the sum field vanishes, at which point there is no further force on the charges and the system becomes static. Therefore, $\vec{E} = 0$ inside any conductor. Note the lack of distinction between the applied field and the field created by the charges: each charge is only sensitive to the *total* field, so it is the *total* field that must vanish inside the conductor. The charges arrange themselves so their contribution to the total field cancels that of the applied field.

Derived Properties of a Conductor

We may derive the following conductor properties from the fact that $\vec{E} = 0$ inside a conductor everywhere:

▶ **ρ also vanishes inside a conductor**
   This follows directly from Gauss's Law: because $\vec{E} = 0$ everywhere in the interior, then $\vec{\nabla} \cdot \vec{E} = \rho/\epsilon_o$ also vanishes.

   Another way of seeing this, at least for a conductor with no net charge, is that, if there were a nonzero ρ, then there must be an equal and opposite amount of charge elsewhere in the conductor because the conductor is neutral overall. An electric field would appear between the oppositely signed charge distributions, contradicting the $\vec{E} = 0$ condition. Alternatively, the opposite charge will be attracted to the nonzero ρ by the field and move to cancel it until the field vanishes.

▶ **Any net charge or induced charge resides on the surface**
The picture we described before, of charge separation being induced by the external field, does imply that there may be such *induced charge* on the surface. This does not violate Gauss's Law because $\vec{E}$ may be nonzero outside the conductor and thus one has to be careful in calculating $\vec{\nabla} \cdot \vec{E}$ at the conductor boundary (we must resort to the boundary conditions we derived, Equations 2.59 and 2.61).

Also, if we intentionally add charge to a conductor, it must also move to the surface by the same Gauss's Law argument. An alternative, microscopic way of seeing this is that, if we add charge to a neutral conductor, which has no electric field or charge density in its interior, the added charge repels itself, pushing itself to the exterior (as far as it can go without leaving the conductor). Or, equivalently, the added charge attracts charge from the surface to cancel it, leaving net charge on the surface. Regardless, the added charge that now appears on the surface arranges itself so there is no net field in the interior.

Aside: As Griffiths notes in a footnote, this property can be interpreted to be a consequence of the fact that the electric field obeys the Coulomb's Law $1/r^2$ dependence in three dimensions (from which we derived Gauss's Law, which we used above in the proof). In a different number of dimensions, or with a different dependence on r, we would not have been able to derive Gauss's Law! There will be a homework problem considering conductors when Coulomb's Law is modified.

▶ **A conductor has the same electric potential everywhere**
That is, a conductor is an *equipotential*. This occurs because $\vec{E}$ vanishes
everywhere in the conductor: any line integral of $\vec{E}$ between two points must
therefore also vanish. The conductor may have a nonzero electric potential, but
the value is the same everywhere.

One can see this using the gradient, too. If V were not constant in the
conductor, there would be a nonzero $\vec{E} = -\vec{\nabla}V$, which we said above is not
allowed.

▶ **The electric field just outside a conductor is always normal to its surface**
This arises from the boundary conditions we derived, Equations 2.59 and 2.61.
Since $\vec{E}$ vanishes inside the conductor, and the tangential component of $\vec{E}$ is
continuous across any interface, the tangential component must vanish just
outside the conductor, too. There is no such condition on the normal component
because there may be an induced or net surface charge density σ on the surface.

Another way of looking at this is is that an electric field tangential to the
surface would cause charge to move along the surface until that tangential
component vanished. No such argument applies to the normal component
because the charge is no longer free to move normal to the surface when it sits
at the surface — it cannot leave the conductor.

Conductors with Cavities

The mental image we have so far is of a conductor that has no cavities inside of it. What additional properties can we derive for a conductor with cavities?

▶ **A charge q inside a cavity in a conductor results in an equal induced charge q on the surface of the conductor**

Gaussian surface
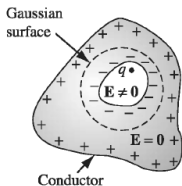
Conductor

© 2013 Griffiths, *Introduction to Electrodynamics*

To see this, construct a surface $\mathcal{S}$ that lies inside the conductor but also contains the cavity. The electric field vanishes on $\mathcal{S}$ because it is in the conductor, so the net charge enclosed must vanish. Since a charge q is inside the cavity, there must be a canceling charge $-q$ inside $\mathcal{S}$. Since $\mathcal{S}$ can be shrunk to be arbitrarily close to the inner surface without changing this statement, the induced charge must lie on the inner surface of the cavity.

Since $-q$ has appeared on the inner surface, we know, by neutrality of the conductor, there must be a charge $+q$ elsewhere on the conductor. If we now expand $\mathcal{S}$ to approach the outer surface, the above statement about $-q$ inside $\mathcal{S}$ continues to hold, so the only place $+q$ can be is on the outer surface.

The exact distribution of q on the surface depends on the geometry. For cases with some symmetry, we may be able to guess the solution easily.

Consider a conductor with a spherical outer surface. Since there are no field lines inside the conductor, there is no way the charge in the cavity or on the inner surface of the conductor can influence the distribution of charge on the outer surface, even if the inner cavity is non-spherical and/or the charge is not placed at the center of the cavity. Thus, the charge must distribute itself on the outer surface of the conductor in the same way as it would if charge $+q$ were added to a spherical conductor with no cavity. By symmetry, that distribution is uniform with surface charge density $\sigma = q/4\pi r^2$.

Note, however, that, in general, the charge on the inner surface of the conductor will **not** be distributed uniformly. It will only be uniform if the inner surface is spherical and the charge in the cavity is at the center of the cavity, as this situation has symmetry. (Note that the shape of the outer surface and the inner cavity's location with respect to the outer surface have no impact, for the same reasons as the inner cavity does not affect the distribution of charge on the outer surface.) In any other case, the field lines from the charge in the cavity will exhibit no symmetry as they terminate on the cavity wall and therefore the surface charge required to cancel those field lines in the conductor will have no symmetry.

▶ **If there is no net charge inside a cavity in a conductor, the electric field inside the cavity vanishes, independent of the external field applied to or net charge added to the conductor**
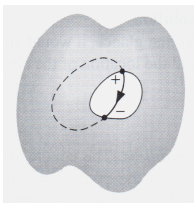
© 2013 Griffiths, *Introduction to Electrodynamics*

We use proof by contradiction. Assume there is a nonzero electric field in the cavity. Since there is no charge in the cavity, the field lines must start and end on charges on the surface of the cavity. Therefore, there is a path through the cavity with $\int d\vec{\ell} \cdot \vec{E} \neq 0$. Now close the path with a segment inside the conductor. This portion of the now-closed loop $\mathcal{C}$ contributes nothing to the line integral $\oint_{\mathcal{C}} d\vec{\ell} \cdot \vec{E}$ over the entire loop because $\vec{E} = 0$ inside the conductor. Since $\oint_{\mathcal{C}} d\vec{\ell} \cdot \vec{E} = 0$, the contribution from inside the cavity must vanish also. Contradiction. So the assumption $\vec{E} \neq 0$ in the cavity must be false.

Aside 1: Note the technique of proof by contradiction, which we will use again in E&M.

Aside 2: This fact is used for shielding of experiments from external electric fields (and also electromagnetic waves) and is called a *Faraday cage*. Note that the conductor can have some net charge on it (and correspondingly sit at some nonzero electric potential with respect to infinity) and this property still holds. As we will see later, it also holds in the presence of external electromagnetic waves, which is the more typical and important application.

Surface Charge and the Force on the Surface of a Conductor

Our boundary condition for the normal component of the electric field combined with the fact that the electric field vanishes inside a conductor tells us that the electric field infinitesimally above the surface of the conductor is

$$\vec{E} = \frac{\sigma}{\epsilon_o}\,\hat{n} \tag{2.77}$$

where $\hat{n}$ points from the inside to the outside of the conductor.

There is a charge density σ at this point, and an electric field above it, so is there a force on the charge? Yes, but the calculation is subtle. The thing to recognize is that the small element of charge $\sigma\,da$ in an infinitesimal area da cannot exert a force on itself. The field to which this element of charge is subject is the field of the charge distribution excluding it. We find this field by finding the field of this charge element and subtracting it from the total field. This is an example of one of the indirect approaches we must apply in E&M: a brute-force approach will not be successful or generic.

We know (Griffiths Example 2.5) that the electric field of a charge sheet in the xy plane is $\vec{E} = \pm(\sigma/2\,\epsilon_o)\,\hat{z}$ where the sign applies depending on whether $z > 0$ or $z < 0$. While the small patch we are considering is not an infinite sheet, it looks like one if we are infinitesimally close to it. We also know $\vec{E}_{other}$ must be continuous at the charge element because, in the absence of that charge element, there is no charge at the boundary and thus no surface charge density to cause a discontinuity in the normal component. (Note that we do not claim we know $\vec{E}_{other}$, only that we know that it has this continuity property!) Thus, we may write the equations

$$\vec{E}_{outside} = \vec{E}_{other} + \frac{\sigma}{2\,\epsilon_o}\,\hat{n} \qquad \vec{E}_{inside} = \vec{E}_{other} - \frac{\sigma}{2\,\epsilon_o}\,\hat{n} \qquad (2.78)$$

where $\vec{E}_{other}$ is the field due to the rest of the charge distribution excepting da and, because of its continuity, the same expression for $\vec{E}_{other}$ appears in both equations. (Note this technique, which you learned doing story problems in middle-school pre-algebra, of writing down an equation in which the knowns are not segregated on one side yet.) Using $\vec{E}_{outside} = (\sigma/\epsilon_o)\,\hat{n}$ and/or $\vec{E}_{inside} = 0$, we find $\vec{E}_{other} = (\sigma/2\,\epsilon_o)\,\hat{n}$. This is the field that acts on the charge $\sigma\,da$ in da. Therefore, the force per unit area is

$$\vec{f} = \frac{\vec{F}}{da} = \frac{\sigma\,da\,\vec{E}_{other}}{da} = \sigma\,\frac{\sigma}{2\,\epsilon_o}\,\hat{n} = \frac{\sigma^2}{2\,\epsilon_o}\,\hat{n} \qquad (2.79)$$

Writing the force per unit area in terms of the field at the surface $\vec{E} = (\sigma/\epsilon_o)\,\hat{n}$:

$$\boxed{\vec{f} = \frac{\sigma^2}{2\,\epsilon_o}\,\hat{n} = \frac{\epsilon_o}{2}\,E^2\,\hat{n}} \qquad (2.80)$$

That is, the surface of a conductor always feels an outward force. Consider what would happen if you put charge on a balloon with a metallized surface.

Note the force per unit area, which has units of energy density, is actually equal to the energy density just above the conductor. We could have in fact used the energy density to derive the force: the force per unit area is the gradient of the energy per unit area, and moving the conductor surface in or out by an infinitesimal distance dz would have changed the total energy per unit area by $u\,dz$.

Note the indirect technique of proof. Again, we did no integral and we did not use Coulomb's Law explicitly.

Lecture 4:

Basic Electrostatics IV:
Capacitance

Date Revised: 2024/01/18 06:45
Revised lecture break
Date Given: 2024/01/16

Capacitance

Consider two conductors (of arbitrary shapes) and suppose we put equal and opposite charges Q and $-Q$ on them. The potential difference ΔV between the two is of course given by the line integral of the electric field from any point on the surface of one to any point on the surface of the other. How does ΔV scale with the charges?

The linear dependence of $\vec{E}$ on the charge density ρ ensures that ΔV is linear in Q. Therefore, we may define the *capacitance*

$$C = \frac{Q}{\Delta V} \qquad (2.81)$$

Capacitance is a *purely geometric quantity*: it does not depend on the amount of charge on the two conductors (as long as equal and opposite charges are given to each, a caveat we will remove soon). It does depend on the shapes of the conductors and their *relative* position and orientation because those determine the shape of the electric field (while Q varies its normalization). The unit of capacitance is Coulombs/volt, which we define to be the Farad, F.

One can talk about the capacitance of a single conductor with charge Q by implicitly assuming there is another conductor at infinity that has charge $-Q$ and is defined to be at $V = 0$.

Now departing from Griffiths and instead following Jackson §1.11, we can generalize capacitance to include multiple conductors by simply assuming a generalized linear relationship between potentials, which we also call voltages, and charges as we argued above must be true:

$$V_i = \sum_{j=1}^{N} D_{ij} Q_j \qquad \text{or} \qquad \underline{V} = \underline{\underline{D}}\, \underline{Q} \qquad (2.82)$$

where $\underline{V}$ and $\underline{Q}$ are N-element column matrices for the voltages and charges on the N conductors and $\underline{\underline{D}}$ is a $N \times N$ matrix that connects the two. It is explicit that any voltage depends linearly on all the charges. The capacitance matrix is then $\underline{\underline{C}} = \underline{\underline{D}}^{-1}$, with

$$\boxed{Q_i = \sum_{j=1}^{N} C_{ij} V_j \qquad \text{or} \qquad \underline{Q} = \underline{\underline{C}}\, \underline{V}} \qquad (2.83)$$

This form serves to make it clear that the capacitance is not just a single quantity between two conductors, but is more general. According to Jackson, the diagonal element C_{ii} is the "capacitance" of electrode i, and the C_{ij} are termed the "coefficients of induction" to convey that they indicate the charge induced on electrode i when a voltage is placed on electrode j. We will show below that neither of these is what one would consider the capacitance of a pair of conductors as we discussed initially.

In all of this, there is an implicit assumption that $V(r \to \infty) = 0$. Without this assumption, we would always need to explicitly include the electrode at ∞ (with an additional index in $\underline{C}$ and $\underline{D}$) in order to get the right offset for V.

To calculate the capacitance or the capacitance matrix, one clearly needs to determine, given a set of charges $\{Q_i\}$, what the voltages $\{V_i\}$ are. To do this trivially, there typically must be a symmetry or approximation that allows one to guess the charge distributions on the conductors (*e.g.*, uniform as for an infinite parallel plate capacitor) and to calculate the field using Gauss's Law and from the field the potential. For more complex geometries, the boundary-value problem techniques we will develop may be sufficient. The total charge on each electrode normalizes the voltage.

For the simple case of two *mirror-symmetric* electrodes with equal and opposite charges $\pm Q$ and voltages $\pm V$, we can relate the elements of the capacitance matrix to the *pair capacitance*, which is what we usually call the capacitance (*e.g.*, in Ph1b). We can assume the following form for the capacitance matrix:

$$\underline{\underline{C}} = \left[\begin{array}{cc} C_s & -C_m \\ -C_m & C_s \end{array} \right] \tag{2.84}$$

Why could we assume the above form? The symmetry of the system implies $C_{11} = C_{22}$. We shall see below that all capacitance matrices are symmetric matrices, so $C_{12} = C_{21}$. We chose the negative sign on $C_{12} = -C_m$ with some foreknowledge of the result, but that's a choice and doesn't affect the value of C_{12}.

The defining condition of the *pair capacitance* is that equal and opposite charges are placed on the two conductors. By symmetry, we can conclude that the conductors carry equal and opposite voltages (not true for a non-mirror-symmetric configuration). Thus

$$Q_1 = C_s V_1 - C_m V_2 = C_s V - C_m(-V) = (C_s + C_m) V \tag{2.85}$$

$$Q_2 = -C_m V_1 + C_s V_2 = -C_m V + C_s(-V) = -(C_s + C_m) V \tag{2.86}$$

which yields $Q_2 = -Q_1 = -Q$ as assumed. Thus, the capacitance of the pair is

$$C = \frac{Q}{\Delta V} = \frac{(C_s + C_m) V}{2 V} = \frac{C_s + C_m}{2} \tag{2.87}$$

After we have discussed energy, we will return to this system for a more detailed analysis of what one can say about C_s and C_m.

Capacitance and Field Lines

Let's also think about capacitance in terms of field lines. The diagonal element D_{ii} tells us the potential of electrode i if we put charge on it and no other electrodes. That potential is the line integral of the field from infinity to the electrode, so it is telling us about the field lines going from the charge on that electrode to infinity (or to/from if the charge is negative). The off-diagonal elements D_{ji} tell us how the potential of electrode j changes when charge is put on electrode i. This makes sense, as that charge on i will change the overall field configuration, also due to the addition of the field lines that must start from or end on its charge, and that change will affect V_j.

The elements of $\underline{\underline{C}}$ are interpreted differently. When we put one electrode i at a voltage while holding the others fixed (possibly at zero), charge must be added to that electrode. The diagonal element C_{ii} tells us how much charge must go onto the electrode, and that charge sources field lines. The off-diagonal elements C_{ji} then tell us how much charge must appear on the other electrodes so their voltages V_j remain fixed. This reflects the fact that some of the new field lines starting (or ending) on electrode i due to the new charge on it must end (start) on some of the other electrodes j, and in fact tells us how much charge must be added to those other electrodes to terminate those new field lines.

As a corollary, an off-diagonal element of $\underline{\underline{D}}$ or $\underline{\underline{C}}$ can only vanish if there is no mutual influence of the two electrodes. For simply connected electrodes (*i.e.*, none of the electrodes have cavities inside them), it is hard to see how this could happen unless they are infinitely far apart!

Electric Potential Energy of a Capacitor

In a simple two-electrode, mirror-symmetric capacitor with charges $\pm q$ on the electrodes and a voltage difference $\Delta V = q/C$ between the two electrodes, the amount of work required to change the charge from q to $q + dq$ is given by the amount of work required to move a charge dq from the negative electrode (which has charge $-q$ and voltage $-\Delta V(q)/2$) to the positive electrode (which has charge $+q$ and voltage $+\Delta V(q)/2$):

$$dU = dq \left[\frac{\Delta V(q)}{2} - \left(-\frac{\Delta V(q)}{2} \right) \right] = \Delta V(q) \, dq = \frac{q}{C} \, dq \qquad (2.88)$$

Note that ΔV is a function of q here: the voltage is not held fixed while the charge is moved; rather, the voltage and charge increase together (linearly).

We integrate this expression from 0 to the final charge Q to find

$$U = \frac{1}{C} \int_0^Q q \, dq = \frac{1}{2} \frac{Q^2}{C} \tag{2.89}$$

Alternatively, using $Q = C \, \Delta V$,

$$\boxed{U = \frac{1}{2} \frac{Q^2}{C} = \frac{1}{2} C \, (\Delta V)^2} \tag{2.90}$$

We could have modeled the above process differently. Our transferral of dq from one electrode to the other is the equivalent of taking charge dq from the negative voltage electrode, carrying it out to infinity (where we set $V = 0$), and bringing it back and putting it on the positive voltage electrode. The equivalence is because the voltage difference between two points is path-independent. This process is, then, equivalent to bringing charges dq and $-dq$ in from infinity and putting them on the positive and negative voltage electrodes, respectively. And the last process is equivalent to bringing the charges in consecutively rather than simultaneously because we proved earlier the potential energy does not depend on the order of assembly of the charge distribution.

The above picture is what we need for considering a multi-electrode system: we build up the charge on each conductor by bringing in charge from infinity and calculating the work done. Consider bringing charge dq_i in from infinity and adding it to electrode i. The change in the electric potential energy of the system due to adding this charge is

$$dU_i = V_i \, dq_i = \sum_{j=1}^{N} D_{ij} \, q_j \, dq_i \qquad (2.91)$$

There are two possible double-countings we must avoid: 1) This infinitesimal element of charge dq_i is moved from $V = 0$ at infinity to $V = V_i$ on the ith electrode, so the voltages of the other electrodes are irrelevant during this infinitesimal charge transfer and we should not bring them into the equation; 2) Because the charges on all the other electrodes $j \neq i$ are physically immobile as dq_i is brought in, no work is done on them, and so there are no other contributions to include (as strange as it may seem given that their voltages change by $dV_j = D_{ji} dq_i$; remember, a force must be exerted over a distance for it to do work).

Now, let's integrate over dq_i. We will later do a sum over i. The ordering of the two steps does not matter because we proved earlier that the electric potential energy does not depend on the order of assembly. But we do need to worry about the order of how we have brought in the charges because we should not calculate cross-terms for charges that do not yet exist. Let's assume that, if we are integrating the ith charge, then the first $i - 1$ charges have already been integrated to their full values $\{Q_j\}$, $j = \{1, \ldots, i - 1\}$, and the remaining $N - i$ electrodes $j = \{i + 1, \ldots, N\}$ have no charge on them yet. Thus, the voltage $V_i(q_i; \{Q_j\}_{j<i})$ is given by

$$V_i(q_i; \{Q_j\}_{j<i}) = \sum_{j=1}^{N} D_{ij}\, q_j = D_{ii}\, q_i + \sum_{j=1}^{i-1} D_{ij}\, Q_j \tag{2.92}$$

because $q_j = Q_j$ has already been achieved for $j = \{1, \ldots, i - 1\}$, $q_j = 0$ for $j = \{i + 1, \ldots, N\}$, and $q_i \neq Q_i$ is still being changed. Therefore,

$$U_i = \int_0^{Q_i} V_i(q_i; \{Q_j\}_{j<i})\, dq_i = \int_0^{Q_i} \left[D_{ii}\, q_i\, dq_i + \sum_{j=1}^{i-1} D_{ij}\, Q_j\, dq_i \right]$$

$$= \frac{1}{2}\, D_{ii} Q_i^2 + \sum_{j=1}^{i-1} D_{ij} Q_j Q_i \tag{2.93}$$

Next, we need to sum over i to account for the charging up of all the electrodes:

$$U = \frac{1}{2} \sum_{i=1}^{N} D_{ii} Q_i^2 + \sum_{i=1}^{N} \sum_{j=1}^{i-1} D_{ij} Q_i Q_j \qquad (2.94)$$

Modifying the second sum to be symmetric (assuming $\underline{\underline{D}}$ is symmetric, which we will prove below) and including a factor of $1/2$ to correct for double-counting, we have

$$U = \frac{1}{2} \sum_{i=1}^{N} D_{ii} Q_i^2 + \frac{1}{2} \sum_{i,j=1, i \neq j}^{N} D_{ij} Q_i Q_j = \frac{1}{2} \sum_{i,j=1}^{N} D_{ij} Q_i Q_j$$

We can write this more succinctly as

$$U = \frac{1}{2}\, \underline{Q}^T \underline{\underline{D}}\, \underline{Q} = \frac{1}{2}\, \underline{Q}^T \underline{\underline{C}}^{-1} \underline{Q} \tag{2.95}$$

Using $\underline{Q} = \underline{\underline{C}}\, \underline{V}$, we can rewrite as

$$U = \frac{1}{2}\, \underline{V}^T \underline{\underline{C}}\, \underline{V} \tag{2.96}$$

Let's check that this gives the correct result for an elementary capacitor with two mirror-symmetric electrodes having equal and opposite charges $\pm Q$ and voltages $\pm V$. Using the capacitance matrix we derived earlier (recall, $C_{11} = C_{22} = C_s$, $C_{12} = C_{21} = -C_m$, and $C = (C_s + C_m)/2$),

$$
\begin{aligned}
U &= \frac{1}{2}\left[C_{11}(+V)^2 + C_{22}(-V)^2 + C_{12}(+V)(-V) + C_{21}(-V)(+V) \right] \\
&= \frac{1}{2}\, V^2 \left[C_s + C_s + C_m + C_m \right] = 2\, C\, V^2 = \frac{1}{2}\, C(\Delta V)^2 \tag{2.97}
\end{aligned}
$$

as expected.

Properties of the Capacitance Matrix and Its Inverse

We can derive a number of useful properties:

▶ **Both $\underline{\underline{C}}$ and $\underline{\underline{D}}$ are symmetric.**
  Let's consider two electrodes, i and j with $i \neq j$. From Equation 2.93, their
  contribution to the potential energy, assuming j has been charged up before i, is

$$U_{ij} = \frac{1}{2} \left(D_{ii} Q_i^2 + D_{jj} Q_j^2 \right) + D_{ij} Q_i Q_j \tag{2.98}$$

What happens if we reverse the charging order? Then we get

$$U_{ji} = \frac{1}{2} \left(D_{ii} Q_i^2 + D_{jj} Q_j^2 \right) + D_{ji} Q_i Q_j \tag{2.99}$$

In our initial discussion of the electric potential energy, we argued that the
charging order does not matter. So we may equate the two, $U_{ij} = U_{ji}$.
Recognizing that Q_i and Q_j are arbitrary then implies

$$D_{ij} = D_{ji} \quad \Longleftrightarrow \quad \underline{\underline{D}}^T = \underline{\underline{D}} \quad \Longleftrightarrow \quad \boxed{\underline{\underline{C}}^T = \underline{\underline{C}}} \tag{2.100}$$

▶ **The self-capacitances C_{ii} are positive.**
We need only consider the energy in the case that all other electrodes are held at zero potential. Then the energy is

$$U_i(\text{all others grounded}) = \frac{1}{2}\, C_{ii}\, V_i^2 \qquad (2.101)$$

Since the energy should be positive (it takes work to add charge dq_i in the presence of the same-sign charge q_i, as is done when charging up the electrode), C_{ii} must be positive.

▶ **The diagonal elements of the inverse capacitance matrix, $C_{ii}^{-1} = D_{ii}$ are positive.**
Now, we consider the energy in the case that all other electrodes are kept neutral. Then the energy is

$$U_i(\text{all others neutral}) = \frac{1}{2}\, D_{ii}\, Q_i^2 \qquad (2.102)$$

Again, since the energy should be positive, D_{ii} must be positive.

▶ **The off-diagonal elements of the inverse capacitance matrix $C_{ij}^{-1} = D_{ij}$ are positive.**

Now, let's consider two electrodes i, j in a multi-electrode configuration, with all the other electrodes uncharged. Let's suppose electrode i is already raised to its final charge, and now we want to consider the work needed to increment electrode j's charge:

$$dU_{ij} = D_{jj}\, q_j\, dq_j + D_{ij}\, Q_i\, dq_j \qquad (2.103)$$

(The self-terms and cross-terms vanish for all the electrodes $k \neq i, j$ because they have $Q_k = 0$.) If we consider the case of Q_i, q_j positive, and if we bring in more positive charge dq_j, it is obvious that both the change in the jth self-energy and the energy cross-term should be positive: we are bringing positive charges in proximity to existing positive charges. (While the existing charge might move around on the electrodes, those electrodes are conductive and so are equipotentials: no work is done.) We already know the self-energy terms are positive. In order for the energy cross-term to be positive, D_{ij} must be positive. In the mirror-symmetric electrode case, we would see via explicit inversion of $\underline{\underline{C}}$ that $\underline{\underline{D}}$'s off-diagonal elements are positive.

Another way to see that the cross-terms must be positive is to recall that the entire expression must be consistent with our original expression for the electric potential energy, Equation 2.72. That expression could be broken down into three integrals, one for each self-energy term and one for the cross-term. When the charge density is positive, all contributions to that expression are manifestly positive.

▶ **The off-diagonal elements of the capacitance matrix C_{ij} are negative.**
Let's consider the same multi-electrode system with electrodes $k \neq i, j$ grounded
(i.e., $V_k = 0$), electrode i at its final positive voltage V_i, and electrode j's
voltage being incremented from v_j to $v_j + dv_j$, both positive. The change in
energy is

$$dU_{ij} = C_{jj} \, v_j \, dv_j + C_{ij} \, V_i \, dv_j \qquad (2.104)$$

We already know the first term is positive. The second term is more
challenging. If we want to increment a positive voltage v_j by a positive amount
dv_j, we need to put positive charge on it. This positive charge will draw
negative charge out of the battery holding V_i constant: some of the field lines
of that new charge on electrode j have to terminate on electrode i if C_{ij} is
non-zero. Again, from Equation 2.72, we know that contribution to the electric
potential energy must be negative even if V_i is positive. Thus, the energy
cross-term must be negative, which requires C_{ij} to be negative. (If V_i is
negative, that implies Q_i is negative. It takes positive work to add negative
charge to an electrode that already has negative charge on it, so $C_{ij} < 0$ ensures
the cross-term becomes positive, as it should.)

▶ $|\sum_{i \neq j} C_{ij}| \leq |C_{jj}|$: **for a given electrode, the sum of the off-diagonal elements of the capacitance matrix is no larger in magnitude than the corresponding diagonal element.**
Just consider the same situation as just considered. The change in the charge on the jth electrode is $dq_j = C_{jj} \, dv_j$. The field lines from those added charges will terminate either on other electrodes or infinity, so the total negative charge added to all the other electrodes can be no larger in magnitude than $|dq_j|$. Therefore,

$$\left| \sum_{i \neq j} dq_i = \sum_{i \neq j} C_{ij} \, dv_j \right| \leq \left| dq_j = C_{jj} \, dv_j \right| \qquad \implies \qquad \left| \sum_{i \neq j} C_{ij} \right| \leq \left| C_{jj} \right| \quad (2.105)$$

Capacitance Matrix of a Mirror-Symmetric Configuration Revisited

Considering again a mirror-symmetric two-electrode configuration, we now know
$C_s > C_m > 0$, and we know the pair capacitance we are familiar with is related to
them by $C = (C_s + C_m)/2$, but can we determine C_s and C_m explicitly?

If we consider the case $V_1 = V$ and $V_2 = 0$, we find $Q_1 = C_s V$ and $Q_2 = -C_m V$, so
we can determine C_s and C_m if we know the full field configuration, with the boundary
condition $V = 0$ at infinity: we obtain the surface charge density from the normal
component of the field at the electrode surfaces and integrate it to get Q_1 and Q_2 and
thus C_s and C_m. (Remember, if $V \neq 0$ at infinity, we need to include infinity explicitly
as an electrode of the system.)

Maybe we can then do this for the one mirror-symmetric case whose full electric field
configuration we can calculate trivially, the infinite parallel-plate capacitor? No! The
infinite parallel-plate capacitor violates the condition $V = 0$ at infinity because, if
either plate has non-zero potential, that plate's non-zero equipotential surface extends
off to infinity in the transverse direction. We violate the assumption that allowed us to
ignore the electrode at infinity. Moreover, infinity is no longer even an equipotential
surface in this configuration! On the equipotentials defined by the two electrodes (at,
e.g., $z = \pm d/2$), the potential at infinity is the potential of the corresponding
electrode. If the two plates have equal and opposite potentials, then the field outside
the plates vanishes and the potential on the surface of that volume at infinity is zero.
The potential on the line $z = 0$ is also zero. And then, for $0 < |z| < d/2$ and
$x, y \to \infty$, the potential is the same linear function of z that it would be at $x, y = 0$.
Clearly, our assumptions are violated!

We note that, formally, C is infinite for this mirror-symmetric configuration, anyways: the mirror-symmetric potential configuration requires infinite charge on each electrode! The pair capacitance per unit area, however, is finite and trivially calculated.

So, we are stymied. In order for the $V = 0$ at infinity condition to be satisfied, our electrodes must be finite in extent. But, for electrodes finite in extent, we cannot calculate the potential in a trivial fashion, so we cannot determine C_s and C_m, or even C, trivially. We need to develop the full machininery for solving Poisson's and Laplace''s Equations, which we will begin to do soon.

Section 3
Advanced Electrostatics

Lecture 5:

Advanced Electrostatics I:
Laplace's Equation
Uniqueness Theorem
Method of Images

Date Revised: 2024/01/23 07:00
Revised lecture break
Date Given: 2024/01/18

Intuitive Approach to Laplace's Equation

As we mentioned earlier, the integral forms for the electric field or the potential

$$\vec{E}(\vec{r}) = \frac{1}{4\pi\epsilon_o} \int_{\mathcal{V}} d\tau' \rho(\vec{r}') \frac{\vec{r} - \vec{r}'}{|\vec{r} - \vec{r}'|^3} \qquad \text{and} \qquad V(\vec{r}) = \frac{1}{4\pi\epsilon_o} \int_{\mathcal{V}} d\tau' \frac{\rho(\vec{r}')}{|\vec{r} - \vec{r}'|} \quad (3.1)$$

are always correct but can be difficult to deal with in practice. Most systems will not have symmetries that make the integrals easily doable (or avoidable via Gauss's Law). Moreover, and this is the greater problem, it is rare that one completely specifies $\rho(\vec{r})$ in setting up a problem. Experimentally, what we can easily control are the shapes, positions, and potentials (voltages) of conductors. We do not control how the charge arranges itself on the conductors. Thus, we need to seek alternate ways to solve for the potential and field over all of space. Laplace's and Poisson's Equations are the key.

Laplace's Equation in One Dimension

In one dimension, Laplace's Equation takes the simple form

$$\frac{d^2 V}{dx^2} = 0 \tag{3.2}$$

We can solve this by direct integration to obtain

$$V(x) = m x + b \tag{3.3}$$

where m and b are two constants of integration. We determine m and b by *boundary conditions*: specification of V or dV/dx at specific point(s). In the one dimensional case, there are two options for how to specify the boundary conditions:

► Specify V at two points.
► Specify V at one point and dV/dx at one point (possibly the same point).

Note that these are the only choices in one dimension. Specifying dV/dx at two points either yields a contradiction (if two different values of dV/dx are given) or insufficient information (if the same value is given). There are no other quantities to specify: all higher derivatives vanish thanks to Laplace's Equation.

Let us note two important characteristics of the solutions of Laplace's Equation:

▶ **Averaging Property:** $V(x)$ is equal to the average of any pair of points $V(x + a)$ and $V(x - a)$ for any a such that $x \pm a$ belong to the region being considered:

$$\frac{1}{2}\left[V(x + a) + V(x - a)\right] = \frac{1}{2}\left[(m(x + a) + b) + (m(x - a) + b)\right]$$
$$= mx + b = V(x) \qquad (3.4)$$

▶ $V(x)$ **has no nontrivial local maxima or minima**
We already mentioned this property for the three-dimensional Laplace's Equation. The proof is straightforward in one dimension. Suppose x_0 is a local maximum or minimum. Then we have $dV/dx = 0$ at this point x_0. Then, for any other point x_1:

$$\left.\frac{dV}{dx}\right|_{x_1} = \left.\frac{dV}{dx}\right|_{x_0} + \int_{x_0}^{x_1} \frac{d^2V}{dx^2}\, dx = 0 + 0 = 0 \qquad (3.5)$$

Therefore, if dV/dx vanishes anywhere, then dV/dx vanishes *everywhere* and thus $V(x)$ is a constant. This is a trivial local maximum/minimum. If dV/dx vanishes nowhere, then the endpoints of the region give the maximum and minimum of $V(x)$ or, if there are no endpoints, there are no maxima or minima at all. Consider, for example, a uniform electric field $\vec{E}_0$ over all of space.

Laplace's Equation in Multiple Dimensions

We quote the analogues of the above two properties for arbitrary numbers of dimensions and prove them for three dimensions:

▶ **Averaging Property:** The value $V(\vec{r})$ of a solution to Laplace's Equation at any point is equal to the average of its value on any sphere centered on that point in the region of interest:

$$V(\vec{r}) = \langle V(\vec{r}) \rangle_a \equiv \frac{\int_{\mathcal{S}_a(\vec{r})} da' \, V(\vec{r}\,')}{\int_{\mathcal{S}_a(\vec{r})} da'} \tag{3.6}$$

where $\mathcal{S}_a(\vec{r})$ is the sphere of radius a centered on $\vec{r}$. This is straightforward to show (Griffiths Problem 3.37). Let's integrate Laplace's Equation over the volume enclosed by $\mathcal{S}_a(\vec{r})$, $\mathcal{V}_a(\vec{r})$, and use the divergence theorem:

$$0 \overset{\substack{\text{Laplace's} \\ \text{Equation}}}{=} \int_{\mathcal{V}_a(\vec{r})} d\tau' \, \nabla_{\vec{r}\,'}^2 \, V(\vec{r}\,') \tag{3.7}$$

$$\overset{\substack{\text{divergence} \\ \text{theorem}}}{=} \int_{\mathcal{S}_a(\vec{r})} da' \, \widehat{n}(\vec{r}\,') \cdot \vec{\nabla}_{\vec{r}\,'} V(\vec{r}\,') \quad = \int_{\mathcal{S}_a(\vec{r})} da' \, \widehat{n}(\vec{r}\,') \cdot \vec{\nabla}_{\vec{r}\,'-\vec{r}} V(\vec{r}\,')$$

In the last step, we have used the fact that $\vec{\nabla}$ does not care about the location of the origin (since it is just an offset).

Now, we can define $\vec{s} = \vec{r}\,' - \vec{r}$. In this coordinate system, where $\vec{r}$ is at the origin, $\hat{n}(\vec{r}\,' \in S_a(\vec{r})) = \hat{s}$, the radial unit vector in the $\vec{s}$ coordinate system. So, we have (inserting a factor $1/4\,\pi\,a^2$):

$$0 = \frac{1}{4\,\pi\,a^2} \int_{S_a(\vec{s}=\vec{0})} s^2 d\Omega_s \left.\frac{\partial V}{\partial s}\right|_{s=a} \tag{3.8}$$

where $S_a(\vec{s} = \vec{0})$ is the sphere of radius a centered on the origin of the $\vec{s}$ system (i.e., the same as $S_a(\vec{r})$, the sphere of radius a centered on $\vec{r}$ in the $\vec{r}\,'$ coordinate system). Because the integration is over a sphere of radius a, $s = a$ is fixed in the integrand. This permits us to both pull a factor $s^2 = a^2$ outside the integral, canceling the factor of a^2 in the prefactor, and to pull the radial derivative outside the integral and turn it into a derivative with respect to a. Thus:

$$0 = \frac{1}{4\,\pi\,a^2} a^2 \int_{S_a(\vec{s}=\vec{0})} d\Omega_s \left.\frac{\partial V}{\partial s}\right|_{s=a} = \frac{1}{4\,\pi} \frac{\partial}{\partial a} \int_{S_a(\vec{s}=\vec{0})} d\Omega_s V(\vec{s}) \tag{3.9}$$

Note that $\partial/\partial s$ becomes $\partial/\partial a$ when we move it outside the integral because the limits of integration, $S_a(\vec{r})$, imply s should be evaluated at a when the integral is done: s no longer exists once the integral is done.

Thus, the integral must be a constant

$$C = \frac{1}{4\,\pi} \int_{\mathcal{S}_a(\vec{s}=\vec{0})} d\Omega_s\, V(\vec{s}) = \frac{1}{4\,\pi\,a^2} \int_{\mathcal{S}_a(\vec{r})} da'\, V(\vec{r}\,') \qquad (3.10)$$

where we switched the variable of integration back to $\vec{r}\,'$ and we reinserted a^2. The right side is just the average of V over the sphere of radius a centered at $\vec{r}$. Since this holds for any a, it must hold as $a \to 0$, which tells us $C = V(\vec{r})$. So, we have

$$V(\vec{r}) = \frac{1}{4\,\pi\,a^2} \int_{\mathcal{S}_a(\vec{r})} da'\, V(\vec{r}\,') \qquad (3.11)$$

▶ **V can have no local maxima or minima in the region of interest**
The averaging property makes the proof of this property trivial: if there were such a candidate maximum (minimum), simply draw a sphere around it. Because the point is a maximum (minimum) there must be some radius of the sphere for which the values of all the points on the sphere are less than (greater than) the value at the candidate maximum (minimum). The average over this sphere is therefore less than (greater than) the value at the candidate maximum (minimum). This contradicts the above averaging property.

One could also prove this by a technique similar to the 1D case, calculating $\vec{\nabla} V$ at any point $\vec{r}\,'$ in the region by doing a line integral of Laplace's Equation from the candidate extremum $\vec{r}$ to that point. Since $\vec{\nabla} V$ vanishes at the candidate extremum (because it is an extremum of V), and the integrand ($\nabla^2 V$) of the line integral vanishes by Laplace's Equation, $\vec{\nabla} V$ vanishes at $\vec{r}\,'$.

Uniqueness Theorem

Before obtaining a solution of Laplace's and Poisson's Equations, we prove some uniqueness theorems we will need. This section draws from Jackson §1.8 and §1.9.

Green's Identities and Theorem

First, some mathematical preliminaries. Let us apply the divergence theorem to the function $\phi\vec{\nabla}\psi$ where $\phi(\vec{r})$ and $\psi(\vec{r})$ are arbitrary functions:

$$\oint_{\mathcal{S}} da\,\widehat{n}\cdot\left(\phi\vec{\nabla}\psi\right) = \int_{\mathcal{V}(\mathcal{S})} d\tau\,\vec{\nabla}\cdot\left(\phi\vec{\nabla}\psi\right)$$

This yields *Green's First Identity*:

$$\oint_{\mathcal{S}} da\,\phi\,\widehat{n}\cdot\vec{\nabla}\psi = \int_{\mathcal{V}(\mathcal{S})} d\tau\left[\phi\nabla^2\psi + \vec{\nabla}\phi\cdot\vec{\nabla}\psi\right] \tag{3.12}$$

The function $\widehat{n}\cdot\vec{\nabla}\psi$ is the *normal gradient* of ψ because it is the projection of the gradient of ψ along the direction normal to the surface. If we exchange ϕ and ψ and then difference the two versions, we have *Green's Second Identity* or *Green's Theorem*:

$$\oint_{\mathcal{S}} da\left[\phi\,\widehat{n}\cdot\vec{\nabla}\psi - \psi\,\widehat{n}\cdot\vec{\nabla}\phi\right] = \int_{\mathcal{V}(\mathcal{S})} d\tau\left[\phi\nabla^2\psi - \psi\nabla^2\phi\right] \tag{3.13}$$

Types of Boundary Conditions

We shall see in the proof of the Uniqueness Theorem that three types of boundary conditions are permitted:

▶ *Dirichlet boundary condition*
  In this case, the value of the potential $V(\vec{r})$ is specified on all bounding surfaces. This is the most typical experimentally realized situation, where we attach a number of conductors to voltage sources to set their voltages.

▶ *Neumann boundary condition*
  In this case, the value of the normal derivative of the potential, $\hat{n} \cdot \vec{\nabla} V(\vec{r})$, is specified on the boundary. An example of such a condition is specification of the electric field (or, equivalently, the surface charge density) at the surfaces of a set of conductors; since the tangential electric field vanishes at these surfaces, the normal electric field fully defines the electric field at the conductors.

▶ *Mixed boundary conditions*
  Dirichlet in some places, Neumann in others, is allowed as long as both are not specified at the same place.

If the volume under consideration is not bounded by a surface on which we specify the boundary conditions, then we must also specify a boundary condition at infinity.

The proof of the Uniqueness Theorem will not show why only one of these types of boundary conditions may be specified. That proof will be provided soon, in §3.4.1.

Generic Uniqueness Proof for Poisson's Equation

We will use proof by contradiction.

Suppose we have specified one of the above three types of boundary conditions. Assume that, for a particular given charge distribution $\rho(\vec{r})$, there are two independent solutions $V_1(\vec{r})$ and $V_2(\vec{r})$ of Poisson's Equation that both satisfy the boundary condition. Let $V_3 = V_1 - V_2$. Since the charge distribution is the same, $\nabla^2 V_1 = -\rho/\epsilon_o = \nabla^2 V_2$ and thus $\nabla^2 V_3 = 0$: V_3 satisfies Laplace's Equation. By a similar differencing argument, V_3 either satisfies the Dirichlet boundary condition $V_3(\vec{r} \in \mathcal{S}) = 0$, the Neumann boundary condition $\hat{n} \cdot \vec{\nabla} V_3(\vec{r} \in \mathcal{S}) = 0$, or a mixed boundary condition of these types. If we apply Green's first identity with $\phi = \psi = V_3$, we have

$$\oint_{\mathcal{S}} da \, V_3 \, \hat{n} \cdot \vec{\nabla} V_3 = \int_{\mathcal{V}(\mathcal{S})} d\tau \left(V_3 \nabla^2 V_3 + \vec{\nabla} V_3 \cdot \vec{\nabla} V_3 \right) \qquad (3.14)$$

The left side vanishes because of the boundary condition (any type). The first term on the right side vanishes by Laplace's Equation. Thus, we have

$$\int_{\mathcal{V}(\mathcal{S})} d\tau \, |\vec{\nabla} V_3|^2 = 0 \quad \Longrightarrow \quad \vec{\nabla} V_3(\vec{r}) = 0 \quad \Longrightarrow \quad V_3 = \text{constant} \qquad (3.15)$$

where we take the second step because the integrand is nonnegative. This result implies that our two candidate solutions $V_1(\vec{r})$ and $V_2(\vec{r})$ differ by at most a constant. Hence, uniqueness is proven.

Special Cases of Uniqueness Theorem

Given the above, we may state/prove three special cases of the uniqueness theorem, the ones given in Griffiths:

▶ *The solution to Laplace's Equation in some volume $\mathcal{V}$ is uniquely specified if V is specified on the boundary surface $\mathcal{S}(\mathcal{V})$.*
This is the above uniqueness theorem with $\rho = 0$ in $\mathcal{V}$ and a Dirichlet boundary condition on $\mathcal{S}(\mathcal{V})$.

▶ *The solution to Poisson's Equation in some volume $\mathcal{V}$ is uniquely specified if $\rho(\vec{r})$ is specified throughout the region and V is specified on the boundary surface $\mathcal{S}(\mathcal{V})$.*
This is the above uniqueness theorem with arbitrary $\rho(\vec{r})$ in $\mathcal{V}$ and a Dirichlet boundary condition on $\mathcal{S}(\mathcal{V})$.

▶ *In a volume $\mathcal{V}$ surrounded by conductors at the surface(s) $\mathcal{S}(\mathcal{V})$ and containing a specified charge density $\rho(\vec{r})$, the electric field is uniquely determined if the total charge on each conductor is specified.*
This one is not as obvious, but we can show that this BC yields the same input to the Uniqueness Theorem derivation as the other BCs we have specified.

Let each conductor i have surface $\mathcal{S}_i$ and charge Q_i. Since we know the surface charge density on each conductor is related to the normal component of the electric field at that conductor, we may see

$$\oint_{\mathcal{S}_i} da\,\widehat{n}(\vec{r}) \cdot \vec{E}(\vec{r}) = \frac{1}{\epsilon_o} \oint_{\mathcal{S}_i} da\,\sigma(\vec{r}) = \frac{1}{\epsilon_o} Q_i \qquad (3.16)$$

Now, as before, let's assume that there are two different solutions $V_1(\vec{r})$ and $V_2(\vec{r})$ and their difference is $V_3 = V_2 - V_1$. Let's evaluate the left-hand side of Equation 3.14 for the BC we are specifying here:

$$\oint_{\mathcal{S}} da\,V_3\,\widehat{n} \cdot \vec{\nabla} V_3 = -\sum_i \oint_{\mathcal{S}_i} da\,V_3\,\widehat{n} \cdot \vec{E}_3 = -\sum_i V_{3,i} \oint_{\mathcal{S}_i} da\,\widehat{n} \cdot \vec{E}_3 \qquad (3.17)$$

where we were able to pull V_3 out of the integrals because V_1 and V_2 have equipotentials on each surface and so therefore does V_3 (with values $V_{3,i}$, which we do not need to know). The surface integral of the normal component of $\vec{E}_3$ over each $\mathcal{S}_i$ vanishes because, as we indicated above, specifying Q_i specifies this surface integral to be the same for $\vec{E}_1$ and $\vec{E}_2$, so the surface integral vanishes for $\vec{E}_3 = \vec{E}_2 - \vec{E}_1$. Thus, the LHS of Equation 3.14 also vanishes for this BC, and so the remainder of the proof of uniqueness carries through.

Note how this proof relied on the boundary surfaces being conductors! Knowing the total charges on nonconducting boundary surfaces would not be sufficient.

Method of Images

Overview: The Basic Idea of Method of Images

The method of images uses the concept of uniqueness of solutions to Poisson's Equation. Basically, given a physical setup involving a true charge distribution $\rho(\vec{r})$ and Dirichlet boundary conditions for some volume $\mathcal{V}$, one tries to replace the region *outside* of $\mathcal{V}$ with an *image charge* distribution $\rho_{image}(\vec{r})$ such that, when the image charge's potential is summed with that of $\rho(\vec{r})$, the potential on the boundary is the same as that specified by the Dirichlet BC.

The technique works because of the uniqueness theorem: since the potential due to the image and original charges matches the boundary conditions and satisfies Poisson's Equation with the same source term inside $\mathcal{V}$, it is *the* solution to Poisson's Equation for that source term, that volume $\mathcal{V}$, and that choice of boundary conditions.

The imagined charge distribution is called "image charge" because, at least in the example of the boundary condition being imposed by the presence of a conductor, the image charge is a (possibly distorted) mirror image, through the boundary, of the original charge distribution $\rho(\vec{r})$. "Image charge" is also used (somewhat erroneously) to refer to the surface charge induced on a conducting boundary that sources the potential that one models as due to the image charge.

Note that the image charge must be placed outside the volume $\mathcal{V}$ because we may not change $\rho(\vec{r})$ inside $\mathcal{V}$; that would change the problem we are trying to solve.

We will see later how the potential due to the image charge distribution (the induced surface charge) is a component of the particular problem's Green Function.

A Point Charge near a Grounded Infinite Conducting Plane

For a system with the point charge q at $d\,\hat{z}$ above a conducting plane at $z = 0$ with $V = 0$, and considering the volume $\mathcal{V}$ consisting of the $z > 0$ half-space, the appropriate image charge is $-q$ at $-d\,\hat{z}$. By symmetry, the (Dirichlet) boundary condition $V = 0$ at $z = 0$ is met. Thus, the solution for $V(\vec{r})$ for $\vec{r} \in \mathcal{V}$ (the $z > 0$ half-space) is

$$V(\vec{r}) = \frac{1}{4\pi\,\epsilon_o}\left[\frac{q}{\sqrt{x^2 + y^2 + (z-d)^2}} - \frac{q}{\sqrt{x^2 + y^2 + (z+d)^2}}\right] \qquad (3.18)$$

The potential clearly satisfies $V(z = 0) = 0$ (and $V(r \to \infty) \to 0$). Let's use this solution to do some other calculations:

▶ *Induced surface charge*
  This we can calculate by recognizing that it is given by the change in the normal component of the electric field at the conducting boundary. Since $\vec{E} = -\vec{\nabla}V$,

$$\sigma = -\epsilon_o\left.\frac{\partial V}{\partial z}\right|_{z=0} = \frac{q}{4\pi}\left[\frac{z-d}{(x^2+y^2+(z-d)^2)^{3/2}} - \frac{z+d}{(x^2+y^2+(z+d)^2)^{3/2}}\right]\Bigg|_{z=0}$$

$$= -\frac{q}{2\pi}\frac{d}{(x^2+y^2+d^2)^{3/2}} \qquad (3.19)$$

We will treat the surface charge density and the normal component of the electric field (the normal gradient of the potential) as almost equivalent going forward.

We can calculate the total induced surface charge:

$$Q_{ind} = \int_0^\infty r\, dr \int_0^{2\pi} d\phi\, \frac{-q\, d}{2\pi}\, \frac{1}{(r^2 + d^2)^{3/2}} = q\, d\, \frac{1}{\sqrt{r^2 + d^2}}\Big|_0^\infty = -q \quad (3.20)$$

This is an example of an important general theorem: *The total induced surface charge is equal to the image charge, or to the negative of the real charge, or to some combination of the two, depending on the geometry, by Gauss's Law.* Because of the mirror symmetry of this problem, the two cases are degenerate, so this is not a particularly illustrative example of the theorem. Furthermore, because the volumes and surfaces one must integrate over are infinite, Gauss's Law cannot be applied to such a geometry. We'll return to this theorem in our next example where there is no such issue.

▶ *Force on the point charge*
The induced charge is opposite in sign to the real charge, so the two are attracted to each other. We can calculate the force by taking the gradient of the potential due to the *image charge only* (because the real charge does not feel a force due to its own potential). Since the image charge's potential is just that of a point charge, calculating the force is straightforward:

$$\vec{F} = q\, \vec{E}_{image\ charge}(d\,\hat{z}) = -\frac{1}{4\pi\epsilon_o}\, \frac{q^2}{(2d)^2}\, \hat{z} \quad (3.21)$$

This is equivalent to just calculating the force on the real charge exerted by the image charge, which is in general a valid approach. Whether to calculate the image charge potential and take the gradient or calculate the image charge force is a matter of choice and convenience.

▶ *Electric potential energy*
Here we have to be more careful because potential energy is not linear in charge, and, moreover, because the induced charge depends on the original point charge. Let's figure this out by calculating the work one would have to do *against* the electric force (*i.e.*, the mechanical force F_m doing the work is opposite in sign to the attractive electric force F_e) to bring q from $z = d$ to $z = \infty$.

$$U = -\int_d^\infty (-F_e(z)) \, dz = -\frac{1}{4\pi\epsilon_o} \frac{q^2}{4} \int_d^\infty \frac{dz}{z^2} = -\frac{1}{4\pi\epsilon_o} \frac{q^2}{4d} \qquad (3.22)$$

Note that this result is half what one would get for the potential energy of two equal and opposite point charges separated by a distance $2d$:

$$U_{alt} = -\frac{1}{4\pi\epsilon_o} \frac{q^2}{2d} \qquad (3.23)$$

There are two ways to understand this. The first is to recognize that, unlike in the case of two point charges, no energy is gained or lost in moving the negative charge because it is in the conductor, where $V = 0$ and thus $q\,V = 0$ everywhere. The second is to recognize that the above expression is the energy stored in all of space in the field of two point charges, but, in this case, the field is only real in the $z > 0$ half-space and so the integrated energy is reduced by a factor of 2.

Lecture 6:

Advanced Electrostatics II:
Method of Images (cont.)
Green Functions for Poisson's Equation

Date Revised: 2024/01/25 05:30
Revised lecture break
Date Given: 2024/01/23

A Point Charge near a Grounded, Conducting Sphere

Consider a conducting sphere of radius R centered on the origin and held at $V = 0$. Place a point charge at $a\,\hat{z}$ with $a > R$ so the point charge is outside the sphere. We would like to know the potential in the volume $\mathcal{V}$ *outside* the conducting sphere, which is the volume in which the point charge sits.

By symmetry, the appropriate image charge must be on the z axis. Let its value be q' and its position be $b\,\hat{z}$, where b may be positive or negative. We can find q' and b by requiring that $V = 0$ at $\vec{r} = \pm R\,\hat{z}$:

$$0 = V(+R\,\hat{z}) = \frac{1}{4\,\pi\,\epsilon_o}\left[\frac{q}{a - R} + \frac{q'}{R - b}\right]$$

$$0 = V(-R\,\hat{z}) = \frac{1}{4\,\pi\,\epsilon_o}\left[\frac{q}{a + R} + \frac{q'}{R + b}\right]$$

$$\implies \qquad q' = -q\,\frac{R}{a} \neq -q \qquad\qquad b = \frac{R^2}{a} \qquad\qquad (3.24)$$

(This is an example of how one does not always need to consider the generic case; these special cases at the two poles give us the information we need.) We see that both values are always physically reasonable because $R < a$. In particular, $b < R$ so the image charge remains outside $\mathcal{V}$ (*i.e.*, inside the sphere), as we expect. Note that $q' \neq -q$!

The potential at a point $(r \geq R, \theta, \phi)$ is found by summing the potentials of the real charge and the image charge:

$$V(r \geq R, \theta, \phi) = \frac{q}{4\pi\epsilon_o}\left[\frac{1}{|\vec{r} - a\,\hat{z}|} - \frac{R/a}{|\vec{r} - \frac{R^2}{a}\,\hat{z}|}\right] \tag{3.25}$$

$$= \frac{q}{4\pi\epsilon_o}\left[\frac{1}{\sqrt{r^2\sin^2\theta + (a - r\cos\theta)^2}} - \frac{R/a}{\sqrt{r^2\sin^2\theta + (\frac{R^2}{a} - r\cos\theta)^2}}\right] \tag{3.26}$$

We can use the above expression to see that the boundary condition $V(r = R) = 0$ is satisfied in full generality:

$$V(r = R, \theta, \phi) = \frac{q}{4\pi\epsilon_o}\left[\frac{1}{\sqrt{R^2 sin^2\theta + (a - R\cos\theta)^2}} - \frac{R/a}{\sqrt{R^2 sin^2\theta + (\frac{R^2}{a} - R\cos\theta)^2}}\right]$$

$$= \frac{q}{4\pi\epsilon_o}\left[\frac{1}{\sqrt{R^2\sin^2\theta + (a - R\cos\theta)^2}} - \frac{1}{\sqrt{a^2\sin^2\theta + (R - a\cos\theta)^2}}\right]$$

$$= 0 \tag{3.27}$$

Let's calculate the induced surface charge from $\widehat{n} \cdot \vec{\nabla} V = \partial V / \partial r$:

$$\sigma = -\,\epsilon_o \frac{\partial V}{\partial r}\bigg|_{r=R} \qquad (3.28)$$

$$= \frac{q}{4\,\pi}\left[\frac{R\sin^2\theta - (a - R\cos\theta)\cos\theta}{(R^2\sin^2\theta + (a - R\cos\theta)^2)^{3/2}} - \frac{R}{a}\frac{R\sin^2\theta - (\frac{R^2}{a} - R\cos\theta)\cos\theta}{\left(R^2\sin^2\theta + (\frac{R^2}{a} - R\cos\theta)^2\right)^{3/2}}\right]$$

$$= \frac{q}{4\,\pi}\left[\frac{R - a\cos\theta}{(R^2 + a^2 - 2\,a\,R\cos\theta)^{3/2}} - \frac{a^2}{R^2}\frac{R - \frac{R^2}{a}\cos\theta}{(a^2 + R^2 - 2\,a\,R\cos\theta)^{3/2}}\right]$$

$$= \frac{q}{4\,\pi}\frac{R(1 - \frac{a^2}{R^2})}{(R^2 + a^2 - 2\,a\,R\cos\theta)^{3/2}} = -\frac{q}{4\,\pi R^2}\frac{R}{a}\frac{1 - \frac{R^2}{a^2}}{\left(1 + \frac{R^2}{a^2} - 2\,\frac{R}{a}\cos\theta\right)^{3/2}}$$

One can show by integration that the total induced charge is q'. In this geometry, this makes sense because the volume enclosed by a surface integral of electric field flux at the boundary *encloses* the volume containing the image charge. This example illustrates one case of the theorem stated earlier; in this case, the total induced surface charge is equal to the image charge. We will see other cases illustrated in the next example.

The force on the point charge and the electric potential energy can be calculated in a manner similar to that used for the conducting plane.

Calculate the force by taking gradient of electric field:

$$\vec{F} = q\,\vec{E}_{image\ charge}(z = a) = -q\,\vec{\nabla}\,\frac{q}{4\,\pi\,\epsilon_o}\,\frac{-R/a}{\left|\vec{r} - \frac{R^2}{a}\,\hat{z}\right|}\Bigg|_{\vec{r}=a\,\hat{z}}$$

$$= \frac{q^2}{4\,\pi\,\epsilon_o}\,\hat{z}\,\frac{\partial}{\partial z}\,\frac{R/a}{\left[x^2 + y^2 + \left(z - \frac{R^2}{a}\right)^2\right]^{1/2}}\Bigg|_{z=a}$$

$$= -\frac{q^2}{4\,\pi\,\epsilon_o}\,\frac{R}{a^3}\,\frac{1}{\left[1 - \frac{R^2}{a^2}\right]^2}\,\hat{z}$$

Again, this is the same we would have obtained by directly calculating the force on the real charge from the image charge:

$$\vec{F} = \frac{q\left(-q\,\frac{R}{a}\right)}{4\,\pi\,\epsilon_o}\,\frac{1}{\left[a - \frac{R^2}{a}\right]^2}\,\hat{z} = -\frac{q^2}{4\,\pi\,\epsilon_o}\,\frac{R}{a^3}\,\frac{1}{\left[1 - \frac{R^2}{a^2}\right]^2}\,\hat{z}$$

Both of these match Jackson Equation 2.6.

We calculate the potential energy by line integral of the force:

$$U = \int_{\infty}^{a\,\hat{z}} d\vec{\ell} \cdot \vec{F}(\vec{r}) = \int_{\infty}^{a} (-dz) \frac{-q^2}{4\,\pi\,\epsilon_o} \frac{R}{z^3} \frac{1}{\left[1 - \frac{R^2}{z^2}\right]^2}$$

$$= \frac{q^2}{4\,\pi\,\epsilon_o} R \int_{\infty}^{a} \frac{z\,dz}{[z^2 - R^2]^2} = -\frac{q^2}{8\,\pi\,\epsilon_o} \frac{R}{a^2 - R^2}$$

Note that, if we calculate the potential energy from the image charge and real charge, we get the same factor of two error we saw above for the point charge and the plane:

$$U_{alt} = \frac{q\left(-q\frac{R}{a}\right)}{4\,\pi\,\epsilon_o} \frac{1}{\left|a\,\hat{z} - \frac{R^2}{a}\,\hat{z}\right|} = -\frac{q^2}{4\,\pi\,\epsilon_o} \frac{R}{a^2 - R^2} = 2\,U \qquad (3.29)$$

which is incorrect for the same reasons as given before.

Some Related Examples

These are drawn from Jackson Chapter 2.

Example 3.1: Point charge inside a spherical volume with a conducting boundary

The geometry of this problem is like the last one, except the point charge is *inside* the spherical boundary, $a < R$, and everything outside the boundary is conductor. One can show that the solution is identical: same formula for image charge value and position, same induced surface charge density. *However,* strangely enough, the total surface charge is now just $-q$!

Mathematically, this is because the evaluation of the integral depends on whether $R < a$ or $R > a$. (There is a power series expansion involved, which must be done differently in the two cases.)

Physically, this is because the calculation of the total induced surface charge via Gauss's Law must be done differently. One method is to use a spherical surface just outside the boundary, so it is in the conducting volume where the field vanishes. This implies that the sum of the real and induced charge vanishes, so the induced charge is the negative of the real charge.

The other method is to put the surface just inside the boundary. Now, the charge enclosed is only the real charge. As the surface approaches the boundary, though, the flux integral is equal to the negative of the integral of the surface charge density (up to ϵ_o) because the electric field near a conductor is σ/ϵ_o (with the negative because the field is pointed inward). So this tells us the total induced surface charge is the negative of the real charge too.

Thus, we see illustrated another case of the theorem we stated earlier, that the total induced surface charge is the image charge, the negative of the real charge, or some combination of the two. Which one depends on the geometry: is the boundary outside the volume of interest, inside, or some combination of the two?

In the case of the point charge outside the conducting sphere, we noted that the Gauss's Law calculation, with the Gaussian sphere just inside the volume $\mathcal{V}$ (i.e., having radius infinitesimally larger than a), yields $q' \neq -q$. The distinction is whether the volume $\mathcal{V}$ of interest is "outside" the boundary (neglecting the boundary at infinity) as in the previous case or "inside" the boundary as in this case.

(In the previous case, the Gauss's Law calculation outside $\mathcal{V}$ (i.e., using a Gaussian sphere of radius less than a) yields no useful information because the sphere doesn't contain the induced surface charge. The flux through such a sphere vanishes because the field is zero inside the conductor, which just tells us that all the induced surface charge resides, well, on the surface.)

Example 3.2: Point charge in the presence of a conducting sphere at fixed potential V_0

We can treat this by superposition. Consider first bringing the sphere up to the desired potential in the absence of the point charge, then bringing the point charge in from infinity to its final position $a\hat{z}$. We can use the grounded-case solution for the latter part because it has $V = 0$ on the sphere and $V \to 0$ at infinity, so the sum of it and the solution for the $V \neq 0$ sphere alone satisfies the boundary condition of the problem of the point charge near the $V \neq 0$ sphere, and thus it must be the correct solution. (Note the use of the principle of superposition for the potential.)

What is the solution for the $V \neq 0$ sphere on its own? Certainly, the sphere is an equipotential with the desired value V_0. By symmetry (remember, the point charge is not present in this problem), the charge is uniformly distributed on the surface. Thus, we can apply Gauss's Law to the problem, which tells us that the potential of the sphere, for $r > R$, is identical to that of a point charge at the origin. To figure out the value of the point charge, we require that the point charge's potential match the boundary condition:

$$\frac{q_0}{4\pi\epsilon_o R} = V_0 \quad \implies \quad q_0 = 4\pi\epsilon_o V_0 R \quad \implies \quad V(r) = V_0 \frac{R}{|\vec{r}|} \tag{3.30}$$

Finally, we add the two solutions together:

$$V(r \geq R, \theta, \phi) = \frac{q}{4\pi\epsilon_o} \left[\frac{1}{|\vec{r} - a\hat{z}|} - \frac{R/a}{|\vec{r} - \frac{R^2}{a}\hat{z}|} \right] + V_0 \frac{R}{|\vec{r}|} \tag{3.31}$$

Example 3.3: Point charge in the presence of a charged, insulated, conducting sphere

We can solve this using the solution we just calculated along with the principle of superposition (again!). Suppose we want to have a charge Q on the sphere. This is the same as first bringing the point charge q in while the sphere is grounded, disconnecting the grounding wire, adding $Q - q'$ ($> Q$ for $q > 0$), which causes the sphere to float to some nonzero voltage, and then connecting to a voltage source with that voltage. This situation is identical to the situation we just studied if we require

$$q_0 = Q - q' \quad \Longrightarrow \quad V_0 = \frac{q_0}{4\pi\epsilon_o R} = \frac{Q - q'}{4\pi\epsilon_o R} = \frac{Q + q\frac{R}{a}}{4\pi\epsilon_o R} \tag{3.32}$$

Plugging this into solution for the sphere held at V_0 gives

$$V(r \geq R, \theta, \phi) = \frac{q}{4\pi\epsilon_o}\left[\frac{1}{|\vec{r} - a\hat{z}|} - \frac{R/a}{|\vec{r} - \frac{R^2}{a}\hat{z}|}\right] + \frac{Q + q\frac{R}{a}}{4\pi\epsilon_o|\vec{r}|} \tag{3.33}$$

Notice that this reduces to our original point charge near a sphere solution not when $Q = 0$ but rather when $Q = q' = -qR/a$, which is the charge that must flow onto the sphere for it to stay at $V = 0$ (*i.e.*, grounded).

Formal Solution to Poisson's Equation: Green Functions

The remaining material in this section of the notes is based on Jackson §1.10.

Integral Equation for the Electric Potential

Can we solve Poisson's Equation? Sort of. We can convert it from a differential equation for V in terms of ρ (with boundary conditions separately specified) to an integral equation for V in terms of ρ with the need for the boundary conditions quite explicit. It is still not a closed-form solution for V in terms of ρ and the boundary conditions, but it helps us to frame the problem of finding solutions for V in a different manner that is helpful.

We obtain this equation by applying Green's Theorem (Equation 3.13) with $\phi(\vec{r}') = V(\vec{r}')$ and $\psi(\vec{r}') = |\vec{r} - \vec{r}'|^{-1}$. Note that $\vec{r}'$ is the variable we integrate over; $\vec{r}$ is considered a constant for the purposes of the Green's Theorem integrals.

$$
\int_{\mathcal{V}(\mathcal{S})} d\tau' \left[V(\vec{r}') \, \nabla^2_{\vec{r}'} \frac{1}{|\vec{r} - \vec{r}'|} - \frac{1}{|\vec{r} - \vec{r}'|} \, \nabla^2_{\vec{r}'} V(\vec{r}') \right]
$$
$$
= \oint_{\mathcal{S}} da \left[V(\vec{r}') \, \hat{n}(\vec{r}') \cdot \vec{\nabla}_{\vec{r}'} \frac{1}{|\vec{r} - \vec{r}'|} - \frac{1}{|\vec{r} - \vec{r}'|} \, \hat{n}(\vec{r}') \cdot \vec{\nabla}_{\vec{r}'} V(\vec{r}') \right] \quad (3.34)
$$

We reduce this by making use of the very important relation

$$\nabla^2_{\vec{r}'} \frac{1}{|\vec{r}' - \vec{r}|} = -4\pi\,\delta(\vec{r} - \vec{r}')$$ (3.35)

which is seen by combining Equations 2.53 and 2.33 with $\vec{r} \leftrightarrow \vec{r}'$:

$$\vec{\nabla}_{\vec{r}'} \frac{1}{|\vec{r}' - \vec{r}|} = -\frac{\vec{r}' - \vec{r}}{|\vec{r}' - \vec{r}|^3} \quad \text{and} \quad \vec{\nabla}_{\vec{r}'} \cdot \frac{\vec{r}' - \vec{r}}{|\vec{r}' - \vec{r}|^3} = 4\pi\,\delta(\vec{r}' - \vec{r}) = 4\pi\,\delta(\vec{r} - \vec{r}')$$

Using the above expression for the Laplacian of $|\vec{r} - \vec{r}'|^{-1}$, doing the integral over the delta function, applying Poisson's Equation, moving the second term on the right side to the left side, and multiplying everything by $-\frac{1}{4\pi}$ yields, *now only for $\vec{r} \in \mathcal{V}(\mathcal{S})$:*

$$V(\vec{r} \in \mathcal{V}(\mathcal{S})) = \frac{1}{4\pi\,\epsilon_o} \int_{\mathcal{V}(\mathcal{S})} d\tau' \frac{\rho(\vec{r}')}{|\vec{r} - \vec{r}'|}$$ (3.36)
$$+ \frac{1}{4\pi} \oint_{\mathcal{S}} da' \left[\frac{1}{|\vec{r} - \vec{r}'|}\,\hat{n}(\vec{r}') \cdot \vec{\nabla}_{\vec{r}'} V(\vec{r}') - V(\vec{r}')\,\hat{n}(\vec{r}') \cdot \vec{\nabla}_{\vec{r}'} \frac{1}{|\vec{r} - \vec{r}'|} \right]$$

(The left side vanishes for $\vec{r} \notin \mathcal{V}(\mathcal{S})$ because the integral was over $\vec{r}' \in \mathcal{V}(\mathcal{S})$).

This is a formal equation for the electric potential. The boundary conditions are present on the right side: in the case of Dirichlet, we specify $V(\vec{r}')$ for $\vec{r}' \in \mathcal{S}$, while in the case of Neumann, we specify $\hat{n}(\vec{r}') \cdot \vec{\nabla}_{\vec{r}'} V(\vec{r}')$ for $\vec{r}' \in \mathcal{S}$. Our Uniqueness Theorem says we should only need to specify one or the other at any given point on the boundary. In fact, since the Uniqueness Theorem says that knowing one specifies the other (knowing one gives the full solution, which determines the other), we don't have the freedom to specify both independently! Knowing both essentially requires knowing the solution to the problem. For example, if we consider the simplest possible case of specifying an equipotential on the boundary, then knowing the other boundary term requires knowing the normal gradient of the potential at the boundary, which is equivalent to knowing the surface charge density on the boundary. We would not be able to guess this except in cases with sufficient symmetry.

Therefore, this is not a closed-form solution but rather an integral equation for $V(\vec{r}')$ for $\vec{r}' \in \mathcal{V}(\mathcal{S}) \cup \mathcal{S}$: the boundary condition does not provide everything on the right side, but, if we know the solution, it will satisfy the equation.

Note that, in the limit of $\mathcal{S} \to \infty$ and $V(r \to \infty) \propto 1/r \to 0$, the integrand of the surface integral falls off as r^{-3} and so the surface term vanishes and we recover the usual Coulomb's Law expression for $V(\vec{r})$, Equation 2.51. That is, in a situation where we know the behavior of both surface terms is trivial, the equation does provide a closed-form expression for $V(\vec{r})$ in terms of $\rho(\vec{r})$.

So far, however, this integral equation is not very useful. Once we have introduced the concept of Green Functions, we will see its utility.

The Concept of Green Functions

Suppose we have the generalization of Poisson's Equation, the linear partial differential equation

$$O_{\vec{r}} f(\vec{r}) = g(\vec{r}) \tag{3.37}$$

where $O_{\vec{r}}$ is a linear partial differential operator taking derivatives with respect to the coordinate $\vec{r}$, f is a generalized potential, and g is a generalized source function. Poisson's Equation is an example, with $O_{\vec{r}} = -\epsilon_o \nabla^2$, $f(\vec{r}) = V(\vec{r})$, and $g(\vec{r}) = \rho(\vec{r})$. Is there a general approach for finding f given g?

Yes, there is, it is called the Green Function approach. The basic idea is to find the "impulse" response function for the differential equation: the generalized potential one gets if one has a point-like source. Given the impulse response function, and the linearity of $O_{\vec{r}}$, one can obtain the generalized potential for an arbitrary source function by convolving the impulse response function with that source function.

Mathematically, the impulse response function, or Green Function, is the function $G(\vec{r}, \vec{r}')$ that solves the equation

$$O_{\vec{r}} G(\vec{r}, \vec{r}') = \delta(\vec{r} - \vec{r}') \tag{3.38}$$

meaning that $G(\vec{r}, \vec{r}')$ calculates the generalized potential at the point $\vec{r}$ for a point source of size $q = 1$ at the position $\vec{r}'$ (i.e., the total source charge recovered by integrating over the source function is 1). If such a G exists, then, for an arbitrary source function $g(\vec{r})$, G gives us the following solution $f(\vec{r})$ to the generalized linear partial differential equation, Equation 3.37:

$$f(\vec{r}) = \int d\tau' \, G(\vec{r}, \vec{r}') \, g(\vec{r}') \tag{3.39}$$

We can check that Equation 3.37 is satisfied by this solution by applying the operator:

$$O_{\vec{r}} f(\vec{r}) = O_{\vec{r}} \int d\tau' \, G(\vec{r}, \vec{r}') \, g(\vec{r}') = \int d\tau' \left[O_{\vec{r}} G(\vec{r}, \vec{r}') \right] g(\vec{r}') \tag{3.40}$$

$$= \int d\tau' \delta(\vec{r} - \vec{r}') \, g(\vec{r}') = g(\vec{r}) \tag{3.41}$$

Note how this check relied on the linearity of $O_{\vec{r}}$, which allowed us to bring it inside the integral. Assuming solutions to the generalized linear partial differential equation are unique (true for Poisson's Equation), the Green Function is the *only* solution we need to find.

General Discussion of Green Functions for Poisson's Equation

Let's consider the simplest possible case, that in which there is no bounding surface and the potential vanishes at infinity. We can read the Green Function off by rewriting our usual expression for the potential for this boundary condition, Equation 2.51, in the same form as Equation 3.39:

$$V(\vec{r}) = \frac{1}{4\pi\epsilon_o} \int_{\mathcal{V}} d\tau' \frac{\rho(\vec{r}')}{|\vec{r} - \vec{r}'|} = \int_{\mathcal{V}} d\tau' \, G(\vec{r}, \vec{r}') \, \rho(\vec{r}') \tag{3.42}$$

Therefore, the Green Function for Poisson's Equation is

$$G(\vec{r}, \vec{r}') = \frac{1}{4\pi\epsilon_o} \frac{1}{|\vec{r} - \vec{r}'|} \qquad \text{if} \qquad \mathcal{V} = \text{all space}, V(r \to \infty) \to 0 \tag{3.43}$$

More generally — *i.e.*, for a more complex boundary condition — Poisson's Equation implies that its Green Function must decompose into the form

$$G(\vec{r}, \vec{r}') = \frac{1}{4\,\pi\,\epsilon_o} \frac{1}{|\vec{r} - \vec{r}'|} + F(\vec{r}, \vec{r}') \qquad \text{with} \qquad \nabla^2_{\vec{r}} F(\vec{r}, \vec{r}') = 0 \qquad (3.44)$$

where the first term provides the right side of Poisson's Equation but the second term is not only allowed by Poisson's Equation but, we will see, is crucial for satisfying the boundary conditions for any situation except the trivial one noted above, that of the potential vanishing at infinity. The F term plays multiple roles, depending on the type of boundary condition, and we will explain those roles later. Finding G thus consists of finding F.

We note that both G and F are symmetric in their arguments, $G(\vec{r}', \vec{r}) = G(\vec{r}, \vec{r}')$ and $F(\vec{r}', \vec{r}) = F(\vec{r}, \vec{r}')$, for reasons we will explain later.

Lecture 7:

Advanced Electrostatics III:
Green Functions for Poisson's Equation (cont.)
Obtaining Green Functions from the Method of Images
Separation of Variables: General Considerations
Separation of Variables in Cartesian Coordinates

Date Revised: 2024/01/30 09:30
Revised lecture break
Date Given: 2024/01/25

Green Functions for Poisson's Equation with Dirichlet or Neumann Boundary Conditions

To apply the concept of Green Functions to Poisson's Equation, we start by taking $\phi(\vec{r}\,') = V(\vec{r}\,')$ and $\psi(\vec{r}\,') = -\epsilon_o G(\vec{r}, \vec{r}\,')$ in Green's Theorem (Equation 3.13) and assuming

$$-\epsilon_o \nabla^2_{\vec{r}\,'} G(\vec{r}, \vec{r}\,') = \delta(\vec{r} - \vec{r}\,') \tag{3.45}$$

Note that this equation does not match Equation 3.38, which had the Laplacian acting on $\vec{r}$, not $\vec{r}\,'$. We will recover Equation 3.38 later. We then apply the same kinds of manipulations we did to obtain the integral equation for the potential, Equation 3.36 (these manipulations rely on Equation 3.45), giving

$$V(\vec{r}) = \int_{\mathcal{V}} d\tau' \, \rho(\vec{r}\,') \, G(\vec{r}, \vec{r}\,') \tag{3.46}$$
$$+ \epsilon_o \oint_{\mathcal{S}(\mathcal{V})} da' \left[G(\vec{r}, \vec{r}\,') \, \hat{n}(\vec{r}\,') \cdot \vec{\nabla}_{\vec{r}\,'} V(\vec{r}\,') - V(\vec{r}\,') \, \hat{n}(\vec{r}\,') \cdot \vec{\nabla}_{\vec{r}\,'} G(\vec{r}, \vec{r}\,') \right]$$

As noted earlier, a differential equation is not alone sufficient; we need boundary conditions to make G unique. For a particular type of boundary condition on V, we can make the choice to impose a condition on G such that the integrand involving the other type of boundary condition on V vanishes. If we do so, then our integral equation for V reduces to an integration over the source distribution with the Green Function and over the boundary condition with the Green Function (Neumann) or its normal gradient (Dirichlet).

We also see that, even though we assumed Equation 3.45 instead of Equation 3.38 for the equation defining the Green Function, the result we obtain above is consistent with Equation 3.39, which states that the source function $\rho(\vec{r}')$ should be convolved with the Green Function, integrating over its second argument, to obtain the potential function in its first argument. We will resolve this apparent inconsistency shortly.

Note that the equation we obtain for $V(\vec{r})$ is different from the integral equation for $V(\vec{r})$, Equation 3.36, because there we could not impose such a condition on $V(\vec{r})$, since it is set by the situation under consideration, or on $|\vec{r} - \vec{r}'|^{-1}$ (obviously). $G(\vec{r}, \vec{r}')$ is, on the other hand, our tool for solving that integral equation, so we may design the tool — by choosing boundary conditions for it — to do its job as long as it respects its defining equation. (Again, the differential equation is insufficient — we need to impose a boundary condition on G in order to make G's defining equation solvable.)

We can be more specific about what we mean by "forcing the other BC term to vanish" by picking a type of boundary condition:

▶ *Dirichlet boundary condition*

In this case, $V(\vec{r})$ is specified for $\vec{r} \in \mathcal{S}$. Therefore, $\hat{n}(\vec{r}) \cdot \vec{\nabla}_{\vec{r}} V(\vec{r})$ should be left unspecified — it should be determined by the solution itself — so we need for it to not appear in the integral equation. We can eliminate the term containing this normal derivative if we require the Dirichlet Green Function, $G_D(\vec{r}, \vec{r}')$, to satisfy the boundary condition (in $\vec{r}'$, the variable for the defining PDE)

$$G_D(\vec{r}, \vec{r}') = 0 \qquad \text{for} \quad \vec{r}' \in \mathcal{S}, \vec{r} \in \mathcal{V}, \mathcal{S} \tag{3.47}$$

The boundary condition must be defined for all valid $\vec{r}$, which consists of $\vec{r} \in \mathcal{S}$ and $\vec{r} \in \mathcal{V}$. This has the benefit of making Equation 3.46 usable for calculating $V(\vec{r} \in \mathcal{S})$ to check the solution is consistent with the boundary condition.

Using the interpretation implied by the convolution of the charge density with the Green Function in Equation 3.46 (admittedly, an interpretation not obviously consistent with the defining equation, Equation 3.45), the above condition is equivalent to requiring that charge on the boundary ($\vec{r}' \in \mathcal{S}$), given by the normal gradient of V on the boundary, yield no contribution to the potential elsewhere on the boundary ($\vec{r} \in \mathcal{S}$) or in the volume ($\vec{r} \in \mathcal{V}$). In one sense, this is what we expect, as the Dirichlet boundary condition specifies $V(\vec{r})$ on the boundary, so any charge that appears on the boundary to enforce that boundary condition had better do so in a way that does not modify the boundary condition.

However, in another sense, it is the opposite of what we expect: how can the induced surface charge on the boundary not affect the potential on the surface or in the volume? Wasn't that the whole idea behind the method of images, that one calculates the additional potential of the induced surface charge on the boundary by replacing it with an image charge? We resolve this confusion below.

With the above condition, the solution for $V(\vec{r})$ reduces to

$$V(\vec{r}) = \int_{\mathcal{V}} d\tau'\, \rho(\vec{r}')\, G_D(\vec{r}, \vec{r}') - \epsilon_o \oint_{\mathcal{S}(\mathcal{V})} da'\, V(\vec{r}')\, \hat{n}(\vec{r}') \cdot \vec{\nabla}_{\vec{r}'}\, G_D(\vec{r}, \vec{r}') \qquad (3.48)$$

This form allows us to resolve our confusion above:

▶ The first term calculates the potential due to the real charge, including the potential due to the "image" charge induced by it on the boundary. (We'll start being sloppy about the use of the word "image" and drop the quotes.) The latter contribution must come from this term (*and not the surface term*) because the image charge and its potential ought to be linear in the real charge density: there is no image charge without real charge. The defining condition does not contradict this: $G_D(\vec{r}, \vec{r}') \neq 0$ is allowed for $\vec{r}, \vec{r}' \in \mathcal{V}$, $G_D(\vec{r}, \vec{r}') = 0$ is only required for $\vec{r}' \in \mathcal{S}$ (and $\vec{r} \in \mathcal{V}, \mathcal{S}$).

▶ The second term adds a contribution to the potential for surface charge that appears on the boundary in order for the boundary to sit at the nonzero potential given by the boundary condition. This is not image charge because it is not induced by real charge and it appears even if there is no real charge in $\mathcal{V}$ (this term's presence does not depend on whether ρ is present or not). In the case of the point charge near the sphere, this is the charge $q_0 = 4\pi\epsilon_o V_0 R$ that appears so the sphere sits at $V = V_0$. It has nothing to do with the point charge q. The condition $G_D(\vec{r}, \vec{r}') = 0$ for $\vec{r}' \in \mathcal{S}$ is the sensible condition that this additional surface charge does not induce its own image charge. It is sort of amazing that this simple term does all that work — figures out the surface charge required to realize the Dirichlet boundary condition and calculates its potential in $\mathcal{V}$.

For a Dirichlet boundary condition, the symmetry of G_D in its arguments can be proven by applying Green's Theorem with $\phi = G_D(\vec{r}, \vec{x})$ and $\psi = G_D(\vec{r}\,', \vec{x})$, where $\vec{x}$ is the variable that is integrated over, and using the defining equation, Equation 3.45, and the defining boundary condition $G_D(\vec{r}, \vec{x}) = 0$ for $\vec{x}$ on the boundary and $\vec{r}$ in the volume and on the boundary (which also implies the same for $G_D(\vec{r}\,', \vec{x})$). Symmetry of G_D implies symmetry of F_D given that their difference is symmetric in $\vec{r}$ and $\vec{r}\,'$.

When this symmetry property is applied to Equation 3.45, and we also use the symmetry of the delta function, Equation 3.38 is recovered (after relabeling $\vec{r} \leftrightarrow \vec{r}\,'$). This resolves the apparent inconsistency between wanting the Green Function to satisfy Equation 3.38 but having to assume Equation 3.45 at the start to get Equation 3.46.

We can use the symmetry requirement to reinterpret the condition $G_D(\vec{r}, \vec{r}\,') = 0$ for $\vec{r}\,' \in \mathcal{S}$. We can now think of the unit charge as being at $\vec{r} \in \mathcal{V}, \mathcal{S}$ and the potential as being calculated at $\vec{r}\,' \in \mathcal{S}$. This condition requires that G_D yields zero contribution to the potential on the *boundary* from charges *in the volume*. The first half of this statement is the requirement that image charge appear such that the sum of the potentials of the real charge in the volume and its image charge do not modify the boundary condition. (We do not talk about real charge on the boundary because we are considering only the Dirichlet problem right now.)

We can also now provide an interpretation of $F_D(\vec{r}, \vec{r}\,')$ in the Dirichlet case. Because 1) $F_D(\vec{r}, \vec{r}\,')$ satisfies Laplace's Equation in the volume $\mathcal{V}$, and 2) when added to the potential of a unit point charge at $\vec{r}\,'$ (the first term in our expression relating G_D and F_D, Equation 3.44), the sum satisfies the specified boundary condition on $\mathcal{S}$, $F_D(\vec{r}, \vec{r}\,')$ can be interpreted as the potential function in the volume due to the image charge induced on the boundary by the real charges in the volume with the boundary grounded. This image charge depends on where the charges in the volume are, hence the integration over $\vec{r}\,' \in \mathcal{V}$ to calculate *this* effect of this term.

What remains a bit mysterious or magical is how the second term in Equation 3.48 works. Clearly, that term calculates the surface charge density on the boundary needed for the Dirichlet boundary condition to be satisfied and then calculates the potential in the volume due to that surface charge density. It requires both terms in G_D (*i.e.*, $|\vec{r} - \vec{r}\,'|^{-1}$ and F_D) to do that. It seems this part just falls out of the mathematics.

▶ *Neumann boundary condition*

In this case, $\widehat{n} \cdot \vec{\nabla} V(\vec{r})$ is specified for $\vec{r} \in \mathcal{S}$, so we need to render irrelevant the term containing $V(\vec{r})$ because we should not have to simultaneously specify it. While we might be inclined to require $\widehat{n}(\vec{r}') \cdot \vec{\nabla}_{\vec{r}'} G_N(\vec{r}, \vec{r}') = 0$ for $\vec{r}' \in \mathcal{S}$ to make this happen, this requirement is not consistent with Equation 3.45 defining G: if one integrates this equation for G_N over $\vec{r}' \in \mathcal{V}(\mathcal{S})$, and turns it into a surface integral using the divergence theorem, one obtains the requirement

$$-\epsilon_o \oint_{\mathcal{S}(\mathcal{V})} da' \, \widehat{n}(\vec{r}') \cdot \vec{\nabla}_{\vec{r}'} G_N(\vec{r}, \vec{r}') = 1 \qquad \text{for} \quad \vec{r} \in \mathcal{V}, \mathcal{S}$$

Thus, the simplest condition we can impose on G_N is

$$\widehat{n}(\vec{r}') \cdot \vec{\nabla}_{\vec{r}'} G_N(\vec{r}, \vec{r}') = - \left[\epsilon_o \oint_{\mathcal{S}(\mathcal{V})} da' \right]^{-1} \qquad \text{for} \quad \vec{r} \in \mathcal{V}, \mathcal{S}, \vec{r}' \in \mathcal{S} \quad (3.49)$$

Applying this condition, the solution for $V(\vec{r})$ reduces to

$$V(\vec{r}) = \int_{\mathcal{V}} d\tau' \, \rho(\vec{r}') \, G_N(\vec{r}, \vec{r}') + \epsilon_o \oint_{\mathcal{S}(\mathcal{V})} da' \, G_N(\vec{r}, \vec{r}') \, \hat{n}(\vec{r}') \cdot \vec{\nabla}_{\vec{r}'} \, V(\vec{r}') + \langle V(\vec{r}) \rangle_{\mathcal{S}(\mathcal{V})}$$

with $\quad \langle V(\vec{r}) \rangle_{\mathcal{S}(\mathcal{V})} \equiv \dfrac{\oint_{\mathcal{S}(\mathcal{V})} da' \, V(\vec{r}')}{\oint_{\mathcal{S}(\mathcal{V})} da'}$ (3.50)

While $V(\vec{r})$ on the boundary has not been completely eliminated, its only appearance is via its average value on the boundary. This makes sense, as the Neumann boundary condition does not specify the potential offset since it only specifies derivatives of the potential. The appearance of this term reflects the freedom we have to set the potential offset for problems with Neumann boundary conditions. Recall that the Uniqueness Theorem only showed uniqueness up to an overall offset.

What is the interpretation of a Neumann Green Function? Since $\hat{n}(\vec{r}') \cdot \vec{\nabla}_{\vec{r}'} \, V(\vec{r}')$ specifies the surface charge density on the boundary, $G_N(\vec{r}, \vec{r}')$ simply calculates the potential at a point $\vec{r}$ in the volume due to this boundary surface charge density at $\vec{r}'$. Note that G_N is convolved with the volume charge density and the surface charge density in the same way, reinforcing this interpretation. A Neumann Green Function thus has a simpler interpretation than a Dirichlet Green Function. There is no interpretation of G_N or F_N as calculating contributions from image charge.

What is the interpretation of $F_N(\vec{r}, \vec{r}'')$ for the Neumann case? One can show that it has no effect (one needs to make use of symmetry of F_N in its arguments, see below). Not that it is identically zero, but that all terms involving it vanish. This makes sense: if we specify the surface charge density everywhere in the volume and on the surface, we should be able to just use Coulomb's Law to calculate the potential everywhere, which just requires the Coulomb's Law part of G_N.

The triviality of the Neumann Green Function may seem to render pointless the extended discussion leading to this point. Recall, however, that Dirichlet boundary conditions are far more common: we tend to specify potentials on the boundary in real situations, not the charge density. We derived the Neumann Green Function for completeness, not because it is really needed.

For a Neumann boundary condition, the symmetry of G_N and F_N is not a result of the boundary condition, but it may be assumed without loss of generality; see K.-J. Kim and J. D. Jackson, *Am. J. Phys.* **61**:1144 (1993). As with the Dirichlet Green Function, this symmetry property allows Equation 3.38 to be obtained from the assumed defining equation, Equation 3.45, closing the loop on that apparent inconsistency.

To make further progress in obtaining a functional form for the Green Function, we must specify the boundary conditions in more detail. We will consider examples of this next.

Obtaining Green Functions from the Method of Images

We mentioned earlier that the component $F_D(\vec{r}, \vec{r}\,')$ of the full Dirichlet Green Function $G_D(\vec{r}, \vec{r}\,')$ can be determined by the method of images in some cases. Let's see how this works for the two cases we have considered:

▶ *Point charge near grounded conducting plane*

The full potential at a point $\vec{r}$ for the point charge at $d\,\hat{z}$ is

$$V(\vec{r}) = \frac{1}{4\,\pi\,\epsilon_o}\left[\frac{q}{|\vec{r} - d\,\hat{z}|} - \frac{q}{|\vec{r} + d\,\hat{z}|}\right] \tag{3.51}$$

We can see by inspection that the Dirichlet Green Function is given by taking $q = 1$ and by replacing $d\,\hat{z}$ in the first term with $\vec{r}\,'$ and $-d\,\hat{z}$ in the second term with $\vec{r}\,'$ mirrored through the $x'y'$ plane:

$$G_D(\vec{r}, \vec{r}\,') = \frac{1}{4\,\pi\,\epsilon_o}\left[\frac{1}{|\vec{r} - \vec{r}\,'|} - \frac{1}{|\vec{r} - (x'\hat{x} + y'\hat{y} - z'\hat{z})|}\right] \tag{3.52}$$

One can test this by plugging into Equation 3.47 with $\rho(\vec{r}\,') = q\,\delta(\vec{r}\,' - d\,\hat{z})$.

The second term accounts for the fact that induced charge appears on the grounded conducting plane and calculates the contribution to the potential due to it; it is the $F(\vec{r}, \vec{r}\,')$ term while the first term is the usual Coulomb's Law term. The first term solves Poisson's Equation while the second term solves Laplace's Equation. Both terms depend on the position of the point charge at $\vec{r}\,'$.

This G_D is not manifestly symmetric under exchange of $\vec{r}$ and $\vec{r}'$, but one can rewrite it so it is:

$$G_D(\vec{r}, \vec{r}') = \frac{1}{4\,\pi\,\epsilon_o} \left[\frac{1}{[(x-x')^2 + (y-y')^2 + (z-z')^2]^{1/2}} \right.$$
$$\left. - \frac{1}{[(x-x')^2 + (y-y')^2 + (z+z')^2]^{1/2}} \right]$$

One can now also see how $G(z = 0, \vec{r}') = 0$ always: the two terms become identical in this case.

It is also important to notice that, for our boundary condition $V(z = 0) = 0$, there is no term in $V(\vec{r})$ for the surface term because it vanishes in this case. That is, in the Dirichlet case, we expect a surface term from Equation 3.48

$$-\epsilon_o \oint_{\mathcal{S}(\mathcal{V})} da'\, V(\vec{r}')\, \widehat{n}(\vec{r}') \cdot \vec{\nabla}_{\vec{r}'}\, G_D(\vec{r}, \vec{r}') \tag{3.53}$$

Since the Dirichlet boundary condition is $V(z = 0) = 0$, this integral vanishes and we indeed only have the volume integral term from Equation 3.48 convolving the original charge distribution with G_D.

▶ *Point charge near conducting plane held at V_0*

Suppose our boundary condition had instead been $V(z = 0) = V_0$, a constant (and also $V(r \to \infty) = V_0$ for consistency; we will elaborate on this later). Is the above Green Function still valid? *Yes!* We have not changed the charge distribution in $\mathcal{V}$ or the *type* of boundary condition; all we have done is change the *value* of the boundary condition. We can check that the new *value* of the Dirichlet boundary condition is respected when we apply G_D derived on the basis of the $V_0 = 0$ case.

This is an important point about the Dirichlet Green Function: while one may find it using a special case, it is, by construction, valid for any Dirichlet boundary condition for the same geometry. It does not care about the details of *either* the charge distribution or the boundary condition. Of course, the special case used must be general enough that one can find the entire Green Function. When we later do an example using Separation of Variables in Cartesian coordinates to solve Laplace's Equation, we will see how that example determines a portion of the Dirichlet Green Function but not all of it.

Returning to the matter at hand: because $V(\vec{r}') = V_0$ for $\vec{r}' \in \mathcal{S}(\mathcal{V})$, we can pull it outside the integral, so we just have the surface integral of the normal gradient of G_D over the surface:

$$-\epsilon_o \oint_{\mathcal{S}(\mathcal{V})} da' \, V(\vec{r}') \, \hat{n}(\vec{r}') \cdot \vec{\nabla}_{\vec{r}'} \, G_D(\vec{r}, \vec{r}') = -\epsilon_o V_0 \oint_{\mathcal{S}(\mathcal{V})} da' \, \hat{n}(\vec{r}') \cdot \vec{\nabla}_{\vec{r}'} \, G_D(\vec{r}, \vec{r}')$$

We recall that, by definition, $G_D(\vec{r}, \vec{r}')$ is the potential at the point $\vec{r}$ due to a point charge of unit magnitude ($q = 1$) at $\vec{r}'$. By the symmetry of its arguments, it is also the potential at the point $\vec{r}'$ due to a unit point charge at $\vec{r}$. Earlier, when we did the method of images solution for the grounded conducting plane, we calculated the surface charge density at $\vec{r}$ due to the point charge at $d\,\hat{z}$ from $-\epsilon_o\,\vec{\nabla}_{\vec{r}}V(\vec{r}, d\,\hat{z})$. In this case, $-\epsilon_o\,\vec{\nabla}_{\vec{r}'}\,G_D(\vec{r}, \vec{r}')$ is the surface charge density at $\vec{r}'$ due to a unit charge at $\vec{r}$. Since V_0 has come outside the integral, our surface integral is now just the integral of this surface charge density over the boundary, or the total induced charge on the boundary. We calculated this when we did the method of images and found it was $Q_{ind} = -q$, so, in this case, it will be -1. That is:

$$-\epsilon_o \oint_{\mathcal{S}(\mathcal{V})} da'\, V(\vec{r}')\,\hat{n}(\vec{r}')\cdot\vec{\nabla}_{\vec{r}'}\,G_D(\vec{r}, \vec{r}') = -V_0 \oint_{\mathcal{S}(\mathcal{V})} da'\,\sigma_{ind}(\vec{r}', q = 1)$$

$$= -V_0\,Q_{ind}(q = 1) = V_0 \qquad (3.54)$$

So, we see that the surface term serves to add the potential offset that the boundary condition $V(z = 0) = V_0$ requires. Therefore, the solution is now

$$V(\vec{r}) = \frac{1}{4\,\pi\,\epsilon_o}\left[\frac{q}{|\vec{r} - d\,\hat{z}|} - \frac{q}{|\vec{r} + d\,\hat{z}|}\right] + V_0 \qquad (3.55)$$

This solution has $V(z = 0) = V_0$ and $V(r \to \infty) = V_0$.

This example serves to highlight the fact that one has to be careful about the self-consistency of boundary conditions, especially when they involve a condition at infinity. Consider two alternative, invalid BCs:

▶ One cannot set $V(z = 0) = V_0$ and $V(r \to \infty) = 0$ because that is not self-consistent for $z = 0$, $(x, y) \to \infty$: should the BC be V_0 or 0 for this part of the boundary?

▶ One cannot even require $V(z = 0) = V_0$ and $V(z \to \infty) = 0$ because it leaves unspecified the boundary condition for $V(z, \sqrt{x^2 + y^2} \to \infty)$. If one then thinks about what type of BC to specify there, one finds that it should be impossible to specify something that is consistent with $V(z \to \infty) = 0$. Think about the case of the conductor held at V_0 and no point charge. We know the solution is a uniform sheet of surface charge on the conductor, and we know that the field is then a constant $\vec{E}(\vec{r}) = (\sigma/\epsilon_o)\,\hat{z}$ and the potential is $V(\vec{r}) = -(\sigma/\epsilon_o)\,z$. This potential does not vanish as $z \to \infty$. If one knows that a set of boundary conditions is not self-consistent for the case of no point charge, then linearity/superposition tells us there is no way to fix the inconsistency by adding charges to $\mathcal{V}$: one would have to add a potential that is also not self-consistent to cancel out the self-inconsistency of the $q = 0$ potential!

▶ *Point charge near grounded conducting sphere*

The full potential at a point $\vec{r}$ for the point charge at $a\,\hat{z}$ was (Equation 3.25):

$$V(\vec{r}) = \frac{1}{4\pi\epsilon_o}\left[\frac{q}{|\vec{r} - a\,\hat{z}|} - \frac{q\,\frac{R}{a}}{|\vec{r} - \frac{R^2}{a}\,\hat{z}|}\right] \tag{3.56}$$

Thus, the Dirichlet Green Function is given by letting $\vec{r}\,' = a\,\hat{z}$ and taking $q = 1$:

$$G_D(\vec{r}, \vec{r}\,') = \frac{1}{4\pi\epsilon_o}\left[\frac{1}{|\vec{r} - \vec{r}\,'|} - \frac{R/r'}{\left|\vec{r} - \vec{r}\,'\frac{R^2}{(r')^2}\right|}\right] \tag{3.57}$$

Again, the second term accounts for the potential due to the charge induced on the surface of the sphere and is the term that solves Laplace's Equation in this situation (the $F_D(\vec{r}, \vec{r}\,')$ term). And again, one can this test form for G_D by plugging into Equation 3.47 with $\rho(\vec{r}\,') = q\,\delta(\vec{r}\,' - a\,\hat{z})$.

It is perhaps not so obvious that the second term in this Green Function is symmetric in its arguments. Let's rewrite it:

$$\frac{R/r'}{|\vec{r} - \vec{r}' \frac{R^2}{(r')^2}|} = \frac{R}{|\hat{r} \, r \, r' - R^2 \hat{r}'|} = \frac{R}{\sqrt{(r \, r')^2 + R^4 - 2 \, r \, r' R^2 \hat{r} \cdot \hat{r}'}} \quad (3.58)$$

Now the symmetry is manifest.

The same point about the surface integral term as for the conducting plane holds here: that term vanishes because $V(\vec{r}') = 0$ for $\vec{r}' \in \mathcal{S}$.

▶ *Point charge near conducting sphere held at fixed potential*

In this case, we can see the effect of the surface integral term in Equation 3.48 because $V(\vec{r})$ on the boundary does not vanish. The integral term is, from Equation 3.48:

$$-\epsilon_o \oint_{\mathcal{S}(\mathcal{V})} da' \, V(\vec{r}') \, \hat{n}(\vec{r}') \cdot \vec{\nabla}_{\vec{r}'} \, G_D(\vec{r}, \vec{r}') \tag{3.59}$$

When we encountered this nonvanishing surface term for the prior case of a point charge near a conducting plane, we recognized that $V(\vec{r}') = V_0$ could be pulled outside the integral and that the integral of the normal gradient of the Green Function gives the total charge induced on the boundary for a unit charge at $\vec{r}$. To calculate that total induced charge, we invoke the theorem (based on Gauss's Law) we discussed earlier. In this case, the surface encloses the image charge, so the total induced charge is equal to the image charge. That is:

$$-\epsilon_o \oint_{\mathcal{S}(\mathcal{V})} da' \, V(\vec{r}') \, \hat{n}(\vec{r}') \cdot \vec{\nabla}_{\vec{r}'} \, G_D(\vec{r}, \vec{r}') = -V_0 Q_{ind} = -V_0 q_{image} = V_0 \, \frac{R}{r} \tag{3.60}$$

This is again just the potential due to a point charge at the origin whose magnitude is such that the potential at radius R is V_0.

With this integral evaluated, the full solution for $V(\vec{r})$ is given by summing the term that involves the integral with ρ, which we calculated already for the grounded sphere case, with the boundary term:

$$V(\vec{r}) = \frac{q}{4\pi\epsilon_o} \left[\frac{1}{|\vec{r} - a\hat{z}|} - \frac{R/a}{|\vec{r} - \frac{R^2}{a}\hat{z}|} \right] + V_0 \frac{R}{r}$$

This is what we found earlier when we discussed the same problem using the method of images.

▶ *Point charge in the presence of a charged, insulated, conducting sphere*

The prior situation is identical to this one: specifying the charge on a conductor is the same as specifying its potential. So the result for $V(\vec{r})$ is the same, where we must take $V_0 = (Q + (R/a)q)/(4\pi\epsilon_o R)$. Note that, even though we are talking about a boundary condition in which charge is specified, it is not a Neumann boundary condition because we do not specify $\sigma(\vec{r}' \in \mathcal{S})$, we are still effectively specifying $V(\vec{r}' \in \mathcal{S})$. This case is like the third special case of the Uniqueness Theorem we discussed earlier.

Introduction to Separation of Variables

General Points on Separation of Variables

Griffiths makes this seem harder than it is. In separation of variables, we assume that the solution of Laplace's Equation factors into functions of single coordinates. This allows us to reduce the partial differential equation to a set of ordinary differential equations, which can be solved by standard techniques. Constants of integration appear that help to define the solutions. We apply the boundary conditions as defined by the voltages and/or the charge densities (normal derivative of voltage) at the boundaries. Once we find a set of solutions, we know from Sturm-Liouville theory that they form a *complete* set, so we are assured that we can write any solution to Laplace's Equation for the given boundary conditions in terms of these solutions.

We will only develop separation of variables for Laplace's Equation and, in the near term, we will only apply it to solving problems with specific types of boundary conditions rather than trying to use it to find the *F* piece of the Green Function. (Recall, *F* satisfies Laplace's Equation while *G* satisfies Poisson's Equation.) We will see later, at the tail end of our discussion of separation of variables in spherical coordinates, that this technique will actually be sufficient to obtain the Green Function for an arbitrary geometry, which then provides us the solution to Poisson's Equation. (One will be able to see that it is not feasible to do separation of variables for Poisson's Equation in the same way we do it for Laplace's Equation: the process very much relies on the vanishing of one side of the equation!)

Digression on Orthonormal Functions

The general topic of the properties of solutions to second-order linear differential equations is beyond the scope of this course; it falls under the name *Sturm-Liouville theory*, and it is covered in ACM95/100. We will simply quote some results that are important for this course.

Sturm-Liouville theory consists of recognizing that the second-order linear ordinary differential equations we encounter in many places in this course are self-adjoint (Hermitian) operators on the Hilbert space of functions that satisfy the differential equation. You know from linear algebra that Hermitian operators are guaranteed to have a set of eigenvalues and eigenvectors (in this case, eigenfunctions), and that the eigenvectors form an orthonormal basis for the space under consideration (here, again, the space of functions that satisfy the differential equation). The same results apply here. What this means is that, for such equations, there are a set of solution functions $\{f_p(w)\}$ that are the eigenfunctions of the operator, and there are corresponding eigenvalues $\{\lambda_p\}$. These eigenfunctions form a *complete, orthonormal set*. (Note: w is intended to represent any coordinate, one- or multi-dimensional.) The original differential equation (with differential operator O_w and the eigenvalue-eigenvector equation are:

$$O_w\, f(w) = k\, f(w) \qquad O_w\, f_p(w) = \lambda_p(w) \qquad (3.61)$$

where k is initially an undetermined constant; solving the the differential equation determines the allowed values of k, the $\{\lambda_p\}$, and the corresponding solutions $\{f_p(w)\}$.

Orthonormality is written mathematically as

$$\int_s^t dw\, f_p^*(w)\, f_q(w) = \delta_{pq} \tag{3.62}$$

where integration over the interval of interest $[s, t]$ is the Hilbert space inner product.

Completeness is defined to be

$$\sum_p f_p^*(w')\, f_p(w) = \delta(w' - w) \tag{3.63}$$

where the sum is over all eigenfunctions of the differential equation.

Completeness, as its name indicates, enables us to show that any function $g(w)$ on $[s, t]$ can be expanded in terms of the eigenfunctions $\{f_p\}$:

$$g(w) = \int_s^t dw' \, g(w') \, \delta(w' - w) = \int_s^t dw' \, g(w') \sum_p f_p^*(w') f_p(w)$$

$$= \sum_p f_p(w) \int_s^t dw' \, f_p^*(w') g(w')$$

That is, we have the expansion:

$$\boxed{g(w) = \sum_p A_p f_p(w)} \tag{3.64}$$

with coefficients given by

$$\boxed{A_p = \int_s^t dw' \, f_p^*(w') \, g(w')} \tag{3.65}$$

We could have derived Equation 3.65 also by applying orthonormality to the expansion Equation 3.64; this is the usual way we think of finding the $\{A_p\}$ as we will see below. They are of course equivalent derivations.

Separation of Variables in Cartesian Coordinates

We assume that the function $V(\vec{r})$ can be factorized as

$$V(\vec{r}) = X(x)\, Y(y)\, Z(z) \tag{3.66}$$

Plugging this into Laplace's Equation, we obtain

$$Y(y)\, Z(z)\, \frac{d^2 X}{dx^2} + X(x)\, Z(z)\, \frac{d^2 Y}{dY^2} + X(x)\, Y(y)\, \frac{d^2 Z}{dz^2} = 0$$
$$\frac{1}{X(x)}\, \frac{d^2 X}{dx^2} + \frac{1}{Y(y)}\, \frac{d^2 Y}{dY^2} + \frac{1}{Z(z)}\, \frac{d^2 Z}{dz^2} = 0 \tag{3.67}$$

We have three terms, the first a function of x, the second of y, and the third of z. Given these mismatched dependences, the only way the equation can hold is if each term is a constant. That is, it must hold that

$$\frac{1}{X(x)}\, \frac{d^2 X}{dx^2} = K_1 \qquad \frac{1}{Y(y)}\, \frac{d^2 Y}{dY^2} = K_2 \qquad \frac{1}{Z(z)}\, \frac{d^2 Z}{dz^2} = K_3 \tag{3.68}$$

with $K_1 + K_2 + K_3 = 0$.

We know that the solution to these ordinary differential equations are exponentials,

$$X(x) = A \exp(x\sqrt{K_1}) + B \exp(-x\sqrt{K_1}) \qquad (3.69)$$

$$Y(y) = C \exp(y\sqrt{K_2}) + C \exp(-y\sqrt{K_2}) \qquad (3.70)$$

$$Z(z) = E \exp(z\sqrt{-(K_1 + K_2)}) + F \exp(-z\sqrt{-(K_1 + K_2)}) \qquad (3.71)$$

We have not specified which of K_1, K_2, and K_3 are positive and which are negative (clearly, they cannot all be the same sign). That will be determined by the boundary conditions. Note that we are also neglecting linear solutions that also satisfy the individual ordinary differential equations; we will see they are not necessary in the examples we consider here (though they may be needed more generally).

At this point, we cannot make further generic progress; we need to apply a set of boundary conditions. These will place constraints on the allowed values of the exponents and coefficients and restrict the family of solutions. There are a number of examples in Griffiths. To avoid duplication, we use a different one here from Jackson §2.9.

Example 3.4: Empty box with five walls grounded and one held at a potential

Consider a box with side lengths a, b, and c in the x, y, and z dimensions and with one corner at the origin. The boundary conditions are

$$V(x=0) = 0 \qquad V(y=0) = 0 \qquad V(z=0) = 0 \qquad (3.72)$$
$$V(x=a) = 0 \qquad V(y=b) = 0 \qquad V(z=c) = \phi(x,y) \qquad (3.73)$$

where $\phi(x,y)$ is a function that is given. *In SoV, we always apply the homogeneous (vanishing RHS) BCs first because, we will see, they restrict the functional form of the solutions.* The homogeneous BC in the ith dimension (*e.g.*, y) can only be satisfied if the ith function (*e.g.*, $Y(y)$) satisfies it alone because it must be satisfied for all values of the other coordinates. Let's do x, y first for convenience (with foreknowledge of solution):

$$X(0) = A + B = 0 \qquad X(a) = A \exp(a\sqrt{K_1}) + B \exp(-a\sqrt{K_1}) = 0 \qquad (3.74)$$
$$Y(0) = C + D = 0 \qquad Y(b) = C \exp(b\sqrt{K_2}) + D \exp(-b\sqrt{K_2}) = 0 \qquad (3.75)$$

Reducing,

$$A\left[\exp(a\sqrt{K_1}) - \exp(-a\sqrt{K_1})\right] = 0 \qquad (3.76)$$

$$C\left[\exp(b\sqrt{K_2}) - \exp(-b\sqrt{K_2})\right] = 0 \qquad (3.77)$$

There is no solution to these equations for $K_1 > 0$ and $K_2 > 0$: the unit-normalized decaying and rising exponentials are only equal when their arguments both vanish, and they do not. Therefore, let's take $K_1 = -\alpha^2$ and $K_2 = -\beta^2$ so these become oscillating exponentials. We thus obtain the conditions

$$\sin(\alpha\, a) = 0 \qquad \sin(\beta\, b) = 0 \tag{3.78}$$

This places conditions on the allowed values of α and β:

$$\alpha_n = \frac{n\,\pi}{a} \qquad \beta_m = \frac{m\,\pi}{b} \qquad n,\ m \text{ positive integers} \tag{3.79}$$

where n and m may only be positive integers because negative values are redundant with the positive ones and $n = 0$ and $m = 0$ yield vanishing functions. Thus, we have

$$X(x) = \sum_{n=1}^{\infty} A_n \sin \alpha_n x \qquad Y(y) = \sum_{m=1}^{\infty} C_m \sin \beta_m y \tag{3.80}$$

where the $\{A_n\}$ and $\{C_m\}$ are constants to be determined. These solutions clearly respect the $V = 0$ boundary conditions at $x = 0, a$ and $y = 0, b$ because they vanish at those points. There is no relationship between n and m or between α_n and β_m at this point.

Now, let's apply the remaining homogeneous BC to $Z(z)$. At $z = 0$, we have

$$Z(0) = E + F = 0 \qquad \Longrightarrow \qquad F = -E \qquad (3.81)$$

Therefore, $Z(z)$ is of the form

$$Z(z) = E_{nm}\left[\exp(z\sqrt{\alpha_n^2 + \beta_m^2}) - \exp(-z\sqrt{\alpha_n^2 + \beta_m^2})\right] \qquad (3.82)$$

$$= E'_{nm}\sinh(\gamma_{nm}z) \qquad \text{with} \qquad \gamma_{nm} = \sqrt{\alpha_n^2 + \beta_m^2} \qquad (3.83)$$

(sinh not sin because we know $\alpha_n^2 + \beta_m^2 > 0$.) Note how the last BC only determined the form of $Z(z)$ while its eigenvalues were determined by the prior two BC.

Our full solution thus has the form

$$V(x, y, z) = \sum_{n,m=1}^{\infty} V_{nm}(x, y, z) \qquad (3.84)$$

$$= \sum_{n,m=1}^{\infty} A_{nm}\sin(\alpha_n x)\sin(\beta_m y)\sinh(\gamma_{nm}z) \qquad \text{with} \quad \gamma_{nm} = \sqrt{\alpha_n^2 + \beta_m^2}$$

where we have combined all the arbitrary coefficients A_m, C_n, and E'_{nm} into a single coefficient A_{nm}. Each $V_{nm}(\vec{r})$ satisfies all five homogeneous BCs, thus the solution so far (without the last BC applied) is the arbitrary sum over all such $V_{nm}(\vec{r})$.

Lecture 8:

Advanced Electrostatics IV:
Separation of Variables in Cartesian Coordinates (cont.)
Separation of Variables in Spherical Coordinates: General Theory
Separation of Variables in Spherical Coordinates
with Azimuthal Symmetry

Date Revised: 2024/02/01 07:15
Explicitly discuss which BC are
homogeneous and inhomogeneous in Example 3.6
Date Given: 2024/01/30

Now, we want to apply the last boundary condition, $V(x, y, z = c) = \phi(x, y)$. How? Not the same way as we applied the previous ones. The prior boundary conditions were *homogeneous*, meaning that they forced the solution to vanish on some boundary (not all boundaries, otherwise the solution would vanish by the "no extrema" property). The remaining one is *inhomogeneous* because it requires the solution to take on a particular functional form on a boundary. It must be treated differently, for two reasons.

▶ The first involves linearity and uniqueness. Because the right-hand side of a homogeneous BC is zero, the BC is satisfied by any linear combination of functions that satisfy the BC. The same is not true of inhomogeneous BC. If it were possible for two *different* functions to satisfy the inhomogeneous BC, then only a subset of linear combinations of them would satisfy the same BC: the linear combinations in which the coefficients sum to unity. This condition violates linearity. The only resolution is for there to be precisely one solution to the inhomogeneous BC. This requirement is consistent with uniqueness: the inhomogeneous BC is applied last, and it completes the application of the BC, so the solution *should be* unique once it is applied.

▶ From the purely calculational point of view, requiring the solution for a given n, m to satisfy the inhomogeneous boundary condition would imply

$$V_{nm}(x, y, z = c) = \phi(x, y) \tag{3.85}$$

$$A_{nm} \sin(\alpha_n x) \sin(\beta_m y) \sinh(\gamma_{nm} c) = \phi(x, y) \tag{3.86}$$

There simply is not enough freedom in the functional form on the left to satisfy the boundary condition for arbitrary $\phi(x, y)$.

The only way to have enough freedom to satisfy the inhomogeneous boundary condition is to consider a linear combination of the individual n, m:

$$V(\vec{r}) = \sum_{n,m=1}^{\infty} A_{nm} \sin(\alpha_n x) \sin(\beta_m y) \sinh(\gamma_{nm} z) \tag{3.87}$$

where A_{nm} are now constants to find based on requiring the above linear combination solution satisfies the inhomogeneous boundary condition at $z = c$, which now becomes

$$\phi(x, y) = V(x, y, z = c) = \sum_{n,m=1}^{\infty} A_{nm} \sin(\alpha_n x) \sin(\beta_m y) \sinh(\gamma_{nm} c) \tag{3.88}$$

This condition will let us determine the A_{nm}, but how, and why are we certain they exist? We make use of the theory of orthonormal functions we cited earlier.

We will use the fact (not proven here) that the functions $\{\sqrt{2/a}\sin(\alpha_n x)\}$ for $n \geq 1$ form a complete, orthonormal set on the $x \in [0, a]$ interval (*with the given boundary conditions at $x = 0, a$*), as do $\{\sqrt{2/b}\sin(\beta_n y)\}$ for $m \geq 1$ on $y \in [0, b]$ (again, with BC). Therefore, we may recover the A_{nm} by multiplying by them and integrating:

$$\int_0^a dx \int_0^b dy \, \phi(x, y) \sqrt{\frac{2}{a}}\sin(\alpha_p x) \sqrt{\frac{2}{b}}\sin(\beta_q y)$$

$$= \int_0^a dx \int_0^b dy \sum_{n,m=1}^{\infty} A_{nm}\sinh(\gamma_{nm}c)\sin(\alpha_n x)\sqrt{\frac{2}{a}}\sin(\alpha_n x)\sin(\beta_m y)\sqrt{\frac{2}{b}}\sin(\beta_q y)$$

$$= \sum_{n,m=1}^{\infty} A_{nm}\sinh(\gamma_{nm}c)\sqrt{\frac{a}{2}}\delta_{pn}\sqrt{\frac{b}{2}}\delta_{qm} = \frac{\sqrt{a\,b}}{2}A_{pq}\sinh(\gamma_{pq}c) \qquad (3.89)$$

Now, be aware that we did more work than necessary above. Once we are told that the $\{\sqrt{2/a}\sin(\alpha_n x)\sqrt{2/b}\sin(\beta_m y)\}$ form an orthonormal set, we do not need to do the integrals on the right-hand side! We only need write the right-hand side of the original equation in terms of the orthonormal functions, then use orthonormality (Equation 3.65) to obtain the equations for the individual coefficients; *i.e.*:

$$\phi(x, y) = \sqrt{\frac{a\,b}{4}} \sum_{n,m=1}^{\infty} A_{nm}\sqrt{\frac{2}{a}}\sin(\alpha_n x)\sqrt{\frac{2}{b}}\sin(\beta_m y)\sinh(\gamma_{nm}c) \qquad (3.90)$$

$$\implies \quad \int_0^a dx \int_0^b dy \sqrt{\frac{2}{a}}\sin(\alpha_p x)\sqrt{\frac{2}{b}}\sin(\beta_q y)\,\phi(x, y) = \sqrt{\frac{a\,b}{4}}A_{pq}\sinh(\gamma_{pq}c) \qquad (3.91)$$

Next, we move the coefficients to one side to obtain (replacing pq with mn):

$$A_{nm} = \frac{1}{\sinh(\gamma_{nm}c)} \int_0^a dx \int_0^b dy\, \frac{2}{a}\, \sin(\alpha_n x)\, \frac{2}{b}\, \sin(\beta_m y)\, \phi(x,y) \tag{3.92}$$

Our full solution for the applied set of boundary conditions is

$$V(\vec{r}) = \frac{4}{a\,b} \sum_{n,m=1}^{\infty} \sin(\alpha_n x) \sin(\beta_m y)\, \frac{\sinh(\gamma_{nm}z)}{\sinh(\gamma_{nm}c)} \int_0^a dx' \int_0^b dy'\, \phi(x',y')\, \sin(\alpha_n x')\, \sin(\beta_m y') \tag{3.93}$$

Summary: The homogeneous boundary conditions restricted the solutions to a specific orthonormal set, and the single inhomogeneous boundary condition sets the coefficients of the appropriate linear combination of that orthonormal set.

A good exercise is to write down the solutions for the five other inhomogeneous boundary condition cases (especially the ones with the inhomogeneous condition on the x, y, or $z = 0$ planes) "by inspection" — i.e., by simply changing the solution we already have by replacing z with x, y, or $a - x$, $b - y$, or $c - z$ — rather than by rederiving. Clearly, these other problems are not different in any conceptual way, they are only different calculationally, and only barely. There is no reason to redo all that calculation from scratch!

If we had used a more general boundary condition, specifying V to be nonzero on all six sides of the box, then we could solve the similar problem for each of the six faces independently (*i.e.*, let V be nonzero and arbitrary on that face and zero on all the other faces) and then sum the solutions since each individual solution does not affect the capability of the other solutions to satisfy their boundary conditions. (Of course, the boundary conditions themselves must be consistent with each other at the edges and corners where they meet.) *In fact, we would <u>have</u> to do this; the separation of variables technique provides no way to satisfy two generic, independent inhomogeneous boundary conditions simultaneously.* Rather, to solve problems involving multiple inhomogeneous boundary conditions, one must use the property that an inhomogeneous boundary condition solution can always be summed with an arbitrary number of homogeneous boundary condition solutions and still satisfy the inhomogeneous boundary condition.

It is interesting to consider the intermediate case, consisting of the same geometry with constant potentials ϕ_0 at the $z = c$ face and $-\phi_0$ at the $z = 0$ face. As stated above, one can solve the two cases of ϕ_0 and $-\phi_0$ separately and add them. One can also solve the problem directly by simultaneously applying the two boundary conditions, and one can show that the two solutions are the same (using some hyperbolic trigonometry identities). This is possible because the double-inhomogeneous boundary condition in this case is very simple, having only one free parameter, ϕ_0. A generic double-inhomogeneous boundary condition problem cannot be solved in this way.

Referring back to our discussion of Green Functions, the above solution is the surface term in Equation 3.48 for the particular boundary condition we have applied. By comparison of the two expressions, we infer (not derive!)

$$- \epsilon_o \hat{n}(\vec{r}') \cdot \vec{\nabla}_{\vec{r}'} G_D(\vec{r}, \vec{r}' = x'\hat{x} + y'\hat{y} + c\hat{z}) \tag{3.94}$$
$$= \frac{4}{a\,b} \sum_{n,m=1}^{\infty} \sin(\alpha_n x)\, \sin(\beta_m y)\, \frac{\sinh(\gamma_{nm} z)}{\sinh(\gamma_{nm} c)}\, \sin(\alpha_n x')\, \sin(\beta_m y')$$

Note that this expression does not fully specify G_D (or F_D)! The above information is sufficient for the particular physical situation we have set up, which consists of no physical charge in the volume and the above boundary condition, because:

▶ The term consisting of the integral of the charge density in the volume convolved with G_D is zero in this case because the charge density vanishes *in the volume*. Therefore, we do not need to know G_D (or F_D) completely.

▶ The above surface term is the only one needed because $V = 0$ on the other boundaries.

For the more general problem of an arbitrary charge distribution in the volume and arbitrary Dirichlet boundary conditions on the surfaces, we would need to find the full G_D. It may seem like one could do as suggested earlier, finding the solution for each option for which face is held at nonzero potential, then using the results analogous to the above as six Neumann boundary conditions on G_D, and applying separation of variables to find G_D. But one would have to require that G_D solve *Poisson's Equation* for a unit point charge, not Laplace's Equation. This, as we noted earlier, is not feasible with separation of variables because of the nonzero right side of the equation. There is a way to deal with this, which we will show a bit later when we develop the spherical harmonic expansion for the Green Function in spherical coordinates.

Another approach that does work would be the method of images with the condition $V = 0$ on all the surfaces. It is left as an exercise for the reader to think about what set of image charges is appropriate; the situation gets complicated for a charge at an arbitrary position in the box, but it is solvable. Certainly, from the resulting G_D, we could compute the normal gradient of G_D on any surface and thus obtain the general solution for V in the volume for any Dirichlet boundary condition. We should find that the normal gradient of G_D on the $z = c$ surface is what is given above.

It may seem like separation of variables is unsatisfactory for this reason — the procedure does not give you the full Green Function, while the method of images does. But, as we have seen, the method of images is not a systematic procedure — one has to guess the correct image charge distribution. By contrast, separation of variables is an entirely algorithmic procedure to give you a solution if a separable one exists for the particular boundary condition you are applying. It is less general but more reliable. *More importantly, we will show later how, by applying separation of variables in a more sophisticated way, we can in fact find the full Green Function.*

There is, nevertheless, no guarantee that there will be a separable solution; this depends on the geometry of the boundary conditions. The boundary conditions need to respect the separability assumed. For example, a boundary condition on a spherical boundary would not likely yield a solution via separation of variables in Cartesian coordinates!

Note also that the method of images technique is not appropriate for a Neumann boundary condition because the method of images solution generally solves the $V = 0$ Dirichlet BC problem. One needs a technique like separation of variables for such cases.

Separation of Variables in Spherical Coordinates: General Theory

Doing the Separation in Spherical Coordinates

We do this in a slightly more general manner than Griffiths, dropping the assumption of azimuthal symmetry until it is time to solve the separated differential equations.

Laplace's Equation in spherical coordinates is:

$$\frac{1}{r^2}\frac{\partial}{\partial r}\left(r^2\frac{\partial V}{\partial r}\right) + \frac{1}{r^2\sin\theta}\frac{\partial}{\partial\theta}\left(\sin\theta\frac{\partial V}{\partial\theta}\right) + \frac{1}{r^2\sin^2\theta}\frac{\partial^2 V}{\partial\phi^2} = 0 \qquad (3.95)$$

If we assume a separable form

$$V(r,\theta,\phi) = R(r)\,\Theta(\theta)\,\Phi(\phi) \qquad (3.96)$$

then, after dividing through by $V(r,\theta,\phi)$ and multiplying by $r^2\sin^2\theta$, we have

$$\sin^2\theta\left[\frac{1}{R(r)}\frac{d}{dr}\left(r^2\frac{dR}{dr}\right) + \frac{1}{\Theta(\theta)}\frac{1}{\sin\theta}\frac{d}{d\theta}\left(\sin\theta\frac{d\Theta}{d\theta}\right)\right] + \frac{1}{\Phi(\phi)}\frac{d^2\Phi}{d\phi^2} = 0 \quad (3.97)$$

We see that the first term depends only on r and θ while the second term depends only on ϕ, so we can immediately assume they are each equal to a constant:

$$\frac{1}{\Phi(\phi)} \frac{d^2\Phi}{d\phi^2} = -m^2 \tag{3.98}$$

The choice of the form of the constant is motivated by what will come next, but we can see why it needs to be of this form. As we saw in Cartesian coordinates, the above differential equation is solved either by growing/decaying exponentials (right side positive) or oscillating exponentials (right side negative). Since ϕ is a coordinate that repeats on itself ($\phi = 2\,n\,\pi$ are the same physical coordinate) the solutions $\Phi(\phi)$ must also be periodic, forcing the choice of the oscillating exponential. (For the same reason, the linear solutions we ignored in the Cartesian case are disallowed here.) We saw before that it is convenient to define the constant to incorporate a squaring.

The solutions of this equation are straightforward:

$$\Phi(\phi) = A \exp(i\,m\,\phi) + B \exp(-i\,m\,\phi) \tag{3.99}$$

Periodicity in ϕ with period $2\,\pi$ requires m be an integer. One can either require $m \geq 0$ and keep the $\{A_m\}$ and $\{B_m\}$ or allow m to be any integer and drop the $\{B_m\}$ (which would be redundant with the $\{A_m\}$ for $m < 0$). In either case, only one of A_0 or B_0 is required.

Returning to the other term, we now have

$$\sin^2\theta \left[\frac{1}{R(r)} \frac{d}{dr}\left(r^2 \frac{dR}{dr}\right) + \frac{1}{\Theta(\theta)} \frac{1}{\sin\theta} \frac{d}{d\theta}\left(\sin\theta \frac{d\Theta}{d\theta}\right) \right] = m^2 \qquad (3.100)$$

$$\frac{1}{R(r)} \frac{d}{dr}\left(r^2 \frac{dR}{dr}\right) + \left[\frac{1}{\Theta(\theta)} \frac{1}{\sin\theta} \frac{d}{d\theta}\left(\sin\theta \frac{d\Theta}{d\theta}\right) - \frac{m^2}{\sin^2\theta} \right] = 0 \qquad (3.101)$$

Now, we see that the first term depends only on r and the second only on θ, so we can separate again by setting the two terms equal to constants that sum to zero. Here, we rely on prior knowledge of the result to choose the constant to be $\ell(\ell + 1)$ so that

$$\frac{1}{R(r)} \frac{d}{dr}\left(r^2 \frac{dR}{dr}\right) = \ell(\ell + 1) \qquad (3.102)$$

$$\frac{1}{\Theta(\theta)} \frac{1}{\sin\theta} \frac{d}{d\theta}\left(\sin\theta \frac{d\Theta}{d\theta}\right) - \frac{m^2}{\sin^2\theta} = -\ell(\ell + 1) \qquad (3.103)$$

Note that the radial equation does not depend on m. This implies that the $R(r)$ functions will not depend on the azimuthal properties of the problem, in particular whether it has azimuthal symmetry. But $R(r)$ depends on ℓ, so it will depend on the polar properties of the problem. $\Theta(\theta)$ depends on ℓ and m, so its behavior depends on both the polar and azimuthal properties of the problem. $\Phi(\phi)$ looks like it may only depend on the azimuthal properties because it depends only on m, but m is tied to ℓ through the polar equation, so there will be some relationship.

Solving the Radial Equation

Here, we add another item to our "bag of tricks" and define $U(r)$ by $R(r) = U(r)/r$ and plug in. (This is motivated by the r^2 that the second d/dr must act on: assuming this dependence gets rid of the extra terms arising because of that factor.) We find

$$\frac{d^2 U}{dr^2} - \frac{\ell(\ell+1)}{r^2} U(r) = 0 \qquad (3.104)$$

Since the two derivatives would reduce the exponent of a power-law solution by 2, and the second term does the same by dividing by r^2, the above equation suggests $U(r)$ is a power law in r. (Or, try making it work with a transcendental function: you can't.) If we plug in such a form $U(r) = r^a$, we find

$$a(a-1)r^{a-2} - \ell(\ell+1)r^{a-2} = 0 \qquad \Longrightarrow \qquad a_1 = \ell+1 \quad \text{or} \quad a_2 = -\ell \quad (3.105)$$

$$\Longrightarrow \qquad \boxed{R(r) = \frac{U(r)}{r} = A\, r^{a_1-1} + B\, r^{a_2-1} = A\, r^\ell + \frac{B}{r^{\ell+1}}} \qquad (3.106)$$

There is no constraint on ℓ yet.

The Polar Equation and the Generalized Legendre Equation

We may rewrite the polar angle equation as

$$\frac{1}{\sin\theta}\frac{d}{d\theta}\left(\sin\theta\frac{d\Theta}{d\theta}\right) + \left[\ell(\ell+1) - \frac{m^2}{\sin^2\theta}\right]\Theta(\theta) = 0 \qquad (3.107)$$

Motivated by the fact that $\sin\theta\,d\theta = -d(\cos\theta)$, we add another trick to our bag of tricks by writing

$$x = \cos\theta \qquad \Theta(\theta) = P(\cos\theta) = P(x) \qquad 1 - x^2 = \sin^2\theta \qquad (3.108)$$

Then we may rewrite the polar differential equation as

$$\frac{d}{dx}\left[(1-x^2)\frac{dP}{dx}\right] + \left[\ell(\ell+1) - \frac{m^2}{1-x^2}\right]P(x) = 0 \qquad (3.109)$$

This is called the *generalized Legendre equation*.

As you have seen in ACM95/100, differential equations of this type can be solved by assuming the solution is a polynomial in x and requiring termination after a finite number of terms. That is, one assumes

$$P_\ell^m(x) = \sum_{k=1}^{\infty} a_k x^k \qquad (3.110)$$

and then, plugging the above form into the differential equation, one requires the series to terminate ($a_k = 0$ for some k). This condition forces ℓ to be a nonnegative integer and $-\ell \leq m \leq \ell$. (We already know m is an integer to ensure $\Phi(\phi)$ is single-valued.) These polynomials are the *associated Legendre polynomials*.

Mathematically, there should be a second solution for each ℓ, m because the equation is second order. These are the solutions one finds by not requiring termination but simply convergence for $-1 < x < 1$ (corresponding to $0 < \theta < \pi$). If one has a geometry that excludes the z-axis (where these solutions diverge), these solutions must be considered. If the z-axis is in the space, then these solutions are unphysical and can be discarded.

Separation of Variables in Spherical Coordinates with Azimuthal Symmetry

The Polar Equation Solution with Azimuthal Symmetry: the Legendre Equation and Legendre Polynomials

Consider the special case of azimuthal symmetry, for which $m = 0$ and $\Phi(\phi) =$ constant. The generalized Legendre Equation reduces to the *Legendre Equation*:

$$\frac{d}{dx}\left[(1-x^2)\frac{dP}{dx}\right] + \ell(\ell+1)\,P(x) = 0 \qquad (3.111)$$

The same series solution applies here with $m = 0$, so ℓ must still be a nonnegative integer. These solutions are the *Legendre Polynomials*. One can show they obey *Rodrigues' Formula*:

$$P_\ell(x) = \frac{1}{2^\ell \ell!}\left(\frac{d}{dx}\right)^\ell (x^2-1)^\ell \qquad (3.112)$$

Properties of the Legendre Polynomials

One can see by inspection or prove the following properties:

- $P_\ell(x)$ is a ℓth-order polynomial in x.
- $P_\ell(x)$ has only even powers of x if ℓ is even and only odd powers if ℓ is odd.
  $\implies P_\ell(x)$ is an even function of x for ℓ even and an odd function for ℓ odd.
- The Legendre polynomials are a complete, orthonormal set: any function that remains finite on the interval $[-1, 1]$ can be written in terms of them. Their orthonormality relation is

$$\int_{-1}^{1} dx \sqrt{\frac{2\ell + 1}{2}}\, P_\ell(x) \sqrt{\frac{2\ell' + 1}{2}}\, P_{\ell'}(x) = \delta_{\ell\,\ell'} \qquad (3.113)$$

and their completeness relation is

$$\sum_{\ell=0}^{\infty} \frac{2\ell + 1}{2} P_\ell(x) P_\ell(x') = \delta(x - x') \qquad (3.114)$$

- $P_\ell(1) = 1$ and $P_\ell(-1) = (-1)^\ell$.
- $P_\ell(0) = [(-1)^n (2n-1)!!]/2^n\, n!$ for even $\ell = 2\,n$. $P_\ell(0) = 0$ for odd ℓ.

Full Solution to Laplace's Equation with Azimuthal Symmetry

Combining our radial and polar equation solutions, we have that, for any problem with azimuthal symmetry and in which the z-axis is included, the potential must have the form

$$V(r, \theta) = \sum_{\ell=0}^{\infty} \left(A_\ell \, r^\ell + \frac{B_\ell}{r^{\ell+1}} \right) P_\ell(\cos\theta) \tag{3.115}$$

The coefficients $\{A_\ell\}$ and $\{B_\ell\}$ are set by the boundary conditions. If the volume includes the origin and the boundary conditions imply the potential must be finite there, the $\{B_\ell\}$ may be eliminated, and, if the volume includes infinity and the boundary conditions require the potential be finite (usually zero) there, the $\{A_\ell\}$ may be eliminated. In other cases, some or all of the $\{A_\ell\}$ and $\{B_\ell\}$ can be nonzero. Usually, application of the boundary conditions on V will require use of the orthonormality relations for the Legendre polynomials.

We note that, in the process of doing separation of variables, we have proven that the angular solution satisfies the *eigenvalue-eigenfunction equation*

$$\nabla^2 P_\ell(\cos\theta) = -\frac{\ell(\ell+1)}{r^2} P_\ell(\cos\theta) \tag{3.116}$$

For the angular equation, r acts as a constant and so appears in the eigenvalue.

Examples of Separation of Variables with Azimuthal Symmetry

We will start first with a case in which the boundary condition is quite obviously Dirichlet and the application is very much like what we did in Cartesian coordinates. Generally speaking, however, boundary conditions are not always so obvious. One has to use whatever information one is given and turn it into boundary conditions of the type that we know provides uniqueness.

Example 3.5: Dirichlet Boundary Condition on a Spherical Boundary with Azimuthal Symmetry

Suppose $V(R, \theta)$, the potential as a function of θ on a sphere of radius R, is specified, where the sphere is either the outer boundary or the inner boundary of the space. What is the explicit form for the resulting potential?

Let's consider the two cases together. If the space is $r < R$, then we require the $\{B_\ell\}$ to vanish to ensure a finite potential at the origin. (There is no charge in the volume, so we are assured that the potential cannot be infinite there.) If the space is $r > R$, then we require the $\{A_\ell\}$ to vanish so the potential goes to zero at infinity. That is:

$$V(r, \theta) = \sum_{\ell=0}^{\infty} A_\ell \, r^\ell \, P_\ell(\cos \theta) \qquad \text{or} \qquad V(r, \theta) = \sum_{\ell=0}^{\infty} \frac{B_\ell}{r^{\ell+1}} \, P_\ell(\cos \theta) \qquad (3.117)$$

To apply the boundary condition at R, we evaluate the above equations at that value:

$$V(R, \theta) = \sum_{\ell=0}^{\infty} A_\ell \, R^\ell \, P_\ell(\cos \theta) \qquad \text{or} \qquad V(R, \theta) = \sum_{\ell=0}^{\infty} \frac{B_\ell}{R^{\ell+1}} \, P_\ell(\cos \theta) \qquad (3.118)$$

Then, to find the coefficients, we apply orthonomality to both sides, as we did for separation of variables in Cartesian coordinates. For the case of $r < R$, we have:

$$\frac{2\ell+1}{2} \int_0^\pi \sin\theta \, d\theta \, V(R,\theta) \, P_\ell(\cos\theta) \tag{3.119}$$

$$= \sum_{\ell'=0}^\infty A_{\ell'} \, R^{\ell'} \int_0^\pi \sin\theta \, d\theta \, \frac{2\ell+1}{2} \, P_\ell(\cos\theta) \, P_{\ell'}(\cos\theta) \tag{3.120}$$

$$= \sum_{\ell'=0}^\infty A_{\ell'} \, R^{\ell'} \delta_{\ell\ell'} = A_\ell \, R^\ell \tag{3.121}$$

which we can solve for A_ℓ. Or, based on the orthonormality relation Equation 3.113, we can just state by inspection (yielding the same result as the above calculation):

$$A_\ell = \frac{2\ell+1}{2} \, \frac{1}{R^\ell} \int_0^\pi \sin\theta \, d\theta \, V(R,\theta) \, P_\ell(\cos\theta) \tag{3.122}$$

Notice how R^ℓ appears in the formula for A_ℓ. This is analogous to the same way that $\sinh(\gamma_{nm} c)$ appeared in the solution for the coefficients A_{nm} in the Cartesian case (Equation 3.92).

Similarly, for the case $r > R$,

$$B_\ell = \frac{2\ell + 1}{2} R^{\ell+1} \int_0^\pi \sin\theta\, d\theta\, V(R, \theta)\, P_\ell(\cos\theta) \tag{3.123}$$

Therefore, the solutions are

$$V(r < R, \theta) = \sum_{\ell=0}^\infty \frac{2\ell + 1}{2} \frac{r^\ell}{R^\ell} P_\ell(\cos\theta) \int_0^\pi \sin\theta'\, d\theta'\, V(R, \theta')\, P_\ell(\cos\theta') \tag{3.124}$$

$$V(r > R, \theta) = \sum_{\ell=0}^\infty \frac{2\ell + 1}{2} \frac{R^{\ell+1}}{r^{\ell+1}} P_\ell(\cos\theta) \int_0^\pi \sin\theta'\, d\theta'\, V(R, \theta')\, P_\ell(\cos\theta') \tag{3.125}$$

Notice how the units of the coefficients cancel the powers of r in the solution so our result has the same units of electrostatic potential as the boundary condition.

Let's make some other observations, connecting to separation of variables in Cartesian coordinates.

▶ In our Cartesian example, we had five homogeneous boundary conditions and one inhomogeneous one. The five homogeneous ones determined the form of the individual terms in the solution: they created relationships between the coefficients, and also imposed quantization requirements, that reduced the form from being a product of three sums of two exponentials with six arbitrary argument coefficients and four arbitrary normalization coefficients to being a product of two sines and a hyperbolic sine with quantized argument coefficients with one overall arbitrary normalization coefficient. The same happened here: the homogeneous boundary condition at $r = 0$ or $r \to \infty$ eliminated one of the two coefficients in each term. (Why five homogeneous boundary conditions in the Cartesian case and only one here? Requiring single-valued behavior in ϕ and at the poles imposes another three boundary conditions, and azimuthal symmetry is a fourth. So we effectively already applied four in the form for the solution we assumed.) In the Cartesian case, those conditions had the effect of both "quantizing" the argument coefficients (restricting the freedom in the arguments of the exponentials) and restricting the normalization coefficients (showing we had only sines and hyperbolic sines, eliminating cosines and hyperbolic cosines). In this case, the "quantization" is imposed by the geometry and azimuthal symmetry from the start, yielding the "already-quantized" form we started with.

▶ In our Cartesian example, we applied the homogeneous boundary conditions term-by-term and then finally we were forced to consider a sum of them to match the inhomogeneous boundary condition. In this case, we started off with the sum and applied the homogeneous boundary conditions to the sum. But one can see that, by use of orthonormality, this process really was applied term-by-term. In the Cartesian case, we could not write down such a sum so early because we had not yet obtained the quantization conditions on the argument coefficients: in Cartesian coordinates, those conditions come from the specific geometry of the problem and its homogeneous boundary conditions rather than from the coordinate system. At the end of the general derivation, we did not even know whether the argument coefficients were purely real or purely imaginary numbers! Any sum would have had to be written down as an integral over an unspecified domain. So, we had to apply the homogeneous boundary conditions first to even be able to write down a sum.

▶ In both cases, the application of the inhomogeneous boundary condition is done to the entire sum, and the result even looks quite similar, involving an integration of the inhomogeneous boundary condition over the surface with the orthonormal functions of which the solution is composed.

Example 3.6: Dirichlet Boundary Conditions at $r = 0$ and ∞, Neumann Boundary Condition at $r = R$

Griffiths does an example in which a surface charge density is specified at $r = R$ and the potential has to be found over all of space. This is almost a Neumann boundary condition, but not quite, since the surface charge density specifies the change in the normal derivative of V at r, not the normal derivative of V itself. By solving for V over all of space, one effectively turns it into a Neumann boundary condition by using the solution in one region to specify the condition on the normal derivative as one approaches the surface from the other side. One writes down different solutions for the two regions: the $\{B_\ell\}$ vanish for the $r < R$ solution to avoid a divergence at the origin, and the $\{A_\ell\}$ vanish for the $r > R$ solution to ensure the potential vanishes at infinity (as we saw above). Then, one applies the conditions that the potential must be continuous at R (which is a homogeneous boundary condition — the difference of the potentials on the two sides of the boundary equals zero) and that the normal derivative must change by the surface charge density (divided by $-\epsilon_o$; this is the inhomogeneous boundary condition because the charge density appears on one side of the equation). The first condition is effectively the specification of $\langle V \rangle_R$, which we recall from our generic discussion of Green Functions for Neumann boundary conditions. The second condition is the actual Neumann boundary condition. The first condition relates the $\{A_\ell\}$ and $\{B_\ell\}$ at each ℓ. With now just a single set of coefficients to determine, the Neumann boundary condition can be used with the orthonormality relation to find a formula for the coefficient for each ℓ.

Note the use of two different solutions in the two regions: this is a generally useful technique.

Example 3.7: Uncharged Metal Sphere in a Uniform Field: Unusual Dirichlet Boundary Conditions

Griffiths does the example of an uncharged metal sphere in a uniform electric field in the z direction, $\vec{E} = E_0 \hat{z}$. The boundary condition is a bit mixed again. Because the sphere is metal, it is an equipotential. But that doesn't specify the value of V on the sphere. Since the field is uniform, we cannot set V to vanish at infinity. Instead, $V(z = 0) = 0$ is chosen. From that choice and the fact that the equipotential sphere is in contact with $z = 0$, we can conclude that the sphere satisfies $V = 0$. But now V at infinity is not specified, so we don't yet have a Dirichlet boundary condition. The sensible thing to do is to require the potential approach $V(\vec{r}) = -E_0 z$ at infinity: whatever induced charge the sphere picks up, its contribution to the potential and field must fall off at infinity, leaving only the uniform field. Now we have a Dirichlet boundary condition. Because the potential is allowed to diverge at infinity, we cannot eliminate the $\{A_\ell\}$ in this case. But it is easy to see that only A_1 is nonzero: for $\ell > 0$, the behavior goes like r^ℓ, and since the potential must go like $z = r \cos\theta$ at large r, all the $\ell > 1$ terms must vanish. This large r behavior sets $A_1 = -E_0$. $A_0 = 0$ because the potential has no offset. That leaves the $\{B_\ell\}$ to be determined.

Applying the boundary condition $V = 0$ at $r = R$ gives:

$$0 = A_1 R \cos\theta + \sum_{\ell=0}^{\infty} \frac{B_\ell}{R^{\ell+1}} P_\ell(\cos\theta) \tag{3.126}$$

$$-A_1 R \cos\theta = \sum_{\ell=0}^{\infty} \frac{B_\ell}{R^{\ell+1}} P_\ell(\cos\theta) \tag{3.127}$$

Since the left side has a $\ell = 1$ term, and the Legendre polynomials are orthonormal, there can also be only a $\ell = 1$ term on the right side, implying $B_\ell = 0$ for $\ell \neq 1$ and $B_1/R^2 = -A_1 R$ or $B_1 = E_0 R^3$. Thus, the solution is

$$V(\vec{r}) = -E_0 \left(r - \frac{R^3}{r^2} \right) \cos\theta \tag{3.128}$$

Note the use of a nontrivial boundary condition at infinity and the need to realize that the sphere has the same potential as the $z = 0$ plane; without these boundary conditions, it would have been impossible to start the problem.

Lecture 9:

Advanced Electrostatics V:
Separation of Variables in Spherical Coordinates
with Azimuthal Symmetry (cont.)
Separation of Variables in Spherical Coordinates
without Azimuthal Symmetry
Spherical Harmonic Expansion of Green Functions

Date Revised: 2024/02/06 10:00
Revised lecture break
Date Given: 2024/02/01

Example 3.8: Separation of Variables for a Point Charge near a Grounded Conducting Sphere

Let's reconsider the situation we looked at before via method of images, the point charge near the conducting sphere. The setup is as before, with the point charge at $a\hat{z}$ and the sphere centered on the origin with radius R and $V = 0$ on its surface. One difficulty is that the presence of the point charge implies Laplace's equation is not satisfied in the full volume! It is, however, satisfied separately in the regions $R < r < a$ and $a < r < \infty$, and we have the charge density at $r = a$, so we should somehow solve separately in the two regions and then join the solutions together (*as we did before for the spherical shell of charge, which we recognized was a Neumann boundary condition (Example 3.6)*).

Since we have seen how the method of images can provide the Green Function for a system, the aforementioned equivalence suggests that we may be able to use separation of variables to find the full Green Function for a system in the "sum over orthonormal functions" form rather than in the "system of point charges form." This is indeed true and we will do this in general fashion for spherical coordinates later in §3.9.4 using a technique similar to the one we use for this example.

We may guess that the appropriate way to write the charge density at $r = a$ is

$$\sigma(\theta, \phi) = \frac{q}{2 \pi a^2 \sin \theta} \delta(\theta) \tag{3.129}$$

The rationale for this guess is that $a^2 \sin \theta$ cancels the $r^2 \sin \theta$ portion of the volume element and 2π cancels the ϕ integral. It has the right units, too, surface charge density, charge/length^2; remember, $\delta(\theta)/\sin \theta$ is unitless because θ is unitless. One can see the form is correct because integration returns q:

$$\int_0^\pi \int_0^{2\pi} da \, \sigma(\theta, \phi) = \int_0^\pi \sin \theta \, d\theta \int_0^{2\pi} d\phi \, a^2 \frac{q}{2 \pi a^2 \sin \theta} \delta(\theta) \tag{3.130}$$

$$= \frac{1}{2\pi} \int_0^\pi d\theta \int_0^{2\pi} d\phi \, q \, \delta(\theta) = q \tag{3.131}$$

Notice that no $\delta(\phi)$ is required.

We can largely apply what we did in the case of the Example 3.6 except that we cannot eliminate the $\{B_\ell\}$ for $r < a$ because the inner boundary is at $r = R$, not $r = 0$. Let's apply the homogeneous boundary condition $V(r = R) = 0$ first:

$$0 = \sum_{\ell=0}^{\infty} \left(A_\ell^{in} R^\ell + \frac{B_\ell^{in}}{R^{\ell+1}} \right) P_\ell(\cos\theta) \qquad (3.132)$$

where we use the ^{in} superscript to indicate these are the coefficients for the solution in the region *inside* of the charge at $r = a$; *i.e.*, the $R < r < a$ region. Since this a homogeneous boundary condition, we know from prior discussion we can apply it term-by-term. Perhaps easier to remember/justify is to apply orthonormality to the sum, which forces the coefficent of P_ℓ at each ℓ to vanish independently:

$$A_\ell^{in} R^\ell = -\frac{B_\ell^{in}}{R^{\ell+1}} \quad \implies \quad V(r < a, \theta) = \sum_{\ell=0}^{\infty} A_\ell^{in} \left(r^\ell - \frac{R^{2\ell+1}}{r^{\ell+1}} \right) P_\ell(\cos\theta) \quad (3.133)$$

For $r > a$, we start with the same form for the solution, but of course now with different coefficients $\{A_\ell^{out}\}$ and $\{B_\ell^{out}\}$. *Do not confuse these coefficients with the* $\{A_\ell^{in}\}$ *and* $\{B_\ell^{in}\}$ *determined above: these are solutions in different regions, so they are different functions and there is no reason to expect the coefficients are the same!* The $\{A_\ell^{out}\}$ must all vanish so the potential vanishes at infinity (homogeneous boundary condition). So we have

$$V(r > a, \theta) = \sum_{\ell=0}^{\infty} \frac{B_\ell^{out}}{r^{\ell+1}} P_\ell(\cos\theta) \qquad (3.134)$$

Next, we join the solutions at the boundary between them by applying the Neumann boundary condition there, which requires that V be continuous at $r = a$ and that $\partial V/\partial r$ be continuous there except at $\theta = 0$, where it has a discontinuity specified by $\sigma(0)$. We apply the first (homogeneous) condition, term-by-term like any homogeneous boundary condition or via the orthonormality of the P_ℓ:

$$A_\ell^{in}\left(a^\ell - \frac{R^{2\ell+1}}{a^{\ell+1}}\right) = \frac{B_\ell^{out}}{a^{\ell+1}} \qquad \Longrightarrow \qquad B_\ell^{out} = A_\ell^{in}\left(a^{2\ell+1} - R^{2\ell+1}\right) \qquad (3.135)$$

Let's put everything we have so far together in a suggestive form:

$$V^{in}(r,\theta) \equiv V(r < a, \theta) = \sum_{\ell=0}^{\infty} A_\ell^{in}\, a^{\ell+1}\left(\frac{r^\ell}{a^{\ell+1}} - \frac{\frac{R}{a}\left(\frac{R^2}{a}\right)^\ell}{r^{\ell+1}}\right)P_\ell(\cos\theta) \qquad (3.136)$$

$$V^{out}(r,\theta) \equiv V(r > a, \theta) = \sum_{\ell=0}^{\infty} A_\ell^{in}\, a^{\ell+1}\left(\frac{a^\ell}{r^{\ell+1}} - \frac{\frac{R}{a}\left(\frac{R^2}{a}\right)^\ell}{r^{\ell+1}}\right)P_\ell(\cos\theta) \qquad (3.137)$$

Notice the length^{-1} units of the portion in parentheses, implying that A_ℓ^{in} will have units of $\epsilon_o^{-1}(\text{length})^{-(\ell+1)}$. Next, we apply the derivative matching (Neumann) condition:

$$\left(\frac{\partial V^{out}}{\partial r} - \frac{\partial V^{in}}{\partial r}\right)\Bigg|_{r=a} = -\frac{\sigma(\theta)}{\epsilon_o} \qquad (3.138)$$

The derivatives are

$$\frac{\partial V^{in}}{\partial r} = \sum_{\ell=0}^{\infty} A_\ell^{in} a^{\ell+1} \left(\frac{\ell\, r^{\ell-1}}{a^{\ell+1}} + (\ell+1)\, \frac{\frac{R}{a}\left(\frac{R^2}{a}\right)^\ell}{r^{\ell+2}} \right) P_\ell(\cos\theta) \tag{3.139}$$

$$\frac{\partial V^{out}}{\partial r} = \sum_{\ell=0}^{\infty} A_\ell^{in} a^{\ell+1} (\ell+1) \left(-\frac{a^\ell}{r^{\ell+2}} + \frac{\frac{R}{a}\left(\frac{R^2}{a}\right)^\ell}{r^{\ell+2}} \right) P_\ell(\cos\theta) \tag{3.140}$$

Evaluating at $r = a$ gives

$$\left.\frac{\partial V^{in}}{\partial r}\right|_{r=a} = \sum_{\ell=0}^{\infty} A_\ell^{in} a^{\ell+1} \left(\frac{\ell}{a^2} + (\ell+1)\, \frac{\frac{R}{a}\left(\frac{R^2}{a}\right)^\ell}{a^{\ell+2}} \right) P_\ell(\cos\theta) \tag{3.141}$$

$$\left.\frac{\partial V^{out}}{\partial r}\right|_{r=a} = \sum_{\ell=0}^{\infty} A_\ell^{in} a^{\ell+1} (\ell+1) \left(-\frac{1}{a^2} + \frac{\frac{R}{a}\left(\frac{R^2}{a}\right)^\ell}{a^{\ell+2}} \right) P_\ell(\cos\theta) \tag{3.142}$$

When we difference the two, the second terms in the expressions cancel, leaving

$$-\sum_{\ell=0}^{\infty}(2\ell+1)A_{\ell}^{in}a^{\ell-1}P_{\ell}(\cos\theta)=-\frac{q\,\delta(\theta)}{2\,\pi\,a^2\epsilon_o\,\sin\theta} \tag{3.143}$$

This is our inhomogeneous boundary condition so, as usual, we must use orthonormality to obtain a formula for the coefficients in terms of an integral of the boundary condition with the orthonormal functions. We can multiply by $P_{\ell'}(\cos\theta)\sin\theta$ and integrate over θ, or we can just apply orthonormality. (Recall the orthonormality relation: $[2/(2\ell+1)]\int_0^{\pi}\sin\theta\,d\theta\,P_{\ell}(\cos\theta)\,P_{\ell'}(\cos\theta)=\delta_{\ell\ell'}$). This extracts the $A_{\ell'}^{in}$ term we want, and it also simplifies the right-hand side:

$$-2\,A_{\ell'}^{in}a^{\ell'-1}=-\frac{q}{2\,\pi\,a^2\epsilon_o}\int_0^{\pi}\sin\theta\,d\theta\,\frac{\delta(\theta)\,P_{\ell'}(\cos\theta)}{\sin\theta} \tag{3.144}$$

$$=-\frac{q}{2\,\pi\,a^2\epsilon_o}\,P_{\ell'}(\cos(\theta=0))=-\frac{q}{2\,\pi\,a^2\epsilon_o} \tag{3.145}$$

$$A_{\ell'}^{in}=\frac{1}{a^{\ell'+1}}\,\frac{q}{4\,\pi\,\epsilon_o} \tag{3.146}$$

Writing the full solution, we have

$$V(r < a, \theta) = \frac{q}{4\pi\epsilon_o} \sum_{\ell=0}^{\infty} \left(\frac{r^\ell}{a^{\ell+1}} - \frac{\frac{R}{a}\left(\frac{R^2}{a}\right)^\ell}{r^{\ell+1}} \right) P_\ell(\cos\theta) \qquad (3.147)$$

$$V(r > a, \theta) = \frac{q}{4\pi\epsilon_o} \sum_{\ell=0}^{\infty} \left(\frac{a^\ell}{r^{\ell+1}} - \frac{\frac{R}{a}\left(\frac{R^2}{a}\right)^\ell}{r^{\ell+1}} \right) P_\ell(\cos\theta) \qquad (3.148)$$

The form is hardly one we would have guessed! Separation of variables is more algorithmic than method of images, but it is also less intuitive. We will connect the two next.

Recognize that the integral over the inhomogeneous boundary condition that we expect from past experience with separation of variables has already been done on the prior page, so it is no longer explicit here; it yielded the $a^{-(\ell+1)}$ factor that is present in all the terms (in the first term for the $r > a$ solution, it is multiplied by $a^{2\ell+1}$, a factor that came from the homogeneous boundary conditions, so it is no longer obvious). Also, that integral did not include an integral over ϕ as we might have expected. We could have integrated over ϕ on both sides if we wanted, yielding a closer analogy to Equation 3.92, but it would have just yielded a common factor of 2π on the two sides since neither side has ϕ dependence. We did not need to do this because the problem is azimuthally symmetric and thus we know the solution must include only the $m = 0$ term. We will see such an integral when we consider cases without azimuthal symmetry.

Connecting Method of Images to Separation of Variables via a Useful Expansion in Legendre Polynomials

We have two techniques — method of images and separation of variables — that we can evidently use for the same problem. By the Uniqueness Theorem, the solutions must be the same. Comparing Equations 3.147 and 3.148 that we just obtained via separation of variables to Equation 3.25 obtained via method of images, the connection is hardly obvious! To see it, we must first prove a theorem.

We will show

$$\boxed{\frac{1}{|\vec{r} - \vec{r}\,'|} = \sum_{\ell=0}^{\infty} \frac{r_<^\ell}{r_>^{\ell+1}} P_\ell(\cos\gamma)} \tag{3.149}$$

with $\qquad r_< = \min(|\vec{r}|, |\vec{r}\,'|) \qquad r_> = \max(|\vec{r}|, |\vec{r}\,'|) \qquad \cos\gamma = \hat{r} \cdot \hat{r}\,'$

This will let us go back and forth between separation-of-variables solutions and functions that look like the Coulomb potential (*e.g.*, point charge near the grounded sphere!). Griffiths *sort of* derives this, using a far less interesting and powerful technique. He also does it in §3.4.1, after the discussion of separation of variables, so he is unable to use this theorem to connect the method of images and separation of variables solutions for the point charge near the grounded, conducting sphere.

To prove this, orient the coordinate system so $\vec{r}\,' = r'\,\hat{z}$. The function on the left-hand side of Equation 3.149 is the potential at $\vec{r}$ of a point charge $q = 4\,\pi\,\epsilon_o$ in magnitude (not units!) at r' along the z-axis. It satisfies azimuthal symmetry and thus is expandable in terms of the above solutions of Laplace's Equation in spherical coordinates with azimuthal symmetry (because these solutions form a complete, orthonormal set!):

$$\frac{1}{|\vec{r} - \vec{r}\,'|} = \sum_{\ell=0}^{\infty} \left(A_\ell\, r^\ell + \frac{B_\ell}{r^{\ell+1}} \right) P_\ell(\cos\theta) \tag{3.150}$$

Consider two cases separately:

▶ $r < r'$

We must eliminate the B_ℓ coefficients to keep the function finite as $r \to 0$. To find the A_ℓ, let's consider the point $\vec{r} = r\,\hat{z}$ (i.e., $\cos\gamma = 1$), which implies

$$\frac{1}{r' - r} = \frac{1}{|\vec{r} - \vec{r}\,'|} = \sum_{\ell=0}^{\infty} A_\ell\, r^\ell \tag{3.151}$$

(Recall, $P_\ell(1) = 1$.) Thus, the A_ℓ are just the coefficients of the power series expansion of the left side, which we know (recall: $(1-x)^{-1} = 1 + x + x^2 + \cdots$ for $0 < x < 1$) is

$$\frac{1}{r' - r} = \frac{1}{r'}\, \frac{1}{1 - \frac{r}{r'}} = \frac{1}{r'} \sum_{\ell=0}^{\infty} \left(\frac{r}{r'} \right)^\ell \tag{3.152}$$

The series converges because $x = r/r' < 1$. Thus, $A_\ell = 1/(r')^{\ell+1}$. This now sets the $\{A_\ell\}$ for arbitrary $\vec{r}$ (i.e., arbitrary $\cos\gamma$ rather than the special case $\cos\gamma = 1$ we have considered).

▶ $r > r'$

We must eliminate the A_ℓ coefficients to keep the function finite as $r \to \infty$. Again, consider $\vec{r} = r\,\hat{z}$, which implies

$$\frac{1}{r - r'} = \frac{1}{|\vec{r} - \vec{r}'|} = \sum_{\ell=0}^{\infty} \frac{B_\ell}{r^{\ell+1}} \tag{3.153}$$

For this case, we consider an expansion in r'/r rather than r/r' because now $0 < r'/r < 1$ while, above, $0 < r/r' < 1$. Again, the B_ℓ are just the coefficients of the power series expansion of the left side, which we know is

$$\frac{1}{r - r'} = \frac{1}{r}\,\frac{1}{1 - \frac{r'}{r}} = \frac{1}{r}\sum_{\ell=0}^{\infty}\left(\frac{r'}{r}\right)^{\ell} \tag{3.154}$$

Thus, $B_\ell = (r')^{\ell}$.

Combining the above two cases, and generalizing back from $\cos\theta$ to $\cos\gamma$, yields Equation 3.149.

A few notes on the above derivation:

▶ Note some elements of technique: without loss of generality, we: a) set $\vec{r}' = r'\,\hat{z}$ so $\cos\gamma = \cos\theta$; and b) evaluated the expression at $\cos\theta = 1$, similar to the manner in which we applied the boundary conditions for the point charge near the grounded sphere. These are useful techniques to keep in mind for the future.

▶ Note that the process was effectively separation of variables, separately in the $r < r'$ and $r > r'$ spaces (like our separate consideration of $r < R$ and $r > R$ in the previous example) but with an unusual boundary condition: Rather than specifying a condition on the function (the "potential") on the boundary $r = r'$, we used the fact that we knew the solution along the line $\vec{r} = r\,\hat{r}'$ (which we took to be $\vec{r} = r\,\hat{z}$ in this case). That is, we specified the potential for a locus of points in the volume $\mathcal{V}$ rather than on the surface $\mathcal{S}(\mathcal{V})$. We do not have a general theorem about such boundary conditions because the derivation of the Uniqueness Theorem used Green's Theorem, which involves $\mathcal{S}$. Evidently, though, appropriate specification of the potential on some locus of points in $\mathcal{V}$ is also sufficient to yield a unique solution!

▶ We could have derived the result instead by treating this as a separation of variables problem with a Neumann boundary condition at $r = r'$ due to the singularity there, which is the same singularity one gets for the potential of a point charge of value $q = 4\pi\epsilon_o$ as we explained above. Explicitly, we would combine the solution to Example 3.6 (a spherical shell of charge $\sigma(R, \theta)$ in free space) with the charge density we developed for the point charge near the conducting sphere. In fact, we can read off the result from Equations 3.147 and 3.148: let $R \to 0$, replace a by r', and replace q by $4\pi\epsilon_o$! That approach is a bit more cumbersome but benefits from the Uniqueness Theorem and the full separation-of-variables machinery. We did not do that here because we knew ahead of time the solution on the $\vec{r} = r\hat{r}'$ locus, and this approach introduced you to a new technique.

With that theorem proven, we can make the advertised connection. If we compare Equations 3.147 and 3.148 from separation of variables for the point charge near the conducting sphere to Equation 3.149, we see that all four terms in the former are of the form used in the latter. The first term of the first equation has $r_< = r$ and $r_> = a$ as appropriate for $r < a$, while the first term of the second equation has $r_< = a$ and $r_> = r$ as needed for $r > a$. The second terms of both equations are of the same form with $r_< = R^2/a$, $r_> = r$ and the charge multiplied by $-R/a$. Thus, we recover

$$V(\vec{r}) = \frac{q}{4\pi\epsilon_o}\left[\frac{1}{|\vec{r} - a\hat{z}|} - \frac{R/a}{\left|\vec{r} - \frac{R^2}{a}\hat{z}\right|}\right] \tag{3.155}$$

which matches Equation 3.25. Remarkable! This is a case where we were able to use separation of variables to recover the full potential and thus the full method of images solution, which we know then gives us the Green Function: it is possible!

Could we have done a similar thing if we had a point charge in the five-sides-grounded box problem? There is no reason to think it would not work.

In fact, we will later show how to use a similar technique to find the Green Function in spherical coordinates for systems without azimuthal symmetry.

Separation of Variables in Spherical Coordinates without Azimuthal Symmetry

The Full Polar Equation Solution: the Associated Legendre Polynomials

There is a relation yielding the associated Legendre polynomials for $m \geq 0$ from the Legendre polynomials:

$$P_\ell^m(x) = (-1)^m (1 - x^2)^{m/2} \frac{d^m}{dx^m} P_\ell(x) \tag{3.156}$$

which, using Rodrigues' Formula (Equation 3.112), implies

$$P_\ell^m(x) = \frac{(-1)^m}{2^\ell \, \ell!} (1 - x^2)^{m/2} \frac{d^{\ell+m}}{dx^{\ell+m}} (x^2 - 1)^\ell \tag{3.157}$$

which is now valid for all m.

Some properties of the P_ℓ^m:

- $P_\ell^0 = P_\ell$ ($m = 0$ in Equation 3.156).
- The parity in x (evenness/oddness) of the associated Legendre functions is given by $(-1)^{\ell+m}$ (where -1 implies oddness): the parity of P_ℓ is given by $(-1)^\ell$, and each derivative changes the parity by a factor of -1 (note that the powers of $(1 - x^2)$ have no effect on the parity because it is an even function).
- The $\{P_\ell^m\}$ for a given m form an orthogonal set. We can see this by going back to the polar equation (Equation 3.103 or 3.109) and recognizing that it is literally a *different* differential equation for every value of m. The solutions at different m are related to each other by Equation 3.156 because the differential equations are related to each other, but they are different equations and thus their solutions form different orthogonal sets. We will address their normalization imminently.

There are a number of other properties of these functions, but it is more useful to consider them together with the ϕ solutions.

The Full Solution to the Angular Piece of Laplace's Equation: the Spherical Harmonics

When one combines the $P_\ell^m(\cos\theta)$ and the $e^{im\phi}$ solutions of the polar and azimuthal equations, one obtains the *Spherical Harmonics*

$$Y_{\ell m}(\theta,\phi) = \sqrt{\frac{2\ell+1}{4\pi}\frac{(\ell-m)!}{(\ell+m)!}}\, P_\ell^m(\cos\theta)\, e^{im\phi} \tag{3.158}$$

They are an orthonormal, complete basis for functions on the sphere (θ,ϕ) (assuming the z-axis is part of the sphere; recall our comment about a second set of solutions to the Legendre equation if it is not). The $(\ell-m)!/(\ell+m)!$ factors come from normalizing the P_ℓ^m, while the 2π comes from normalizing the $e^{im\phi}$. They satisfy numerous important and useful conditions:

▶ Conjugation:

$$Y_{\ell(-m)}(\theta,\phi) = (-1)^m Y_{\ell m}^*(\theta,\phi) \tag{3.159}$$

▶ Orthonormality:

$$\int_0^{2\pi} d\phi \int_0^\pi \sin\theta\, d\theta\, Y_{\ell'm'}^*(\theta,\phi)\, Y_{\ell m}(\theta,\phi) = \delta_{\ell\ell'}\delta_{mm'} \tag{3.160}$$

▶ Completeness ($\cos\theta$ is the argument because the differential is $\sin\theta\,d\theta = -d(\cos\theta)$):

$$\sum_{\ell=0}^{\infty}\sum_{m=-\ell}^{\ell} Y_{\ell m}^*(\theta',\phi')Y_{\ell m}(\theta,\phi) = \delta(\phi-\phi')\,\delta(\cos\theta-\cos\theta') \qquad (3.161)$$

▶ $m = 0$ devolves to Legendre polynomials:

$$Y_{\ell 0}(\theta,\phi) = \sqrt{\frac{2\ell+1}{4\pi}}\,P_\ell(\cos\theta) \qquad (3.162)$$

This should be obvious from Equation 3.156, the relation between the Legendre and the associated Legendre polynomials.

▶ The $\theta = 0$ behavior is simple given Equation 3.156 (the $(1-x^2)$ factor):

$$P_\ell^{m\neq0}(\pm 1) = 0 \quad \implies \quad Y_{\ell\,m\neq0}(\theta=0,\phi) = Y_{\ell\,m\neq0}(\theta=\pi,\phi) = 0 \qquad (3.163)$$

This condition ensures the $Y_{\ell\,m\neq0}$ are single-valued at the poles. (Single-valuedness is automatic for $m = 0$ because $e^{i(0)\phi} = 1$.) Recall that we also stated $P_\ell(1) = 1$, $P_\ell(-1) = (-1)^\ell$, which implies

$$Y_{\ell 0}(\theta=0,\phi) = \sqrt{\frac{2\ell+1}{4\pi}} \qquad Y_{\ell 0}(\theta=\pi,\phi) = (-1)^\ell\sqrt{\frac{2\ell+1}{4\pi}} \qquad (3.164)$$

▶ The above implies that any expansion in terms of $Y_{\ell m}$ simplifies at $\theta = 0$, π:

$$\text{given} \quad g(\theta, \phi) = \sum_{\ell=0}^{\infty} \sum_{m=-\ell}^{\ell} A_{\ell m} Y_{\ell m}(\theta, \phi) \tag{3.165}$$

$$\text{then} \quad g(\theta = 0, \phi) = \sum_{\ell=0}^{\infty} \sqrt{\frac{2\ell+1}{4\pi}} A_{\ell 0} \tag{3.166}$$

$$\text{and} \quad g(\theta = \pi, \phi) = \sum_{\ell=0}^{\infty} (-1)^{\ell} \sqrt{\frac{2\ell+1}{4\pi}} A_{\ell 0} \tag{3.167}$$

▶ The *Addition Theorem for Spherical Harmonics*: Given $\widehat{r}$ and $\widehat{r}'$ pointing in the directions (θ, ϕ) and (θ', ϕ'), respectively, then

$$\boxed{P_{\ell}(\widehat{r} \cdot \widehat{r}') = \frac{4\pi}{2\ell+1} \sum_{m=-\ell}^{\ell} Y_{\ell m}^{*}(\theta', \phi') \, Y_{\ell m}(\theta, \phi)} \tag{3.168}$$

where $\widehat{r} \cdot \widehat{r}' = \cos\gamma = \cos\theta \cos\theta' + \sin\theta \sin\theta' \cos(\phi - \phi')$. The proof of this can be found in Jackson §3.6.

▶ An important corollary of the Addition Theorem can be obtained by combining the above with Equation 3.149, the formula for the inverse of the relative distance between two points in terms of the Legendre polynomials:

$$\frac{1}{|\vec{r} - \vec{r}\,'|} = \sum_{\ell=0}^{\infty} \frac{r_<^\ell}{r_>^{\ell+1}} P_\ell(\cos\gamma)$$

Plugging in the Addition Theorem gives us

$$\frac{1}{|\vec{r} - \vec{r}\,'|} = 4\pi \sum_{\ell=0}^{\infty} \sum_{m=-\ell}^{\ell} \frac{1}{2\ell+1} \frac{r_<^\ell}{r_>^{\ell+1}} Y_{\ell m}^*(\theta',\phi') Y_{\ell m}(\theta,\phi) \qquad (3.169)$$

The utility of this relation is even more obvious than that of Equation 3.149, especially for doing integrals over charge distributions with the relative distance function (*i.e.*, calculating the potential due to Coulomb's Law): decompose the charge distribution in terms of spherical harmonics and integrate the charge distribution in a particular spherical harmonic $Y_{\ell m}$ over r' with weighting by $(r')^\ell$ to obtain the component of the potential at a distance r from the origin with spatial dependence $Y_{\ell m}(\theta,\phi)/r^{\ell+1}$.

The Full Solution of Laplace's Equation in Spherical Coordinates

Putting it all together, we see that the most general solution to Laplace's Equation in spherical coordinates is r

$$V(r, \theta, \phi) = \sum_{\ell=0}^{\infty} \sum_{m=-\ell}^{\ell} \left(A_{\ell m} \, r^\ell + \frac{B_{\ell m}}{r^{\ell+1}} \right) Y_{\ell m}(\theta, \phi) \tag{3.170}$$

Again, the coefficients $\{A_{\ell m}\}$ and $\{B_{\ell m}\}$ are set by the volume under consideration and one or the other entire set may vanish. As well, application of the boundary conditions will require the orthonormality relations for the spherical harmonics.

As with the case of azimuthal symmetry, we note that, in the process of doing separation of variables, we have proven that the angular solution satisfies the *eigenvalue-eigenfunction equation*

$$\nabla^2 Y_{\ell m}(\theta, \phi) = -\frac{\ell(\ell+1)}{r^2} Y_{\ell m}(\theta, \phi) \tag{3.171}$$

As before, the appearance of r^2 on the right side is not surprising. Note also that m does not appear in the angular equation. This is because Laplace's Equation itself is spherically (and therefore azimuthally) symmetric. The charge distribution and boundary conditions are what may break the spherical symmetry.

Expansion of the Green Function in Spherical Coordinates in Terms of the Spherical Harmonics

The fact that the spherical harmonics combined with the usual power laws in radius solve Laplace's Equation for problems that are separable in spherical coordinates can be used to show that the Green Function for such problems will have a convenient expansion in terms of the radial solutions and spherical harmonics, like Equation 3.170. It is convenient to recall at this point that a Green Function is specified (is unique) once one specifies the geometry and the type of boundary condition; the value of the boundary condition *does not* affect the Green Function. So, once we have specified a geometry and *type* of boundary condition, the expansion can be determined and is unique. Alternatively, one can think of this expansion as a generalization of the corollary of the Addition Theorem, Equation 3.169. It is shown by using the completeness property of the spherical harmonics and the eigenvalue-eigenfunction equation for the angular solution. But let's see that this is true explicitly for a couple example geometries first:

▶ *Free space*
  The corollary of the Addition Theorem above is the desired expansion of the Green Function for charge in free space with no finite radius boundaries and with the condition $V \to 0$ as $r \to \infty$.

▶ *Point charge near a grounded, conducting sphere*
For this geometry, we saw that the Green Function can be written as sum of the
Coulomb potential of two point charges, the original one at $r'\,\widehat{z}$ and the image
charge $q' = -q\,R/r'$ at $\widehat{z}\,R^2/r'$:

$$G(\vec{r}, \vec{r}') = \frac{1}{4\,\pi\,\epsilon_o} \left[\frac{1}{|\vec{r} - \vec{r}'|} - \frac{R/r'}{\left|\vec{r} - \vec{r}'\left(\frac{R}{r'}\right)^2\right|} \right] \tag{3.172}$$

Using the same corollary of the Addition Theorem, we can immediately write
(using the fact $r'\,(R/r')^2 < r'$ always because the the image charge is always
at radius $< R$ while the true charge is at $r' > R$):

$$G(\vec{r}, \vec{r}') = \frac{1}{\epsilon_o} \sum_{\ell=0}^{\infty} \sum_{m=-\ell}^{\ell} \left[\frac{r_<^{\ell}}{r_>^{\ell+1}} - \frac{R}{r'}\frac{\left[r'\left(\frac{R}{r'}\right)^2\right]^{\ell}}{r^{\ell+1}} \right] \frac{Y_{\ell m}^*(\theta', \phi')\,Y_{\ell m}(\theta, \phi)}{2\,\ell + 1} \tag{3.173}$$

$$= \frac{1}{\epsilon_o} \sum_{\ell=0}^{\infty} \sum_{m=-\ell}^{\ell} \left[\frac{r_<^{\ell}}{r_>^{\ell+1}} - \frac{1}{R}\left(\frac{R^2}{r\,r'}\right)^{\ell+1} \right] \frac{Y_{\ell m}^*(\theta', \phi')\,Y_{\ell m}(\theta, \phi)}{2\,\ell + 1} \tag{3.174}$$

Note also the symmetry in $\vec{r}$ and $\vec{r}'$ is manifest.

In both cases, we finally have forms for the Green Function that could plausibly come from separation of variables. Note, however, that we did not use separation of variables to obtain it; we used the method of images combined with the corollary of the Addition Theorem.

Earlier, we solved for the potential of the latter configuration using separation of variables with azimuthal symmetry, Equations 3.147 and 3.148 reproduced here but rewritten using the $r_<$, $r_>$ notation:

$$V(r, \theta) = \frac{q}{4\pi\epsilon_o} \sum_{\ell=0}^{\infty} \left(\frac{r_<^\ell}{r_>^{\ell+1}} - \frac{\frac{R}{a}\left(\frac{R^2}{a}\right)^\ell}{r^{\ell+1}} \right) P_\ell(\cos\theta)$$

$$\text{with} \quad r_< = \min(r, a) \qquad r_> = \max(r, a) \tag{3.175}$$

Why was this not enough to give us the full Green Function? Because this solution for the potential in terms of Legendre polynomials assumed the point charge was along the z-axis.

What we can do is generalize this solution by replacing $\cos\theta$ with $\cos\gamma = \hat{r} \cdot \hat{r}'$ and a with r' followed by application of the Addition Theorem. Then the solution would be in a form where we could read off the Green Function expansion in spherical coordinates not assuming azimuthal symmetry. But this approach is not the same as obtaining the solution directly, and clearly the above approach does not generalize to a system that does not have azimuthal symmetry.

The general approach to the problem of finding the Green Function for an arbitrary (spherical) geometry is to go back to the definition of the Green Function:

$$-\epsilon_o \nabla_{\vec{r}}^2 G(\vec{r}, \vec{r}\,') = \delta(\vec{r} - \vec{r}\,') \tag{3.176}$$

and decompose both sides in terms of spherical harmonics. We do not know the Green Function yet, so its expansion is the arbitrary general form, which here we write

$$G(\vec{r}, \vec{r}\,') = \sum_{\ell=0}^{\infty} \sum_{m=-\ell}^{\ell} A_{\ell m}(r|\vec{r}\,') \, Y_{\ell m}(\theta, \phi) \tag{3.177}$$

where the coefficients in the expansion $A_{\ell m}$ depend on r, as usual, and they also depend parametrically on $\vec{r}\,'$ because it is a parameter in the differential equation. (We do not know the solutions for the radial dependence of the $A_{\ell m}$ yet for the general case we are trying to solve (which is not Laplace's Equation!), so we cannot assume they are the power laws we saw for solutions to Laplace's Equation.)

The right side can be rewritten using the breakdown of the delta function into delta functions in each spherical coordinate followed by completeness of the spherical harmonics. The breakdown of the delta function is:

$$\delta(\vec{r} - \vec{r}\,') = \frac{\delta(r - r\,')}{r^2}\,\delta(\phi - \phi\,')\,\delta(\cos\theta - \cos\theta\,') \tag{3.178}$$

The $1/r^2$ on the radial component is required to cancel the r^2 in the volume element in spherical coordinates. The fact that the delta function in θ is a function of $\cos\theta$ and $\cos\theta\,'$ is because the volume element contains $\sin\theta\,d\theta = d(\cos\theta)$. One could have instead written $\delta(\theta - \theta\,')/\sin\theta$ as we did when rewriting the point charge near the grounded, conducting sphere as a surface charge density $\sigma(\theta)$, Equation 3.129. Using completeness of the spherical harmonics, we have

$$\delta(\vec{r} - \vec{r}\,') = \frac{\delta(r - r\,')}{r^2} \sum_{\ell=0}^{\infty} \sum_{m=-\ell}^{\ell} Y_{\ell m}^*(\theta\,', \phi\,')\,Y_{\ell m}(\theta, \phi) \tag{3.179}$$

Thus, our differential equation for the Green Function becomes

$$-\epsilon_o \nabla_{\vec{r}}^2 \sum_{\ell=0}^{\infty} \sum_{m=-\ell}^{\ell} A_{\ell m}(r|\vec{r}') \, Y_{\ell m}(\theta, \phi) = \frac{\delta(r - r')}{r^2} \sum_{\ell=0}^{\infty} \sum_{m=-\ell}^{\ell} Y_{\ell m}^*(\theta', \phi') \, Y_{\ell m}(\theta, \phi)$$

(3.180)

Note that the Laplacian acts on the unprimed coordinates only. When we evaluate the action of the Laplacian, a cross term $\vec{\nabla}_{\vec{r}} A_{\ell m}(r|\vec{r}') \cdot \vec{\nabla}_{\vec{r}} Y_{\ell m}(\theta, \phi)$ appears, but it vanishes because the first term points along $\hat{r}$ while the second is along $\hat{\theta}$ and $\hat{\phi}$, leaving only $\nabla_{\vec{r}}^2$ acting on each factor in the product individually. We wrote down earlier Equation 3.171, the eigenvalue-eigenfunction equation satisfied by the angular solutions of Laplace's Equation, which we use here to evaluate $\nabla_{\vec{r}}^2 Y_{\ell m}(\theta, \phi)$:

$$-\epsilon_o \sum_{\ell=0}^{\infty} \sum_{m=-\ell}^{\ell} \left[\left(\nabla_{\vec{r}}^2 - \frac{\ell(\ell+1)}{r^2} \right) A_{\ell m}(r|\vec{r}') \right] Y_{\ell m}(\theta, \phi)$$

(3.181)

$$= \frac{\delta(r - r')}{r^2} \sum_{\ell=0}^{\infty} \sum_{m=-\ell}^{\ell} Y_{\ell m}^*(\theta', \phi') \, Y_{\ell m}(\theta, \phi)$$

Note that the Laplacian on the left side is now acting with its radial derivatives only on $A_{\ell m}$; its action on the spherical harmonics has yielded the $\ell(\ell+1)/r^2$ term.

Lecture 10:

Advanced Electrostatics VI:

Spherical Harmonic Expansion of Green Functions (cont.)
Examples of Using the Spherical Harmonic Expansion
for Green Functions

Date Revised: 2024/02/08 06:15
Revised selection of skipped material
Date Given: 2024/02/06

The coefficients of the individual $Y_{\ell m}(\theta, \phi)$ on the two sides must be equal because of the orthonormality relation for the spherical harmonics, implying

$$-\epsilon_o \left[\left(\nabla_{\vec{r}}^2 - \frac{\ell(\ell+1)}{r^2} \right) A_{\ell m}(r | \vec{r}') \right] = \frac{\delta(r - r')}{r^2} Y_{\ell m}^*(\theta', \phi') \qquad (3.182)$$

Now, given that we have $Y_{\ell m}^*(\theta', \phi')$ on the right side (from applying completeness), and again the spherical harmonics are orthonormal functions, the dependence of $A_{\ell m}(r | \vec{r}')$ on its $\vec{r}'$ angular coordinates must be proportional to $Y_{\ell m}^*(\theta', \phi')$. Therefore, we may write (with $g_\ell(r, r')$ still to be determined)

$$A_{\ell m}(r | r', \theta', \phi') = g_\ell(r, r') Y_{\ell m}^*(\theta', \phi') \qquad (3.183)$$

Plugging in this form to the above reduced version of Laplace's Equation and canceling $Y_{\ell m}^*(\theta', \phi')$, we get:

$$-\epsilon_o \left(\nabla_{\vec{r}}^2 - \frac{\ell(\ell+1)}{r^2} \right) g_\ell(r, r') = \frac{\delta(r - r')}{r^2} \qquad (3.184)$$

Only the Laplacian's radial derivatives yield a nonzero contribution here, so we have (also multiplying both sides by $-r^2/\epsilon_o$):

$$\frac{d}{dr} \left[r^2 \frac{d}{dr} g_\ell(r, r') \right] - \ell(\ell+1) g_\ell(r, r') = -\frac{\delta(r - r')}{\epsilon_o} \qquad (3.185)$$

We see that, when $r \neq r'$ (r' is a parameter, not a variable, here), $g_\ell(r, r')$ satisfies the radial ODE in r from separation of variables in spherical coordinates, Equation 3.102. Therefore, in the two separate regions $r < r'$ and $r > r'$, the solutions to that ODE are also our solutions here:

$$g_\ell(r, r') = \begin{cases} A_\ell^{in}(r') \, r^\ell + B_\ell^{in}(r') \, r^{-(\ell+1)} & r < r' \\ A_\ell^{out}(r') \, r^\ell + B_\ell^{out}(r') \, r^{-(\ell+1)} & r > r' \end{cases} \tag{3.186}$$

Because r' is a parameter of the differential equation, the coefficients and therefore the solutions depend on it parametrically. Therefore, the general form for the expansion of the Green Function in spherical harmonics is

$$\boxed{r < r' : G(\vec{r}, \vec{r}') = \sum_{\ell=0}^{\infty} \sum_{m=-\ell}^{\ell} \left[A_\ell^{in}(r') \, r^\ell + \frac{B_\ell^{in}(r')}{r^{\ell+1}} \right] Y_{\ell m}(\theta, \phi) \, Y_{\ell m}^*(\theta', \phi')} \tag{3.187}$$

$$\boxed{r > r' : G(\vec{r}, \vec{r}') = \sum_{\ell=0}^{\infty} \sum_{m=-\ell}^{\ell} \left[A_\ell^{out}(r') \, r^\ell + \frac{B_\ell^{out}(r')}{r^{\ell+1}} \right] Y_{\ell m}(\theta, \phi) \, Y_{\ell m}^*(\theta', \phi')}$$

$$\tag{3.188}$$

To determine the coefficients, we need to apply boundary conditions. Since we have not yet specified the geometry and boundary conditions, the only generic boundary condition we can write down is the one at $r = r'$, which we obtain by integrating Equation 3.185 from $r = r' - \epsilon$ to $r = r' + \epsilon$ and letting $\epsilon \to 0$:

$$\int_{r'-\epsilon}^{r'+\epsilon} dr \left\{ \frac{d}{dr} \left[r^2 \frac{d}{dr} g_\ell(r, r') \right] - \ell(\ell+1)\, g_\ell(r, r') \right\} = - \int_{r'-\epsilon}^{r'+\epsilon} dr \frac{\delta(r - r')}{\epsilon_o}$$
(3.189)

The first term is the integral of a total differential, so it is trivially integrated. For the second term, the form of $g_\ell(r, r')$, where it is sum of two terms, each of which includes a power law in r and some function of r' not dependent on r, ensures it cannot diverge at $r = r'$. Therefore, the second term is an integral of a function with no singularity at $\epsilon = 0$ (*i.e.*, at $r = r'$) and thus, as $\epsilon \to 0$, that integral vanishes. The right side gives $-1/\epsilon_0$ when integrated. Therefore, we have

$$\lim_{\epsilon \to 0} \left[r^2 \frac{d}{dr} g_\ell(r, r') \right]\Bigg|_{r=r'-\epsilon}^{r=r'+\epsilon} = -\frac{1}{\epsilon_o}$$

$$\frac{d}{dr} g_\ell^{out}(r, r')\bigg|_{r=r'} - \frac{d}{dr} g_\ell^{in}(r, r')\bigg|_{r=r'} = -\frac{1}{\epsilon_0 (r')^2}$$
(3.190)

where $g_\ell^{out}(r, r')$ is the $r > r'$ solution and $g_\ell^{in}(r, r')$ is the $r < r'$ solution. This is a Neumann-type boundary condition as we had for the examples of the arbitrary charge density on a sphere $\sigma(R, \theta)$ and for the point charge near the conducting sphere $\sigma(a, \theta) = \delta(\cos\theta)/2\pi a^2$, but this form is equivalent to what we obtained in those cases *after* we used orthonormality in θ to isolate the expansion coefficients.

We note, as an aside, that we derived the above matching condition Equation 3.190 in a somewhat different way here than when we considered the above examples. In the examples, we used the fact that we knew the boundary condition on the normal derivative of the potential from Gauss's Law. Here, we effectively rederived that boundary condition for the special case of a radial boundary because we have only to determine the radial function $g(r, r')$. We could have gone back a step and written down the boundary condition on the normal derivative of the potential and derived the same condition above, but it would have required going back to the full potential and applying orthonormality and completeness again. We circumvented that step by rederiving the boundary condition considering only the radial function.

Evaluating the above condition explicitly using the $r < r'$ and $r > r'$ pieces of the solution, and multiplying both sides by $(r')^2$, we obtain

$$\ell \left[A_\ell^{in}(r') - A_\ell^{out}(r') \right] (r')^{\ell+1} + (\ell+1) \left[B_\ell^{out}(r') - B_\ell^{in}(r') \right] (r')^{-\ell} = \frac{1}{\epsilon_o} \quad (3.191)$$

Since A_ℓ^{in}, B_ℓ^{in}, A_ℓ^{out}, and B_ℓ^{out} all depend on r', all the powers of r' match up. The above is an inhomogeneous boundary condition on $g_\ell(r, r')$.

The finite discontinuity in the radial derivative of $g_\ell(r, r')$ implies that $g_\ell(r, r')$ itself must be continuous at $r = r'$: the derivative would need to have a singularity in order for there to be a discontinuity in $g_\ell(r, r')$. Therefore, we also have the condition

$$g_\ell^{out}(r = r', r') - g_\ell^{in}(r = r', r') = 0 \qquad (3.192)$$

Explicitly evaluating this condition, again using the two portions of the solution, yields

$$\left[A_\ell^{in}(r') - A_\ell^{out}(r') \right] (r')^{2\ell+1} + \left[B_\ell^{in}(r') - B_\ell^{out}(r') \right] = 0 \qquad (3.193)$$

The above is a homogeneous boundary condition on $g_\ell(r, r')$.

The above two matching conditions, along with application of the boundary conditions that define Dirichlet or Neumann Green Functions (Equations 3.47 and 3.49), provide four conditions for the four unknowns $A_\ell^{in}(r')$, $B_\ell^{in}(r')$, $A_\ell^{out}(r')$, and $B_\ell^{out}(r')$, which should fully specify them. We finally have a completely algorithmic way to obtain the full Green Function! What a powerful technique! This general approach can be applied for any coordinate system in which Laplace's Equation and the boundary conditions are separable.

We also note that the above two equations imply the solutions $A_\ell^{in}(r')$, $B_\ell^{in}(r')$, $A_\ell^{out}(r')$, and $B_\ell^{out}(r')$ will be power laws in r'. This is sensible: because we expect the Green Function to be symmetric in $\vec{r}$ and $\vec{r}'$, the functional dependences on $\vec{r}$ and $\vec{r}'$, and thus on r and r', must be the same, and, so, because G has power-law dependence on r, it must also for r'.

We note that the above procedure is almost identical to what we would have done had we used separation of variables more explicitly. We certainly assumed a SoV-style expansion for the Green Function, Equation 3.177. As noted above, we could have applied a Neumann-style boundary condition at $r = r'$ by arguing that $G(\vec{r}, \vec{r}')$ is the potential at $\vec{r}$ due to a point charge at $\vec{r}'$. That would have yielded the same conditions on $g_\ell(r, r')$ that integrating the ODE, Equation 3.185, yielded. We chose to do it the way we did so we would have to rely less on such analogies. But one can see that the approaches are equivalent.

Example 3.9: Expansion in spherical harmonics for the Green Function for $R < r < \infty$ with Dirichlet boundary conditions at $r = R$ and $r \to \infty$

These boundary conditions impose the requirement $G_D(\vec{r}, \vec{r}') = 0$ for $\vec{r} \in \mathcal{S}, \mathcal{V}$, $\vec{r}' \in \mathcal{S}$. We use the symmetry of the Dirichlet Green Function to convert this to the requirement $G_D(\vec{r}, \vec{r}') = 0$ for $\vec{r} \in \mathcal{S}, \vec{r}' \in \mathcal{S}, \mathcal{V}$ because we do not know the dependence of the coefficients on r' and we want to obtain relations between the coefficients of the expansion that are valid at all r', not just values on the boundary, because those full dependences are needed to use the matching conditions at $r = r'$ that we just derived. One can check that applying these conditions at $\vec{r}' \in \mathcal{S}$ does not result in useful information.

Our condition implies

$$0 = G_D(\vec{r} \in \mathcal{S}, \vec{r}' \in \mathcal{S}, \mathcal{V}) = \sum_{\ell=0}^{\infty} \sum_{m=-\ell}^{\ell} g_\ell(r \in \mathcal{S}, r' \in \mathcal{S}, \mathcal{V}) \, Y_{\ell m}^*(\theta', \phi') \, Y_{\ell m}(\theta, \phi) \tag{3.194}$$

Applying orthonormality of the $Y_{\ell m}(\theta', \phi')$, we obtain

$$0 = g_\ell(r \in \mathcal{S}, r' \in \mathcal{S}, \mathcal{V}) \, Y_{\ell m}(\theta, \phi) \tag{3.195}$$

By applying orthonormality of the $Y_{\ell m}(\theta, \phi)$, or simply noting $Y_{\ell m}(\theta, \phi)$ is in general nonzero, we obtain

$$g_\ell(r \in \mathcal{S}, r' \in \mathcal{S}, \mathcal{V}) = 0 \tag{3.196}$$

We will apply the above homogeneous boundary condition first, then the matching conditions (one homogeneous and one inhomogeneous boundary condition) at $r = r'$, because the Dirichlet BC are simpler algebraically (this is the same order of steps we used when we solved this problem using separation of variables).

First, consider the boundary at $r = R$. Since $r' > r = R$ for all $\vec{r}' \in \mathcal{V}$, this implies that we should require $g_\ell(r = R, r') = 0$ for the $r < r'$ solution, yielding:

$$A_\ell^{in}(r')\, R^\ell + B_\ell^{in}(r')\, R^{-(\ell+1)} = 0 \qquad \Longrightarrow \qquad B_\ell^{in}(r') = -R^{2\,\ell+1}\, A_\ell^{in}(r') \quad (3.197)$$

Next, consider the boundary at $r \to \infty$: $g_\ell(r \to \infty, r' \in \mathcal{V}) = 0$. Here, it is the $r > r'$ solution that applies, which implies $A_\ell^{out}(r') = 0$ for all r'.

Next, we apply the matching conditions at $r = r'$. First, the homogeneous one, continuity of $g_\ell(r, r')$ at $r = r'$ (Equation 3.193), which implies

$$A_\ell^{in}(r')(r')^{2\ell+1} + \left\{ \left[-A_\ell^{in}(r')\, R^{2\ell+1} \right] - B_\ell^{out}(r') \right\} = 0$$

$$\Longrightarrow \qquad B_\ell^{out}(r') = A_\ell^{in}(r') \left[(r')^{2\ell+1} - R^{2\ell+1} \right] \qquad (3.198)$$

The condition on the change in the radial derivative at $r = r'$, the inhomogeneous boundary condition, yielded Equation 3.191, which we plug into to obtain

$$\ell\, A_\ell^{in}(r')\, (r')^{\ell+1} + (\ell+1) A_\ell^{in}(r') \left[(r')^{2\ell+1} - R^{2\ell+1} + R^{2\ell+1} \right] (r')^{-\ell} = \frac{1}{\epsilon_o}$$

$$\Longrightarrow \qquad A_\ell^{in}(r') = \frac{1}{2\,\ell+1} \frac{1}{\epsilon_o} \frac{1}{(r')^{\ell+1}} \qquad (3.199)$$

Putting it all together, we have that the Green Function for this Dirichlet boundary condition, expanded in terms of spherical harmonics, is

$$
G(\vec{r}, \vec{r}\,') =
\begin{cases}
\dfrac{1}{\epsilon_o} \displaystyle\sum_{\ell=0}^{\infty} \sum_{m=-\ell}^{\ell} \left[\dfrac{r^\ell}{(r')^{\ell+1}} - \dfrac{1}{R}\left(\dfrac{R^2}{r\,r'}\right)^{\ell+1} \right] \dfrac{Y_{\ell m}(\theta,\phi)\,Y_{\ell m}^*(\theta\,',\phi\,')}{2\,\ell+1} & r < r' \\[3em]
\dfrac{1}{\epsilon_o} \displaystyle\sum_{\ell=0}^{\infty} \sum_{m=-\ell}^{\ell} \left[\dfrac{(r')^\ell}{r^{\ell+1}} - \dfrac{1}{R}\left(\dfrac{R^2}{r\,r'}\right)^{\ell+1} \right] \dfrac{Y_{\ell m}(\theta,\phi)\,Y_{\ell m}^*(\theta\,',\phi\,')}{2\,\ell+1} & r > r'
\end{cases}
\tag{3.200}
$$

$$
= \frac{1}{\epsilon_o} \sum_{\ell=0}^{\infty} \sum_{m=-\ell}^{\ell} \left[\frac{r_<^\ell}{r_>^{\ell+1}} - \frac{1}{R}\left(\frac{R^2}{r\,r'}\right)^{\ell+1} \right] \frac{Y_{\ell m}(\theta,\phi)\,Y_{\ell m}^*(\theta\,',\phi\,')}{2\,\ell+1}
\tag{3.201}
$$

This solution is of course consistent with Equation 3.174, where we used the Addition Theorem for Spherical Harmonics to rewrite the Green Function for this geometry and type of boundary conditions in terms of the spherical harmonics, except now that we used separation of variables from the start rather than relying on the method of images and the Addition Theorem.

Note that, as predicted, the solution consists of sums of power laws in r' as well as r and is of course symmetric under exchange of $\vec{r}$ and $\vec{r}\,'$.

Interesting exercises would be to see that the above expression approaches $|\vec{r} - \vec{r}\,'|^{-1}$ as $\vec{r} \to \vec{r}\,'$ (use the Addition Theorem to recover the method of images solution) and also to recover the defining differential equation, Equation 3.176.

Examples of Using the Expansion of the Green Function in Terms of the Spherical Harmonics

We did a lot of gymnastics to get the expansion of the Green Function in terms of spherical harmonics. Let's see how it can be used. For each of the examples we will consider, it would be possible to solve for the potential without explicitly using our expansion by splitting the volume into regions on two sides of $\vec{r}\,'$ and using separation of variables with application of boundary conditions (including matching conditions at the chosen internal boundary). The advantage of using the Green Function is that it obviates re-solving the same kind of problem many times by simply providing integrals that need to be done.

Another point: in separation of variables, we always end up using orthonormality of the specific set of solutions to Laplace's Equation for the geometry to obtain the solution coefficients from the inhomogeneous boundary condition(s), Dirichlet or Neumann, and the matching conditions, if any. That general approach will become codified here in the way the Green Function is integrated with the charge distribution and boundary conditions in Equations 3.47 and 3.49. In particular, the Green Function connects particular spherical harmonic modes of the charge distribution and/or the voltage (Dirichlet) and/or charge (Neumann) boundary conditions to the corresponding spherical harmonic modes of the potential. This correspondence makes the structure of the solution much easier to understand. The effect of a spherical harmonic mode in charge distribution and/or the boundary conditions at one radius r' on the potential at another radius r is just a function of the two radii, the $g(r, r')$ function (charge distribution in volume or Neumann boundary condition) or its radial derivative (Dirichlet boundary condition).

The application of the Green Function is like a propagator in QM, propagating from the initial condition to later times. We have to do less work to obtain the QM propagator because the solution to the time piece of Schrödinger's Equation is trivial once one has the eigenvalues of the space piece.

For our examples, we will consider charge distributions inside a conducting sphere. We quote the general result from Jackson for the Green Function expansion in spherical harmonics for a geometry consisting of the volume between two spheres at $r = a$ and $r = b$ with Dirichlet BC on the two surfaces:

$$G_D(\vec{r}, \vec{r}\,')\tag{3.202}$$
$$= \frac{1}{\epsilon_o} \sum_{\ell=0}^{\infty} \sum_{m=-\ell}^{\ell} \left[r_<^{\ell} - \frac{1}{a} \left(\frac{a^2}{r_<} \right)^{\ell+1} \right] \left[\frac{1}{r_>^{\ell+1}} - \frac{1}{b} \left(\frac{r_>}{b^2} \right)^{\ell} \right] \frac{Y_{\ell m}^*(\theta', \phi')\, Y_{\ell m}(\theta, \phi)}{\left[1 - \left(\frac{a}{b} \right)^{2\ell+1} \right] (2\ell+1)}$$

where, as usual, $r_< = \min\{r, r'\}$ and $r_> = \max\{r, r'\}$. Obtaining this more general result is a matter of doing the same thing as we did to obtain the result for a spherical conducting boundary at $r = R$ except that the A_ℓ^{out} term cannot be assumed to vanish.

Next, taking the limit $a \to 0$, we get the result we will need for our work below where we want to solve for the potential inside a sphere at $r = b$ with Dirichlet BC:

$$G_D(\vec{r}, \vec{r}\,') = \frac{1}{\epsilon_o} \sum_{\ell=0}^{\infty} \sum_{m=-\ell}^{\ell} r_<^{\ell} \left[\frac{1}{r_>^{\ell+1}} - \frac{1}{b} \left(\frac{r_>}{b^2} \right)^{\ell} \right] \frac{Y_{\ell m}^*(\theta', \phi')\, Y_{\ell m}(\theta, \phi)}{2\ell+1} \tag{3.203}$$

You will also be able to read off this simpler result from a method of images problem you will do in homework. On to our examples!

Example 3.10: Potential inside a conducting sphere of radius b due to an arbitrary Dirichlet boundary condition potential at b but no charge in the volume

With no charge in the volume, we just need to calculate the surface term in Equation 3.48, for which we need the normal gradient of G_D at the surface (remember, $\widehat{n}$ points out of $\mathcal{V}$):

$$
\widehat{n}(\vec{r}\,') \cdot \vec{\nabla}_{\vec{r}\,'}\, G_D(\vec{r}, \vec{r}\,')\Big|_{\vec{r}\,' \in \mathcal{S}}
$$

$$
= \frac{1}{\epsilon_o} \sum_{\ell=0}^{\infty} \sum_{m=-\ell}^{\ell} \frac{Y_{\ell m}^*(\theta', \phi')\, Y_{\ell m}(\theta, \phi)}{2\,\ell + 1}\, r^\ell \frac{d}{dr'} \left[\frac{1}{(r')^{\ell+1}} - \frac{1}{b} \left(\frac{r'}{b^2} \right)^{\ell} \right]\Bigg|_{r'=b}
$$

$$
= -\frac{1}{\epsilon_o}\, \frac{1}{b^2} \sum_{\ell=0}^{\infty} \sum_{m=-\ell}^{\ell} \left(\frac{r}{b} \right)^{\ell} Y_{\ell m}^*(\theta', \phi')\, Y_{\ell m}(\theta, \phi) \tag{3.204}
$$

Therefore, the potential in the volume for the Dirichlet B.C. $V(b, \theta, \phi)$ is

$$
V(\vec{r}) = \sum_{\ell=0}^{\infty} \sum_{m=-\ell}^{\ell} \left(\frac{r}{b} \right)^{\ell} Y_{\ell m}(\theta, \phi) \int d\Omega'\, Y_{\ell m}^*(\theta', \phi')\, V(b, \theta', \phi') \tag{3.205}
$$

We see that the spherical harmonic component ℓm of the potential at r is determined by the spherical harmonic component ℓm of the potential on the boundary: very simple and consistent with the QM propagator picture.

Example 3.11: Potential inside a grounded spherical conductor with a ring of charge of radius a in the xy plane

This time, we do the volume integral but there is no integral over the surface. The charge density due to the ring is

$$\rho(\vec{r}\,') = \frac{Q}{2\,\pi\,a^2}\,\delta(r'-a)\,\delta(\cos\theta\,') \tag{3.206}$$

Again, one can check that the charge density is correct by integrating it: the a^{-2} cancels the $(r')^2$ factor in the volume element and the argument of the θ' delta function is $\cos\theta'$ because the volume element contains $d(\cos\theta\,')$.

We use Equation 3.48 as usual, in this case with no surface term because the boundary has $V = 0$. The potential is then

$$\begin{aligned}
V(\vec{r}) &= \int_{\mathcal{V}} d\tau'\rho(\vec{r}\,')\,G_D(\vec{r},\vec{r}\,')\\[4pt]
&= \frac{Q}{2\,\pi\,\epsilon_o\,a^2}\sum_{\ell=0}^{\infty}\sum_{m=-\ell}^{\ell} Y_{\ell m}(\theta,\phi) \tag{3.207}\\[4pt]
&\quad \times \int_{\mathcal{V}} d\tau'\delta(r'-a)\,\delta(\cos\theta\,')\,r_<^\ell\left[\frac{1}{r_>^{\ell+1}}-\frac{1}{b}\left(\frac{r_>}{b^2}\right)^\ell\right]\frac{Y_{\ell m}^*(\theta\,',\phi\,')}{2\,\ell+1}
\end{aligned}$$

Because the charge density has no azimuthal dependence, the ϕ' integral picks out the $m = 0$ term. Recall that $Y_{\ell 0} = \sqrt{(2\ell + 1)/4\pi}\, P_\ell$, so we may rewrite as

$$V(\vec{r}) = \frac{Q}{4\pi\,\epsilon_o\, a^2} \sum_{\ell=0}^{\infty} P_\ell(\cos\theta) \int_{-1}^{1} d(\cos\theta')\, \delta(\cos\theta')\, P_\ell(\cos\theta') \qquad (3.208)$$

$$\times \int_{0}^{b} (r')^2 dr'\, \delta(r' - a)\, r_<^\ell \left[\frac{1}{r_>^{\ell+1}} - \frac{1}{b}\left(\frac{r_>}{b^2}\right)^\ell \right]$$

$$= \frac{Q}{4\pi\,\epsilon_o} \sum_{\ell=0}^{\infty} P_\ell(\cos\theta)\, P_\ell(0)\, r_<^\ell \left[\frac{1}{r_>^{\ell+1}} - \frac{1}{b}\left(\frac{r_>}{b^2}\right)^\ell \right] \qquad (3.209)$$

where now $r_< = \min\{r, a\}$ and $r_> = \max\{r, a\}$ because the δ function does the r' integral for us, effectively replacing r' with a. (The next example will show the case of a more complex charge distribution for which the radial integral is not done so easily.) Now, recall $P_\ell(0) = 0$ for odd ℓ and $P_\ell(0) = [(-1)^n (2n - 1)!!]/2^n\, n!$ for even $\ell = 2n$, so we may reduce the above further to (replacing ℓ with $2n$ so n runs over all nonnegative integers rather than ℓ running over all nonnegative even integers):

$$V(\vec{r}) = \frac{Q}{4\pi\,\epsilon_o} \sum_{n=0}^{\infty} \frac{(-1)^n (2n - 1)!!}{2^n\, n!}\, r_<^{2n} \left[\frac{1}{r_>^{2n+1}} - \frac{1}{b}\left(\frac{r_>}{b^2}\right)^{2n} \right] P_{2n}(\cos\theta) \qquad (3.210)$$

where $r_< = \min(r, a)$ and $r_> = \max(r, a)$ again: *i.e.*, not surprisingly, the solution has a different form depending on whether one wants to know the potential inside the ring ($r < a$) or outside the ring ($r > a$). This is now the complete solution.

To get some intuition for the solution, let's calculate the induced surface charge density at $r = b$. We obtain it from the normal gradient of V, which, recall, is just $\vec{E}$, and the change in its normal component at a boundary gives the surface charge density. Since the normal gradient is just d/dr for this particular geometry, it does not act at all on P_{2n}. In calculating this gradient, $r_< = a$ and $r_> = r$ since we will in the end evaluate at $r = b > a$. Therefore:

$$\sigma(\vec{r}) = \epsilon_o \left. \frac{dV}{dr} \right|_{r_> = r = b, r_< = a}$$

$$= -\frac{Q}{4\pi b^2} \sum_{n=0}^{\infty} \frac{(4n+1)(-1)^n (2n-1)!!}{2^n \, n!} \left(\frac{a}{b}\right)^{2n} P_{2n}(\cos\theta)$$

$$= -\frac{Q}{4\pi b^2} \left[1 + \sum_{n=1}^{\infty} \frac{(4n+1)(-1)^n (2n-1)!!}{2^n \, n!} \left(\frac{a}{b}\right)^{2n} P_{2n}(\cos\theta) \right] \qquad (3.211)$$

The expression is written in the above suggestive form on the last line so that it is easy to obtain the total induced surface. Since $P_0(\cos\theta) = 1$, the integral of the $n > 0$ terms over $\cos\theta$ can be viewed as integrating P_{2n} with P_0; by orthonormality of the Legendre polynomials, these terms all yield zero. The first term yields $-Q$ when integrated over the sphere. This is what we would expect from Gauss's Law applied just inside the $r = b$ boundary. All the other terms yield θ dependence in the charge density but their averages vanish (easily seen by using orthonormality with P_0, which is a constant).

We have seen in this example how the integration of the charge density with the Green Function breaks the charge density down into its spherical harmonic components, calculates the potential due to each component individually (and fairly trivially, just multiplying by a function of the radius of the source charge and the radius at which the potential is desired) and then sums up those components. The same kind of correspondence clearly holds for the induced surface charge density.

Note that the additional $4n + 1$ factor implies the θ dependence of the induced surface charge density is different from that of the original ring charge; *i.e.*, the induced surface charge is not just a ring.

To get some intuition about the surface charge distribution, let's go back to the potential and rewrite it into a method of images solution. Using Equation 3.149 (the expansion of $|\vec{r} - \vec{r}\,'|^{-1}$ in Legendre polynomials), we can imagine that the first term arises from the convolution of $|\vec{r} - \vec{r}\,'|^{-1}$ with the ring charge distribution (though we won't prove it explicitly). What about the second term? Let's manipulate it a bit:

$$-r_<^{2n} \frac{1}{b} \left(\frac{r_>}{b^2} \right)^{2n} \Bigg|_{r_<=a, r_>=r} = -a^{2n} \frac{r^{2n}}{b^{4n+1}} = -\frac{b}{a} \frac{r^{2n} a^{2n+1}}{b^{4n+2}} = -\frac{b}{a} \frac{r^{2n}}{(b^2/a)^{2n+1}} \qquad (3.212)$$

$$= -\frac{b}{a} \frac{r_<^{2n}}{r_>^{2n+1}} \Bigg|_{r_<=r, r_>=b^2/a} \qquad (3.213)$$

note: meaning of $r_<$ and $r_>$ changed!

Thus, we see the second term has the right form for the potential at $r_< = r$ due to an image charge at radius $r_> = b^2/a$ and normalization $-b/a$ relative to the true charge. (Note that the meaning of $r_<$ and $r_>$ change between the initial and final expression above.) The ring shape comes from the weighted sum over Legendre polynomials, which is the same as the corresponding sum for the potential of the true charge, the first term.

Seeing that the image charge is a ring at radius b^2/a explains the induced surface charge density distribution via its proportionality to the field lines from the true charge to the image charge at the $r = b$ surface. Drawing a picture using the image charge configuration should make this clear.

Example 3.12: Potential inside a grounded spherical conductor with a line charge density along the z axis

This is done in Jackson Section 3.10. We reproduce it here because it has some calculational twists.

The first twist is figuring out how to write down the charge density in spherical coordinates. One could probably rigorously derive the form by writing down the charge density trivially in Cartesian or cylindrical coordinates and then applying Jacobian transformation to convert it to spherical coordinates, but there is an easier, more intuitive way.

It is all present at $\cos\theta = 1$ and $\cos\theta = -1$, so clearly delta functions for these positions need to be included. It has azimuthal symmetry, so there will be no ϕ dependence, only a factor of $1/2\pi$. The charge is distributed in radius, so there is some to-be-determined radial dependence $f(r)$. To figure out $f(r)$, let's write down the requirement that the integral be the total charge Q:

$$\rho(\vec{r}) = \frac{Q}{2\pi} f(r) \left[\delta(\cos\theta - 1) + \delta(\cos\theta + 1)\right] \tag{3.214}$$

$$Q = \int_{\mathcal{V}} d\tau\, \rho(\vec{r})$$

$$= \frac{Q}{2\pi} \int_0^b dr\, r^2 f(r) \int_{-1}^1 d(\cos\theta) \left[\delta(\cos\theta - 1) + \delta(\cos\theta + 1)\right] \int_0^{2\pi} d\phi$$

$$= 2Q \int_0^b dr\, r^2 f(r)$$

If we choose $f(r) = c/r^2$ where c is a constant to be determined, then the remaining integral becomes trivial and yields b, which we can use to find c:

$$Q = 2\,Q\,c\,b \qquad \Longrightarrow \qquad c = \frac{1}{2b} \tag{3.215}$$

$$\Longrightarrow \qquad \rho(\vec{r}) = \frac{Q}{4\,\pi\,b\,r^2}\,[\delta(\cos\theta - 1) + \delta(\cos\theta + 1)] \tag{3.216}$$

Now, since the sphere is grounded, we just need to do the integral of the charge density with the Dirichlet Green Function:

$$V(\vec{r}) = \frac{1}{\epsilon_o} \int_{\mathcal{V}} d\tau'\,\rho(\vec{r}\,')\,G_D(\vec{r}, \vec{r}\,') \tag{3.217}$$

$$= \frac{Q}{4\,\pi\,\epsilon_o\,b} \sum_{\ell=0}^{\infty} \sum_{m=-\ell}^{\ell} \int_{\mathcal{V}} d\tau'\,\frac{\delta(\cos\theta'-1) + \delta(\cos\theta'+1)}{(r')^2}$$

$$\times\,r_<^\ell \left[\frac{1}{r_>^{\ell+1}} - \frac{1}{b}\left(\frac{r_>}{b^2}\right)^\ell\right] \frac{Y_{\ell m}^*(\theta', \phi')\,Y_{\ell m}(\theta, \phi)}{2\ell + 1}$$

We apply azimuthal symmetry as we did in the previous example, selecting the $m = 0$ terms that we can write as Legendre polynomials. The normalization of the spherical harmonics cancels the factor of $2\ell + 1$ in the denominator but adds a factor of 4π in the denominator. The ϕ integral cancels a factor of 2π in the denominator. The θ' integrals can be done trivially, selecting $P_\ell(1)$ and $P_\ell(-1)$. Note also that the $(r')^2$ from the $d\tau'$ cancels the $(r')^2$ in the denominator from the charge density. Thus, we have

$$V(\vec{r}) = \frac{Q}{8\pi\,\epsilon_0\,b} \sum_{\ell=0}^{\infty} P_\ell(\cos\theta) \int_0^b dr'\, r_<^\ell \left[\frac{1}{r_>^{\ell+1}} - \frac{1}{b}\left(\frac{r_>}{b^2}\right)^\ell \right] [P_\ell(1) + P_\ell(-1)]$$

We know $P_\ell(1) = 1$ and $P_\ell(-1) = (-1)^\ell$, so the term containing these two factors yields 2 for even ℓ and 0 for odd ℓ. Thus, the above reduces to

$$V(\vec{r}) = \frac{Q}{4\pi\,\epsilon_0\,b} \sum_{\ell\ \text{even}} P_\ell(\cos\theta) \int_0^b dr'\, r_<^\ell \left[\frac{1}{r_>^{\ell+1}} - \frac{1}{b}\left(\frac{r_>}{b^2}\right)^\ell \right]$$

The integral over radius must be broken into two pieces, one for $r' < r$ and one for $r' > r$, because the $r_<$ and $r_>$ variables take on different values for these two regions (by definition!). Doing so, and doing the integrals (they are straightforward) yields

$$\int_0^b dr' \, r_<^\ell \left[\frac{1}{r_>^{\ell+1}} - \frac{1}{b} \left(\frac{r_>}{b^2} \right)^\ell \right] = \frac{2\ell+1}{\ell+1} \frac{1}{\ell} \left[1 - \left(\frac{r}{b} \right)^\ell \right] \qquad (3.218)$$

The second portion of the above quantity is well-defined for $\ell \neq 0$, but not for $\ell = 0$. We need to use L'Hôpital's rule to evaluate it for $\ell = 0$:

$$\lim_{\ell \to 0} \frac{1}{\ell} \left[1 - \left(\frac{r}{b} \right)^\ell \right] = \lim_{\ell \to 0} \frac{\frac{d}{d\ell} \left[1 - \left(\frac{r}{b} \right)^\ell \right]}{\frac{d}{d\ell} \ell} = - \lim_{\ell \to 0} \frac{\left(\frac{r}{b} \right)^\ell \left(\ln \frac{r}{b} \right) \frac{d}{d\ell} \ell}{\frac{d}{d\ell} \ell} = \ln \frac{b}{r} \quad (3.219)$$

Therefore, we may write the full solution as, separating out the $\ell = 0$ term and rewriting in terms of $\ell = 2\,n$,

$$V(\vec{r}) = \frac{Q}{4\pi\epsilon_o b} \left[\ln \frac{b}{r} + \sum_{n=1}^{\infty} \frac{4n+1}{2n(2n+1)} \left[1 - \left(\frac{r}{b} \right)^{2n} \right] P_{2n}(\cos\theta) \right] \qquad (3.220)$$

Let's calculate the induced surface charge density and the total induced charge again:

$$\sigma(\theta) = \left. \epsilon_o \, \frac{\partial V}{\partial r} \right|_{r=b} = -\frac{Q}{4 \, \pi \, b^2} \left[1 + \sum_{n=1}^{\infty} \frac{4n+1}{2n+1} P_{2n}(\cos \theta) \right] \qquad (3.221)$$

Note again how the surface charge density has a different n-dependent weighting than the potential. Finally, integrating over the sphere to get the total induced charge, all $n \geq 1$ terms vanish, yielding

$$Q_{ind} = \int_{r=b} b^2 \, d\phi \, d\cos\theta \, \sigma(\theta) = -Q \qquad (3.222)$$

as we expect from Gauss's Law. It would again be interesting to rewrite the solution in the form of a method of images solution, which you have the tools to do. It clearly should look like a line charge at the north and south poles. Its density will presumably fall off as $1/z^2$ because the true charge density is uniform (in linear units, z) and the image charge magnitude and position both scale as $1/z$.

Lecture 11:

Advanced Electrostatics VII:
Multipoles

Date Revised: 2024/02/13 05:30
Revised lecture break
Date Given: 2024/02/08

Multipole Expansions

Dipoles: Quick Review

Recall from Ph1b the idea of an *electric dipole*: two charges of equal and opposite magnitude $\pm q$ spaced very close together at $\vec{r}_+$ and $\vec{r}_-$. The net charge cancels almost perfectly, so, rather than the potential falling off like $1/r$ at large radius, it falls off as $1/r^2$ with functional form

$$V(\vec{r}) = \frac{1}{4\pi\epsilon_o} \frac{\vec{p}\cdot\hat{r}}{r^2} \qquad \text{as} \qquad \frac{r}{|\vec{r}_+|}, \frac{r}{|\vec{r}_-|}, \frac{r}{|\vec{r}_+ - \vec{r}_-|} \to \infty \qquad (3.223)$$

where $\vec{p} = q(\vec{r}_+ - \vec{r}_-)$ is the dipole moment.

This idea generalizes. When one has a charge distribution with vanishing net charge, but inside of which there is a variation in the charge density, that variation is still noticeable at large distance as a set of potentials that fall off more quickly than $1/r$. The first additional term is the *dipole*, falling as $1/r^2$, the second is the *quadrupole*, falling as $1/r^3$, the third is the *octupole*, falling as $1/r^4$, and so on. The nomenclature comes from the minimum number of different source charges one must have to obtain that moment: one for monopole, two for dipole, four for quadrupole, etc.

Multipoles: Full Derivation

We derive the full form by considering the potential due to *a charge distribution near the origin as viewed at a point $\vec{r}$ such that r is much larger than the extent of the charge distribution*. This the key assumption! We begin with

$$V(\vec{r}) = \frac{1}{4\pi\epsilon_o} \int_{\mathcal{V}} d\tau' \frac{\rho(\vec{r}')}{|\vec{r} - \vec{r}'|} \qquad (3.224)$$

We now use Equation 3.149, taking $r_< = r'$ and $r_> = r$ because r is outside the charge distribution. Thus,

$$V(\vec{r}) = \frac{1}{4\pi\epsilon_o} \int_{\mathcal{V}} d\tau' \, \rho(\vec{r}') \sum_{\ell=0}^{\infty} \frac{(r')^\ell}{r^{\ell+1}} \, P_\ell(\cos\gamma) \qquad (3.225)$$

where $\cos\gamma = \hat{r} \cdot \hat{r}'$ is the angle between the two vectors. There is a common $1/r$ we can factor out, leaving

$$\boxed{V(\vec{r}) = \frac{1}{4\pi\epsilon_o} \frac{1}{r} \sum_{\ell=0}^{\infty} \frac{1}{r^\ell} \int_{\mathcal{V}} d\tau' \rho(\vec{r}') \, (r')^\ell \, P_\ell(\cos\gamma)} \qquad (3.226)$$

This is the *multipole expansion* of the potential of the charge distribution. One can see that the successive terms fall off as successively higher powers of $1/r$. The angular dependence is given by the Legendre polynomials in $\cos\gamma = \hat{r} \cdot \hat{r}'$.

The Monopole, Dipole, and Quadrupole Terms

Let's write out the first three terms more explicitly to get some physical intuition:

▶ *Monopole term*
   The first term is

$$V_1(\vec{r}) = \frac{1}{4\pi\epsilon_o} \frac{1}{r} \int_{\mathcal{V}} d\tau' \rho(\vec{r}') = \frac{1}{4\pi\epsilon_o} \frac{Q}{r} \qquad (3.227)$$

   This is the standard Coulomb's Law term due to the total charge. Far enough away, all charge distributions look pointlike. But, if $Q = 0$, this term vanishes identically and the higher-order terms must be considered. Even if $Q \neq 0$, if one is close enough to the charge distribution to see its non-pointlike nature, the higher-order terms will be important corrections to the monopole term.

▶ *Dipole term*
The second term is

$$V_2(\vec{r}) = \frac{1}{4\,\pi\,\epsilon_o}\,\frac{1}{r^2}\int_{\mathcal{V}} d\tau'\rho(\vec{r}\,')\,r\,'\cos\gamma = \frac{1}{4\,\pi\,\epsilon_o}\,\frac{1}{r^2}\int_{\mathcal{V}} d\tau'\rho(\vec{r}\,')\,r\,'\,\widehat{r}\,'\cdot\widehat{r}$$

$$= \frac{1}{4\,\pi\,\epsilon_o}\,\frac{1}{r^2}\,\widehat{r}\cdot\int_{\mathcal{V}} d\tau'\rho(\vec{r}\,')\,\vec{r}\,' \qquad (3.228)$$

or
$$\boxed{V_2(\vec{r}) = \frac{1}{4\,\pi\,\epsilon_o}\,\frac{1}{r^2}\,\widehat{r}\cdot\vec{p} \qquad \text{where} \qquad \vec{p} = \int_{\mathcal{V}} d\tau'\rho(\vec{r}\,')\,\vec{r}\,'} \qquad (3.229)$$

is the dipole moment vector. It is the generalization of $\vec{p} = q\,(\vec{r}_+ - \vec{r}_-)$. It can be written in component form (which is how you would actually calculate it — recall our discussion during the first lecture of how to break vector integrals into a set of scalar integrals!) as

$$\boxed{p_j = \int_{\mathcal{V}} d\tau'\,\rho(\vec{r}\,')\,r_j' = \widehat{r}_j\cdot\vec{p}} \qquad (3.230)$$

▶ *Quadrupole term*
The third term is

$$V_3(\vec{r}) = \frac{1}{4\pi\epsilon_o} \frac{1}{r^3} \int_{\mathcal{V}} d\tau' \rho(\vec{r}') (r')^2 \frac{1}{2} \left(3\cos^2\gamma - 1\right)$$

$$= \frac{1}{4\pi\epsilon_o} \frac{1}{r^3} \int_{\mathcal{V}} d\tau' \rho(\vec{r}') (r')^2 \frac{1}{2} \left(3 \left(\hat{r}\cdot\hat{r}'\right)\left(\hat{r}'\cdot\hat{r}\right) - 1\right)$$

$$= \frac{1}{4\pi\epsilon_o} \frac{1}{r^3} \hat{r} \cdot \left[\int_{\mathcal{V}} d\tau' \rho(\vec{r}') (r')^2 \frac{1}{2} \left(3\hat{r}'\hat{r}' - \underline{\underline{1}}\right)\right] \cdot \hat{r} \qquad (3.231)$$

or $$\boxed{V_3(\vec{r}) = \frac{1}{4\pi\epsilon_o} \frac{1}{r^3} \frac{1}{2} \hat{r} \cdot \underline{\underline{Q}} \cdot \hat{r} \quad \text{where} \quad \underline{\underline{Q}} = \int_{\mathcal{V}} d\tau' \rho(\vec{r}') \left[3\vec{r}'\vec{r}' - (r')^2\underline{\underline{1}}\right]}$$
(3.232)

is the quadrupole moment and $\underline{\underline{1}} = \mathrm{diag}(1,1,1)$ is the identity tensor with ones along the diagonal. Because it is composed of $\vec{r}'\vec{r}'$ and $\underline{\underline{1}}$, $\underline{\underline{Q}}$ is a tensor, implying that one can take a dot product with a vector on each side. Written out in component form (which is again how you would calculate it):

$$\boxed{Q_{jk} = \int_{\mathcal{V}} d\tau' \rho(\vec{r}') \left[3 r_j' r_k' - (r')^2\delta_{jk}\right] = \hat{r}_j \cdot \underline{\underline{Q}} \cdot \hat{r}_k}$$
(3.233)

It is now obvious that Q_{jk} is symmetric in its indices.

Origin Dependence of the Dipole Moment

Suppose we take a charge distribution and shift the origin by a vector $\vec{a}$ such that the charge distribution is now centered around $\vec{a}$. Then the new dipole moment is

$$\vec{p}\,' = \int d\tau'\, \rho'(\vec{r}\,')\, \vec{r}\,' = \int d\tau\, \rho(\vec{r})\, (\vec{a} + \vec{r}) = \vec{a}\, Q + \vec{p} \qquad (3.234)$$

where we define the charge distribution in the new coordinate system $\rho'(\vec{r}\,')$ in terms of the original charge distribution $\rho(\vec{r})$ to be such that $\rho'(\vec{r}\,') = \rho(\vec{r} = \vec{r}\,' - \vec{a})$ when $\vec{r}\,' = \vec{r} + \vec{a}$. Thus, an origin shift can induce an artificial dipole moment for a charge distribution that has a monopole moment. This part of the dipole moment is not real: it is a reflection of the fact that the multipole potentials are written in terms of distance from the origin under the assumption that the charge distribution is centered around the origin. When it is not, this is an unnatural coordinate system to use, requiring corrections of order $1/r^2$ to the standard monopole term ($\propto Q/r$) to handle the fact that the charge distribution is displaced. The above tells us the correction term has the same form as a dipole term. Obviously, one should choose the origin wisely to avoid such complications.

Note also the somewhat counterintuitive implication that, if $Q = 0$, then the dipole moment is independent of origin! Our assumption that the charge distribution, including its displacement from the origin, is small compared to the to the observation point implies that a must also be small so that any corrections, which are of order a^3/r^3, are small compared to the leading $1/r^2$ dependence for the dipole term.

Field of an Electric Dipole

This is simply a matter of taking the gradient. If we let $\vec{p} = p\,\hat{z}$, then this is easy:

$$V_2(\vec{r}) = \frac{p\,\cos\theta}{4\,\pi\,\epsilon_o\,r^2} \tag{3.235}$$

$$\implies \quad E_r(\vec{r}) = -\frac{\partial V_2}{\partial r} = -\frac{2\,p\,\cos\theta}{4\,\pi\,\epsilon_o\,r^3} \tag{3.236}$$

$$E_\theta(\vec{r}) = -\frac{1}{r}\,\frac{\partial V_2}{\partial \theta} = \frac{p\,\sin\theta}{4\,\pi\,\epsilon_o\,r^3} \tag{3.237}$$

$$E_\phi(\vec{r}) = -\frac{1}{r\,\sin\theta}\,\frac{\partial V_2}{\partial \phi} = 0 \tag{3.238}$$

$$\text{or} \quad \vec{E}(\vec{r}) = \frac{p}{4\,\pi\,\epsilon_o\,r^3}\left(2\,\hat{r}\,\cos\theta + \hat{\theta}\,\sin\theta\right) \tag{3.239}$$

To generalize this result for an arbitrary orientation of $\vec{p}$ requires some vector algebra. We have Equation 3.229 for the dipole potential in generic form, which we write out as

$$V_2(\vec{r}) = \frac{1}{4\pi\epsilon_o} \frac{1}{r^3} \vec{r} \cdot \vec{p} = \frac{1}{4\pi\epsilon_o} \frac{1}{r^3} \sum_i r_i \, p_i \qquad (3.240)$$

Now, we take the gradient, first noting

$$\frac{\partial}{\partial r_j} \frac{r_i}{r^3} = \frac{\partial}{\partial r_j} \frac{r_i}{(r^2)^{3/2}} = -\frac{3}{2} \frac{r_i}{(r^2)^{5/2}} \frac{\partial r^2}{\partial r_j} + \frac{\delta_{ij}}{r^3} = -\frac{3}{2} \frac{r_i}{r^5} (2\,r_j) + \frac{\delta_{ij}}{r^3} \qquad (3.241)$$

Where we used $r^3 = \left(r^2\right)^{3/2}$ and $r^2 = \sum_k r_k^2$ to more easily calculate the partial derivative. Therefore, with r_j and $\hat{r}_j$ being the jth Cartesian coordinate and unit vector,

$$\vec{E}_2(\vec{r}) = -\vec{\nabla} V_2(\vec{r}) = -\sum_j \hat{r}_j \frac{\partial V_2}{\partial r_j} = \frac{1}{4\pi\epsilon_o \, r^5} \sum_{ij} \hat{r}_j \, p_i \left[3\, r_i\, r_j - \delta_{ij}\, r^2\right]$$

$$= \frac{1}{4\pi\epsilon_o \, r^5} \sum_j \left[r_j \hat{r}_j \left(3 \sum_i p_i\, r_i\right) - p_j \hat{r}_j\, r^2 \right] = \frac{1}{4\pi\epsilon_o \, r^5} \left[3\,(\vec{p}\cdot\vec{r})\,\vec{r} - r^2\vec{p}\right]$$

$$\implies \boxed{\vec{E}_2(\vec{r}) = \frac{1}{4\pi\epsilon_o \, r^3} \left[3\,(\vec{p}\cdot\hat{r})\,\hat{r} - \vec{p}\right]} \qquad (3.242)$$

Electrostatic Potential Energy of a Multipole Distribution in an External Potential

The general expression for the potential energy of a charge distribution in an external potential is

$$U = \int_{\mathcal{V}} \rho(\vec{r}\,') \, V(\vec{r}\,') \qquad (3.243)$$

Under the assumption that $V(\vec{r}\,')$ varies slowly (but need not be constant!) over the spatial extent of the charge distribution, we can rewrite this in terms of moments of the charge distribution and derivatives of the potential. To do so, we need to expand $V(\vec{r})$ about some point in the distribution. To make the calculation easier, assume the charge distribution is centered on the origin, around which we will expand. (We will generalize later.) We use the multidimensional Taylor expansion of $V(\vec{r})$:

$$V(\vec{r}\,') = V(\vec{r}\,' = \vec{0}) + \sum_{j=1}^{3} r_j' \left.\frac{\partial V}{\partial r_j}\right|_{\vec{r}\,'=\vec{0}} + \frac{1}{2} \sum_{j,k=1}^{3} r_j' r_k' \left.\frac{\partial^2 V}{\partial r_j \, \partial r_k}\right|_{\vec{r}\,'=\vec{0}} + \cdots \qquad (3.244)$$

We can already foresee how integrating the above form for $V(\vec{r}\,')$ with $\rho(\vec{r}\,')$ is going to result in a dipole moment in the first term and quadrupole moment in the second.

Using $E_j = -\frac{\partial V}{\partial r_j}$, we may simplify

$$V(\vec{r}\,') = V(\vec{0}) - \vec{r}\,' \cdot \vec{E}(\vec{0}) + \frac{1}{6} \sum_{j,k=1}^{3} 3\, r_j'\, r_k' \left. \frac{\partial^2 V}{\partial r_j\, \partial r_k} \right|_{\vec{0}} + \cdots \qquad (3.245)$$

$$= V(\vec{0}) - \vec{r}\,' \cdot \vec{E}(\vec{0}) + \frac{1}{6} \sum_{j,k=1}^{3} \left(3\, r_j'\, r_k' - (r\,')^2 \delta_{jk} \right) \left. \frac{\partial^2 V}{\partial r_j\, \partial r_k} \right|_{\vec{0}} + \cdots \qquad (3.246)$$

where we were able to add the $(r\,')^2 \delta_{jk}$ term because

$$\sum_{j,k} (r\,')^2 \delta_{jk} \left. \frac{\partial^2 V}{\partial r_j\, \partial r_k} \right|_{\vec{0}} = (r\,')^2 \nabla^2 V(\vec{r}\,' = 0) = 0 \qquad (3.247)$$

because the charge distribution *sourcing V* is not present near the origin. Remember, $\rho(\vec{r})$ is not the distribution sourcing V; V is provided to us and is due to some charge distribution far away from the origin.

With the above expansion, the electrostatic potential energy is now (note that $\vec{E}(\vec{0})$ and $\partial^2 V/\partial r_j \partial r_k \big|_{\vec{0}}$ are constant with respect to r', so they come outside of the r' integral)

$$U = V(\vec{0}) \int_{\mathcal{V}} d\tau' \rho(\vec{r}') - \vec{E}(\vec{0}) \cdot \int_{\mathcal{V}} d\tau' \, \rho(\vec{r}') \, \vec{r}' \tag{3.248}$$

$$+ \frac{1}{6} \sum_{j,k=1}^{3} \frac{\partial^2 V}{\partial r_j \, \partial r_k} \bigg|_{\vec{0}} \int_{\mathcal{V}} d\tau' \rho(\vec{r}') \left[3 \, r_j' r_k' - \delta_{jk}(r')^2 \right] + \cdots$$

$$= Q \, V(\vec{0}) - \vec{p} \cdot \vec{E}(\vec{0}) + \frac{1}{6} \sum_{j,k=1}^{3} Q_{jk} \, \frac{\partial^2 V}{\partial r_j \, \partial r_k} \bigg|_{\vec{0}} + \cdots \tag{3.249}$$

We generalize for a charge distribution centered around $\vec{r}$ instead of the origin:

$$U(\vec{r}) = Q \, V(\vec{r}) - \vec{p} \cdot \vec{E}(\vec{r}) + \frac{1}{6} \sum_{j,k=1}^{3} Q_{jk} \, \frac{\partial^2 V}{\partial r_j \, \partial r_k} \bigg|_{\vec{r}} + \cdots \tag{3.250}$$

$$= Q \, V(\vec{r}) - \vec{p} \cdot \vec{E}(\vec{r}) + \frac{1}{6} \vec{\nabla}_{\vec{r}} \cdot \underline{\underline{Q}} \cdot \vec{\nabla}_{\vec{r}} V(\vec{r}) + \cdots \tag{3.251}$$

where we have written the last term in tensor dot product form. There are now contributions to the potential energy from the relative alignment of $\vec{p}$ and $\vec{E}$ and from the orientation of $\underline{\underline{Q}}$'s principal axes relative to the principal axes of the potential's curvature matrix. Note that $\vec{\nabla}_{\vec{r}}$ acts on the spatial dependence of $V(\vec{r})$; $\vec{r}'$ has already been integrated over to obtain $\underline{\underline{Q}}$.

Force on a Multipole Distribution in an External Field

We can calculate the force on the charge distribution by taking the derivative of U with respect to the *charge distribution's* nominal position $\vec{r}$, now replacing one derivative of V with the electric field $\vec{E}$ in the quadrupole term:

$$\vec{F}(\vec{r}) = -\vec{\nabla} U(\vec{r}) = Q \left(-\vec{\nabla} V(\vec{r}) \right) + \vec{\nabla} \left(\vec{p} \cdot \vec{E}(\vec{r}) \right) + \frac{1}{6} \sum_{j,k,m=1}^{3} \widehat{r}_m \, Q_{jk} \frac{\partial^2 E_j}{\partial r_m \, \partial r_k} + \cdots$$

$$= Q \, \vec{E}(\vec{r}) + \left(\vec{p} \cdot \vec{\nabla} \right) \vec{E}(\vec{r}) + \frac{1}{6} \sum_{j,k,m=1}^{3} \widehat{r}_m \, Q_{jk} \frac{\partial^2 E_j}{\partial r_m \, \partial r_k} + \cdots$$

$$= Q \, \vec{E}(\vec{r}) + \left(\vec{p} \cdot \vec{\nabla} \right) \vec{E}(\vec{r}) + \frac{1}{6} \vec{\nabla} \left[\vec{\nabla} \cdot \left(\underline{\underline{Q}} \cdot \vec{E}(\vec{r}) \right) \right] + \cdots \tag{3.252}$$

In going from the first to the second row, we used the vector identity $\vec{\nabla} \left(\vec{a} \cdot \vec{f}(\vec{r}) \right) = \left(\vec{a} \cdot \vec{\nabla} \right) \vec{f}(\vec{r})$ when $\vec{a}$ is a constant vector and $\vec{f}(\vec{r})$ has no curl. Note that all $\vec{\nabla}$ are with respect to $\vec{r}$ (since $\vec{r}\,'$ has been integrated over already).

We see that the total force is a sum of contributions from the interaction of the monopole with the electric field, the dipole with gradients in the electric field and, the quadrupole with the local curvature (second derivatives) of the electric field.

Lecture 12:

Advanced Electrostatics VIII:
Multipoles (cont.)

Electrostatics in Matter I:
Polarizability and Polarization
Bound Charges and their Potential
Electric Displacement Field
Linear Dielectrics

Date Revised: 2024/02/15 06:15
Revised lecture break
Date Given: 2024/02/13

Torque on a Multipole in an External Field

Let's also calculate the torque. As you know from Ph106a, to calculate a torque, we need to take the gradient of the potential energy in spherical coordinates with respect to the orientation of the charge distribution relative to the electric field.

The monopole term yields no torque because there is no orientation angle involved: Q and $V(\vec{r})$ are scalars.

Considering the dipole term, we understand that there are only two vectors involved, $\vec{p}$ and $\vec{E}$, and the potential energy only depends on the angle between them. So the torque will be given by the derivative with respect to this angle, which we call θ_p to differentiate it from the θ coordinate of the system in which we consider $\vec{E}$. This angle will be measured from $\vec{E}$ to $\vec{p}$. Then,

$$\vec{N}_{elec} = -\frac{\partial}{\partial \theta_p} \left(-\vec{p} \cdot \vec{E}(\vec{r}) \right) \tag{3.253}$$

$$= \frac{\partial}{\partial \theta_p} \, p \left| \vec{E}(\vec{r}) \right| \cos \theta_p = -p \left| \vec{E}(\vec{r}) \right| \sin \theta_p$$

$$= \vec{p} \times \vec{E}(\vec{r}) \tag{3.254}$$

This is a result you are familiar with from Ph1b, indicating the torque acts in a direction to align the dipole moment with the field direction.

Moving on to the quadrupole term, we recognize from Ph106a that any symmetric tensor can be diagonalized via a rotation. Let's write

$$\underline{\underline{Q}} = \mathcal{R}(\phi_Q, \theta_Q, \psi_Q) \, \underline{\underline{\mathcal{Q}}} \, [\mathcal{R}(\phi_Q, \theta_Q, \psi_Q)]^T \qquad \text{with} \quad \underline{\underline{\mathcal{Q}}} = \text{diag}(Q_1, Q_2, Q_3) \quad (3.255)$$

where the Q_i are quadrupole moments along the principal axes of the quadrupole tensor and $\mathcal{R}(\phi_Q, \theta_Q, \psi_Q)$ is the rotation matrix that rotates from the frame in which the coordinate axes align with the quadrupole tensor's principal axes to the arbitrary frame we started in, with the three Euler angles $(\phi_Q, \theta_Q, \psi_Q)$ defining the orientation of the principal axes of $\underline{\underline{\mathcal{Q}}}$ relative to the this arbitrary frame. This kind of diagonalization should be familiar to you from Ph106a, with $\mathcal{R}$ rotating from the "body" frame (the one fixed to the charge distribution's quadrupole principal axes) to the "space" frame. The quadrupole potential energy term is then

$$U_3 = -\frac{1}{6} \vec{\nabla}_{\vec{r}} \cdot \left\{ \mathcal{R}(\phi_Q, \theta_Q, \psi_Q) \, \underline{\underline{\mathcal{Q}}} \, [\mathcal{R}(\phi_Q, \theta_Q, \psi_Q)]^T \right\} \cdot \vec{E}(\vec{r}) \qquad (3.256)$$

To calculate the torque, we need to take the gradient of U_3 with respect to the orientation of the quadrupole. This amounts to taking gradients of $\mathcal{R}$ and $\mathcal{R}^T$ with respect to this orientation. As you know from the case of the symmetric top, the Euler angles are particularly useful angles with respect to which these derivatives can be taken. $\partial/\partial\phi_Q$ gives the torque about the z-axis of the space frame, which causes precession around that axis. $\partial/\partial\theta_Q$ gives the torque that causes motion in the polar angle direction with respect to the same space-frame z-axis, which is like nutation in the case of a top. And $\partial/\partial\psi_Q$ calculates the torque about one particular principal axis of the quadrupole, chosen at will, which accelerates or decelerates the rotation about that axis. You are familiar with symmetric tops, with $I_1 = I_2$. Here, we can have symmetric quadrupoles, with $Q_1 = Q_2$. In this case, the ψ_Q angle is the angle about the 3 axis of the quadrupole (the principal axis that aligns with the z-axis in the body frame). We do not take this further because, as you know from the study of tops in Ph106a, the phenomenology can be quite rich.

Section 4
Electrostatics in Matter

Polarizability and Polarization

Review of Polarizability of Materials

Griffiths §4.1 does a good job of providing physical motivation for the study of the polarizability of materials, and also reviews material you saw in Ph1b, so we only summarize the basics here.

▶ Atoms and molecules are *polarizable*, meaning that they can acquire a dipole moment when an external electric field is applied because of the separation of the positive and negative charge in response to the applied field. The charge distribution that results is such that its field is in the opposite direction as the applied field at the location of the atom or molecule.

▶ We assume that this polarizability is *a linear process*, so that the induced dipole moment is linear in the applied electric field, though the response may be anisotropic. The *polarizability tensor* $\underline{\underline{\alpha}}$ relates the induced dipole moment to the applied field:

$$\vec{p} = \underline{\underline{\alpha}} \cdot \vec{E} \tag{4.1}$$

▶ As we showed in our discussion of multipoles, dipoles can experience torques and forces in an electric field. If a dipole is placed in an electric field, it feels a torque (Equation 3.254)

$$\vec{N} = \vec{p} \times \vec{E} \tag{4.2}$$

If the electric field is nonuniform, the dipole feels a force (Equation 3.252)

$$\vec{F} = \left(\vec{p} \cdot \vec{\nabla} \right) \vec{E} \tag{4.3}$$

▶ If a medium consists of polarizable atoms or molecules, then that medium can become polarized under the application of an electric field. The *polarization* (or *polarization density*) of the medium is

$$\vec{P} = n \vec{p} \tag{4.4}$$

where *n* is the density of polarizable atoms or molecules and $\vec{p}$ is the induced dipole per atom or molecule.

Bound Charges and the Potential of a Polarizable Material

When a medium is polarized and acquires a polarization vector $\vec{P}$, then it can generate its own electric field. This comes from the superposition of the dipole fields of the individual polarized atoms or molecules. In Ph1b, you saw how the polarization could be interpreted as yielding *bound charge* densities: when the medium polarizes, the positive components of some dipoles are cancelled by the negative components of nearby dipoles, but there can appear a net effective charge: on the boundaries, where the cancellation fails, and in the bulk if the dipole density is not uniform, also causing the cancellation to fail. This argument was made in Purcell in Ph1b to derive the bound charge densities, and Griffiths makes it in §4.2.2. Here we derive the relationship between the polarization vector and the bound charge density in rigorous fashion.

The total electric potential generated by a polarizable medium is found by summing up the dipole potentials of the individual dipoles:

$$V(\vec{r}) = \frac{1}{4\pi\epsilon_o} \int_{\mathcal{V}} d\tau' \, \frac{\vec{P}(\vec{r}') \cdot (\vec{r} - \vec{r}')}{|\vec{r} - \vec{r}'|^3} \tag{4.5}$$

We use the identity $(\vec{r} - \vec{r}')/|\vec{r} - \vec{r}'|^3 = \vec{\nabla}_{\vec{r}'}(1/|\vec{r} - \vec{r}'|)$ (note: no minus sign because this is $\vec{\nabla}_{\vec{r}'}$, not $\vec{\nabla}_{\vec{r}}$, and we have $\vec{r} - \vec{r}'$ in the numerator, not $\vec{r}' - \vec{r}$) to rewrite this as

$$V(\vec{r}) = \frac{1}{4\pi\epsilon_o} \int_{\mathcal{V}} d\tau' \, \vec{P}(\vec{r}') \cdot \vec{\nabla}_{\vec{r}'}\left(\frac{1}{|\vec{r} - \vec{r}'|}\right) \tag{4.6}$$

We can integrate by parts to obtain

$$V(\vec{r}) = \frac{1}{4\pi\epsilon_o} \left[\int_{\mathcal{V}} d\tau' \vec{\nabla}_{\vec{r}'} \cdot \left(\frac{\vec{P}(\vec{r}')}{|\vec{r}-\vec{r}'|} \right) - \int_{\mathcal{V}} d\tau' \frac{1}{|\vec{r}-\vec{r}'|} \left(\vec{\nabla}_{\vec{r}'} \cdot \vec{P}(\vec{r}') \right) \right] \quad (4.7)$$

The first term can be converted to a surface integral via the divergence theorem:

$$V(\vec{r}) = \frac{1}{4\pi\epsilon_o} \left[\int_{\mathcal{S}(\mathcal{V})} da' \frac{\widehat{n}(\vec{r}') \cdot \vec{P}(\vec{r}')}{|\vec{r}-\vec{r}'|} - \int_{\mathcal{V}} d\tau' \frac{1}{|\vec{r}-\vec{r}'|} \left(\vec{\nabla}_{\vec{r}'} \cdot \vec{P}(\vec{r}') \right) \right] \quad (4.8)$$

We thus see that the potential appears to be that of a surface charge density $\sigma_b(\vec{r}')$ on $\mathcal{S}(\mathcal{V})$ and a volume charge density $\rho_b(\vec{r}')$ in $\mathcal{V}$ with ($\widehat{n}$ is the outward normal from the polarizable material):

$$\sigma_b(\vec{r}') = \widehat{n}(\vec{r}') \cdot \vec{P}(\vec{r}') \qquad \rho_b(\vec{r}') = -\vec{\nabla}_{\vec{r}'} \cdot \vec{P}(\vec{r}') \quad (4.9)$$

$$V(\vec{r}) = \frac{1}{4\pi\epsilon_o} \left[\int_{\mathcal{S}(\mathcal{V})} da' \frac{\sigma_b(\vec{r}')}{|\vec{r}-\vec{r}'|} + \int_{\mathcal{V}} d\tau' \frac{\rho_b(\vec{r}')}{|\vec{r}-\vec{r}'|} \right] \quad (4.10)$$

These charges are called "bound charges" because they are bound to the polarizable medium.

Example 4.1: Potential and Field of a Uniformly Polarized Sphere

This problem from Ph1b is much easier to solve with our knowledge of solutions to
Laplace's Equation than it was without such techniques. The polarization density is a
constant $\vec{P} = P\,\hat{z}$. The bound volume charge density vanishes because $\vec{P}$ is constant.
The bound surface charge density on the surface at radius R is

$$\sigma_b = \hat{n}(\vec{r}) \cdot \vec{P} = \hat{r} \cdot P\,\hat{z} = P\,\cos\theta \tag{4.11}$$

Therefore, Laplace's Equation holds everywhere except at $r = R$, so we apply the same
techniques as we developed previously. Recall that we discussed this same problem in
Example 3.6 (Griffiths Example 3.9) for a generic $\sigma(\theta)$. The generic solution was

$$V(r < R, \theta) = \sum_{\ell=0}^{\infty} A_\ell r^\ell P_\ell(\cos\theta) \qquad V(r > R, \theta) = \sum_{\ell=0}^{\infty} \frac{B_\ell}{r^{\ell+1}} P_\ell(\cos\theta) \tag{4.12}$$

$$\text{with} \quad A_\ell = \frac{1}{2\,\epsilon_o\,R^{\ell-1}} \int_0^\pi d\theta'\,\sin\theta'\,\sigma(\theta')\,P_\ell(\cos\theta') \qquad B_\ell = A_\ell\,R^{2\,\ell+1} \tag{4.13}$$

Since $\sigma(\theta) = P\cos\theta = P\,P_1(\cos\theta)$, the orthonormal functions do their job and we get
(making sure to include the normalization factor $2/(2\,\ell+1) = 2/3$):

$$V(r < R, \theta) = \frac{P\,r\,\cos\theta}{3\,\epsilon_o} \qquad V(r > R, \theta) = \frac{P\,R^3\,\cos\theta}{3\,\epsilon_o\,r^2} \tag{4.14}$$

We can write these more simply. We recognize $z = r \cos\theta$ and that the total dipole moment of the sphere is $\vec{p} = 4\pi R^3 P \hat{z}/3$, yielding

$$V(r < R, \theta) = \frac{P z}{3 \epsilon_o} \qquad V(r > R, \theta) = \frac{\vec{p} \cdot \hat{r}}{4 \pi \epsilon_o r^2} \qquad (4.15)$$

Thus, the field inside the sphere is *uniform*, $\vec{E} = -\vec{P}/3 \epsilon_o$, and the field outside the sphere is that of a dipole $\vec{p}$. Note that the field outside the sphere is a perfect dipole field all the way to $r = R$; this is not an approximation (until you get so close to the surface that you can see the discretization of the dipoles).

We remind the reader of the Ph1b technique, where we obtained this same result by treating the sphere as two spheres of uniform charge density $\rho = q/(4\pi R^3/3)$ with their centers displaced by $\vec{d} = \vec{p}/q$. The field inside a uniform sphere of charge is proportional to the radial vector outward from its center, so the two vectors $\vec{r} - \vec{d}/2$ and $\vec{r} + \vec{d}/2$ end up differencing (because the two spheres have opposite charge) to yield $\vec{d}$, yielding the uniform internal field. Outside the spheres, they look like point charges, so the system looks like a point dipole $\vec{p}$.

One could also use this argument to figure out that the charge density on the surface is $\sigma = P \cos\theta$ and evaluate the potential and field of that charge distribution.

The Electric Displacement Field

Definition of the Electric Displacement Field

We proved earlier that the potential due to a polarization density $\vec{P}(\vec{r})$ is

$$V(\vec{r}) = \frac{1}{4\pi\epsilon_o}\left[\int_{\mathcal{S}(\mathcal{V})} da' \frac{\hat{n}(\vec{r}') \cdot \vec{P}(\vec{r}')}{|\vec{r} - \vec{r}'|} + \int_{\mathcal{V}} d\tau' \frac{-\vec{\nabla}_{\vec{r}'} \cdot \vec{P}(\vec{r}')}{|\vec{r} - \vec{r}'|}\right] \qquad (4.16)$$

These are analogues of Coulomb's law for ρ_b, so the potential and field due to the polarization density satisfy

$$\nabla^2 V_b = -\frac{1}{\epsilon_o}\rho_b \qquad \vec{\nabla} \cdot \vec{E}_b = \frac{1}{\epsilon_o}\rho_b = -\frac{1}{\epsilon_o}\vec{\nabla} \cdot \vec{P} \qquad (4.17)$$

If there is a *free* charge density ρ_f (which will contribute to V and $\vec{E}$!), then we see that the total potential and field satisfy

$$\nabla^2 V = -\frac{1}{\epsilon_o}(\rho_f + \rho_b) \qquad \vec{\nabla} \cdot \vec{E} = \frac{1}{\epsilon_o}\left(\rho_f - \vec{\nabla} \cdot \vec{P}\right) \qquad (4.18)$$

We will see later that it will be convenient to have a field that depends primarily on the free charge density. Thus, we define the *electric displacement field* by

$$\vec{D} = \epsilon_o \vec{E} + \vec{P} \tag{4.19}$$

We immediately see that Gauss's Law can be written as

$$\vec{\nabla} \cdot \vec{D} = \rho_f \qquad \Longleftrightarrow \qquad \oint_{\mathcal{S}} da \, \widehat{n} \cdot \vec{D} = Q_{free,encl} \tag{4.20}$$

The Helmholtz Theorem tells us that any vector field can be written as the sum of a curl-free component (sourced by the divergence of the field) and a divergence-free component (sourced by the curl of the field). Thus, to fully understand $\vec{D}$, we also need to determine its curl:

$$\vec{\nabla} \times \vec{D} = \epsilon_o \vec{\nabla} \times \vec{E} + \vec{\nabla} \times \vec{P} = \vec{\nabla} \times \vec{P} \tag{4.21}$$

Because the right side may not vanish, the left side may not vanish. This possibly nonzero curl is an important distinction between $\vec{D}$ and $\vec{E}$.

While Gauss's Law does indeed hold for $\vec{D}$, the possibility that $\vec{\nabla} \times \vec{D} \neq 0$ implies that the standard symmetry assumptions we make to apply Gauss's Law to find the field may not apply.

However, if one knows that, due to symmetry or some other consideration, $\vec{\nabla} \times \vec{P} = 0$, then *one can apply the standard techniques for using Gauss's Law combined with symmetry to calculate the displacement field.* ($\vec{\nabla} \times \vec{P} = 0$ should be interpreted as also requiring that any boundaries be normal to $\vec{P}$ because we will see below that, unlike for $\vec{E}$, the tangential component of $\vec{D}$ is not continuous if $\vec{P}$ has a tangential component.)

When the above is true, $\vec{D}$ provides a calculational convenience: if a free charge density ρ_f and a polarization field $\vec{P}$ are specified, then we should calculate $\vec{D}$ from the free charge density using Gauss's Law and then obtain the electric field from $\vec{E} = (\vec{D} - \vec{P})/\epsilon_o$. This simplification is possible only because of the particular form of the bound charge density, $\rho_b = -\vec{\nabla} \cdot \vec{P}$, which parallels the mathematical form of Gauss's Law, along with the condition $\vec{\nabla} \times \vec{P} = 0$.

Note the extra condition $\vec{\nabla} \times \vec{P} = 0$; this reflects the fact that $\vec{P}$ has more degrees of freedom than a scalar field ρ_b, so those extra degrees of freedom need to be specified via the curl-free condition for $\vec{D}$ to be derivable from ρ_f alone.

The situation will simplify somewhat when we consider linear, uniform dielectrics where $\vec{P} \propto \vec{E}$; then $\vec{\nabla} \times \vec{P} = 0$ is guaranteed, though the requirement that $\vec{P}$ be normal to any boundaries may still create complications.

Boundary Conditions on the Displacement Field

We derived boundary conditions on $\vec{E}$ earlier, Equations 2.59 and 2.61:

$$\widehat{n} \cdot \left(\vec{E}_2 - \vec{E}_1 \right) = \frac{1}{\epsilon_o}\, \sigma \qquad \widehat{s} \cdot \left(\vec{E}_2 - \vec{E}_1 \right) = 0 \tag{4.22}$$

where $\widehat{n}$ is the normal vector pointing from region 1 into region 2 and $\widehat{s}$ is any tangential vector (i.e., $\widehat{s} \cdot \widehat{n} = 0$). We derived the equation for the normal component using the divergence of $\vec{E}$. So, here, we can use the fact that $\vec{\nabla} \cdot \vec{D} = \rho_f$, which yields

$$\boxed{\widehat{n} \cdot \left(\vec{D}_2 - \vec{D}_1 \right) = \sigma_f} \tag{4.23}$$

Note that, by definition, we have $\sigma_b = \widehat{n} \cdot \vec{P}$ where $\widehat{n}$ is the outward normal going from a region with a polarization density to vacuum. Therefore, by superposition,

$$\boxed{\widehat{n} \cdot \left(\vec{P}_2 - \vec{P}_1 \right) = -\sigma_b} \tag{4.24}$$

We could also have used $\rho_b = -\vec{\nabla} \cdot \vec{P}$ and followed the same type of derivation as used for $\vec{E}$ and $\vec{D}$. The sign on the right side of the boundary condition enters because of the sign in $\vec{\nabla} \cdot \vec{P} = -\rho_b$.

In general, we know nothing about $\vec{\nabla} \times \vec{P}$, so the boundary condition on the tangential component of $\vec{D}$ just reflects the fact that its curl is the curl of the polarization field. We obtain this condition by inserting the relation between $\vec{E}$, $\vec{D}$, and $\vec{P}$ into the above tangential condition:

$$\hat{s} \cdot \left(\vec{D}_2 - \vec{D}_1 \right) = \hat{s} \cdot \left(\vec{P}_2 - \vec{P}_1 \right) \tag{4.25}$$

Note that, even in the case of linear dielectrics, the right side can be nonzero, as we will see below.

What Sources $\vec{D}$? When Does It Vanish?

Consider the uniformly polarized sphere we just discussed in Example 4.1. In that example, the displacement field is

$$\vec{D}(r < R) = \epsilon_o \vec{E} + \vec{P} = -\frac{\vec{P}}{3} + \vec{P} = \frac{2}{3}\,\vec{P} \tag{4.26}$$

$$\vec{D}(r > R) = \epsilon_o \vec{E} = \epsilon_o \left(\text{field of an electric dipole } \vec{p} = \frac{4\,\pi}{3}\,R^3\,\vec{P} \right) \tag{4.27}$$

There is no free charge in the problem, yet we have a nonzero $\vec{D}$! The nonzero nature of $\vec{\nabla} \times \vec{D}$ implies that $\vec{D}$ has another sourcing term that is not captured by Gauss's Law for $\vec{D}$. In this case, this sourcing term manifests as a discontinuity of the tangential component of $\vec{P}$ at $r = R$: $\vec{P} \cdot \hat{\theta} = P\hat{z} \cdot \hat{\theta} = -P\sin\theta$. This nonzero value of $\hat{s} \cdot \vec{P}$ is what makes $\hat{s} \cdot \vec{D} \neq 0$ and thus we cannot apply the usual symmetry arguments to use Gauss's Law for $\vec{D}$ (which would otherwise tell us $\vec{D} = 0$ because $\rho_f = 0$ everywhere). For $\vec{D}$ to vanish completely, one not only needs the free charge to vanish but there must also be no nontrivial boundary conditions on $\vec{D}$.

Linear Dielectrics

So far, we have considered situations where $\vec{P}$ has been specified for us. But, it is usually caused by an external field, and so what we really want to do is figure out what observed potential and field arise by summing the externally applied potential/field and that due to the polarization of the dielectric in response to that external potential/field. For most substances, at least at low fields, the relation between the two is linear and there is a simple scalar constant of proportionality:

$$\vec{P} = \epsilon_o \, \chi_e \, \vec{E} \tag{4.28}$$

where χ_e is the *electric susceptibility*. Such materials are called *linear dielectrics*. An immediate implication of the above is:

$$\vec{D} = \epsilon_o \, \vec{E} + \vec{P} = \epsilon_o \, (1 + \chi_e) \, \vec{E} \equiv \epsilon \, \vec{E} \tag{4.29}$$

where $\epsilon \equiv \epsilon_o \, (1 + \chi_e)$ is the *permittivity* of the material and $\epsilon_r \equiv 1 + \chi_e$ is the *relative permittivity* or *dielectric constant* of the material.

A very important point is that $\vec{E}$ above is the *total* field, not just the externally applied field. You can think of polarization as an iterative process: an applied field $\vec{E_0}$ causes polarization $\vec{P_0}$, which creates its own field $\vec{E_1}$, which the polarization responds to by adding a contribution $\vec{P_1}$, which creates its own field $\vec{E_2}$, and so on. The process converges to the final total electric field $\vec{E}$ and polarization $\vec{P}$.

Example 4.2: Conducting sphere with dielectric shell around it

Consider a conducting sphere of radius a with (free) charge Q on it surrounded by a (thick) shell of dielectric ϵ with inner and outer radii a and b. Because the system is spherically symmetric and contains a linear dielectric, we know that $\vec{E}$, $\vec{D}$, and $\vec{P}$ all have the form

$$\vec{E} = E(r)\,\hat{r} \qquad \vec{D} = D(r)\,\hat{r} \qquad \vec{P} = P(r)\,\hat{r} \qquad (4.30)$$

This ensures that the curl of all three vanish and that, at the boundaries, we have no tangential components of $\vec{D}$ and $\vec{P}$. *We have now satisfied all the conditions required for us to be able to derive $\vec{D}$ directly from the free charge by Gauss's Law*, which yields

$$\vec{D}(\vec{r}) = \frac{Q}{4\,\pi\,r^2}\,\hat{r} \qquad r > a \qquad (4.31)$$

($\vec{D} = \vec{E} = \vec{P} = 0$ for $r < a$.) Then we just apply the relation between $\vec{D}$ and $\vec{E}$:

$$\vec{E}(\vec{r}) = \frac{Q}{4\,\pi\,\epsilon(r)\,r^2}\,\hat{r} = \left\{ \begin{array}{ll} \left(Q/4\,\pi\,\epsilon\,r^2\right)\hat{r} & a < r < b \\ \left(Q/4\,\pi\,\epsilon_o\,r^2\right)\hat{r} & b < r \end{array} \right. \qquad (4.32)$$

The electric field is *screened* (reduced) inside the dielectric and unchanged outside.

Let's calculate the polarization vector and bound charge density:

$$\vec{P}(\vec{r}) = \epsilon_o \chi_e(r) \vec{E}(\vec{r}) = (\epsilon(r) - \epsilon_o) \vec{E}(\vec{r}) = \frac{\epsilon(r) - \epsilon_o}{\epsilon(r)} \frac{Q}{4\pi r^2} \hat{r}$$

$$= \begin{cases} \frac{\epsilon - \epsilon_o}{\epsilon} \frac{Q}{4\pi r^2} \hat{r} & a < r < b \\ 0 & b < r \end{cases} \tag{4.33}$$

$$\rho_b = -\vec{\nabla} \cdot \vec{P} = 0 \tag{4.34}$$

$$\sigma_b = \begin{cases} -\hat{r} \cdot \vec{P}(r = a) = -\frac{\epsilon - \epsilon_o}{\epsilon} \frac{Q}{4\pi a^2} & r = a \\ \hat{r} \cdot \vec{P}(r = b) = \frac{\epsilon - \epsilon_o}{\epsilon} \frac{Q}{4\pi b^2} & r = b \end{cases} \tag{4.35}$$

Note the ϵ in the denominator! We see that $\vec{P}$ is radially outward and decreasing with r like $1/r^2$ as $\vec{E}$ does. Note that, even though $\vec{P}$ is position-dependent, its divergence vanishes, so there is no bound charge density. There is surface charge density, negative at $r = a$ and positive at $r = b$. This is to be expected, as the dielectric polarizes so the negative ends of the dipoles are attracted to Q on the conducting sphere and the positive ends are repelled, leaving uncancelled layers of negative charge on the inner boundary and positive charge on the outer boundary.

The electric field is reduced inside the dielectric because the negative charge on the inner boundary screens (generates a field that partially cancels) the field of the free charge on the conducting sphere: the total surface charge density $\sigma_f + \sigma_b$ at $r = a$ is less than $Q/4\pi a^2$, and it is the *total* charge that determines $\vec{E}$.

Note that, because of the neutrality of the dielectric, the total surface charge on the outer boundary cancels that on the inner boundary, so the net charge enclosed inside a sphere of radius $r > b$ is just Q: outside the dielectric, no screening effect is present.

It is worth thinking about the above a bit: it occurs both because the dielectric has no net charge *and* the problem is spherically symmetric. In contrast, we will see a dielectric sphere can polarize in an external field and generate a field outside itself in spite of having no net charge, which is possible because spherical symmetry is broken in that case. But there is no monopole field, only a dipole field.

Note also that, once you have calculated σ_b and ρ_b, you can ignore the presence of the dielectric: as we stated earlier, the total field is sourced by the sum of the free and bound charge densities and the dielectric has no further effect, which one can see here from noticing that $\vec{E}$ in the dielectric is what one would have calculated if one had been given $\sigma_f + \sigma_b$ at $r = a$.

Finally, let's calculate the electric potential from $\vec{E}$ (not $\vec{D}$!):

$$V(\vec{r}) = -\int_{\infty}^{\vec{r}} d\vec{s}' \cdot \vec{E}(\vec{r}') = -\int_{\infty}^{r} dr' E(r')$$

$$V(r > b) = -\frac{Q}{4\pi}\left[\int_{\infty}^{r} dr' \frac{1}{\epsilon_o r^2}\right] = \frac{Q}{4\pi}\frac{1}{\epsilon_o r} \tag{4.36}$$

$$V(a < r < b) = -\frac{Q}{4\pi}\left[\int_{\infty}^{b} dr' \frac{1}{\epsilon_o r^2} + \int_{b}^{r} dr' \frac{1}{\epsilon r^2}\right]$$

$$= \frac{Q}{4\pi}\left[\frac{1}{\epsilon_o r}\Big|_{\infty}^{b} + \frac{1}{\epsilon r}\Big|_{b}^{r}\right] = \frac{Q}{4\pi}\left[\frac{1}{b}\left(\frac{1}{\epsilon_o} - \frac{1}{\epsilon}\right) + \frac{1}{\epsilon r}\right] \tag{4.37}$$

$$V(r < a) = V(r = a) = \frac{Q}{4\pi}\left[\frac{1}{b}\left(\frac{1}{\epsilon_o} - \frac{1}{\epsilon}\right) + \frac{1}{\epsilon a}\right] \tag{4.38}$$

where V is constant for $r < a$ because $r < a$ is occupied by a conductor.

A final comment: if one takes the $\epsilon \to \infty$ limit, one can see that one recovers the behavior one would have if the entire region $r < b$ were filled with conductor. A conductor can be considered to be an infinitely polarizable dielectric, with $\vec{E} = 0$ inside, which requires $\chi_e \to \infty$.

Example 4.3: Parallel plate capacitor with dielectric

You all know from Ph1b that filling the volume between the plates of a parallel-plate capacitor increases the capacitance to $C = \epsilon_r C_{vac}$ where C_{vac} is the capacitance with vacuum between the plates. We remind you why this is true.

Let the capacitor plates lie parallel to the xy-plane at $z = 0$ (negative plate) and $z = a$ (positive plate) so $\hat{z}$ is the unit vector pointing from the negative plate to the positive one. In such a geometry, we know from symmetry that $\vec{E}$, $\vec{D}$, and $\vec{P}$ are all parallel to $\hat{z}$ and independent of xy, assuming we ignore the capacitor edges. Thus, at the interfaces at $z = 0$ and $z = a$, all these vectors are normal to the interface and so no tangential components are present. These features of the fields imply that we can apply Gauss's Law to the free charge density to find $\vec{D}$.

The free charge density is $\sigma_f = \pm Q/A$ where Q is the charge on the plates ($+Q$ at $z = a$ and $-Q$ at $z = 0$) and A is the plate area. Gauss's Law for an infinite sheet of charge (Griffiths Example 2.5) tells us that the field of a single sheet is $E = \sigma/2\epsilon_o$. Therefore, we have for this case

$$\vec{D} = \left\{ \begin{array}{ll} -\frac{Q}{A}\hat{z} & 0 < z < a \\ 0 & z < 0, z > a \end{array} \right. \tag{4.39}$$

because the fields of the two plates cancel for $z < 0$ and $z > a$ but add for $0 < z < a$, and there is no ϵ_o because we are calculating $\vec{D}$, not $\vec{E}$.

This implies:

$$\vec{E} = \begin{cases} -\frac{1}{\epsilon} \frac{Q}{A} \hat{z} & 0 < z < a \\ 0 & z < 0, z > a \end{cases} \qquad \vec{P} = \begin{cases} -\frac{\epsilon - \epsilon_o}{\epsilon} \frac{Q}{A} \hat{z} & 0 < z < a \\ 0 & z < 0, z > a \end{cases} \qquad (4.40)$$

$$\rho_b = -\vec{\nabla} \cdot \vec{P} = 0 \qquad (4.41)$$

$$\sigma_b = \hat{n} \cdot \vec{P} = \begin{cases} \hat{z} \cdot \vec{P}(z = a) & z = a \\ -\hat{z} \cdot \vec{P}(z = 0) & z = 0 \end{cases} = \begin{cases} -\frac{\epsilon - \epsilon_o}{\epsilon} \frac{Q}{A} & z = a \\ \frac{\epsilon - \epsilon_o}{\epsilon} \frac{Q}{A} & z = 0 \end{cases} \qquad (4.42)$$

The electric field is, like in Example 4.2, screened inside the dielectric, with its amplitude reduced by a factor $\epsilon_r = \epsilon/\epsilon_o$. The bound surface charge densities near the plates are the source of this screening, contributing a field opposite to the field of the free charge. The electric field inside the dielectric is the field one expects from surface charge densities $\sigma_f + \sigma_b = \pm(\epsilon_o/\epsilon)\,(Q/A)$. Finally, the voltage is

$$V(0 < z < a) = -\int_0^z d\vec{s}' \cdot \vec{E}(\vec{r}') = -\int_0^z dz' \left(-\frac{1}{\epsilon} \frac{Q}{A} \right) = \frac{1}{\epsilon} \frac{Q}{A} z \qquad (4.43)$$

From this, we can calculate the capacitance, which comes out as expected:

$$C = \frac{Q}{\Delta V} = \frac{Q}{(1/\epsilon)\,(Q/A)\,a} = \epsilon \frac{A}{a} = \epsilon_r C_{vac} \qquad (4.44)$$

C is increased because ΔV is reduced because of the screening inside the dielectric.

Example 4.4: Parallel plate capacitor with two-layer dielectric

Let's repeat, but now with a capacitor that has two slabs of dielectric with different ϵ: ϵ_1 for $0 < z < a$ and ϵ_2 for $a < z < b$, where the top plate is now at $z = b$. Because the interface is normal to $\vec{P}$, we can apply Gauss's Law for $\vec{D}$ as we did before, yielding no change in $\vec{D}$, but now the ϵ quantities in $\vec{E}$ and $\vec{P}$ depend on z.

The volume bound charge density vanishes again. The surface charge density at the top and bottom has the same expression, but again with ϵ being evaluated for the particular value of z. The surface bound charge density at the $z = a$ interface is

$$\sigma_b(z = a) = \widehat{n}_1 \cdot \vec{P}_1 + \widehat{n}_2 \cdot \vec{P}_2 = \widehat{z} \cdot \vec{P}_1 - \widehat{z} \cdot \vec{P}_2 = \frac{Q}{A}\left(-\frac{\epsilon_1 - \epsilon_o}{\epsilon_1} + \frac{\epsilon_2 - \epsilon_o}{\epsilon_2}\right) \quad (4.45)$$

Depending on which dielectric constant is greater, this can be positive or negative. Of course, it vanishes if $\epsilon_1 = \epsilon_2$. The potential and capacitance are

$$V(0 < z < a) = \frac{1}{\epsilon_1}\frac{Q}{A}z \qquad V(a < z < b) = \frac{1}{\epsilon_1}\frac{Q}{A}a + \frac{1}{\epsilon_2}\frac{Q}{A}(z - a) \qquad (4.46)$$

$$C = \frac{Q}{\Delta V} = \left(\frac{a}{\epsilon_1} + \frac{b - a}{\epsilon_2}\right)^{-1}A = \epsilon_{eff}\frac{A}{b} = \epsilon_{eff,r}\,C_{vac} \qquad (4.47)$$

where $1/\epsilon_{eff} = [a/\epsilon_1 + (b - a)/\epsilon_2]/b$ is the thickness-weighted inverse mean of the dielectric constants and $\epsilon_{eff,r} = \epsilon_{eff}/\epsilon_o$. This is the same as two capacitors in series, which is not surprising since that problem has the same equipotential surfaces. The total field is that of three sheets of surface charge $\sigma_f + \sigma_b$, with $\sigma_f = 0$ at the interface between the two dielectrics.

Lecture 13:

Electrostatics in Matter II:
Linear Dielectrics (cont.)
Boundary Value Problems with Linear Dielectrics
Electrostatic Energy for Linear Dielectrics

Date Revised: 2024/02/20 06:30
Revised lecture break
Date Given: 2024/02/15

Example 4.5: Capacitor with two side-by-side (parallel) dielectrics

Now, allow the capacitor to have plate spacing a but with two different dielectrics side-by-side, with ϵ_1 occupying A_1 and $\mathcal{V}_1$ and ϵ_2 occupying A_2 and $\mathcal{V}_2$. It is a reasonable guess that one should treat this as two capacitors in parallel so that

$$C = C_1 + C_2 = \frac{1}{a}\left(\epsilon_1 A_1 + \epsilon_2 A_2\right) \tag{4.48}$$

But let's derive this from scratch, appreciating the subtlety at the interface.

Because the voltage difference between the two plates is independent of ϵ (they are equipotentials), it is reasonable to guess that $\vec{E}$ is the same in ϵ_1 and ϵ_2: this is the key insight! Because the dielectrics are uniform in z, it is also reasonable to assume the field is independent of z as one would have in the single-dielectric case. So, our guess for the form of the fields is:

$$\vec{E} = -E_0\,\hat{z} \quad \vec{D} = \left\{ \begin{array}{l} -\epsilon_1\,E_0\,\hat{z} \quad \text{in } \mathcal{V}_1 \\ -\epsilon_2\,E_0\,\hat{z} \quad \text{in } \mathcal{V}_2 \end{array} \right. \quad \vec{P} = \left\{ \begin{array}{l} -\left(\epsilon_1 - \epsilon_o\right)E_0\,\hat{z} \quad \text{in } \mathcal{V}_1 \\ -\left(\epsilon_2 - \epsilon_o\right)E_0\,\hat{z} \quad \text{in } \mathcal{V}_2 \end{array} \right. \tag{4.49}$$

We see this form respects the tangential boundary conditions at the interface between the two dielectrics, as it has to:

$$\hat{z}\cdot\left(\vec{E}_2 - \vec{E}_1\right) = 0 \qquad \hat{z}\cdot\left(\vec{D}_2 - \vec{D}_1\right) = \left(\epsilon_1 - \epsilon_2\right)E_0 = \hat{z}\cdot\left(\vec{P}_2 - \vec{P}_1\right) \tag{4.50}$$

Because $\vec{D}$ and $\vec{P}$ are different in the two volumes, we must allow the free (and bound) charge densities to be different. This provides us a set of equations to solve for E_0:

$$\epsilon_1 E_0 = \sigma_{f,1} \qquad \epsilon_2 E_0 = \sigma_{f,2} \qquad A_1 \sigma_{f,1} + A_2 \sigma_{f,2} = Q \tag{4.51}$$

$$\implies \qquad E_0 = \frac{1}{\epsilon_{eff}} \frac{Q}{A} \qquad \epsilon_{eff} = \frac{\epsilon_1 A_1 + \epsilon_2 A_2}{A_1 + A_2} \qquad A = A_1 + A_2 \tag{4.52}$$

$$C = \frac{Q}{\Delta V} = \frac{Q}{a E_0} = \epsilon_{eff} \frac{A}{a} = \epsilon_{eff,r} C_{vac} \tag{4.53}$$

which matches our parallel-capacitor expectation. The displacement field, polarization field, and free and bound charge densities are

$$\vec{D} = \begin{cases} -\frac{\epsilon_1}{\epsilon_{eff}} \frac{Q}{A} \hat{z} & \text{in } \mathcal{V}_1 \\ -\frac{\epsilon_2}{\epsilon_{eff}} \frac{Q}{A} \hat{z} & \text{in } \mathcal{V}_2 \end{cases} \qquad \vec{P} = \begin{cases} -\frac{\epsilon_1 - \epsilon_o}{\epsilon_{eff}} \frac{Q}{A} \hat{z} & \text{in } \mathcal{V}_1 \\ -\frac{\epsilon_2 - \epsilon_o}{\epsilon_{eff}} \frac{Q}{A} \hat{z} & \text{in } \mathcal{V}_2 \end{cases} \qquad \rho_b = -\vec{\nabla} \cdot \vec{P} = 0 \tag{4.54}$$

$$|\sigma_f| = \begin{cases} \frac{\epsilon_1}{\epsilon_{eff}} \frac{Q}{A} & \text{in } \mathcal{V}_1 \\ \frac{\epsilon_2}{\epsilon_{eff}} \frac{Q}{A} & \text{in } \mathcal{V}_2 \end{cases} \qquad |\sigma_b| = \begin{cases} \frac{\epsilon_1 - \epsilon_o}{\epsilon_{eff}} \frac{Q}{A} \hat{z} & \text{in } \mathcal{V}_1 \\ \frac{\epsilon_2 - \epsilon_o}{\epsilon_{eff}} \frac{Q}{A} \hat{z} & \text{in } \mathcal{V}_2 \end{cases} \tag{4.55}$$

σ_b always has the opposite sign as σ_f. For $Q > 0$, the sign of σ_f is positive at $z = a$ and negative at $z = 0$. Note that, because $\vec{P}$ is different in $\mathcal{V}_1$ and $\mathcal{V}_2$, so too does σ_b differ between the two dielectrics.

Finally, if one calculates the total charge density $\sigma_f + \sigma_b$ at $z = 0$ or $z = a$, one gets

$$\sigma_{t,1} = \sigma_{f,1} + \sigma_{b,1} = \left(\frac{\epsilon_1}{\epsilon_{eff}} - \frac{\epsilon_1 - \epsilon_o}{\epsilon_{eff}} \right) \frac{Q}{A} = \frac{\epsilon_o}{\epsilon_{eff}} \frac{Q}{A} \qquad (4.56)$$

$$\sigma_{t,2} = \sigma_{f,2} + \sigma_{b,2} = \left(\frac{\epsilon_2}{\epsilon_{eff}} - \frac{\epsilon_2 - \epsilon_o}{\epsilon_{eff}} \right) \frac{Q}{A} = \frac{\epsilon_o}{\epsilon_{eff}} \frac{Q}{A} \qquad (4.57)$$

This makes sense: since the electric field is the same in $\mathcal{V}_1$ and $\mathcal{V}_2$, the total (free + bound) surface charge density sourcing it should be the same. The total charge density is a factor $\epsilon_o/\epsilon_{eff}$ smaller than would be present in the absence of dielectrics because the bound charge density screens the free charge density. The free charge density is different in the two regions because the opposite-sign bound charge density is different because of the different dielectric constants. *In contrast to our naive expectation, the free charge density is not uniform on the conductor; rather, it redistributes itself so the fundamental condition, that the conductors be equipotentials, is satisfied when one includes the effect of the dielectric. Instead, the total charge density is uniform, which yields a field independent of (x, y), which is what ensures the equipotential condition is met.*

Boundary Value Problems with Linear Dielectrics

General Conditions for Linear, Homogeneous Dielectrics

In linear, homogeneous dielectrics,

$$\rho_b = -\vec{\nabla} \cdot \vec{P} = -\vec{\nabla} \cdot \left(\frac{\epsilon - \epsilon_0}{\epsilon} \vec{D} \right) = - \left(\frac{\epsilon - \epsilon_0}{\epsilon} \right) \vec{\nabla} \cdot \vec{D} = - \left(\frac{\epsilon - \epsilon_0}{\epsilon} \right) \rho_f \quad (4.58)$$

(Homogeneity is required so the gradient does not act on ϵ.) Therefore, if there is no free charge density in a linear, homogeneous dielectric, there is no bound charge density either. Thus, the dielectric volume satisfies Laplace's Equation. All our machinery for solving Laplace's Equation applies here.

We always need boundary conditions, though, and we can use the ones we derived earlier (the tangential $\vec{E}$ and $\vec{D}$ conditions will yield the same condition on V, so we start with the simpler one):

$$\hat{n} \cdot \left[\vec{D}_2 - \vec{D}_1\right] = \sigma_f \qquad \hat{s} \cdot \left[\vec{E}_2 - \vec{E}_1\right] = 0 \qquad (4.59)$$

Writing this in terms of the potential, we have

$$\hat{n} \cdot \left[\epsilon_2 \vec{\nabla} V_2 - \epsilon_1 \vec{\nabla} V_1\right] = -\sigma_f \qquad \hat{s} \cdot \left[\vec{\nabla} V_2 - \vec{\nabla} V_1\right] = 0 \qquad (4.60)$$

And, we always require $V_1 = V_2$: the potential must be continuous. While we have three conditions, in general the continuity and tangential gradient conditions will be redundant: the normal gradient condition must be independent because it depends on the free surface charge density while the two others do not. The continuity condition is simpler and so is the one that should be used.

Example 4.6: Spherical cavity in a dielectric medium with uniform field applied

Let's apply the above to a spherical cavity of radius R in a medium with permittivity ϵ with a uniform field $\vec{E} = E_0 \hat{z}$ applied. There is no free charge anywhere. Our boundary conditions therefore are

$$V(r \to \infty) = -E_0\, z = -E_0\, r\, P_1(\cos\theta) \tag{4.61}$$

and, with $V_{in}(r) = V(r < R)$ and $V_{out}(r) = V(r > R)$,

$$\epsilon_o \left. \frac{\partial V_{in}}{\partial r} \right|_{r=R} = \epsilon \left. \frac{\partial V_{out}}{\partial r} \right|_{r=R} \qquad \text{and} \qquad V_{in}(r = R) = V_{out}(r = R) \tag{4.62}$$

We also choose the zero of the potential to be at $z = 0$, $V(z = 0) = 0$, by symmetry as in the case of the conducting sphere in a uniform electric field.

As usual, we begin by writing our generic solutions to Laplace's Equation in spherical coordinates:

$$V_{in}(r) = \sum_{\ell=0}^{\infty} A_\ell^{in} r^\ell P_\ell(\cos\theta) \qquad V_{out}(r) = \sum_{\ell=0}^{\infty} \left(A_\ell^{out} r^\ell + \frac{B_\ell^{out}}{r^{\ell+1}} \right) P_\ell(\cos\theta) \quad (4.63)$$

where we have applied the requirement that V be finite at the origin to eliminate the $1/r^{\ell+1}$ terms for V_{in}. Recall that we cannot eliminate the r^ℓ terms for V_{out} because the potential does not vanish at infinity.

Let's first apply the $r \rightarrow \infty$ condition. We did this before in the case of a metal sphere in a uniform field, and we found

$$A_1^{out} = -E_0 \qquad A_{\ell \neq 1}^{out} = 0 \tag{4.64}$$

Next, we apply the continuity condition at $r = R$, making use of orthonormality of the P_ℓ:

$$A_1^{in} R = -E_0 R + \frac{B_1^{out}}{R^2} \qquad A_{\ell \neq 1}^{in} R^\ell = \frac{B_{\ell \neq 1}^{out}}{R^{\ell+1}} \tag{4.65}$$

Finally, let's take the radial derivative and apply the matching condition on it, again using orthonormality:

$$\epsilon_0 A_1^{in} = -\epsilon \left(E_0 + \frac{2}{R^3} B_1^{out} \right) \qquad \epsilon_0 A_{\ell \neq 1}^{in} \, \ell \, R^{\ell-1} = -\epsilon \, \frac{B_{\ell \neq 1}^{out}}{R^{\ell+2}} \, (\ell + 1) \qquad (4.66)$$

Doing the algebra, we find

$$A_{\ell \neq 1}^{in} = B_{\ell \neq 1}^{out} = 0 \qquad B_1^{out} = -\frac{\epsilon - \epsilon_o}{2\,\epsilon - \epsilon_o} \, E_0 \, R^3 \qquad A_1^{in} = -\frac{3\,\epsilon}{2\,\epsilon + \epsilon_o} \, E_0 \qquad (4.67)$$

Thus, the potential is

$$V_{in}(r) = V(r < R) = -\frac{3\,\epsilon}{2\,\epsilon + \epsilon_o} \, E_0 \, r \cos\theta = -\frac{3\,\epsilon}{2\,\epsilon + \epsilon_o} \, E_0 \, z \qquad (4.68)$$

$$V_{out}(r) = V(r > R) = -E_0 \, r \cos\theta - \frac{\epsilon - \epsilon_o}{2\,\epsilon + \epsilon_o} \, E_0 \, \frac{R^3}{r^2} \, \cos\theta \qquad (4.69)$$

$$= -E_0 \, z + \frac{\vec{p} \cdot \widehat{r}}{4\,\pi\,\epsilon_o \, r^2} \quad \text{with} \quad \vec{p} = -\frac{4\,\pi}{3} \, R^3 \, E_0 \frac{3\,\epsilon_o}{2\,\epsilon + \epsilon_o} \, (\epsilon - \epsilon_o) \, \widehat{z}$$

The potential inside the cavity is that of a uniform electric field in the same direction as the applied field but multiplied by the factor $3\epsilon/(2\epsilon + \epsilon_o) > 1$, while the potential outside is that of the uniform field plus that of a dipole whose orientation is *opposite* the uniform field and whose magnitude is given above. It is as if the *cavity* acquired a polarization density in the negative z direction, though of course that cannot happen because $\chi_e(r < R) = 0$ there and thus $\vec{P}(r < R) = \epsilon_o\chi_e(r < R)\vec{E}(r < R) = 0$. The polarization density outside the cavity is just the *total* (not the applied uniform) field times $\epsilon - \epsilon_o$ (which is not particularly illuminating).

The (bound) surface charge density is

$$\sigma_b = \widehat{n} \cdot \vec{P}(r = R) = \widehat{n} \cdot (\epsilon - \epsilon_o)\,\vec{E}_{out}(r = R)$$

$$= (\epsilon - \epsilon_o)\left(-\widehat{r} \cdot E_0\,\widehat{z} - \frac{\partial}{\partial r}\frac{\epsilon - \epsilon_o}{2\epsilon + \epsilon_o}\,E_0\,\frac{R^3}{r^2}\,\cos\theta\bigg|_{r=R}\right) = -3\,\epsilon_o\,\frac{\epsilon - \epsilon_o}{2\epsilon + \epsilon_o}\,E_0\cos\theta$$

(Notice that $\widehat{n} = -\widehat{r}$ because $\widehat{n}$ is taken to point *out of* the dielectric medium in the definition of σ_b.) We see the boundary of the cavity acquires a surface charge density with the same magnitude and cosine dependence as the bound charge on the surface of a uniformly polarized sphere, though with *opposite* sign (so there is negative charge at the $+z$ end and positive charge at the $-z$ end). The sign follows naturally from our arguments about cancellation of dipole charge.

The field is enhanced in the cavity for two reasons: first, there is no polarizable material to screen the electric field, and, second there is surface charge density on the cavity's boundary that creates an additional field in the direction of the uniform field.

For reference, we note that the solution for a dielectric sphere (Griffiths Example 4.7) in a uniform field looks very similar:

$$V(r < R) = -\frac{3\,\epsilon_o}{2\,\epsilon_o + \epsilon}\,E_0 z \qquad V(r > R) = -E_0 z + \frac{\vec{p} \cdot \hat{r}}{4\,\pi\,\epsilon_o\,r^2} \tag{4.70}$$

$$\text{with} \quad \vec{p} = \frac{4\,\pi}{3}\,R^3\,E_0\,\frac{3\,\epsilon_o}{2\,\epsilon_o + \epsilon}\,(\epsilon - \epsilon_o)\,\hat{z} \equiv \frac{4\,\pi}{3}\,R^3\,\vec{P}(r < R) \tag{4.71}$$

$$\sigma_b = 3\,\epsilon_o\,\frac{\epsilon - \epsilon_o}{2\,\epsilon_o + \epsilon}\,E_0\cos\theta \tag{4.72}$$

Basically, exchange ϵ_o and ϵ everywhere to go between the two results. In this case, the sphere acquires a polarization density $3\,\epsilon_o(\epsilon - \epsilon_o)/(2\,\epsilon_o + \epsilon)$, now in the direction of the applied field. The surface charge density is also of same form as the cavity case with the $\epsilon \leftrightarrow \epsilon_o$ exchange. That exchange flips the sign so that the $+z$ end acquires a positive charge, again as expected from the dipole charge cancellation argument. The field amplitude is reduced (screened) in the dielectric.

From the polarized sphere, one can recover the case of a conducting sphere in an external uniform field by taking $\epsilon \to \infty$ as noted earlier.

Electrostatic Energy in and Forces on Linear Dielectrics

Electrostatic Potential Energy due to an Assembly of Free Charge in the Presence of Dielectrics

It turns out that electrostatic potential energy in the presence of dielectrics is a subtle topic because of the existence of the charges forming the dielectric. There are different kinds of electrostatic potential energy: that needed to assemble the free and bound charge distributions versus that needed to assemble the free charge distribution and polarize the preexisting dielectric. It is generally the latter we are interested in, so we consider that case.

Suppose we have a system in which an electric field $\vec{E}(\vec{r})$ and its potential $V(\vec{r})$ have already been set up and we want to bring in additional *free charge* $\delta\rho_f$ from infinity (assuming the potential vanishes at infinity). In this case, the change in potential energy is

$$\delta U = \int_{\mathcal{V}} d\tau' \left[\delta\rho_f(\vec{r}')\right] V(\vec{r}') \tag{4.73}$$

The free charge density is related to the displacement field by $\vec{\nabla} \cdot \vec{D} = \rho_f$, so a change $\delta\rho_f$ corresponds to a change in the divergence of the displacement field $\delta\left(\vec{\nabla} \cdot \vec{D}\right)$. Linearity of the divergence lets us rewrite this as $\delta\rho_f = \vec{\nabla} \cdot \delta\vec{D}$.

Then, we may integrate by parts and apply the divergence theorem:

$$
\begin{aligned}
\delta U &= \int_{\mathcal{V}} d\tau' \left[\vec{\nabla} \cdot \delta \vec{D}(\vec{r}') \right] V(\vec{r}') \\
&= \int_{\mathcal{V}} d\tau' \vec{\nabla} \cdot \left[V(\vec{r}') \, \delta \vec{D}(\vec{r}') \right] - \int_{\mathcal{V}} d\tau' \left[\delta \vec{D}(\vec{r}') \right] \cdot \vec{\nabla} V(\vec{r}') \\
&= \oint_{\mathcal{S}(\mathcal{V})} da' \, \widehat{n}(\vec{r}') \cdot \left[V(\vec{r}') \, \delta \vec{D}(\vec{r}') \right] + \int_{\mathcal{V}} d\tau' \left[\delta \vec{D}(\vec{r}') \right] \cdot \vec{E}(\vec{r}')
\end{aligned}
\tag{4.74}
$$

Assuming the potential falls off at infinity, the surface term can be taken out to infinity to vanish. (Recall, $V \sim 1/r$ and $D \sim 1/r^2$ while $da' \, r^2$, so the integral falls off as $1/r$.) So, we are then left with

$$
U = \int_0^{\vec{D}} \int_{\mathcal{V}} d\tau' \, \vec{E}(\vec{r}') \cdot d\vec{D}(\vec{r}')
\tag{4.75}
$$

There are two integrals here, one over volume and one over the value of $\vec{D}$ from zero to its final value. $\vec{E}$ is of course tied to $\vec{D}$ and they vary together.

For the case of a linear (but perhaps not homogeneous) dielectric, we may use $\vec{D}(\vec{r}) = \epsilon(\vec{r})\vec{E}(\vec{r})$ and therefore

$$
\begin{aligned}
U &= \int_0^{\vec{E}} \int_{\mathcal{V}} d\tau'\, \epsilon(\vec{r}')\, \vec{E}(\vec{r}') \cdot d\vec{E}(\vec{r}') \\
&= \frac{1}{2} \int_0^{\vec{E}} \int_{\mathcal{V}} d\tau' \epsilon(\vec{r}')\, d\left[\vec{E}(\vec{r}') \cdot \vec{E}(\vec{r}') \right] \\
&= \frac{1}{2} \int_{\mathcal{V}} d\tau'\, \epsilon(\vec{r}')\, E^2(\vec{r}') = \frac{1}{2} \int_{\mathcal{V}} d\tau'\, \vec{E}(\vec{r}') \cdot \vec{D}(\vec{r}')
\end{aligned}
\tag{4.76}
$$

If the medium is linear *and* homogeneous, one can pull ϵ outside the integral at any point, yielding

$$
U = \frac{\epsilon}{2} \int_{\mathcal{V}} d\tau' \left| \vec{E}(\vec{r}') \right|^2 = \frac{1}{2\epsilon} \int_{\mathcal{V}} d\tau' \left| \vec{D}(\vec{r}') \right|^2
\tag{4.77}
$$

We may infer that the energy density, neglecting the energy density intrinsic to the creation of the dipoles, is

$$
u(\vec{r}) = \frac{\epsilon}{2} \left| \vec{E}(\vec{r}) \right|^2 = \frac{1}{2\epsilon} \left| \vec{D}(\vec{r}) \right|^2
\tag{4.78}
$$

By contrast, if we wanted to know the total electrostatic potential energy stored in the assembly of the free and bound charge, we would just do the usual volume integral of E^2 with ϵ_o instead of ϵ. That energy is smaller because $\epsilon > \epsilon_o$. The reason for this difference is that assembling the medium in the first place, which consists of bringing positive and negative charges together, creates a system with negative potential energy, and thus the total potential energy of the system would be lower if we accounted for the energy of assembling the medium. But we will never pull the dielectric apart, so it is natural to treat that component of the potential energy as an offset that is inaccessible and neglect it in the electrostatic potential energy.

Energy of a Dielectric in an External Field

A topic naturally related to the above is the electrostatic energy of a polarizable material in an external field.

Suppose we start with a system with a free charge distribution ρ_f that sources a field $\vec{E}_1$ in a dielectric medium ϵ_1, yielding a displacement $\vec{D}_1 = \epsilon_1 \vec{E}_1$. The initial energy is

$$U_1 = \frac{1}{2} \int d\tau \, \vec{E}_1 \cdot \vec{D}_1 \tag{4.79}$$

Now, *with the charges sourcing $\vec{E}_1$ held fixed*, let's introduce a piece of dielectric occupying the volume $\mathcal{V}_2$ and having dielectric constant ϵ_2, replacing the dielectric of dielectric constant ϵ_1 there. The remainder of space outside $\mathcal{V}_2$ is occupied by ϵ_1 in both configurations. The electric field and displacement field everywhere change to $\vec{E}_2$ and $\vec{D}_2$, where $\vec{D}_2(\vec{r}) = \epsilon(\vec{r}) \, \vec{E}_2(\vec{r})$. *Note that $\vec{E}_1$ and $\vec{E}_2$ are not identical outside $\mathcal{V}_2$, and the same is true for $\vec{D}_1$ and $\vec{D}_2$. The dielectric affects the field everywhere, not just inside $\mathcal{V}_2$.* The energy is now

$$U_2 = \frac{1}{2} \int d\tau \, \vec{E}_2 \cdot \vec{D}_2 \tag{4.80}$$

The difference in energy between the two configurations is therefore

$$U_2 - U_1 = \frac{1}{2} \int d\tau \left[\vec{E}_2 \cdot \vec{D}_2 - \vec{E}_1 \cdot \vec{D}_1 \right] \tag{4.81}$$

Let us rewrite the energy difference as

$$U_2 - U_1 = \frac{1}{2} \int d\tau \left[\vec{E}_2 \cdot \vec{D}_1 - \vec{E}_1 \cdot \vec{D}_2 \right] + \frac{1}{2} \int d\tau \left[\vec{E}_2 + \vec{E}_1 \right] \cdot \left[\vec{D}_2 - \vec{D}_1 \right] \qquad (4.82)$$

It holds that $\vec{\nabla} \times \left[\vec{E}_2 + \vec{E}_1 \right] = 0$, so it can be derived from a potential V, so the second integral becomes

$$-\frac{1}{2} \int d\tau \left(\vec{\nabla} V \right) \cdot \left[\vec{D}_2 - \vec{D}_1 \right] \qquad (4.83)$$

We integrate by parts (the surface term vanishes because it depends on $\vec{D}_2 - \vec{D}_1$, which should vanish as one goes far from the dielectric) to obtain

$$\frac{1}{2} \int d\tau \, V \, \vec{\nabla} \cdot \left[\vec{D}_2 - \vec{D}_1 \right] \qquad (4.84)$$

This divergence vanishes because the free charge has not changed between the two configurations (recall, $\vec{\nabla} \cdot \vec{D} = \rho_f$).

So the second term in the energy vanishes, leaving

$$U_2 - U_1 = \frac{1}{2} \int d\tau \left[\vec{E}_2 \cdot \vec{D}_1 - \vec{E}_1 \cdot \vec{D}_2 \right] \tag{4.85}$$

Now, outside $\mathcal{V}_2$, it holds that $\vec{D}_2 = \epsilon_1 \vec{E}_2$ (remember, ϵ only changed inside $\mathcal{V}_2$), and recall also $\vec{D}_1 = \epsilon_1 \vec{E}_1$ everywhere, so the two terms cancel each other there and the integrand vanishes outside $\mathcal{V}_2$. Therefore, we can restrict the integral to $\mathcal{V}_2$:

$$U_2 - U_1 = -\frac{1}{2} \int_{\mathcal{V}_2} d\tau \, (\epsilon_2 - \epsilon_1) \, \vec{E}_2 \cdot \vec{E}_1 \tag{4.86}$$

This is already interesting — even though the field changes in all of space, we need only look at the before and after fields in the volume $\mathcal{V}_2$ rather than the entire system. If $\epsilon_1 = \epsilon_o$ (vacuum outside $\mathcal{V}_2$ and in $\mathcal{V}_2$ before the introduction of ϵ_2), then we can use $\vec{P} = (\epsilon_2 - \epsilon_o) \vec{E}_2$ to rewrite as

$$W = U_2 - U_1 = -\frac{1}{2} \int_{\mathcal{V}_2} d\tau \, \vec{P} \cdot \vec{E}_1 \qquad \Longleftrightarrow \qquad w = -\frac{1}{2} \vec{P} \cdot \vec{E}_1 \tag{4.87}$$

where we recall that $\vec{E}_1$ *is the electric field in the absence of the dielectric* and $\vec{P}$ is the polarization density of the dielectric, and w refers to an energy density. This is just like the energy of a dipole in an external electric field, except that the factor of $1/2$ accounts for the integration from zero field to actual field, from the fact that the dielectric polarizes in response to the applied field. *We see that the introduction of the dielectric into an existing electric field in vacuum, holding the source charges fixed, reduces the overall electrostatic energy.*

Why is the integrand not $\vec{P} \cdot \vec{E}_2$ or $\vec{D}_2 \cdot \vec{E}_2$? Because we are asking for the *difference* in energy between the field configurations without and with the dielectric. There was field in $\mathcal{V}_2$ before the dielectric was placed there, so we have to subtract off that original field energy density, and we also need to consider the field energy density difference between the two configurations *outside* the dielectric. It turns out that the above integrand correctly accounts for the differencing relative to the no-dielectric starting condition. We can see this by trying to evaluate the potential alternate expressions:

$$-\frac{1}{2} \int_{\mathcal{V}_2} d\tau \, \vec{P} \cdot \vec{E}_2 = -\frac{1}{2} \int_{\mathcal{V}_2} d\tau \, \left(\vec{D}_2 - \epsilon_o \vec{E}_2 \right) \cdot \vec{E}_2 = \frac{1}{2} \int_{\mathcal{V}_2} d\tau \, \left[\epsilon_o \left| \vec{E}_2 \right|^2 - \vec{D}_2 \cdot \vec{E}_2 \right]$$

$$(4.88)$$

This is some sort of difference between the total electrostatic potential energy in $\mathcal{V}_2$ and the electrostatic potential energy neglecting that associated with the assembly of the dielectric medium. The expression has two problems: there is no differencing with the initial configuration, and it neglects the energy stored in $\mathcal{V}_1$. It is part of the energy difference we are interested in, but not all of. The use of $\vec{D}_2 \cdot \vec{E}_2$ would suffer the same problems.

Lecture 14:

Electrostatics in Matter III:
Electrostatic Forces and Torques for Linear Dielectrics

Magnetostatics I:
Lorentz Force, Biot-Savart Law, Fields and Forces

Date Revised: 2024/02/20 06:30
Date Given: 2024/02/22

Force and Torque on a Linear, Homogeneous Dielectric in an External Field with Free Charge Fixed

Let us first consider the force on the dielectric in the case that the free charge is held fixed. There are no batteries involved, so we need only consider the electrostatic energy of the field. We take the negative of its gradient with respect to some generalized displacement ξ to find the generalized force F_ξ:

$$F_\xi\big|_Q = -\left(\frac{\partial W}{\partial \xi}\right)_Q = -\left(\frac{\partial W}{\partial C}\right)_Q \frac{\partial C}{\partial \xi} \tag{4.89}$$

where we made the second step because, if Q is held fixed, the variation of the system energy is given entirely by the variation of the capacitance. ξ can be a spatial displacement coordinate like x, y, or z, or it can be an angular orientation coordinate, in which case the generalized force is actually a torque.

Any system of conductors can be reduced to a capacitance matrix, so the above can also be written using Equation 2.82 (recall, $\underline{\underline{D}} = \underline{\underline{C}}^{-1}$)

$$F_\xi\big|_Q = -\frac{\partial}{\partial \xi} \frac{1}{2} \sum_{i,j=1}^{N} Q_i\, Q_j\, D_{ij}\Bigg|_Q = -\frac{1}{2} \sum_{i,j=1}^{N} Q_i\, Q_j\, \frac{\partial D_{ij}}{\partial \xi} = -\frac{1}{2}\underline{Q}^T \left[\frac{\partial}{\partial \xi}\underline{\underline{C}}^{-1}\right]\underline{Q} \tag{4.90}$$

(We have intentionally avoided using the confusing notation C_{ij}^{-1}, using D_{ij} instead.)

Example 4.7: Force on a Dielectric Slab in a Parallel Plate Capacitor, Free Charge Fixed

Let's consider a parallel-plate capacitor with plate separation d, plate side dimensions ℓ and w, and with a slab of linear, homogeneous dielectric partially inserted between the plates, with vacuum from 0 to x and dielectric from x to ℓ with $0 < x < \ell$.

Let's do this by calculating the total energy of the slab in the capacitor, with E dependent on the position of the slab. The energy is (using the calculation of C from the earlier example)

$$W = \frac{1}{2} \frac{Q^2}{C} \qquad \text{with} \qquad C = \frac{\epsilon_o\, w\, x + \epsilon\, w\, (\ell - x)}{d} \tag{4.91}$$

Therefore,

$$F_x|_Q = -\left(-\frac{1}{2}\frac{Q^2}{C^2}\right)\frac{dC}{dx} = \frac{1}{2}\frac{Q^2}{C^2}\frac{(\epsilon_o - \epsilon)\,w}{d} = -\frac{1}{2}\,V^2\,(\epsilon - \epsilon_o)\,\frac{w}{d} \tag{4.92}$$

which matches Griffiths Equation 4.65 (recall, $\epsilon_o\chi_e = \epsilon - \epsilon_o$).

Intuitively, the dielectric is pulled in because it lowers the energy of the configuration: the field energy density is proportional to $\epsilon |\vec{E}|^2$, and $|\vec{E}| \propto \epsilon^{-1}$, so the field energy density is $\propto \epsilon^{-1}$: larger ϵ implies lower energy.

Microscopically, what is happening is that the fringing field of the capacitor polarizes the dielectric, leading to bound charge on the surface. The bound charge on the surface is attracted to the free charge on the capacitor plates, causing the dielectric to be pulled in. It's a runaway effect, with the movement of the dielectric into the capacitor leading to greater polarization of the fringing field region, increasing the bound surface charge density and thereby leading to a greater attractive force. The system only reaches equilibrium when the dielectric is maximally contained in the capacitor. (It would be interesting to calculate the trajectory, in particular the harmonic oscillations that would occur around the equilibrium position because the slab will have been accelerated and thus have some kinetic energy when it gets to the equilibrium position.)

Force and Torque on a Linear, Homogeneous Dielectric in an External Field with Voltages Fixed

In general, we do not encounter the above situation. Rather, we hold the voltages constant on a set of electrodes while we consider the work done during a virtual displacement $d\xi$.

Before we get into it, though, let's ask ourselves what we expect to have happen. Should the force change depending on whether we hold the voltage or the charge fixed? No, because the force is due to the arrangement of charges on the conductors and the dielectric at the current instant in time, not at some point in the future that is affected by whether the charges or voltages are kept constant.

Let's model the fixed voltage situation in two steps, first disconnecting the batteries and holding the charge fixed while we move the dielectric as we did above, then reconnecting the batteries so that charge flows on to or off of the electrodes and restores them to their original potentials.

Since we are now focusing on a situation with voltages on electrodes, it makes sense to think about a set of electrodes $i = 1...N$ with voltages V_i and charges Q_i. The electrodes have a capacitance matrix $\underline{\underline{C}}$. Let's first consider the change in electrostatic energy for the first step with the charges held fixed (again, using $\underline{\underline{D}} = \underline{\underline{C}}^{-1}$):

$$dW_{field}\big|_Q = d\left[\frac{1}{2}\sum_{i,j=1}^{N} Q_i Q_j D_{ij}\right]_Q = \frac{1}{2}\sum_{i,j=1}^{N} Q_i Q_j \, dD_{ij} \qquad (4.93)$$

The change in the inverse capacitance matrix results in a change in the voltages on the electrodes given by

$$dV_i\big|_Q = \sum_{j=1}^{N} dD_{ij} Q_j \qquad (4.94)$$

Now, let's return the voltages to their original values by allowing charge to flow on/off the electrodes from batteries while holding the dielectrics fixed (*i.e.*, D_{ij} held constant). The charge transfer required to undo the above voltage changes is

$$dQ_k|_V = \sum_{i=1}^{N} C_{ki} \left(-dV_i\right)_Q = -\sum_{i,j=1}^{N} C_{ki} \, Q_j \, dD_{ij} \qquad (4.95)$$

The change in the electrostatic energy of the configuration (energy flowing out of the battery into the field) due to this flow of charge is

$$dW_{field}^{bat}\bigg|_V = \sum_{k=1}^{N} V_k \, dQ_k|_V = -\sum_{i,j,k=1}^{N} V_k \, C_{ki} \, Q_j \, dD_{ij} = -\sum_{i,j=1}^{N} Q_i \, Q_j \, dD_{ij}$$

$$= -2 \, dW_{field}|_Q \qquad (4.96)$$

where we used $C_{ki} = C_{ik}$ and $\sum_{k=1}^{N} V_k C_{ik} = Q_i$.

Therefore, the total infinitesimal change in energy is

$$dW_{field}|_V = dW_{field}|_Q + dW_{field}^{bat}\Big|_V = dW_{field}|_Q - 2\,dW_{field}|_Q = -\,dW_{field}|_Q \quad (4.97)$$

As we explained earlier, the force cannot depend on whether the charge is held fixed or the voltage is held fixed. To ensure we get the same force in the two cases, we therefore must conclude

$$F_\xi|_V = \left(\frac{\partial W_{field}}{\partial \xi}\right)_V = -\left(\frac{\partial W_{field}}{\partial \xi}\right)_Q = F_\xi|_Q \quad (4.98)$$

That is, when the battery is involved, we must consider the energy of the entire system and take the positive gradient of the field energy rather than considering only the energy of the field and taking the negative gradient of that energy. The reason these two gradients are different, with a sign between them, is because the derivative is calculationally different depending on whether V or Q is held fixed.

We can see this works mathematically by trying it:

$$\left(\frac{\partial W_{field}}{\partial \xi}\right)_V = \frac{\partial}{\partial \xi}\left[\frac{1}{2}\sum_{i,j=1}^{N} V_i\, V_j\, C_{ij}\right]_V = \frac{1}{2}\sum_{i,j=1}^{N} V_i\, V_j\, \frac{\partial C_{ij}}{\partial \xi}$$

$$= \frac{1}{2}\underline{V}^T\left[\frac{\partial}{\partial \xi}\underline{\underline{C}}\right]\underline{V} \tag{4.99}$$

Since $\partial \underline{\underline{C}}^{-1}/\partial \xi = -\underline{\underline{C}}^{-1}[\partial \underline{\underline{C}}/\partial \xi]\underline{\underline{C}}^{-1}$ (one can see this by evaluating $\partial[\underline{\underline{C}}\,\underline{\underline{C}}^{-1}]/\partial \xi = \partial \underline{\underline{1}}/\partial \xi = \underline{0}$), this form yields Equation 4.90 for $F_\xi|_Q$. Thus,

$$F_\xi|_V = \left(\frac{\partial W_{field}}{\partial \xi}\right)_V = -\left(\frac{\partial W_{field}}{\partial \xi}\right)_Q = F_\xi|_Q \tag{4.100}$$

One can check this result using the parallel plate capacitor example by starting with $W = C\,V^2/2$ instead of $W = Q^2/2\,C$. Taking the positive derivative at fixed V gives the same result as taking the negative derivative at fixed Q because C is in the numerator in the first case while C is in the denominator in the second.

Section 5
Magnetostatics

Study Guidelines

As with basic electrostatics, you have seen much of the material in this section before in Ph1c. As with electrostatics, we will use more rigor here. We will also consider some more advanced topics such as the multipole expansion of the magnetic vector potential, off-axis fields for azimuthally symmetric configurations, etc. As with basic electrostatics, we won't do any examples in lecture or the notes where they would duplicate Ph1c. But you should be review the examples in Griffiths Chapter 5 and make sure you are comfortable with them.

Lorentz Forces and Current Densities

Force on a Moving Point Charge in a Magnetic Field

The magnetic force on a point charge q moving with velocity $\vec{v}$ in a magnetic field $\vec{B}$ is given by the *Lorentz Force Law*:

$$\boxed{\vec{F}_{mag} = q\left(\vec{v} \times \vec{B}\right)} \tag{5.1}$$

If an electric field is present, the total electrostatic and magnetostatic force on q is

$$\vec{F} = q\left(\vec{E} + \vec{v} \times \vec{B}\right) \tag{5.2}$$

Note that the electrostatic force on q is not modified by the fact that it is moving.

See the nice examples in Griffiths of cyclotron and cycloid motion (Examples 5.1 and 5.2). These are at the level of Ph1c, so we do not spend time in lecture on them.

Magnetic Forces Do No Work

Because $\vec{F}_{mag} \propto \vec{v} \times \vec{B}$, it holds that $\vec{F}_{mag} \perp \vec{v}$. Since the differential of work done by a force is $dW = \vec{F} \cdot d\vec{\ell} = \vec{F} \cdot \vec{v}\, dt$, we thus see that $dW = 0$ for magnetic forces. This may seem counterintuitive. In cases where it appears work is being done, there is usually a battery involved that is doing the work, while the magnetic force is redirecting the force doing the work (in the same way that a constraint force in mechanics does no work).

The one exception to this is the case of intrinsic magnetic moments of fundamental particles, which emerge from quantum field theory. In such cases, the magnetic moment is not identified with a current loop, it is just an intrinsic property of the particle. Since our proof above requires the Lorentz Force Law, and such moments are not assocated with a current that experiences the Lorentz Force, the proof does not apply. In cases concerning such moments, work can be done by the field of the moment or on the magnetic moment by an external magnetic field because no battery is required to maintain the magnetic moment.

Line Currents

A current carried by a wire can be modeled as a constant line charge density λ that is moving at fixed speed v:

$$\mathrm{I} = \lambda\, v \qquad (5.3)$$

For the sake of the generalizations we will consider below, let us write this as a position-dependent vector

$$\vec{\mathrm{I}}(\vec{r}) = \lambda(\vec{r})\, \vec{v}(\vec{r}) \qquad (5.4)$$

where $\vec{v}(\vec{r})$ is a function of position and its direction follows the wire. By conservation of charge, the only position dependence of $\vec{\mathrm{I}}(\vec{r})$ can be its direction. This implies that any position dependence in $\lambda(\vec{r})$ must be canceled by the position dependence of the magnitude of $\vec{v}(\vec{r})$. If λ is position-independent, then only the direction of $\vec{v}$ may change with position.

For magnetostatics, we assume that such a line current, and the surface and volume current densities that follow below, are time-independent, or **steady**: they were set up an infinitely long time ago and have been flowing at their current values since then. We also ignore the discretization of the charge density (in this case λ) and consider it to be a continuous quantity. This is called the **steady-state** assumption or approximation.

Force on a Line Current

It is straightforward to calculate the force on a line current by integrating the Lorentz Force Law over the wire:

$$\vec{F}_{mag} = \int dq \left[\vec{v}(\vec{r}) \times \vec{B}(\vec{r}) \right] = \int_{\mathcal{C}} d\ell \, \lambda \left[\vec{v}(\vec{r}) \times \vec{B}(\vec{r}) \right] \tag{5.5}$$

$$\boxed{\vec{F}_{mag} = \int_{\mathcal{C}} d\ell \left[\vec{I}(\vec{r}) \times \vec{B}(\vec{r}) \right]} \tag{5.6}$$

where we have used the fact that $d\vec{\ell}$, $\vec{v}$, and $\vec{I}$ are all in the same direction at any point on the wire because the current flows in the wire. Now, realizing that I is independent of position along the wire (due to conservation of charge as noted above), we can pull it out in front of the integral, yielding

$$\vec{F}_{mag} = I \int_{\mathcal{C}} \left[d\vec{\ell} \times \vec{B}(\vec{r}) \right] \tag{5.7}$$

Griffiths Example 5.3 is a nice example of calculating the force on a current loop and also illustrates the point of the battery supplying the energy to do the work that appears to be done by the magnetic field. The magnetic field acts like a constraint force to redirect that work.

Current Densities

Just as we generalized point charges to line, surface, and volume charge densities, we can generalize single moving point charges to line, surface, and volume current densities. We have already made the first generalization, which is straightforward to understand since one intuitively thinks of a current as an ensemble of point charges moving through a wire.

A *surface current density* is a current flowing in a sheet; think of water flowing over the surface of an object. The surface current density $\vec{K}$ is defined by

$$d\vec{\mathrm{I}}(\vec{r}) = \vec{K}(\vec{r})\, d\ell_\perp = \left| \widehat{K}(\vec{r}) \times d\vec{\ell} \right| \vec{K}(\vec{r}) \tag{5.8}$$

where $d\ell_\perp$ is an infinitesimal length perpendicular to $\vec{K}$ and $d\vec{\ell}$ is an arbitrary infinitesimal length. The cross-product takes the projection of $d\vec{\ell}$ perpendicular to $\vec{K}$.

If one thinks about the surface current density as a moving distribution of a surface charge density, then

$$\vec{K}(\vec{r}) = \sigma(\vec{r})\, \vec{v}(\vec{r}) \tag{5.9}$$

where $\sigma(\vec{r})$ is the surface charge density at $\vec{r}$ and $\vec{v}(\vec{r})$ is the velocity of the surface charge density at $\vec{r}$.

A *volume current density* is a current flowing in a bulk volume; think of water flowing in a pipe or in a river. The volume current density $\vec{J}$ is defined by

$$d\vec{I}(\vec{r}) = \vec{J}(\vec{r})\, da_\perp = \left|\widehat{J}(\vec{r}) \cdot \widehat{n}\right| da\, \vec{J}(\vec{r}) \tag{5.10}$$

where $\widehat{n}$ is the normal to the area element *da*. (If we had defined a normal $\widehat{n}$ to the line element $d\vec{\ell}$ in the plane of the sheet, we could have used a dot product instead of a cross product in the definition of the surface current density. But it is conventional to do it as we have done it.)

If one thinks about the volume current density as a moving distribution of a volume charge density, then

$$\vec{J}(\vec{r}) = \rho(\vec{r})\, \vec{v}(\vec{r}) \tag{5.11}$$

where $\rho(\vec{r})$ is the volume charge density at $\vec{r}$ and $\vec{v}(\vec{r})$ is the velocity of the volume charge density at $\vec{r}$.

Forces on Current Densities

We can integrate the force over the current densities just as we did for the line current:

$$\vec{F}_{mag} = \int dq \left[\vec{v}(\vec{r}) \times \vec{B}(\vec{r}) \right] = \int_{\mathcal{S}} da \, \sigma(\vec{r}) \left[\vec{v}(\vec{r}) \times \vec{B}(\vec{r}) \right] \tag{5.12}$$

$$\boxed{\vec{F}_{mag} = \int_{\mathcal{S}} da \left[\vec{K}(\vec{r}) \times \vec{B}(\vec{r}) \right]} \tag{5.13}$$

$$\vec{F}_{mag} = \int dq \left[\vec{v}(\vec{r}) \times \vec{B}(\vec{r}) \right] = \int_{\mathcal{V}} d\tau \, \rho(\vec{r}) \left[\vec{v}(\vec{r}) \times \vec{B}(\vec{r}) \right] \tag{5.14}$$

$$\boxed{\vec{F}_{mag} = \int_{\mathcal{V}} d\tau \left[\vec{J}(\vec{r}) \times \vec{B}(\vec{r}) \right]} \tag{5.15}$$

It should be clear that we could have considered Equation 5.15 to be the fundamental statement of the Lorentz Force Law and derived the lower-dimensional versions by inclusion of appropriate delta functions in the definition of ρ or $\vec{J}$. Such a reduction would be cumbersome because the sheet or line carrying the current may not be easy to parameterize, but the reduction is conceptually straightforward.

Conservation of Charge and the Continuity Equation

We defined the current densities above in terms of the infinitesimal current passing through an infinitesimal line element (for a surface current density) or through an infinitesimal area element (for a volume current density). Let's integrate the latter over a surface to obtain the total current passing through that surface:

$$I_{\mathcal{S}} = \int_{\mathcal{S}} da\, \hat{n}(\vec{r}) \cdot \vec{J}(\vec{r}) \tag{5.16}$$

If we take $\mathcal{S}$ to be a closed surface, we may apply the divergence theorem to the above:

$$\oint_{\mathcal{S}} da\, \hat{n}(\vec{r}) \cdot \vec{J}(\vec{r}) = \int_{\mathcal{V}(\mathcal{S})} d\tau\, \vec{\nabla} \cdot \vec{J}(\vec{r}) \tag{5.17}$$

where $\mathcal{V}(\mathcal{S})$ is the volume enclosed by $\mathcal{S}$. By conservation of charge, the current is just the time derivative of the charge enclosed by $\mathcal{S}$, with the sign such that if a positive current is exiting $\mathcal{S}$, then the charge enclosed must be decreasing, assuming that the surface itself is time-independent. With this, we have

$$I_{\mathcal{S}} = -\frac{d}{dt}\, Q_{\mathcal{V}(\mathcal{S})} = -\frac{d}{dt} \int_{\mathcal{V}(\mathcal{S})} d\tau\, \rho(\vec{r}) = -\int_{\mathcal{V}(\mathcal{S})} d\tau\, \frac{\partial \rho(\vec{r})}{\partial t} \tag{5.18}$$

Thus, we have

$$\int_{\mathcal{V}(\mathcal{S})} d\tau \, \vec{\nabla} \cdot \vec{J}(\vec{r}) = -\int_{\mathcal{V}(\mathcal{S})} d\tau \, \frac{\partial \rho(\vec{r})}{\partial t} \tag{5.19}$$

Since the surface $\mathcal{S}$ is arbitrary, it must hold that the integrands are equal everywhere:

$$\boxed{\vec{\nabla} \cdot \vec{J}(\vec{r}) = -\frac{\partial \rho(\vec{r})}{\partial t}} \tag{5.20}$$

This is the *continuity equation* and is effectively the differential version of conservation of charge.

With this equation, we can define our *steady-state assumption* more mathematically: it corresponds to $\partial \rho / \partial t = 0$, which then implies $\vec{\nabla} \cdot \vec{J} = 0$. The interpretation is that the charge density at any point cannot change with time, which implies that the net current flow into or out of any point vanishes.

Fields of and Magnetic Forces between Currents

Biot-Savart Law

For a *steady-state current distribution* — one in which the current densities are time-independent — it is an empirical observation, validated by the Lorentz force that moving charges or currents experience, that the magnetic field at $\vec{r}$ due to the current distribution is given by

$$\vec{B}(\vec{r}) = \frac{\mu_o}{4\pi} \int d\ell' \, \frac{\vec{I}(\vec{r}') \times (\vec{r} - \vec{r}')}{|\vec{r} - \vec{r}'|^3} = \frac{\mu_o}{4\pi} \, I \int \frac{d\vec{\ell}'(\vec{r}') \times (\vec{r} - \vec{r}')}{|\vec{r} - \vec{r}'|^3} \qquad (5.21)$$

$\mu_o = 4\pi \times 10^{-7} \text{ N A}^{-2}$ is the *permeability of free space*. The magnetic field carries units of *teslas*, $T = N/(A \cdot m)$. The Biot-Savart Law is the analogue in magnetostatics of Coulomb's Law in electrostatics, and it has the same $1/r^2$ dependence.

You are well aware of the result that the field of a straight wire along the *z*-axis carrying current I at a transverse distance *s* from the wire is

$$\vec{B}(\vec{r}) = \frac{\mu_o}{2\pi} \frac{I}{s} \, \widehat{\phi} \qquad (5.22)$$

where $\widehat{\phi}$ is the azimuthal unit vector in cylindrical coordinates. The field forms circles around the wire with orientation set by the right-hand rule. This is derived in Griffiths Example 5.5, which we will not repeat here since you saw it in Ph1c.

Force between Two Current-Carrying Wires

We can combine the Lorentz Force Law and the Biot-Savart Law to calculate the force between two current-carrying wires; this force is the empirical basis for magnetostatics, as it is much easier to measure the force between two wires than it is to create ideal test charges and measure their motion in the magnetic field of a wire. We just plug the Biot-Savart Law into the Lorentz Force Law for a line current distribution, Equation 5.6, to find the force on the first wire due to the field of the second wire:

$$\vec{F}_{mag} = I_1 \int_{\mathcal{C}_1} d\vec{\ell} \times \vec{B}(\vec{r}) \tag{5.23}$$

$$= \frac{\mu_o}{4\pi} I_1 I_2 \int_{\mathcal{C}_1} \int_{\mathcal{C}_2} \frac{d\vec{\ell}(\vec{r}) \times \left[d\vec{\ell}'(\vec{r}') \times (\vec{r} - \vec{r}') \right]}{|\vec{r} - \vec{r}'|^3} \tag{5.24}$$

Consider the special case of both wires running parallel to the z axis separated by $s\,\hat{s}$ in the xy-plane, with the first wire on the z-axis itself. Then $d\vec{\ell} = \hat{z}\,dz$, $d\vec{\ell}' = \hat{z}\,dz'$, $\vec{r} = z\,\hat{z}$, $\vec{r}' = s\,\hat{s} + z'\,\hat{z}$.

Therefore,

$$d\vec{\ell}(\vec{r}) \times \left[d\vec{\ell}'(\vec{r}') \times (\vec{r} - \vec{r}') \right] = dz\, dz'\, \hat{z} \times \left[\hat{z} \times \left((z - z')\hat{z} - s\,\hat{s} \right) \right] \tag{5.25}$$

$$= dz\, dz'\, s\,\hat{s} \tag{5.26}$$

$$\text{and} \qquad |\vec{r} - \vec{r}'|^3 = \left[(z - z')^2 + s^2 \right]^{3/2} \tag{5.27}$$

Thus,

$$\vec{F}_{mag} = \frac{\mu_o}{4\pi}\, I_1 I_2\, s\,\hat{s} \int_{-\infty}^{\infty} dz \int_{-\infty}^{\infty} dz' \frac{1}{\left[(z - z')^2 + s^2 \right]^{3/2}} \tag{5.28}$$

$$= \frac{\mu_o}{4\pi}\, I_1 I_2\, s\,\hat{s} \int_{-\infty}^{\infty} dz \left[\frac{z' - z}{s^2 \left[(z - z')^2 + s^2 \right]^{1/2}} \right] \Bigg|_{-\infty}^{\infty} \tag{5.29}$$

$$= \frac{\mu_o}{4\pi}\, I_1 I_2\, s\,\hat{s} \int_{-\infty}^{\infty} dz \frac{2}{s^2} = \frac{\mu_o}{2\pi} \frac{I_1 I_2}{s}\, \hat{s} \int_{-\infty}^{\infty} dz \tag{5.30}$$

where we did the integral using the trigonometric substitution $z' - z = s\tan\theta$. The total force is infinite, but we can abstract out of the above expression the force per unit length on the first wire, which is attractive (pointing towards the second wire) if the currents flow in the same direction:

$$\vec{f}_{mag} = \frac{\mu_o}{2\pi} \frac{I_1 I_2}{s}\, \hat{s} \tag{5.31}$$

General Expressions for Fields due to Current Densities

The obvious generalizations of the Biot-Savart Law are

$$
\vec{B}(\vec{r}) = \frac{\mu_o}{4\pi} \int da' \, \frac{\vec{K}(\vec{r}\,') \times (\vec{r} - \vec{r}\,')}{|\vec{r} - \vec{r}\,'|^3} \qquad \vec{B}(\vec{r}) = \frac{\mu_o}{4\pi} \int d\tau' \, \frac{\vec{J}(\vec{r}\,') \times (\vec{r} - \vec{r}\,')}{|\vec{r} - \vec{r}\,'|^3}
$$

(5.32)

Griffiths notes that a line current distribution is the lowest-dimensional current distribution one can have because the zero-dimensional version — a point charge moving with velocity $\vec{v}$ — does not constitute a steady-state current: the charge passing a given point in space is time-dependent.

As with the Lorentz Force Law, it should also be clear that one could consider the volume version to be the fundamental statement of the Biot-Savart Law and one can derive the lower-dimensional versions by including delta functions in the definition of $\vec{J}$. This does not apply to a reduction to zero dimensionality, as noted above.

There are good examples of the use of the Biot-Savart Law in Griffiths. Again, these are at the level of Ph1c, so we do not spend time in lecture on them.

Another Form for the Biot-Savart Law

We begin by using Equation 2.53 to rewrite the Biot-Savart Law expression for the magnetic field:

$$\vec{B}(\vec{r}) = \frac{\mu_o}{4\pi} \int_{\mathcal{V}} d\tau' \, \frac{\vec{J}(\vec{r}') \times (\vec{r} - \vec{r}')}{|\vec{r} - \vec{r}'|^3} = -\frac{\mu_o}{4\pi} \int_{\mathcal{V}} d\tau' \, \vec{J}(\vec{r}') \times \vec{\nabla}_{\vec{r}} \left(\frac{1}{|\vec{r} - \vec{r}'|} \right) \quad \text{(5.33)}$$

We use one of the product rules for the curl, $\vec{\nabla} \times (f\vec{a}) = f(\vec{\nabla} \times \vec{a}) - \vec{a} \times (\vec{\nabla}f)$, and notice that $\vec{\nabla}_{\vec{r}} \times \vec{J}(\vec{r}') = 0$ because $\vec{J}(\vec{r}')$ is a function of $\vec{r}'$ while $\vec{\nabla}_{\vec{r}}$ is with respect to $\vec{r}$, to obtain

$$\boxed{\vec{B}(\vec{r}) = \vec{\nabla} \times \frac{\mu_o}{4\pi} \int_{\mathcal{V}} d\tau' \, \frac{\vec{J}(\vec{r}')}{|\vec{r} - \vec{r}'|}} \quad \text{(5.34)}$$

where we have brought $\vec{\nabla}_{\vec{r}}$ outside the integral over $\vec{r}'$ because it acts with respect to $\vec{r}$. We also dropped the $\vec{r}$ subscript since now, being outside the integral, it must act only on $\vec{r}$. This form is obviously suggestive of the idea of $\vec{B}$ being derived from a vector potential, which we will return to shortly.

We note that, while our derivation of this equation did not appear to require any assumptions about the way the current behaves at infinity, we will see later that the steady-state assumption does imply the net current through any sphere must vanish.

Lecture 15:

Magnetostatics II:
Ampere's Law
Divergence of B
Vector Potential
Uniqueness Theorem
Magnetostatic Scalar Potential

Date Revised: 2024/02/25 08:15
Date Given: 2024/02/25

Curl and Divergence of the Magnetic Field; Ampere's Law

Curl of the Magnetic Field

From the field of a current-carring wire, Equation 5.22, we get the clear impression that $\vec{B}$ has curl and that the curl is related to the current sourcing the field. Here, we explicitly calculate this curl from the Biot-Savart Law. Griffiths Section 5.3.2 provides one technique for this; we use Jackson's technique instead to avoid duplication.

We take the curl of Equation 5.34 and apply the $BAC - CAB$ rule for the triple vector product, $\vec{\nabla} \times (\vec{\nabla} \times \vec{a}) = \vec{\nabla}(\vec{\nabla} \cdot \vec{a}) - \nabla^2 \vec{a}$, writing the coordinate that $\vec{\nabla}$ acts on explicitly:

$$\vec{\nabla}_{\vec{r}} \times \vec{B}(\vec{r}) = \vec{\nabla}_{\vec{r}} \times \left[\vec{\nabla}_{\vec{r}} \times \frac{\mu_o}{4\pi} \int_{\mathcal{V}} d\tau' \frac{\vec{J}(\vec{r}')}{|\vec{r} - \vec{r}'|} \right] \tag{5.35}$$

$$= \frac{\mu_o}{4\pi} \left[\vec{\nabla}_{\vec{r}} \left(\vec{\nabla}_{\vec{r}} \cdot \int_{\mathcal{V}} d\tau' \left(\frac{\vec{J}(\vec{r}')}{|\vec{r} - \vec{r}'|} \right) \right) - \nabla_{\vec{r}}^2 \int_{\mathcal{V}} d\tau' \left(\frac{\vec{J}(\vec{r}')}{|\vec{r} - \vec{r}'|} \right) \right] \tag{5.36}$$

$$= \frac{\mu_o}{4\pi} \left[\vec{\nabla}_{\vec{r}} \int_{\mathcal{V}} d\tau' \, \vec{\nabla}_{\vec{r}} \cdot \left(\frac{\vec{J}(\vec{r}')}{|\vec{r} - \vec{r}'|} \right) - \int_{\mathcal{V}} d\tau' \, \nabla_{\vec{r}}^2 \left(\frac{\vec{J}(\vec{r}')}{|\vec{r} - \vec{r}'|} \right) \right] \tag{5.37}$$

We were able to bring $\vec{\nabla}_{\vec{r}}$ and $\nabla_{\vec{r}}^2$ inside the integrals because $\vec{\nabla}_{\vec{r}}$ is with respect to $\vec{r}$ and the integral is over $\vec{r}'$. Similarly, because $\vec{\nabla}_{\vec{r}}$ is with respect to $\vec{r}$ and $\vec{J}$ is a function of $\vec{r}'$, $\vec{J}$ passes through the divergence in the first term and the Laplacian in the second one, preserving the necessary dot product in the first term and the vectorial nature of the second term:

$$\vec{\nabla}_{\vec{r}} \times \vec{B}(\vec{r}) = \frac{\mu_o}{4\pi} \left[\vec{\nabla}_{\vec{r}} \int_{\mathcal{V}} d\tau' \, \vec{J}(\vec{r}') \cdot \vec{\nabla}_{\vec{r}} \left(\frac{1}{|\vec{r} - \vec{r}'|} \right) - \int_{\mathcal{V}} d\tau' \, \vec{J}(\vec{r}') \, \nabla_{\vec{r}}^2 \left(\frac{1}{|\vec{r} - \vec{r}'|} \right) \right] \tag{5.38}$$

We know from electrostatics that

$$\vec{\nabla}_{\vec{r}}\left(\frac{1}{|\vec{r}-\vec{r}'|}\right) = -\vec{\nabla}_{\vec{r}'}\left(\frac{1}{|\vec{r}-\vec{r}'|}\right)$$

$$\nabla_{\vec{r}}^2\left(\frac{1}{|\vec{r}-\vec{r}'|}\right) = -4\pi\,\delta(\vec{r}-\vec{r}')$$

The first equation may seem surprising if one considers the exchange $\vec{r}\leftrightarrow\vec{r}'$, but one can see it is true by simply evaluating the gradient on both sides or by defining $\vec{s}=\vec{r}-\vec{r}'$ and applying the offset and inversion techniques we used in electrostatics. The second is Equation 3.35 with the exchange $\vec{r}\leftrightarrow\vec{r}'$ (where here there is no sign flip because the Laplacian is quadratic in the derivatives and the delta function is symmetric in its argument). Applying them, we obtain

$$\vec{\nabla}\times\vec{B}(\vec{r}) = \frac{\mu_o}{4\pi}\left[-\vec{\nabla}_{\vec{r}}\int_{\mathcal{V}}d\tau'\,\vec{J}(\vec{r}')\cdot\vec{\nabla}_{\vec{r}'}\left(\frac{1}{|\vec{r}-\vec{r}'|}\right) + 4\pi\int_{\mathcal{V}}d\tau'\,\vec{J}(\vec{r}')\,\delta(\vec{r}-\vec{r}')\right] \tag{5.39}$$

The second term just becomes $4\pi\,\vec{J}(\vec{r})$, yielding

$$\vec{\nabla}\times\vec{B}(\vec{r}) = \frac{\mu_o}{4\pi}\left[-\vec{\nabla}_{\vec{r}}\int_{\mathcal{V}}d\tau'\,\vec{J}(\vec{r}')\cdot\vec{\nabla}_{\vec{r}'}\left(\frac{1}{|\vec{r}-\vec{r}'|}\right)\right] + \mu_o\,\vec{J}(\vec{r}') \tag{5.40}$$

We can apply the product rule $\vec{\nabla} \cdot (f\vec{a}) = \vec{a} \cdot \vec{\nabla} f + f \vec{\nabla} \cdot \vec{a}$ to rewrite the first term:

$$\int_{\mathcal{V}} d\tau' \, \vec{J}(\vec{r}') \cdot \vec{\nabla}_{\vec{r}'} \left(\frac{1}{|\vec{r} - \vec{r}'|} \right) = \int_{\mathcal{V}} d\tau' \, \vec{\nabla}_{\vec{r}'} \cdot \left(\frac{\vec{J}(\vec{r}')}{|\vec{r} - \vec{r}'|} \right) - \int_{\mathcal{V}} d\tau' \, \frac{\vec{\nabla}_{\vec{r}'} \cdot \vec{J}(\vec{r}')}{|\vec{r} - \vec{r}'|}$$

$$(5.41)$$

$$= \oint_{\mathcal{S}(\mathcal{V})} da' \, \widehat{n}(\vec{r}') \cdot \left(\frac{\vec{J}(\vec{r}')}{|\vec{r} - \vec{r}'|} \right) = 0 \qquad (5.42)$$

The second term vanishes because $\vec{\nabla}_{\vec{r}'} \cdot \vec{J}(\vec{r}') = 0$ under the steady-state assumption by the continuity equation with $\partial\rho/\partial t = 0$. We used the divergence theorem to transform the first term into a surface integral, and then we take the surface to infinity. Assuming the currents are localized, the integrand vanishes on that surface, causing the first term to vanish. Thus, we obtain, *under the steady-state assumption*,

$$\boxed{\vec{\nabla} \times \vec{B}(\vec{r}) = \mu_o \, \vec{J}(\vec{r})} \qquad (5.43)$$

This equation is the differential version of *Ampere's Law*, which we will return to shortly.

Let's discuss some subtleties in the above derivation connected to the vanishing of the $\vec{\nabla}(\vec{\nabla} \cdot \vec{a})$ term. There are two points to make:

▶ When we get to the definition of the vector potential $\vec{A}$, we will be able to interpret the vanishing of that term as implying $\vec{\nabla} \cdot \vec{A} = 0$ *for the form of the vector potential implied by Equation 5.34. $\vec{\nabla} \cdot \vec{A}$ will not vanish for any other form of the vector potential that yields the same field.* Just keep this point in mind, we'll provide more explanation later.

▶ We assumed that the currents are localized (confined to a finite volume) to make the surface term vanish. This is not the minimal condition required. We only need the integral to vanish. If we let the surface go off to infinity while keeping the point $\vec{r}$ at which we want to know the field at finite distance from the origin, then $1/|\vec{r} - \vec{r}'| \to 1/r'$. Thus, we can also make the integral vanish by simply requiring that the net flux of $\vec{J}$ through a surface of radius r' vanishes. Griffiths notes this subtlety in Footnote 14 in §5.3.2. It explains how Ampere's Law works for an infinitely long wire: for any sphere at large radius, as much current flows in as out of that sphere, so the integral vanishes.

Do we have to make this requirement? It may seem that we do not; we would just get a nonstandard Ampere's Law if we did not. But we do have to make it to be self-consistent with our steady-state assumption. If there were a net current through some sphere, then the charge contained in that volume would be changing with time, violating our steady-state assumption. This is the point we made in connection to Equation 5.34.

Divergence of the Magnetic Field

The vector identity $\vec{\nabla} \cdot (\vec{\nabla} \times \vec{a}) = 0$ combined with Equation 5.34 immediately implies

$$\boxed{\vec{\nabla} \cdot \vec{B}(\vec{r}) = 0} \tag{5.44}$$

The magnetic field has no divergence. This immediately implies *there are no magnetic point charges*: magnetic fields are sourced by currents only. It should be realized that this apparent fact is really an assumption inherent in the Biot-Savart Law. If we had added to the Biot-Savart Law a second term that looks like Coulomb's Law, due to magnetic monopoles, then the above divergence would have yielded that density of magnetic charge on the right side. It is *an empirical observation* that there are no magnetic monopoles, and hence we *assume* that magnetic fields are only sourced by currents via the Biot-Savart Law. That magnetic fields are sourced by currents at all is also *an empirical observation*; the Biot-Savart Law simplify codifies that observation.

General Thoughts on the Curl and Divergence of the Electric and Magnetic Field

Considering the corresponding expressions for electrostatics, we recognize that the electric field has divergence equal to the charge density because of the empirical observation of Coulomb's Law describing the electric field. It has a vanishing curl because of the empirical absence of a current that sources electric fields in the way that electric currents source magnetic fields; if there were a Biot-Savart-like term that added to Coulomb's Law, then the electric field would have curl. (In fact, when we consider time-varying fields, we will see that such a term comes into existence, proportional to $\partial \vec{B}/\partial t$.) We can in fact guess that, if magnetic monopoles existed, moving magnetic monopoles would generate an electric field in the same way that moving electric monopoles generate a magnetic field.

The key point in all of the above is that the nature of the divergence and the curl of the electric and magnetic fields reflect *empirical observations* about the way these fields are generated. These are not *derivable* results: they are inherent in the formulae we wrote down for the electric and magnetic fields, which themselves are based on *observations*.

We will see later that we can replace the assumption of Coulomb's Law and the Biot-Savart Law with assumptions about the potentials from which the electric and magnetic fields can be derived. But, again, we can only make those assumptions because they yields the correct empirical relations, Coulomb's Law and the Biot-Savart Law.

Integral form of Ampere's Law

We obtained the differential version of Ampere's Law above by taking the curl of the Biot-Savart Law for the magnetic field. We may obtain the integral form of Ampere's Law from it. We begin by integrating over an open surface $\mathcal{S}$ with normal $\hat{n}(\vec{r})$:

$$\int_{\mathcal{S}} da\,\hat{n}(\vec{r}) \cdot \left[\vec{\nabla} \times \vec{B}(\vec{r}) \right] = \mu_o \int_{\mathcal{S}} da\,\hat{n} \cdot \vec{J}(\vec{r}) \tag{5.45}$$

The left side can be transformed using Stokes' Theorem into a line integral around the edge of $\mathcal{S}$, which we denote by the closed contour $\mathcal{C}(\mathcal{S})$, while the right side is just total current passing through $\mathcal{C}(\mathcal{S})$, I_{encl}:

$$\boxed{\oint_{\mathcal{C}(\mathcal{S})} d\vec{\ell} \cdot \vec{B}(\vec{r}) = \mu_o I_{encl}} \tag{5.46}$$

yielding the integral version of Ampere's Law.

As before, there are a number of examples in Griffiths that are at the level of Ph1c, so we do not spend time on them here.

Magnetic Vector Potential

Form for the Magnetic Vector Potential

We saw (Equations 5.43 and 5.44) that the magnetic field has no divergence and has curl. You know from vector calculus (Griffiths §1.6) that this implies the magnetic field can be written purely as the curl of a *vector potential*. Equation 5.34 gave us its form:

$$\vec{B}(\vec{r}) = \vec{\nabla} \times \vec{A}(\vec{r}) \qquad \vec{A}(\vec{r}) = \frac{\mu_o}{4\,\pi} \int_{\mathcal{V}} d\tau' \, \frac{\vec{J}(\vec{r}')}{|\vec{r} - \vec{r}'|} \tag{5.47}$$

But this form, implied by the Biot-Savart Law, is not the only form. We had freedom with the electrostatic potential to add an offset. Here, we can add any curl-less function to $\vec{A}$ without affecting $\vec{B}$. The form above corresponds to the additional condition

$$\vec{\nabla} \cdot \vec{A}(\vec{r}) = 0 \tag{5.48}$$

If one tries to test this requirement on the above form for $\vec{A}$, one will find oneself doing the same manipulations needed to derieve Ampere's Law, Equation 5.43. In repeating those manipulations, *which is possible for this form of $\vec{A}$ only*, one sees that $\vec{\nabla} \cdot \vec{A} = 0$ is the representation of the steady-state assumption and that the net current through a surface of any radius vanishes (and also how the latter implies the former). For a different choice of $\vec{A}$ (and thus of $\vec{\nabla} \cdot \vec{A}$), the mathematical manifestation of this physical requirement will be different. In fact, it must be, because $\vec{\nabla} \cdot \vec{A} = 0$ is unique to this form.

Explicit Proof that $\vec{\nabla} \cdot \vec{A} = 0$ Can Always Be Obtained

It is interesting to prove "mechanically" that the choice $\vec{\nabla} \cdot \vec{A}$ is possible even if one, for some reason, started out with a form that did not satisfy this condition. Suppose one has a vector potential $\vec{A}_0$ that is not divergenceless. We need to add to it a function that makes the result divergenceless. For reasons we will see below, let's add a function $\vec{\nabla}\lambda(\vec{r})$:

$$\vec{A} = \vec{A}_0 + \vec{\nabla}\lambda \qquad (5.49)$$

Then

$$\vec{\nabla} \cdot \vec{A} = \vec{\nabla} \cdot \vec{A}_0 + \nabla^2 \lambda \qquad (5.50)$$

If we require the left side to vanish, then we have a version of Poisson's Equation:

$$\nabla^2 \lambda = -\vec{\nabla} \cdot \vec{A}_0 \qquad (5.51)$$

One thus sees one of the motivations for the assumed form $\vec{\nabla}\lambda$.

Let's choose boundary conditions that place the boundary at infinity with the field falling off at infinity. For these boundary conditions, we know from Coulomb's Law that the solution to Poisson's Equation is

$$\lambda(\vec{r}) = \frac{1}{4\pi} \int_{\mathcal{V}} d\tau' \frac{\vec{\nabla} \cdot \vec{A}_0(\vec{r}')}{|\vec{r} - \vec{r}'|} \tag{5.52}$$

The vector calculus identity $\vec{\nabla} \times \vec{\nabla} \lambda = 0$ implies that $\vec{\nabla} \times \vec{A} = \vec{\nabla} \times \vec{A}_0$ and thus the magnetic field is the same for the two vector potentials (our second motivation for the choice to add $\vec{\nabla}\lambda$). We thus have an explicit formula for the term that has to be added to $\vec{A}_0$ so that the resulting form $\vec{A}$ is divergenceless while leaving the magnetic field unchanged.

The above explicit formula may not be valid if we assume different boundary conditions, but we know Poisson's Equation always has a solution, so we are guaranteed that the desired function $\lambda(\vec{r})$ exists.

Let us make a final point about how the above relates to the connection between $\vec{\nabla} \cdot \vec{A} = 0$ and the behavior of the currents at infinity. It is *not true* that starting with $\vec{\nabla} \cdot \vec{A}_0 \neq 0$ corresponds to a different physical assumption about the currents at infinity: changing $\vec{\nabla} \cdot \vec{A}$ has no effect on the fields and thus can have no effect on the currents. Our standard formula for $\vec{A}$ is only valid under the assumption $\vec{\nabla} \cdot \vec{A} = 0$, and so the relation between $\vec{\nabla} \cdot \vec{A}$ and the assumption about how the currents behave is only valid for that form. If one assumes a different form for $\vec{A}$, one that has $\vec{\nabla} \cdot \vec{A} \neq 0$, then taking its divergence will not necessarily result in the particular expressions that we encountered before in deriving the differential form of Ampere's Law, so the interpretation of $\vec{\nabla} \cdot \vec{A} = 0$ will be different, and the mathematical manifestation of the currents vanishing at infinity will also change. One benefit of the choice $\vec{\nabla} \cdot \vec{A} = 0$ is that this mathematical manifestation is simple.

Alternate Proof of the Form for the Magnetic Vector Potential

We can arrive at Equation 5.47 via a slightly different path, which makes uses of Ampere's Law and the same triple vector identity we used to prove Ampere's Law, $\vec{\nabla} \times (\vec{\nabla} \times \vec{a}) = \vec{\nabla}(\vec{\nabla} \cdot \vec{a}) - \nabla^2 \vec{a}$:

$$\text{Ampere's Law:} \qquad \vec{\nabla} \times (\vec{\nabla} \times \vec{A}) = \vec{\nabla} \times \vec{B} = \mu_o \vec{J} \qquad (5.53)$$

$$\text{use vector identity:} \qquad \vec{\nabla}(\vec{\nabla} \cdot \vec{A}) - \nabla^2 \vec{A} = \mu_o \vec{J} \qquad (5.54)$$

$$\text{set } \vec{\nabla} \cdot \vec{A} = 0: \qquad \nabla^2 \vec{A} = -\mu_o \vec{J} \qquad (5.55)$$

Note that the vector components of $\vec{A}$ and $\vec{J}$ line up. Thus, the last equation is a component-by-component Poisson's Equation. Again, under the assumption that the currents are localized and for appropriate boundary conditions (as we assumed in providing the alternate version of the Biot-Savart Law that we previously used to define $\vec{A}$), we know the solution:

$$\nabla^2 \vec{A}(\vec{r}) = -\mu_o \vec{J}(\vec{r}) \qquad \overset{\text{localized currents}}{\Longleftrightarrow} \qquad \vec{A}(\vec{r}) = \frac{\mu_o}{4\pi} \int_{\mathcal{V}} d\tau' \, \frac{\vec{J}(\vec{r}')}{|\vec{r} - \vec{r}'|} \qquad (5.56)$$

This is just Equation 5.47 again. Essentially, we can think of the three components of the current density as sourcing the three components of the vector potential in the same way that the electric charge density sources the electric potential.

The Vector Potential for Line and Surface Currents

We can consider the specific cases of line and surface current densities as volume current densities that include delta functions specifying the localization to a line or sheet. When one does the volume integral, the delta function reduces the three-dimensional integral over the volume to one- or two-dimensional integrals over a line or sheet, yielding:

$$\vec{A}(\vec{r}) = \frac{\mu_o}{4\,\pi} \int_{\mathcal{C}} d\ell\, \frac{\vec{\mathrm{I}}(\vec{r}\,')}{|\vec{r} - \vec{r}\,'|} \qquad \vec{A}(\vec{r}) = \frac{\mu_o}{4\,\pi} \int_{\mathcal{S}} da'\, \frac{\vec{K}(\vec{r}\,')}{|\vec{r} - \vec{r}\,'|} \tag{5.57}$$

Note that the units of the vector potential are unchanged: the change in the units of the current densities are canceled by the change in the units of the measure of integration.

Example 5.1: Spinning Sphere of Charge (Griffiths Example 5.11)

The calculation of the vector potential for a spinning spherical shell of charge is a straightforward application of the definition of the vector potential. The only complication is the vector arithmetic.

Center the sphere of radius R at the origin. Let the surface change density be σ and the angular rotation frequency be $\vec{\omega} = \omega\,\hat{z}$. Then the surface current density is

$$\vec{K}(\vec{r}) = \sigma\,\vec{v}(\theta) = \sigma\,\vec{\omega} \times \vec{r} = \hat{\phi}\,\sigma\,\omega\,R\,\sin\theta \tag{5.58}$$

We need to calculate

$$\vec{A}(\vec{r}) = \frac{\mu_o}{4\,\pi}\,R^2 \int_0^{2\,\pi} d\phi' \int_0^{\pi} d\theta'\,\sin\theta'\,\frac{\sigma\,\omega\,\sin\theta'\,\hat{\phi}}{|\vec{r} - \vec{r}\,'|} \tag{5.59}$$

$$= \frac{\mu_o}{4\,\pi}\,\sigma\,\omega\,R^3 \int_0^{2\,\pi} d\phi' \int_0^{\pi} d\theta'\,\sin\theta'\,\frac{\sin\theta'\,(-\hat{x}\,\sin\phi' + \hat{y}\,\cos\phi')}{|\vec{r} - \vec{r}\,'|} \tag{5.60}$$

(The R^2 out front is because an area integral, not just a solid angle integral, needs to be done.) This is done in Griffiths Example 5.11 via explicit integration. We will use a different technique benefiting from our knowledge of the spherical harmonics and the Addition Theorem.

We use Equation 3.169, the Spherical Harmonic Addition Theorem Corollary, which expands $|\vec{r} - \vec{r}'|^{-1}$ in terms of spherical harmonics, recognizing $|\vec{r}'| = R$ because the integral is over the sphere of radius R:

$$\frac{1}{|\vec{r} - \vec{r}'|} = 4\pi \sum_{\ell=0}^{\infty} \sum_{m=-\ell}^{\ell} \frac{1}{2\ell+1} \frac{r_<^\ell}{r_>^{\ell+1}} Y_{\ell m}^*(\theta', \phi') Y_{\ell m}(\theta, \phi) \tag{5.61}$$

Let's consider the $\hat{x}$ piece of the above angular integral; the other term will be similar in spirit. We will write the numerator in terms of spherical harmonics and use the expansion. We abbreviate $\int_0^{2\pi} d\phi' \int_0^\pi d\theta' \sin\theta' = \int d\Omega'$ and recall $Y_{\ell,-m} = (-1)^m Y_{\ell,m}^*$. Applying these facts yields

$$\int d\Omega' \frac{\sin\theta'(-\sin\phi')}{|\vec{r} - \vec{r}'|} = \tag{5.62}$$

$$\int d\Omega' \sqrt{\frac{8\pi}{3}} \frac{Y_{1,1}(\theta', \phi') + Y_{1,-1}(\theta', \phi')}{2i} \sum_{\ell=0}^{\infty} \sum_{m=-\ell}^{\ell} \frac{4\pi}{2\ell+1} \frac{r_<^\ell}{r_>^{\ell+1}} Y_{\ell m}^*(\theta', \phi') Y_{\ell m}(\theta, \phi)$$

The integral over Ω' gives $\delta_{\ell,1}\delta_{m,1}$ and $\delta_{\ell,1}\delta_{m,-1}$, eliminating the sum and yielding

$$\int d\Omega' \frac{\sin\theta'(-\sin\phi')}{|\vec{r}-\vec{r}'|} = \frac{4\pi}{2i}\sqrt{\frac{8\pi}{3}}\frac{1}{3}\frac{r_<^\ell}{r_>^{\ell+1}}\left[Y_{1,1}(\theta,\phi) + Y_{1,-1}(\theta,\phi)\right] \qquad (5.63)$$

$$= -\frac{4\pi}{3}\frac{r_<}{r_>^2}\sin\theta\sin\phi \qquad (5.64)$$

where the $1/3$ came from $1/(2\ell+1)$. We get back the same type of angular dependence, but with the $1/|\vec{r}-\vec{r}'|$ turned into the prefactor shown, which has the correct dimensions.

We can repeat the same kind of manipulation for the $\widehat{y}$ term, yielding

$$\int d\Omega' \frac{\sin\theta'(\cos\phi')}{|\vec{r}-\vec{r}'|} = \frac{4\pi}{3}\frac{r_<}{r_>^2}\sin\theta\cos\phi \qquad (5.65)$$

Therefore,

$$\vec{A}(\vec{r}) = \frac{\mu_o}{4\pi}\sigma\omega R^3 \frac{4\pi}{3}\frac{r_<}{r_>^2}\sin\theta\left[-\widehat{x}\sin\phi + \widehat{y}\cos\phi\right] = \frac{\mu_o}{4\pi}\sigma\omega\frac{4\pi R^3}{3}\frac{r_<}{r_>^2}\sin\theta\,\widehat{\phi} \qquad (5.66)$$

where $r_< = \min(r,R)$ and $r_> = \max(r,R)$. While the vector potential's r dependence change from r inside the sphere to $1/r^2$ outside the sphere, the direction and angular dependence are always $\widehat{\phi}\sin\theta$. One can see that $\vec{A}$ points in the same direction as $\vec{K}$.

Example 5.2: Solenoid (Griffiths Example 5.12)

The calculation of the vector potential for a solenoid, which is the equivalent of a spinning cylinder of charge if one ignores the small axial current contribution, is more interesting because one need not do it by explicit integration as done for the spinning sphere of charge. Instead, one can use some intuition along with the combination of Stokes' Theorem and the relation between $\vec{B}$ and $\vec{A}$:

$$\oint_{\mathcal{C}(\mathcal{S})} d\vec{\ell} \cdot \vec{A} = \int_{\mathcal{S}} da\, \widehat{n} \cdot \vec{\nabla} \times \vec{A} = \int_{\mathcal{S}} da\, \widehat{n} \cdot \vec{B} \qquad (5.67)$$

The intuition part is to recognize that, because $\vec{B}$ is along the z-axis inside the solenoid and vanishing outside and because $\vec{A}$ "wraps around" $\vec{B}$, it is natural to assume $\vec{A}$ is along $\widehat{\phi}$. (This directionality can also be seen from the inherent coalignment of $\vec{A}$ and $\vec{J}$ from the integral formula for $\vec{A}$.) Then one can do the calculation in the same way as one applies the integral form of Ampere's Law, except that instead of current through a surface ("enclosed current"), we have enclosed magnetic flux, and, instead of a line integral of magnetic field around the edge of the surface, we have a line integral of vector potential. Please study the details in Griffiths, as a variant on this problem will be given in homework.

Uniqueness Theorem for Magnetic Fields

This is Griffiths Problem 5.56.

Just as we did for electric fields, we can show that, given a current distribution and a well-defined set of boundary conditions, the magnetic field obtained is unique. We assume that a current distribution $\vec{J}(\vec{r})$ in a volume $\mathcal{V}$ is specified. We will see later how specific we must be about the boundary conditions.

First, we need something analogous to the Green's Identities we used in the case of electrostatics. Using the vector identity $\vec{\nabla} \cdot (\vec{a} \times \vec{b}) = \vec{b} \cdot \vec{\nabla} \times \vec{a} - \vec{a} \cdot \vec{\nabla} \times \vec{b}$, letting $\vec{u}$ and $\vec{v}$ be two arbitrary vector fields, and applying the identity with $\vec{a} = \vec{u}$ and $\vec{b} = \vec{\nabla} \times \vec{v}$, we may write

$$\int_{\mathcal{V}} d\tau \, \vec{\nabla} \cdot (\vec{u} \times (\vec{\nabla} \times \vec{v})) = \int_{\mathcal{V}} d\tau \left[(\vec{\nabla} \times \vec{v}) \cdot (\vec{\nabla} \times \vec{u}) - \vec{u} \cdot (\vec{\nabla} \times (\vec{\nabla} \times \vec{v})) \right] \quad (5.68)$$

Since the expression on the left-hand side is a divergence, we may turn it into a surface integral using the divergence theorem:

$$\oint_{\mathcal{S}(\mathcal{V})} da \, \hat{n} \cdot (\vec{u} \times (\vec{\nabla} \times \vec{v})) = \int_{\mathcal{V}} d\tau \left[(\vec{\nabla} \times \vec{u}) \cdot (\vec{\nabla} \times \vec{v}) - \vec{u} \cdot (\vec{\nabla} \times (\vec{\nabla} \times \vec{v})) \right] \quad (5.69)$$

We will use this below.

Now, suppose that we have two different magnetic field configurations $\vec{B}_1 \neq \vec{B}_2$, derived from two different magnetic vector potentials $\vec{A}_1 \neq \vec{A}_2$, that both satisify Ampere's Law for the same current distribution: $\vec{\nabla} \times \vec{B}_1 = \vec{\nabla} \times \vec{B}_2 = \mu_o \vec{J}$. Let $\vec{A}_3 = \vec{A}_2 - \vec{A}_1$ and $\vec{B}_3 = \vec{B}_2 - \vec{B}_1$. We apply the above vector identity with $\vec{u} = \vec{v} = \vec{A}_3$:

$$\oint_{\mathcal{S}(\mathcal{V})} da\, \widehat{n} \cdot (\vec{A}_3 \times (\vec{\nabla} \times \vec{A}_3)) = \int_{\mathcal{V}} d\tau \left[(\vec{\nabla} \times \vec{A}_3) \cdot (\vec{\nabla} \times \vec{A}_3) - \vec{A}_3 \cdot (\vec{\nabla} \times (\vec{\nabla} \times \vec{A}_3)) \right] \tag{5.70}$$

We have that $\vec{\nabla} \times (\vec{\nabla} \times \vec{A}_3) = \vec{\nabla} \times \vec{B}_3 = \vec{\nabla} \times \vec{B}_2 - \vec{\nabla} \times \vec{B}_1 = \mu_o(\vec{J} - \vec{J}) = 0$ by Ampere's Law and the assumption that both field configurations are sourced by the same current distribution, so the second term on the right side vanishes. Exchanging the two sides, plugging in $\vec{B}_3 = \vec{\nabla} \times \vec{A}_3$, and using the cyclic property of the triple scalar product, $\vec{a} \cdot (\vec{b} \times \vec{c}) = \vec{c} \cdot (\vec{a} \times \vec{b}) = \vec{b} \cdot (\vec{c} \times \vec{a})$, we have

$$\int_{\mathcal{V}} d\tau \left| \vec{B}_3 \right|^2 = \oint_{\mathcal{S}(\mathcal{V})} da\, \widehat{n} \cdot (\vec{A}_3 \times \vec{B}_3) = \oint_{\mathcal{S}(\mathcal{V})} da\, \vec{B}_3 \cdot (\widehat{n} \times \vec{A}_3) \tag{5.71}$$

$$= \oint_{\mathcal{S}(\mathcal{V})} da\, \vec{A}_3 \cdot (\vec{B}_3 \times \widehat{n}) = - \oint_{\mathcal{S}(\mathcal{V})} da\, \vec{A}_3 \cdot (\widehat{n} \times \vec{B}_3) \tag{5.72}$$

From the above equation, we can see what (minimal) boundary condition information we must have to obtain uniqueness of $\vec{B}$: we must have that, at any given point on the surface, $\vec{A}$, $\vec{B}$, $\hat{n} \times \vec{A}$, or $\hat{n} \times \vec{B}$ is specified. If this is true, then $\vec{A}_3 = \vec{A}_2 - \vec{A}_1 = 0$ where $\vec{A}$ is specified, $\vec{B}_3 = \vec{B}_2 - \vec{B}_1 = 0$ where $\vec{B}$ is specified, $\hat{n} \times \vec{A}_3 = \hat{n} \times (\vec{A}_2 - \vec{A}_1) = 0$ where $\hat{n} \times \vec{A}$ is specified, and $\hat{n} \times \vec{B}_3 = \hat{n} \times (\vec{B}_2 - \vec{B}_1) = 0$ where $\hat{n} \times \vec{B}$ is specified. Requiring one of these four conditions at every point on $\mathcal{S}(\mathcal{V})$ ensures the integrand on the right side vanishes at every point on $\mathcal{S}(\mathcal{V})$ and thus the right side vanishes. Since the integrand on the left side is nonnegative, it must therefore vanish everywhere: $\vec{B}_3 = 0$. Hence, $\vec{B}_1 = \vec{B}_2$ and the fields are identical and the field solution is unique.

Specifying $\vec{A}$ is like a Dirichlet boundary condition where we specify the electrostatic potential on the boundary, and specifying $\hat{n} \times \vec{B} = \hat{n} \times (\vec{\nabla} \times \vec{A})$ is a lot like a Neumann boundary condition where we specify the normal gradient of the electrostatic potential $\hat{n} \cdot \vec{\nabla} V$ (which is proportional to the normal component of the electric field, $\hat{n} \cdot \vec{E}$). In fact, we will see via Ampere's Law that this is equivalent to specifying the surface current density flowing on the boundary. The other two types of conditions, specifying $\hat{n} \times \vec{A}$ or specifying $\vec{B}$, have no obvious analogue.

Uniqueness of the Vector Potential?

We have already discussed how the $\vec{A}$ that generates a particular $\vec{B}$ is unique up to the gradient of an additional function if its divergence is left unspecified. The above theorem for the uniqueness of the magnetic field therefore now tells us that specification of $\vec{J}$ in the volume and of $\vec{A}$, $\vec{B}$, $\hat{n} \times \vec{A}$, or $\hat{n} \times \vec{B}$ on the boundary gives a vector potential that is unique up to the gradient of an additional function if its divergence is unspecified. But what do we need to know to completely determine the vector potential?

Obtaining a unique vector potential is the equivalent of being able to also know the λ function (up to an offset). We showed that λ satisfies Poisson's Equation with $\vec{\nabla} \cdot \vec{A}$ as the source, Equation 5.51. So, clearly, to obtain a unique $\vec{A}$, we would need to specify $\vec{\nabla} \cdot \vec{A}$. We also would need appropriate boundary conditions for this Poisson Equation. We may conclude from our proof of the uniqueness of the scalar potential (up to an offset) that we must either specify λ or $\hat{n} \cdot \vec{\nabla}\lambda$ on the boundary to obtain a unique λ (again, up to an offset) and thus a unique $\vec{A}$.

Which of the above conditions provide the necessary boundary condition on λ? Only specification of $\vec{A}$ on the boundary is certain to be sufficient. This gives $\vec{\nabla}\lambda$ and thus $\hat{n} \cdot \vec{\nabla}\lambda$, a Neumann boundary condition for λ and thus sufficient to render λ unique.

We can see specifying $\hat{n} \times \vec{A}$ would only be sufficient in special cases. Doing so specifies $\hat{n} \times \vec{\nabla}\lambda$, which gives the component of $\vec{\nabla}\lambda$ tangent to the boundary. If the boundary is either at infinity or is a single, closed boundary, it seems likely one could then construct λ on the boundary by doing the line integral of $\hat{n} \times \vec{\nabla}\lambda$, much like one constructs the scalar potential from its gradient, the electric field. (It is ok that we would only know the component of $\vec{\nabla}\lambda$ tangent to the boundary, as $\hat{n} \cdot \vec{\nabla}\lambda$ will have zero dot product with the line element $d\vec{\ell}$ involved in the line integral.) As with the scalar potential, the offset is not important. However, if the boundary is not simply connected, then there is no way to connect λ on different pieces of the boundary without specifying its value on at least one point on each of those pieces. But we do not specify λ anywhere if we are given $\hat{n} \times \vec{A}$ and thus $\hat{n} \times \vec{\nabla}\lambda$ on the boundary. So specifying $\hat{n} \times \vec{A}$ (and $\vec{\nabla} \cdot \vec{A}$) is sufficient to make $\vec{A}$ unique only if the boundary is simply connected.

We can be assured that specifying $\vec{B}$ or $\hat{n} \times \vec{B}$ is entirely insufficient: because $\vec{B}$ is unaffected by λ, providing information about $\vec{B}$ cannot give us any information about λ.

Lastly, we remind the reader that, even if $\vec{A}$ is specified on the boundary, one also needs to know $\vec{\nabla} \cdot \vec{A}$ in the volume. Providing the former without the latter is equivalent to having a boundary condition but no differential equation to solve: the source term in the latter is unspecified.

The Magnetostatic Scalar Potential

If one considers current-free regions, then we have $\vec{\nabla} \times \vec{B} = 0$ and the magnetic field should be derivable from a scalar potential:

$$\vec{B}(\vec{r}) = -\vec{\nabla} U(\vec{r}) \tag{5.73}$$

One must take some care, though: in addition to being current-free, the region under consideration must be *simply connected*. Griffiths Problem 5.29 shows a situation where the current in a region may vanish but $\vec{\nabla} \times \vec{B} \neq 0$ because the region is not simply connected and the enclosed volume outside the region contains current.

With the above assumptions, and noting $\vec{\nabla} \cdot \vec{B} = 0$, we can infer that U satisfies Laplace's Equation:

$$\nabla^2 U(\vec{r}) = -\vec{\nabla} \cdot \vec{B}(\vec{r}) = 0 \tag{5.74}$$

Our usual assumption of simple boundary conditions — everything falls off to zero at infinity — yields a trivial result here, $U(\vec{r}) = 0$, so we must assume less trivial boundary conditions to obtain a nonzero U. We will return to the use of the magnetostatic scalar potential in connection with magnetically polarizable materials.

Lecture 16:

Magnetostatics III:
Boundary Conditions on Magnetic Field and Vector Potential
Magnetic Multipoles

Date Revised: 2024/02/27 06:00
Date Given: 2024/02/27

Boundary Conditions on Magnetic Field and Vector Potential

We will use techniques similar to those we used in determining the boundary conditions on the electric field. We will not immediately apply these conditions to boundary value problems for currents in vacuum because there are no nontrivial boundary-value problems of this type. That is because there is no way to directly set the vector potential, unlike for the electrostatic potential. There is also no equivalent to the perfect conductor, which yields equipotential surfaces in electrostatics. One only has Neumann boundary conditions, with current densities on surfaces, from which one can calculate the field directly via the Biot-Savart Law rather than solving Laplace's or Poisson's Equation. We will find the boundary conditions more useful in the context of magnetically polarizable materials.

Boundary Conditions on the Magnetic Field

Recall that Gauss's Law, $\vec{\nabla} \cdot \vec{E} = \rho/\epsilon_o$, implied that the normal component of the electric field satisfied Equation 2.59

$$\hat{n}(\vec{r}) \cdot \left[\vec{E}_2(\vec{r}) - \vec{E}_1(\vec{r}) \right] = \frac{1}{\epsilon_o}\, \sigma(\vec{r}) \qquad (5.75)$$

Since $\vec{\nabla} \cdot \vec{B} = 0$, we can conclude by analogy that

$$\boxed{\hat{n}(\vec{r}) \cdot \left[\vec{B}_2(\vec{r}) - \vec{B}_1(\vec{r}) \right] = 0} \qquad (5.76)$$

That is, the normal component of the magnetic field is continuous at any boundary.

For the tangential component, we return to the derivation leading to Equation 2.61. In that case, we considered a contour $\mathcal{C}$ that consisted of two legs $\mathcal{C}_1$ and $\mathcal{C}_2$ parallel to the interface and to each other and two legs normal to the interface whose length would be shrunk to zero. We saw

$$\oint_{\mathcal{C}} d\vec{\ell} \cdot \vec{E}(\vec{r}) = -\int_{\mathcal{C}_1, \vec{r}_a - \hat{n}(\vec{r})\frac{dz}{2}}^{\vec{r}_b - \hat{n}(\vec{r})\frac{dz}{2}} \vec{E}_1(\vec{r}) \cdot d\vec{\ell} + \int_{\mathcal{C}_2, \vec{r}_a + \hat{n}(\vec{r})\frac{dz}{2}}^{\vec{r}_b + \hat{n}(\vec{r})\frac{dz}{2}} \vec{E}_2(\vec{r}) \cdot d\vec{\ell} \qquad (5.77)$$

$$\stackrel{dz \to 0}{\longrightarrow} \int_{\mathcal{C}_2, \vec{r}_a}^{\vec{r}_b} \left[\vec{E}_2(\vec{r}) - \vec{E}_1(\vec{r}) \right] \cdot d\vec{\ell} \qquad (5.78)$$

where the ends of the loop are near $\vec{r}_a$ and $\vec{r}_b$, $\hat{n}$ is the normal to the surface (parallel to the short legs of the loop), $\hat{t}$ is the normal to the loop area, $\hat{s} = \hat{t} \times \hat{n}$ is the unit vector parallel to the long legs of the loop, and ds is a line element along $\hat{s}$. In the electric field case, the left side of the above expression vanished. In the case of the magnetic field, Ampere's Law tells us that it is the current enclosed flowing in the direction $\hat{t}$. Therefore, the magnetic field version of the above equation is:

$$\mu_o \int_{\mathcal{C}_2} ds \, \hat{t}(\vec{r}) \cdot \vec{K}(\vec{r}) = \int_{\mathcal{C}_2, \vec{r}_a}^{\vec{r}_b} \left[\vec{B}_2(\vec{r}) - \vec{B}_1(\vec{r}) \right] \cdot d\vec{\ell} \qquad (5.79)$$

where $\mathcal{C}_1 \to \mathcal{C}_2$ in the plane of the interface as $dz \to 0$. We neglect any volume current density passing through the area enclosed by the contour $\mathcal{C}$ because the integral of that volume current density vanishes as $dz \to 0$.

Since the contour $\mathcal{C}_2$ is arbitrary, the integrands must be equal

$$\left[\vec{B}_2(\vec{r}) - \vec{B}_1(\vec{r})\right] \cdot \widehat{s}(\vec{r}) = \mu_o\, \widehat{t}(\vec{r}) \cdot \vec{K}(\vec{r}) \tag{5.80}$$

Next, we use $\widehat{t} = \widehat{n} \times \widehat{s}$:

$$\left[\vec{B}_2(\vec{r}) - \vec{B}_1(\vec{r})\right] \cdot \widehat{s}(\vec{r}) = \mu_o\left[\widehat{n}(\vec{r}) \times \widehat{s}(\vec{r})\right] \cdot \vec{K}(\vec{r}) \tag{5.81}$$

Finally, using the cyclic nature of triple vector products ,
$\vec{a} \cdot (\vec{b} \times \vec{c}) = \vec{c} \cdot (\vec{a} \times \vec{b}) = \vec{b} \cdot (\vec{c} \times \vec{a})$:

$$\boxed{\left[\vec{B}_2(\vec{r}) - \vec{B}_1(\vec{r})\right] \cdot \widehat{s}(\vec{r}) = \mu_o\left[\vec{K}(\vec{r}) \times \widehat{n}(\vec{r})\right] \cdot \widehat{s}(\vec{r})} \tag{5.82}$$

Note that this condition holds for any $\widehat{s}$ tangential to the interface. To give some intuition, $\widehat{n} \times \vec{K}$ has the magnitude of $\vec{K}$ (because $\widehat{n} \perp \vec{K}$ always) but points in a direction perpendicular to $\vec{K}$ while still tangent to the interface. The sign is set by the cross-product right-hand rule.

We can combine the conditions on the normal and tangential components of $\vec{B}$ to obtain one compact expression for the boundary condition on the magnetic field. By the definition of the cross product, $\vec{K} \times \hat{n}$ is always perpendicular to $\hat{n}$ and thus has no component along $\hat{n}$. Therefore, the expression

$$\boxed{\vec{B}_2(\vec{r}) - \vec{B}_1(\vec{r}) = \mu_o \, \vec{K}(\vec{r}) \times \hat{n}(\vec{r})} \qquad (5.83)$$

captures both boundary conditions: the projection of $\vec{B}$ normal to the interface (along $\hat{n}$) is continuous because the projection of the right side along that direction vanishes, and the projection of $\vec{B}$ along any $\hat{s}$ parallel to the interface can be discontinuous by the projection of $\mu_o \vec{K} \times \hat{n}$ along that direction. This is a very nice relation: given $\vec{K}$, it provides a way to calculate the change in the entire magnetic field across the interface, not just the change of a component.

We can rewrite the above in another way. Take the cross product of both sides with $\widehat{n}(\vec{r})$ from the left. The right side becomes a triple vector product, which we can rewrite using the $BAC - CAB$ rule, $\vec{a} \times (\vec{b} \times \vec{c}) = \vec{b}(\vec{a} \cdot \vec{c}) - \vec{c}(\vec{a} \cdot \vec{b})$. The second term has $\widehat{n} \cdot \vec{K}$, which vanishes, while the first term has $\widehat{n} \cdot \widehat{n} = 1$. Thus, we have

$$\boxed{\widehat{n}(\vec{r}) \times \left[\vec{B}_2(\vec{r}) - \vec{B}_1(\vec{r})\right] = \mu_o \vec{K}(\vec{r})}$$

(5.84)

The earlier form is more useful when $\vec{K}$ is specified, and the second form would more easily yield $\vec{K}$ if the fields are specified. Note, however, that this form does not preserve the information about the normal component of $\vec{B}$ because the contribution of that component to the left side vanishes.

Boundary Conditions on the Vector Potential

As one might expect by analogy to the electrostatic case, the vector potential itself
has to be continuous across a boundary:

$$\boxed{\vec{A}_2(\vec{r}) - \vec{A}_1(\vec{r}) = 0} \tag{5.85}$$

This is seen easily:

▶ We have chosen the divergence of $\vec{A}$ to vanish, so the normal component of $\vec{A}$
  must be continuous, just as we found the normal component of $\vec{B}$ is continuous
  for the same reason.

▶ The curl of $\vec{A}$ does not vanish, $\vec{\nabla} \times \vec{A} = \vec{B}$. This implies the line integral of $\vec{A}$
  around the contour $\mathcal{C}$ used above is nonzero and equals $\Phi_{\mathcal{S}(\mathcal{C})} = \int_{\mathcal{S}(\mathcal{C})} da\,\hat{n} \cdot \vec{B}$,
  the *magnetic flux* of $\vec{B}$ through the surface $\mathcal{S}(\mathcal{C})$ defined by $\mathcal{C}$. But, as the area
  of the contour is shrunk to zero, the magnetic flux vanishes via an argument
  similar to the one we used to show that the flux of the electric field always goes
  to zero as the area through which it is calculated goes to zero: while the field
  can be quite singular $(1/r^2)$, there are always cancellations that cause the flux
  to vanish. Therefore, the tangential component of $\vec{A}$ is also continuous.

While the vector potential itself is continuous, its derivatives are not necessarily continuous because its derivatives are related to $\vec{B}$, which is not necessarily continuous. Evaluating these discontinuities is a bit harder than in the case of the electric potential because the derivatives are not related in a trivial component-by-component way to the field. We need an expression involving second derivatives of $\vec{A}$ if we want to obtain boundary conditions on the first derivatives of $\vec{A}$. Let's use Equation 5.56:

$$\nabla^2 \vec{A}(\vec{r}) = -\mu_o \, \vec{J}(\vec{r}) \tag{5.86}$$

Consider a projection of this equation in Cartesian coordinates by taking the dot product with a Cartesian unit vector on the left and then passing it through the Laplacian, rewritten so the divergence is clear:

$$\vec{\nabla} \cdot \vec{\nabla} \left(\widehat{x} \cdot \vec{A}(\vec{r}) \right) = -\mu_o \, \widehat{x} \cdot \vec{J}(\vec{r}) \tag{5.87}$$

We have used Cartesian coordinates rather than a coordinate system using $\widehat{n}$, $\widehat{t}$, and $\widehat{s}$ because the latter vary in direction depending on where one is on the surface; their derivatives do not vanish, so we would not have been able to pull them inside the Laplacian as we did with $\widehat{x}$.

Given the above, we now apply the same kind of geometry we used to derive the boundary condition on the normal component of $\vec{E}$. That yields

$$\hat{n} \cdot \left[\vec{\nabla} \left(\hat{x} \cdot \vec{A}_2(\vec{r}) \right) - \vec{\nabla} \left(\hat{x} \cdot \vec{A}_1(\vec{r}) \right) \right] = -\mu_o \, \hat{x} \cdot \vec{K}(\vec{r}) \tag{5.88}$$

$$\hat{n} \cdot \vec{\nabla} \left[\hat{x} \cdot \vec{A}_2(\vec{r}) - \hat{x} \cdot \vec{A}_1(\vec{r}) \right] = \tag{5.89}$$

where $\hat{x} \cdot \vec{K}$ is what is left of $\hat{x} \cdot \vec{J}$ as the Gaussian volume used in that proof shrinks to zero thickness in the direction normal to the interface, just as ρ reduced to σ in the case of the electric field.

The above argument holds for the $\hat{y}$ and $\hat{z}$ projections of $\vec{A}$ and $\vec{K}$ also, so we may combine them to obtain

$$\boxed{\hat{n} \cdot \vec{\nabla} \left[\vec{A}_2(\vec{r}) - \vec{A}_1(\vec{r}) \right] = -\mu_o \, \vec{K}(\vec{r})} \tag{5.90}$$

Thus, we see that the normal derivative of each component of the vector potential has a discontinuity set by the surface current density in the direction of that component of the vector potential. This is a lot like the discontinuity in the normal gradient of the electric potential being determined by the surface charge density at the boundary.

We may derive, from the above, conditions in the normal and tangential directions by recognizing that

$$\left(\widehat{n} \cdot \vec{\nabla}\right) \widehat{n} = 0 \qquad \left(\widehat{n} \cdot \vec{\nabla}\right) \widehat{s} = 0 \qquad (5.91)$$

These relations should be intuitively obvious: the direction of $\widehat{n}$, $\widehat{s}$, and $\widehat{t}$ change as one moves transversely along the surface (along $\widehat{s}$ or $\widehat{t}$), but they simply are not defined off the surface and thus they can have no derivative in that direction. This implies that the normal derivative of the normal component of $\vec{A}$ has no discontinuity since there can be no surface current in that direction:

$$\boxed{\widehat{n} \cdot \vec{\nabla} \left\{ \widehat{n} \cdot \left[\vec{A}_2(\vec{r}) - \vec{A}_1(\vec{r}) \right] \right\} = 0} \qquad (5.92)$$

It also implies that the normal gradient of the vector potential in a particular direction parallel to the interface changes by the surface current density in that direction:

$$\boxed{\widehat{n} \cdot \vec{\nabla} \left\{ \widehat{s} \cdot \left[\vec{A}_2(\vec{r}) - \vec{A}_1(\vec{r}) \right] \right\} = -\mu_o \, \widehat{s} \cdot \vec{K}(\vec{r})} \qquad (5.93)$$

Next, let's consider the tangential derivatives of the vector potential. Here, we use the vector identity

$$\vec{\nabla} \times \vec{\nabla} \vec{A}(\vec{r}) = 0 \tag{5.94}$$

where again we consider each component of $\vec{A}$ as a scalar function and the above equation holds for all three components. If we again project by Cartesian components; e.g.

$$\vec{\nabla} \times \vec{\nabla} \left(\hat{x} \cdot \vec{A}(\vec{r}) \right) = 0 \tag{5.95}$$

then we can apply the same type of argument as we applied for calculating the boundary condition on the tangential components of $\vec{E}$, which in this case yields

$$\hat{s} \cdot \left[\vec{\nabla} \left(\hat{x} \cdot \vec{A}_2(\vec{r}) \right) - \vec{\nabla} \left(\hat{x} \cdot \vec{A}_1(\vec{r}) \right) \right] = 0 \tag{5.96}$$

$$\hat{s} \cdot \vec{\nabla} \left[\hat{x} \cdot \vec{A}_2(\vec{r}) - \hat{x} \cdot \vec{A}_1(\vec{r}) \right] = \tag{5.97}$$

Since the argument again generalizes to any Cartesian component, we may combine the three expressions to obtain

$$\boxed{\hat{s} \cdot \vec{\nabla} \left[\vec{A}_2(\vec{r}) - \vec{A}_1(\vec{r}) \right] = 0} \tag{5.98}$$

for any $\hat{s}$ parallel to the interface: the tangential derivatives of $\vec{A}$ are continuous across an interface.

Magnetic Multipoles

Derivation of Magnetic Multipole Expansion

Since the vector potential is sourced by the current distribution in a manner similar to the way the charge distribution sources the electric potential, it is natural to develop the same multipole expansion. We follow Jackson for the sake of generality and variety; you can of course read the derivation in Griffiths, too. *We continue to make the steady-state assumption, and now we also make the assumption the currents are localized.* We start with the equation for the vector potential in terms of the current distribution:

$$\vec{A}(\vec{r}) = \frac{\mu_o}{4\pi} \int_{\mathcal{V}} d\tau' \, \frac{\vec{J}(\vec{r}')}{|\vec{r} - \vec{r}'|} \tag{5.99}$$

We recall Equation 3.149:

$$\frac{1}{|\vec{r} - \vec{r}'|} = \sum_{\ell=0}^{\infty} \frac{r_<^{\ell}}{r_>^{\ell+1}} P_\ell(\cos\gamma) \tag{5.100}$$

where $r_<$ and $r_>$ and the smaller and larger of r and r'.

As with the multipole expansion for the electrostatic potential, we will take $r \gg r'$: we want to know what the potential looks like far away from the current distribution. Therefore, $r_< = r'$ and $r_> = r$:

$$\vec{A}(\vec{r}) = \frac{\mu_o}{4\pi} \int_{\mathcal{V}} d\tau' \, \vec{J}(\vec{r}') \sum_{\ell=0}^{\infty} \frac{(r')^\ell}{r^{\ell+1}} P_\ell(\cos\gamma) \tag{5.101}$$

where $\cos\gamma = \hat{r} \cdot \hat{r}'$ is the angle between the two vectors.

There is a common $1/r$ we can factor out, leaving

$$\vec{A}(\vec{r}) = \frac{\mu_o}{4\pi} \frac{1}{r} \sum_{\ell=0}^{\infty} \frac{1}{r^\ell} \int_{\mathcal{V}} d\tau' \, \vec{J}(\vec{r}') \, (r')^\ell \, P_\ell(\cos\gamma) \tag{5.102}$$

Now, consider the first term, which is just the volume integral of the current density. Under the steady-state assumption, it is intuitively clear this integral must vanish. To prove this explicitly, we first use the vector identity $\vec{\nabla} \cdot (f\vec{a}) = f\vec{\nabla} \cdot \vec{a} + \vec{a} \cdot \vec{\nabla} f$ with $\vec{a} = \vec{J}$ and $f = r_i$ any of the Cartesian coordinates:

$$\vec{\nabla} \cdot (r_i \vec{J}) = r_i \vec{\nabla} \cdot \vec{J} + \vec{J} \cdot \vec{\nabla} r_i = 0 + \sum_{j=1}^{3} J_j \frac{\partial}{\partial r_j} r_i = \sum_{j=1}^{3} J_j \delta_{ij} = J_i \tag{5.103}$$

where the first term vanishes because of the steady-state assumption and so continuity implies $\vec{\nabla} \cdot \vec{J} = 0$. With this, we can compute the integral using the divergence theorem:

$$\int_{\mathcal{V}} d\tau' \, J_i(\vec{r}') = \int_{\mathcal{V}} d\tau' \, \vec{\nabla}' \cdot \left[r_i' \vec{J}(\vec{r}') \right] = \oint_{\mathcal{S}(\mathcal{V})} da' \, \hat{n}(\vec{r}') \cdot \left[r_i' \vec{J}(\vec{r}') \right] = 0 \tag{5.104}$$

where the surface integral in the last term vanishes because the current distribution is localized.

So, we are left with

$$\vec{A}(\vec{r}) = \frac{\mu_o}{4\pi} \frac{1}{r} \sum_{\ell=1}^{\infty} \frac{1}{r^\ell} \int_{\mathcal{V}} d\tau' \, \vec{J}(\vec{r}') \, (r')^\ell \, P_\ell(\cos\gamma) \qquad (5.105)$$

This is the *multipole expansion* of the vector potential of the current distribution. As with the multipole expansion of the electric potential, one can see that the successive terms fall off as successively higher powers of $1/r$.

It makes sense that there is no monopole term because $\vec{\nabla} \cdot \vec{B} = 0$: if there were a way to make a current distribution look like a monopole from far away, then one would have a field configuration with a nonzero Gauss's law integral of magnetic flux through a closed surface containing the current distribution, which is not allowed by $\vec{\nabla} \cdot \vec{B} = 0$.

The Magnetic Dipole Term

Let's consider the first nonzero term in more detail, which we subscript with a $_2$ because it will look like the electric dipole potential, and let's expand $\vec{J}$ in terms of its components so it is easier to work with:

$$\vec{A}_2(\vec{r}) = \frac{\mu_o}{4\pi} \frac{1}{r^2} \int_{\mathcal{V}} d\tau' \, \vec{J}(\vec{r}') \, r' \, P_2(\cos\gamma) = \frac{\mu_o}{4\pi} \frac{1}{r^3} \int_{\mathcal{V}} d\tau' \, \vec{J}(\vec{r}') \, \vec{r} \cdot \vec{r}' \qquad (5.106)$$

$$= \frac{\mu_o}{4\pi} \frac{1}{r^3} \sum_{i,j=1}^{3} \hat{r}_i \int_{\mathcal{V}} d\tau' \, J_i(\vec{r}') \, r_j \, r_j' \qquad (5.107)$$

We must first prove an identity. We start with the same vector identity as before, now with $f = r_i \, r_j$ and $\vec{a} = \vec{J}$:

$$\vec{\nabla} \cdot (r_i \, r_j \, \vec{J}) = r_i \, r_j \, \vec{\nabla} \cdot \vec{J} + \vec{J} \cdot \vec{\nabla}(r_i \, r_j) = 0 + r_j \, \vec{J} \cdot \vec{\nabla} r_i + r_i \, \vec{J} \cdot \vec{\nabla} r_j \qquad (5.108)$$

$$= r_j \, J_i + r_i \, J_j \qquad (5.109)$$

where we have again used $\vec{\nabla} \cdot \vec{J} = 0$. We apply the same technique of integrating over volume and turning the left side into a surface term that vanishes, so we are left with

$$\int_{\mathcal{V}} d\tau' \left[r_i' \, J_j(\vec{r}') + r_j' \, J_i(\vec{r}') \right] = 0 \qquad (5.110)$$

We can use this identity to rewrite the $\vec{A}_2$ term as:

$$\vec{A}_2(\vec{r}) = \frac{\mu_o}{4\,\pi} \frac{1}{r^3} \sum_{i,j=1}^{3} \widehat{r}_i\, r_j \int_{\mathcal{V}} d\tau'\, \frac{1}{2} \left[J_i(\vec{r}\,')\, r_j' - J_j(\vec{r}\,')\, r_i' \right] \tag{5.111}$$

where we split out half of the $J_i\, r_j'$ factor and used the identity to exchange the indices. You have learned in Ph106a and hopefully elsewhere that the cross-product can be written

$$(\vec{a} \times \vec{b})_k = \sum_{m,n=1}^{3} \epsilon_{kmn}\, a_m\, b_n \quad \text{with} \quad \epsilon_{kmn} = \left\{ \begin{array}{rl} 1 & \text{for cyclic index permutations} \\ -1 & \text{for anticyclic index permutations} \\ 0 & \text{when any two indices are identical} \end{array} \right. \tag{5.112}$$

where ϵ_{kmn} is the *Levi-Civita* symbol. Multiplying this definition by ϵ_{ijk} and summing over k gives

$$\sum_{k=1}^{3} \epsilon_{ijk}(\vec{a} \times \vec{b})_k = \sum_{k,m,n=1}^{3} \epsilon_{ijk}\, \epsilon_{kmn}\, a_m\, b_n = \sum_{k,m,n=1}^{3} \epsilon_{kij}\, \epsilon_{kmn}\, a_m\, b_n \tag{5.113}$$

There is an identity for the Levi-Civita symbol

$$\sum_{k=1}^{3} \epsilon_{kij} \, \epsilon_{kmn} = \delta_{im} \, \delta_{jn} - \delta_{in} \, \delta_{jm} \tag{5.114}$$

(this is the identity that produces the $BAC - CAB$ rule,
$\vec{a} \times (\vec{b} \times \vec{c}) = \vec{b}(\vec{a} \cdot \vec{c}) - \vec{c}(\vec{a} \cdot \vec{b})$) which lets us rewrite the above as

$$\sum_{k=1}^{3} \epsilon_{ijk}(\vec{a} \times \vec{b})_k = \sum_{m,n=1}^{3} a_m \, b_n \left(\delta_{im} \, \delta_{jn} - \delta_{in} \, \delta_{jm} \right) = a_i \, b_j - a_j \, b_i \tag{5.115}$$

This is exactly the expression we have inside the integral above.

Using the above identity, we may rewrite the $\vec{A}_2$ term as

$$\vec{A}_2(\vec{r}) = \frac{\mu_o}{4\,\pi} \frac{1}{r^3} \sum_{i,j,k=1}^{3} \widehat{r}_i \, r_j \int_{\mathcal{V}} d\tau' \, \frac{1}{2} \epsilon_{ijk} \left[J(\vec{r}') \times \vec{r}' \right]_k \tag{5.116}$$

$$= -\frac{\mu_o}{4\,\pi} \frac{1}{r^3} \frac{1}{2} \sum_{i}^{3} \widehat{r}_i \left\{ \vec{r} \times \int_{\mathcal{V}} d\tau' \, \left[\vec{r}' \times J(\vec{r}') \right] \right\}_i \tag{5.117}$$

$$= -\frac{\mu_o}{4\,\pi} \frac{1}{r^3} \frac{1}{2} \vec{r} \times \int_{\mathcal{V}} d\tau' \, \left[\vec{r}' \times J(\vec{r}') \right] \tag{5.118}$$

If we define the *magnetization density* $\vec{\mathcal{M}}(\vec{r})$ and the *magnetic dipole moment* $\vec{m}$ by

$$\vec{\mathcal{M}}(\vec{r}) = \frac{1}{2}\,\vec{r} \times \vec{J}(\vec{r}) \qquad \text{and} \qquad \vec{m} = \int_{\mathcal{V}} d\tau'\,\vec{\mathcal{M}}(\vec{r}') \tag{5.119}$$

then the $_2$ term is the *magnetic dipole vector potential*

$$\vec{A}_2(\vec{r}) = \frac{\mu_o}{4\,\pi}\,\frac{\vec{m} \times \vec{r}}{r^3} \tag{5.120}$$

Interestingly, this form has the same radial dependence as that of the electrostatic potential of a dipole, but the cross-product in the numerator differs from the dot product in the numerator of the electric dipole potential. However, because the magnetic field is obtained from the curl of the vector potential, while the electric field is obtained from the gradient of the electric potential, we will see that the two forms result in the same field configuration (up to normalization)!

Specialization of Magnetic Dipole Potential to a Current Loop

Now, let us consider a current loop. The only assumption we make is that the current throughout the loop is the same so that we can extract it from the integral. The volume integral reduces to a line integral over the loop contour:

$$\vec{A}_2(\vec{r}) = -\frac{\mu_o}{4\pi}\frac{1}{r^3}\frac{1}{2}\,\vec{r}\times\oint_{\mathcal{C}}\vec{r}\,'\times I\,d\vec{\ell}\,'(\vec{r}\,') = -\frac{\mu_o}{4\pi}\frac{1}{r^3}\,\vec{r}\times I\oint_{\mathcal{C}}\frac{\vec{r}\,'\times d\vec{\ell}\,'(\vec{r}\,')}{2} \quad (5.121)$$

The integral is now just a geometric quantity that has units of area. Separating out the magnetic moment, we have

$$\boxed{\;\vec{A}_2^{\,loop}(\vec{r}) = \frac{\mu_o}{4\pi}\frac{\vec{m}_{loop}\times\vec{r}}{r^3} \qquad \vec{m}_{loop} = I\oint_{\mathcal{C}}\frac{\vec{r}\,'\times d\vec{\ell}\,'(\vec{r}\,')}{2}\;} \qquad (5.122)$$

For the case of a loop confined to a plane that contains the origin, the quantity $\vec{r}\,' \times d\vec{\ell}\,'/2$ is the differential area element for the loop: it is the area of the triangle formed by $\vec{r}\,'$, the vector from the origin to a point on the loop, and $d\vec{\ell}\,'$, the line element tangent to the loop at $\vec{r}\,'$ and in the direction of the current, and this cross product has the standard right-hand-rule orientation. The integral thus calculates the area of the loop! Thus, for a planar loop, the above reduces to

$$\vec{A}_2(\vec{r}) = -\frac{\mu_o}{4\,\pi}\,\frac{1}{r^3}\,\vec{r}\times I\,\widehat{n}\,a \tag{5.123}$$

where a is the loop area and $\widehat{n}$ is the normal to the loop with orientation defined by the current via the right-hand rule. Therefore, for this case, we have

$$\boxed{\vec{A}_2^{\,flat\ loop}(\vec{r}) = \frac{\mu_o}{4\,\pi}\,\frac{\vec{m}_{flat\ loop}\times\vec{r}}{r^3} \qquad \vec{m}_{flat\ loop} = I\,\widehat{n}\,a} \tag{5.124}$$

Field of a Magnetic Dipole

If we let $\vec{m} = m\,\widehat{z}$, then the dipole vector potential is

$$\vec{A}_2(\vec{r}) = \frac{\mu_o}{4\pi}\,\frac{m\,\sin\theta}{r^2}\,\widehat{\phi} \equiv A_{2,\phi}\,\widehat{\phi} \tag{5.125}$$

This form offers some intuition about how $\vec{A}_2(\vec{r})$ behaves. In general, $\vec{A}_2$ "circulates" around $\vec{m}$ using the right-hand rule in the same way that $\vec{A}$ "circulates" around $\vec{B}$ or $\vec{B}$ "circulates" around $\vec{J}$ using the right-hand rule. Since we are considering the distribution from far enough away that it is indistinguishable from a simple circular current loop in the xy-plane, the direction of $\vec{A}_2$ just results from the fact that $\vec{A}$ is the convolution of $\vec{J}$ with a scalar function: the direction of $\vec{A}$ always follows that of $\vec{J}$.

If we take the curl of this in spherical coordinates, we obtain

$$B_{2,r}(\vec{r}) = \frac{1}{r\,\sin\theta}\,\frac{\partial}{\partial\theta}(\sin\theta\,A_{2,\phi}) = 2\,\frac{\mu_o}{4\pi}\,\frac{m\,\cos\theta}{r^3} \tag{5.126}$$

$$B_{2,\theta}(\vec{r}) = -\frac{1}{r}\,\frac{\partial}{\partial r}(r\,A_{2,\phi}) = \frac{\mu_o}{4\pi}\,\frac{m\,\sin\theta}{r^3} \tag{5.127}$$

$$B_{2,\phi}(\vec{r}) = 0 \tag{5.128}$$

$$\text{or}\quad \vec{B}_2(\vec{r}) = \frac{\mu_o}{4\pi}\,\frac{m}{r^3}\left(2\,\widehat{r}\,\cos\theta + \widehat{\theta}\,\sin\theta\right) \tag{5.129}$$

which matches the form of Equation 3.239 for an electric dipole.

Let's derive the more generic result by releasing the condition $\vec{m} = m\hat{z}$:

$$\vec{B}_2(\vec{r}) = \vec{\nabla} \times \vec{A} = \sum_{i,j,k=1}^{3} \epsilon_{ijk}\hat{r}_i \frac{\partial A_k}{\partial r_j} = \frac{\mu_o}{4\pi} \sum_{i,j,k,\ell,m=1}^{3} \epsilon_{ijk}\hat{r}_i \frac{\partial}{\partial r_j} \epsilon_{k\ell m} \left(\frac{m_\ell r_m}{r^3} \right) \quad (5.130)$$

$$= \frac{\mu_o}{4\pi} \sum_{i,j,k,\ell,m=1}^{3} \epsilon_{ijk}\epsilon_{k\ell m}\hat{r}_i \left[\frac{m_\ell \delta_{jm}}{r^3} - \frac{3}{2} \frac{m_\ell r_m}{r^5} \left(2r_j \right) \right] \quad (5.131)$$

We use the cyclicity of the Levi-Civita symbol in its indices and the identities
$\sum_{k=1}^{3} \epsilon_{kij}\epsilon_{k\ell m} = \delta_{i\ell}\delta_{jm} - \delta_{im}\delta_{j\ell}$ and $\sum_{j,k=1}^{3} \epsilon_{jki}\epsilon_{jk\ell} = 2\delta_{i\ell}$ to rewrite the above in a
form identical to that of the electric dipole, Equation 3.242:

$$\vec{B}_2(\vec{r}) = \frac{\mu_o}{4\pi} \sum_{i=1}^{3} \hat{r}_i \left[\frac{2m_i}{r^3} - \frac{3}{r^5} \left(m_i \sum_{j=1}^{3} r_j r_j - r_i \sum_{j=1}^{3} m_j r_j \right) \right] \quad (5.132)$$

$$= \frac{\mu_o}{4\pi} \sum_{i=1}^{3} \hat{r}_i \frac{3\,r_i\,(\vec{m}\cdot\vec{r}) - m_i\,(\vec{r}\cdot\vec{r})}{r^5} \quad (5.133)$$

$$\implies \boxed{\vec{B}_2(\vec{r}) = \frac{\mu_o}{4\pi} \frac{3\,(\vec{m}\cdot\hat{r})\,\hat{r} - \vec{m}}{r^3}} \quad (5.134)$$

Lecture 17:

Magnetostatics IV:
Magnetic Multipoles (cont'd)

Magnetostatics in Matter I:
Field of a Magnetized Object
Auxiliary Field $\vec{H}$

Date Revised: 2024/02/29 06:00
Date Given: 2024/02/29

Force on a Magnetic Dipole (à la Jackson)

As we did for electric multipoles, let's consider the problem of the force and torque on a magnetic dipole. However, because there is no magnetic potential energy function, we must begin from the Lorentz Force on the current distribution, which is given by

$$\vec{F}_{mag} = \int_{\mathcal{V}} d\tau \, \vec{J}(\vec{r}) \times \vec{B}(\vec{r}) \tag{5.135}$$

As we did in the case of the force on an electric multipole, we Taylor expand $\vec{B}(\vec{r})$. Again, as we did for electrostatics, we place the multipole at the origin and will generalize the result later. The expansion is

$$B_k(\vec{r}) = B_k(\vec{r} = \vec{0}) + \sum_{m=1}^{3} r_m \left. \frac{\partial B_k}{\partial r_m} \right|_{\vec{r}=\vec{0}} + \cdots \tag{5.136}$$

Thus, the Lorentz Force is

$$\vec{F}_{mag} = \sum_{i,j,k=1}^{3} \epsilon_{ijk} \, \widehat{r}_i \int_{\mathcal{V}} d\tau \, J_j(\vec{r}) \, B_k(\vec{r}) \tag{5.137}$$

$$= \sum_{i,j,k=1}^{3} \epsilon_{ijk} \, \widehat{r}_i \left[B_k(\vec{0}) \int_{\mathcal{V}} d\tau \, J_j(\vec{r}) + \sum_{m=1}^{3} \left(\left. \frac{\partial B_k}{\partial r_m} \right|_{\vec{0}} \right) \int_{\mathcal{V}} d\tau \, J_j \, r_m + \cdots \right] \tag{5.138}$$

We have done both these integrals before. The first one contains the monopole of the current distribution, which vanishes as in Equation 5.104. Since we will see that the second term is in general nonzero and is proportional to the magnetic dipole moment, let's call it $\vec{F}_{dip}$ and focus on it, dropping the higher-order terms. It is very similar in structure to what we encountered in calculating the dipole term in Equation 5.107. Applying the same tricks we used there to obtain Equation 5.116, we may rewrite it as

$$\vec{F}_{dip} = \sum_{i,j,k,m,n=1}^{3} \epsilon_{ijk} \, \widehat{r}_i \left(\left. \frac{\partial B_k}{\partial r_m} \right|_{\vec{0}} \right) \int_{\mathcal{V}} d\tau \, \frac{1}{2} \, \epsilon_{jmn} \left[\vec{J}(\vec{r}) \times \vec{r} \right]_n \tag{5.139}$$

$$= - \sum_{i,j,k,m,n=1}^{3} \epsilon_{ijk} \epsilon_{jmn} \, \widehat{r}_i \left(\left. \frac{\partial B_k}{\partial r_m} \right|_{\vec{0}} \right) m_n \qquad \text{with} \quad \vec{m} = \frac{1}{2} \int_{\mathcal{V}} d\tau \left[\vec{r} \times \vec{J}(\vec{r}) \right] \tag{5.140}$$

We use the vector identity Equation 5.114, $\sum_{j=1}^{3} \epsilon_{jik}\epsilon_{jmn} = \delta_{im}\delta_{kn} - \delta_{in}\delta_{km}$, and also use $\epsilon_{ijk} = -\epsilon_{jik}$ to adjust the indices to match this expression, yielding

$$\vec{F}_{dip} = \sum_{i,k,m,n=1}^{3} (\delta_{im}\delta_{kn} - \delta_{in}\delta_{km}) \, \widehat{r}_i \left(\frac{\partial B_k}{\partial r_m}\bigg|_{\vec{0}} \right) m_n \tag{5.141}$$

$$= \sum_{i,k=1}^{3} \widehat{r}_i \left[\left(\frac{\partial B_k}{\partial r_i}\bigg|_{\vec{0}} \right) m_k - \left(\frac{\partial B_k}{\partial r_k}\bigg|_{\vec{0}} \right) m_i \right] \tag{5.142}$$

$$= \vec{\nabla} \left(\vec{m} \cdot \vec{B} \right)\bigg|_{\vec{0}} - \vec{m} \left(\vec{\nabla} \cdot \vec{B} \right)\bigg|_{\vec{0}} \tag{5.143}$$

The second term vanishes. Generalizing the first term to a dipole at an arbitrary position, we have

$$\boxed{\vec{F}_{dip} = \vec{\nabla} \left[\vec{m} \cdot \vec{B}(\vec{r}) \right] \qquad \text{with} \quad \vec{m} = \frac{1}{2} \int_{\mathcal{V}} d\tau \left[\vec{r}' \times \vec{J}(\vec{r}') \right]} \tag{5.144}$$

The force causes the magnetic dipole to move to a local maximum of $\vec{m} \cdot \vec{B}$. Note how it is identical to the force on an electric dipole in an electric field, Equation 3.252. We'll address below the implication that the magnetic field can do work on the dipole.

Torque on a Magnetic Dipole (à la Jackson)

We may obtain from the Lorentz Force Law on a current distribution the corresponding torque:

$$\vec{N}_{mag} = \int_{\mathcal{V}} d\tau \, \vec{r} \times \left[\vec{J}(\vec{r}) \times \vec{B}(\vec{r}) \right] \tag{5.145}$$

where we have just added up the torque volume element by volume element in the same way we summed the force. When we Taylor expand the magnetic field, we have

$$\vec{N}_{mag} = \int_{\mathcal{V}} d\tau \, \vec{r} \times \left[\vec{J}(\vec{r}) \times \vec{B}(\vec{0}) \right] + \cdots \tag{5.146}$$

Because of the $\vec{r} \times$ inside the integrand, the zeroth-order term no longer vanishes and so we do not need to consider the next order term in the Taylor expansion. We will write the zeroth-order term as $\vec{N}_{dip}$ below for reasons that will become clear.

To get the above expression into a familiar form, we need to repeat the same kinds of vector arithmetic tricks we have used before. First, we apply the $BAC - CAB$ rule, $\vec{a} \times (\vec{b} \times \vec{c}) = \vec{b}(\vec{a} \cdot \vec{c}) - \vec{c}(\vec{a} \cdot \vec{b})$, which we can do without having to write things in terms of indices because there are no derivatives floating around:

$$\vec{N}_{dip} = \int_{\mathcal{V}} d\tau \; \vec{r} \times \left[\vec{J}(\vec{r}) \times \vec{B}(\vec{0}) \right] = \int_{\mathcal{V}} d\tau \; \vec{J}(\vec{r}) \left[\vec{r} \cdot \vec{B}(\vec{0}) \right] - \int_{\mathcal{V}} d\tau \; \vec{B}(\vec{0}) \left[\vec{r} \cdot \vec{J}(\vec{r}) \right] \tag{5.147}$$

We can make the second term vanish by the same kinds of tricks we used earlier during the vector potential multipole expansion:

$$\vec{r} \cdot \vec{J}(\vec{r}) = \left[r \, \vec{\nabla} r \right] \cdot \vec{J}(\vec{r}) = \frac{1}{2} \left[\vec{\nabla} r^2 \right] \cdot \vec{J}(\vec{r}) = \frac{1}{2} \left\{ \vec{\nabla} \cdot \left[r^2 \vec{J}(\vec{r}) \right] - r^2 \, \vec{\nabla} \cdot \vec{J}(\vec{r}) \right\} \tag{5.148}$$

In this expression, the second term vanishes under the steady-state assumption, and the first term can be turned into a surface integral with integrand $r^2 \vec{J}(\vec{r})$. Since we are considering a localized current distribution, the surface can be taken far enough out that $\vec{J}(\vec{r})$ vanishes on the surface.

The first term looks again like the expression we have encountered in Equation 5.107, which becomes apparent when we write it out in component form:

$$\vec{N}_{dip} = \sum_{i,j=1}^{3} \hat{r}_i \, B_j(\vec{0}) \int_{\mathcal{V}} d\tau \, J_i(\vec{r}) \, r_j \tag{5.149}$$

We again apply the same tricks used to arrive at Equation 5.116:

$$\vec{N}_{dip} = \sum_{i,j=1}^{3} \hat{r}_i \, B_j(\vec{0}) \int_{\mathcal{V}} d\tau \, \frac{1}{2} \, \epsilon_{ijk} \left[\vec{J}(\vec{r}) \times \vec{r} \right]_k = -\frac{1}{2} \, \vec{B}(\vec{0}) \times \int_{\mathcal{V}} d\tau \, \vec{r} \times \vec{J}(\vec{r}) \tag{5.150}$$

$$= -\vec{B}(\vec{0}) \times \vec{m} \qquad \text{with} \qquad \vec{m} = \frac{1}{2} \int_{\mathcal{V}} d\tau \left[\vec{r} \times \vec{J}(\vec{r}) \right] \tag{5.151}$$

Generalizing to a multipole distribution centered on an arbitrary point, the zeroth-order term in the torque is (and hence the $_{dip}$ subscript)

$$\boxed{\vec{N}_{dip} = \vec{m} \times \vec{B}(\vec{r}) \qquad \text{with} \qquad \vec{m} = \frac{1}{2} \int_{\mathcal{V}} d\tau' \left[\vec{r}' \times \vec{J}(\vec{r}') \right]} \tag{5.152}$$

The magnetic dipole feels a torque that tends to align it with the magnetic field (the torque vanishes when $\vec{m}$ is aligned with $\vec{B}$), again like the situation for an electric dipole in an electric field.

Potential Energy of a Magnetic Dipole

We can do the line integral of the force or the angular integral of the torque to determine that we can write a potential energy

$$\boxed{U(\vec{r}) = -\vec{m} \cdot \vec{B}(\vec{r})} \tag{5.153}$$

This form for the potential energy expresses two features of magnetic dipoles: they like to be aligned with the local magnetic field, and they seek the region of largest $\vec{m} \cdot \vec{B}$.

The thing that should be concerning about this expression is that we argued earlier that magnetic fields can do no work, yet here we have the possibility of such work. That is because we are assuming $\vec{m}$ is held fixed. For a finite current loop, there must be a battery doing work to keep the current fixed as $\vec{m}$ moves or turns relative to $\vec{B}$: such motion yields changing magnetic fields, which, as you know from Ph1c, generate voltages around the loop in which the current for $\vec{m}$ flows. The battery will be the thing doing the work to counter these voltages and keep the current flowing. If $\vec{m}$ is a property of a fundamental particle, then there is no explicit battery: it is simply an empirical fact that $|\vec{m}|$ cannot change, and one that we must incorporate as a postulate.

Section 6
Magnetostatics in Matter

6.1 Paramagnetism and Diamagnetism
6.2 The Field of a Magnetized Object
6.3 The Auxiliary Field $\vec{H}$ and Magnetic Permeability
6.4 Boundary Value Problems in Magnetic Materials
6.5 Nonlinear Materials and Ferromagnetism

Paramagnetism and Diamagnetism

See Griffiths Sections 6.1.1 and 6.1.3 and Purcell Sections 11.1 and 11.5 for discussions of paramagnetism and diamagnetism. This will be discussed in class briefly, but there is little to add to their discussions, but we briefly summarize that discussion here.

Diamagnetism arises due to the orbital angular momentum of electrons: when they execute their orbits in a magnetic field, the magnetic field adds to ($\vec{B}||\vec{\omega}|| - \vec{m}$) or partially cancels ($\vec{B}|| - \vec{\omega}||\vec{m}$) the Coulomb force that provides the centripetal force. This causes the electron velocity to, respectively, increase or decrease. In both cases, this change in speed yields a change $\Delta \vec{m}|| - \vec{B}$. We will see later that an ensemble of magnetic dipoles that change in this way yield an additional field that partially cancels the applied magnetic field, hence the term diamagnetism.

Paramagnetism arises from an entirely quantum-mechanical phenomenon: electrons have a magnetic moment antiparallel to their spin and, when electrons are unpaired in atomic orbitals, those magnetic moments are uncanceled. We saw in the previous section that magnetic dipoles align with a magnetic field, opposite to the effect seen in diamagnetism. We will see that an ensemble of such aligning magnetic moments yields a field in the same direction as the applied field, hence the term paramagnetism.

We remind you that unpaired electrons are present not just in atoms with odd numbers of electrons but also in electrons with even Z in which there are not enough electrons to fill a particular atomic orbital shell (ℓ) that has multiple ℓ_z values. Because of Coulomb repulsion, such electrons tend to spread out among different ℓ_z values, and then there is an "exchange" term that we will discuss later that causes the lower-energy state to be the one in which the spatial wavefunction is antisymmetric under exchange, which then implies the spin state must be symmetric under exchange in order for the entire state to be antisymmetric under exchange.

The Field of a Magnetized Object

Bound Currents

Suppose we have an object with a position-dependent macroscopic density of magnetic moments, or *macroscopic magnetization density* $\vec{M}(\vec{r})$, where the magnetic moment of an infinitesimal volume $d\tau$ is

$$d\vec{m} = \vec{M}(\vec{r})\, d\tau \tag{6.1}$$

$\vec{M}$ is not to be confused with the magnetization density $\mathcal{M}(\vec{r})$; the latter can be for some arbitrary current distribution, while the former is specifically to be considered to be a density of magnetic dipole moments. $\mathcal{M}(\vec{r})$ should give $\vec{M}(\vec{r})$ for this special case of pure dipoles. We will, confusingly, drop "macroscopic" from here on out. Assuming we are looking at the dipoles from a macroscopic enough scale that the dipole approximation is valid, we may use our expression for the vector potential of a magnetic dipole, Equation 5.120, to calculate the contribution to the vector potential at $\vec{r}$ due to the above infinitesimal volume at $\vec{r}'$:

$$d\vec{A}(\vec{r}) = \frac{\mu_o}{4\pi} \frac{d\vec{m}(\vec{r}') \times (\vec{r} - \vec{r}')}{|\vec{r} - \vec{r}'|^3} = \frac{\mu_o}{4\pi} \frac{d\tau'\, \vec{M}(\vec{r}') \times (\vec{r} - \vec{r}')}{|\vec{r} - \vec{r}'|^3} \tag{6.2}$$

Integrating over the volume containing the magnetization density, we have

$$\vec{A}(\vec{r}) = \frac{\mu_o}{4\pi} \int_{\mathcal{V}} d\tau' \, \frac{\vec{M}(\vec{r}') \times (\vec{r} - \vec{r}')}{|\vec{r} - \vec{r}'|^3} \tag{6.3}$$

Now, we use $(\vec{r} - \vec{r}')/|\vec{r} - \vec{r}'|^3 = \vec{\nabla}_{\vec{r}'}|\vec{r} - \vec{r}'|^{-1}$ (note that the gradient is with respect to $\vec{r}'$, not $\vec{r}$!), which allows us to apply the product rule for curl, $\vec{\nabla} \times (f\vec{a}) = f\vec{\nabla} \times \vec{a} - \vec{a} \times \vec{\nabla} f$:

$$\vec{A}(\vec{r}) = \frac{\mu_o}{4\pi} \int_{\mathcal{V}} d\tau' \, \vec{M}(\vec{r}') \times \vec{\nabla}_{\vec{r}'} \left(\frac{1}{|\vec{r} - \vec{r}'|} \right) \tag{6.4}$$

$$= \frac{\mu_o}{4\pi} \int_{\mathcal{V}} d\tau' \, \frac{\vec{\nabla}_{\vec{r}'} \times \vec{M}(\vec{r}')}{|\vec{r} - \vec{r}'|} - \frac{\mu_o}{4\pi} \int_{\mathcal{V}} d\tau' \, \vec{\nabla}_{\vec{r}'} \times \left(\frac{\vec{M}(\vec{r}')}{|\vec{r} - \vec{r}'|} \right) \tag{6.5}$$

$$= \frac{\mu_o}{4\pi} \int_{\mathcal{V}} d\tau' \, \frac{\vec{\nabla}_{\vec{r}'} \times \vec{M}(\vec{r}')}{|\vec{r} - \vec{r}'|} + \frac{\mu_o}{4\pi} \int_{\mathcal{S}(\mathcal{V})} da' \, \frac{\vec{M}(\vec{r}') \times \hat{n}(\vec{r}')}{|\vec{r} - \vec{r}'|} \tag{6.6}$$

where, in the last step, we have used a vector identity that we will prove on the following slide.

Let's prove the vector identity we just used, which is a corollary of the divergence theorem for the curl. Let $\vec{a}(\vec{r})$ be an arbitrary vector field and let $\vec{c}$ be an arbitrary constant vector. Then the divergence theorem tells us

$$\int_{\mathcal{V}} d\tau \, \vec{\nabla} \cdot [\vec{a}(\vec{r}) \times \vec{c}] = \oint_{\mathcal{S}(\mathcal{V})} da \, \hat{n}(\vec{r}) \cdot [\vec{a}(\vec{r}) \times \vec{c}] \tag{6.7}$$

Now, apply the cyclicity of triple scalar products (along with the fact that $\vec{c}$ is constant and can thus it can be moved past $\vec{\nabla}$) and bring $\vec{c}$ outside the integrals (since it is a constant vector):

$$\vec{c} \cdot \int_{\mathcal{V}} d\tau \left[\vec{\nabla} \times \vec{a}(\vec{r}) \right] = \vec{c} \cdot \oint_{\mathcal{S}(\mathcal{V})} da \, [\hat{n}(\vec{r}) \times \vec{a}(\vec{r})] \tag{6.8}$$

Since $\vec{c}$ is arbitrary, the expression must hold for any $\vec{c}$ and thus:

$$\int_{\mathcal{V}} d\tau \left[\vec{\nabla} \times \vec{a}(\vec{r}) \right] = \oint_{\mathcal{S}(\mathcal{V})} da \, [\hat{n}(\vec{r}) \times \vec{a}(\vec{r})] \tag{6.9}$$

which is what we wanted to prove.

Making some definitions, we recognize that the vector potential can be considered to be sourced by a *bound volume current density* $\vec{J}_b(\vec{r})$ and a *bound surface current density* $\vec{K}_b(\vec{r})$:

$$\boxed{\vec{J}_b(\vec{r}) = \vec{\nabla} \times \vec{M}(\vec{r}) \qquad \vec{K}_b(\vec{r}) = \vec{M}(\vec{r}) \times \hat{n}(\vec{r})}$$ (6.10)

$$\boxed{\vec{A}(\vec{r}) = \frac{\mu_o}{4\pi} \int_{\mathcal{V}} d\tau' \, \frac{\vec{J}_b(\vec{r}')}{|\vec{r} - \vec{r}'|} + \frac{\mu_o}{4\pi} \oint_{\mathcal{S}(\mathcal{V})} da' \, \frac{\vec{K}_b(\vec{r}')}{|\vec{r} - \vec{r}'|}}$$ (6.11)

The way in which these current densities source $\vec{A}$ is identical to the way in which free current densities do. Moreover, we can see the clear analogy to bound volume and surface charges in the case of polarized materials.

Griffiths Section 6.2.2 gives a nice discussion of the physical interpretation of bound currrents that will be presented in class, but there is not much to add here.

Example 6.1: Uniformly Magnetized Sphere

Center the sphere of radius R at the origin. Let $\vec{M} = M\,\hat{z}$. Then

$$\vec{J}_b(\vec{r}) = \vec{\nabla} \times M\,\hat{z} = 0 \qquad \vec{K}_b(\vec{r}) = M\,\hat{z} \times \hat{n} = M\,\hat{z} \times \hat{r} = \hat{\phi}\,M\sin\theta \qquad (6.12)$$

We need to calculate

$$\vec{A}(\vec{r}) = \frac{\mu_o}{4\pi}\,R^2 \int_0^{2\pi} d\phi' \int_0^{\pi} d\theta'\,\sin\theta'\,\frac{M\sin\theta'\,\hat{\phi}}{|\vec{r} - \vec{r}'|} \qquad (6.13)$$

$$= \frac{\mu_o}{4\pi}\,R^2 \int_0^{2\pi} d\phi' \int_0^{\pi} d\theta'\,\sin\theta'\,\frac{M\sin\theta'\,(-\hat{x}\sin\phi' + \hat{y}\cos\phi')}{|\vec{r} - \vec{r}'|} \qquad (6.14)$$

(The R^2 out front is because an area integral, not just a solid angle integral, needs to be done.) Recall that we did this calculation when we considered the magnetic vector potential of a spinning spherical shell with a uniform charge density, Example 5.1. In that case the surface current density was $\vec{K} = \hat{\phi}\,\sigma\,\omega\,R\,\sin\theta$, so we can use that result with $\sigma\,\omega\,R$ replaced by M. The result is

$$\vec{A}(r \leq R, \theta, \phi) = \frac{\mu_o}{3}\,M\,r\,\sin\theta\,\hat{\phi} \qquad \vec{A}(r \geq R, \theta, \phi) = \frac{\mu_o}{4\pi}\left(\frac{4\pi}{3}\,R^3 M\right)\frac{\sin\theta}{r^2}\,\hat{\phi} \quad (6.15)$$

Evaluating the curl of the first term to obtain the magnetic field, we have inside the sphere

$$\vec{B}(r \leq R) = \vec{\nabla} \times \vec{A}(r \leq R) = \frac{1}{3}\,\mu_o\,M\left[2\,\widehat{r}\,\cos\theta - \widehat{\theta}\,\sin\theta\right] = \frac{2}{3}\,\mu_o\,\vec{M} \qquad (6.16)$$

which is a uniform field pointing in the *same* direction as the magnetization.

For $r \geq R$, we have

$$\vec{A}(r \geq R) = \frac{\mu_o}{4\,\pi}\,\frac{\vec{m} \times \widehat{r}}{r^2} \qquad \vec{m} = \frac{4\,\pi}{3}\,R^3\vec{M} \qquad (6.17)$$

which is the vector potential (thus yielding the field of) a pure dipole with magnetic moment given by integrating the uniform magnetization density over the sphere. This form is exact for *all $r \geq R$.*

Let's compare to the case of a uniformly polarized dielectric sphere:

$$r \leq R \qquad \vec{E}(\vec{r}) = -\frac{1}{3\,\epsilon_o}\,\vec{P} \qquad\qquad \vec{B}(\vec{r}) = \frac{2}{3}\,\mu_o\,\vec{M} \qquad (6.18)$$

$$r \geq R \qquad V(\vec{r}) = \frac{1}{4\,\pi\,\epsilon_o}\,\frac{\vec{p}\cdot\hat{r}}{r^2} \qquad\qquad \vec{A}(\vec{r}) = \frac{\mu_o}{4\,\pi}\,\frac{\vec{m}\times\hat{r}}{r^2} \qquad (6.19)$$

$$\vec{p} = \frac{4\,\pi}{3}\,R^3\vec{P} \qquad\qquad \vec{m} = \frac{4\,\pi}{3}\,R^3\vec{M} \qquad (6.20)$$

Inside the sphere, the difference is a factor of -2 and the exchange of $1/\epsilon_o$ for μ_o. Outside the sphere, the two potentials result in fields identical up to the replacement of $\vec{P}/\epsilon_o$ by $\mu_o\,\vec{M}$. The difference in the $r \leq R$ expressions reflects the fact that the magnetic field of the bound surface current (i.e., of $\vec{M}$) is aligned with $\vec{M}$ while the electric field of the surface bound charge density (i.e., of $\vec{P}$) is opposite to $\vec{P}$. This sign difference is a generic phenomenon, resulting in the very different behavior of electrostatic and magnetostatic fields in matter.

The Auxiliary Field $\vec{H}$ and Magnetic Permeability

Definition of the Auxiliary Field

We saw that $\vec{A}$ is sourced by the bound current density $\vec{J_b} = \vec{\nabla} \times \vec{M}$ in the same way it would be sourced by a *free* current density $\vec{J_f}$. Therefore, Ampere's Law is satisfied with the sum of the two currents:

$$\frac{1}{\mu_o} \vec{\nabla} \times \vec{B} = \vec{J_f} + \vec{J_b} = \vec{J_f} + \vec{\nabla} \times \vec{M} \tag{6.21}$$

If we want to write an Ampere's Law in terms of the free currents only, in the same way that we wanted to write Gauss's Law in terms of the free charges only, then we can define the *auxiliary field*

$$\boxed{\vec{H} \equiv \frac{\vec{B}}{\mu_o} - \vec{M}} \tag{6.22}$$

In contrast to electrostatics, where the displacement field was the sum of the electric field and the polarization density, here the auxiliary field is the difference of the magnetic field and the magnetization density. The sign flip comes from the differing signs in the definition of the bound charge and current densities: $\rho_b = -\vec{\nabla} \cdot \vec{P}$ while $\vec{J_b} = \vec{\nabla} \times \vec{M}$, which itself comes from the commutative vs. anticommutative natures of the dot and cross product.

With this definition of $\vec{H}$, we then have

$$\vec{\nabla} \times \vec{H} = \frac{1}{\mu_o} \vec{\nabla} \times \vec{B} - \vec{\nabla} \times \vec{M} = \vec{J_f} + \vec{J_b} - \vec{J_b} = \vec{J_f} \qquad (6.23)$$

Therefore, we have a modified Ampere's Law

$$\boxed{\vec{\nabla} \times \vec{H} = \vec{J_f} \qquad \Longleftrightarrow \qquad \oint_{\mathcal{C}} d\vec{\ell} \cdot \vec{H}(\vec{r}) = \int_{\mathcal{S}(\mathcal{C})} da \, \widehat{n}(\vec{r}) \cdot \vec{J_f}(\vec{r}) = I_{f,enc}} \qquad (6.24)$$

Thus, as intended, we have an Ampere's Law in terms of the free currents only, which (partially) source $\vec{H}$. The fact that $\vec{H}$ satisfies Ampere's Law in the free current leads some to use the name *applied field* for it. That may be misleading, though, because the free current does not tell one everything one must know to determine $\vec{H}$ (in the same way that ρ_f does not completely determine the displacement field $\vec{D}$).

To fully specify $\vec{H}$, we need to know its divergence, which is given by applying $\vec{\nabla} \cdot \vec{B} = 0$:

$$\boxed{\vec{\nabla} \cdot \vec{H} = -\vec{\nabla} \cdot \vec{M}} \qquad (6.25)$$

This nonvanishing of $\vec{\nabla} \cdot \vec{H}$ is analogous to the nonvanishing of $\vec{\nabla} \times \vec{D}$ in electrostatics.

There is an example of how to calculate $\vec{H}$ using the above Ampere's Law in Griffiths Example 6.2.

Boundary Conditions on $\vec{H}$

From the boundary conditions on $\vec{B}$ at an interface, we can derive boundary conditions on $\vec{H}$. The continuity of the normal component of the magnetic field (Equation 5.76) along with Equation 6.22 implies

$$\boxed{\widehat{n}(\vec{r}) \cdot \left[\vec{H}_2(\vec{r}) - \vec{H}_1(\vec{r}) \right] = -\widehat{n}(\vec{r}) \cdot \left[\vec{M}_2(\vec{r}) - \vec{M}_1(\vec{r}) \right]} \tag{6.26}$$

Applying the same arguments using Ampere's Law for $\vec{H}$ as we did using Ampere's Law for $\vec{B}$, we can also conclude the analogy of Equation 5.82:

$$\boxed{\left[\vec{H}_2(\vec{r}) - \vec{H}_1(\vec{r}) \right] \cdot \widehat{s}(\vec{r}) = \left[\vec{K}_f(\vec{r}) \times \widehat{n}(\vec{r}) \right] \cdot \widehat{s}(\vec{r})} \tag{6.27}$$

where $\vec{K}_f$ is the free surface current density at the interface.

Recall that we found alternative forms of the corresponding boundary conditions for $\vec{B}$, Equations 5.83 and 5.84:

$$\vec{B}_2(\vec{r}) - \vec{B}_1(\vec{r}) = \mu_o \, \vec{K}(\vec{r}) \times \widehat{n}(\vec{r})$$

$$\widehat{n}(\vec{r}) \times \left[\vec{B}_2(\vec{r}) - \vec{B}_1(\vec{r}) \right] = \mu_o \, \vec{K}(\vec{r})$$

There is no trivial analogue of the first one because it relied on the normal component of $\vec{B}$ being continuous. However, we can obtain the analogue of the second equation, though we have to do it in a different way because, for $\vec{B}$, we used the first equation above to obtain the second one.

We start by using $\widehat{s} = \widehat{t} \times \widehat{n}$ and then applying the cyclicity of the triple scalar product on both sides:

$$\left[\vec{H}_2 - \vec{H}_1 \right] \cdot \left[\widehat{t} \times \widehat{n} \right] = \left[\widehat{n} \times \widehat{s} \right] \cdot \vec{K}_f \tag{6.28}$$

$$\widehat{t} \cdot \left(\widehat{n} \times \left[\vec{H}_2 - \vec{H}_1 \right] \right) = \widehat{t} \cdot \vec{K}_f \tag{6.29}$$

The same equation holds trivially with $\widehat{t}$ replaced by $\widehat{n}$: the left side vanishes because $\widehat{n}$ is perpendicular to any cross product involving $\widehat{n}$ and the right side vanishes because $\vec{K}_f$ is always perpendicular to $\widehat{n}$. This, combined with the fact that $\widehat{t}$ in the above can be any vector in the plane of the boundary, implies the more general statement

$$\boxed{\widehat{n}(\vec{r}) \times \left[\vec{H}_2(\vec{r}) - \vec{H}_1(\vec{r}) \right] = \vec{K}_f(\vec{r})} \tag{6.30}$$

which is the analogue of the second equation on the previous slide. But note that this equation provides no information about the normal component of $\vec{H}$ because it is related to the normal component of $\vec{M}$.

What Sources $\vec{H}$? When Does It Vanish?

Considering the uniformly magnetized sphere example we just looked at, we see

$$\vec{H}(r < R) = \frac{\vec{B}(r \leq R)}{\mu_o} - \vec{M} = \frac{2}{3}\vec{M} - \vec{M} = -\frac{1}{3}\vec{M} \tag{6.31}$$

$$\vec{H}(r > R) = \frac{\vec{B}(r \geq R)}{\mu_o} = \frac{\text{field of the magnetic dipole } \vec{m} = \frac{4\pi}{3}R^3\vec{M}}{\mu_o} \tag{6.32}$$

This example highlights the importance of the nonvanishing of $\vec{\nabla} \cdot \vec{H}$. There is no free current in this problem, so one might be inclined to think $\vec{H}$ vanishes by analogy to the fact $\vec{B}$ would vanish if there were no total current. But the nonzero nature of $\vec{\nabla} \cdot \vec{H}$ means that $\vec{H}$ has another sourcing term that is not captured by Ampere's Law alone. In this case, this sourcing term manifests as a discontinuity of the normal component of $\vec{M}$ at $r = R$. This is analogous to the way that, even if there is no free charge, there may be a displacement field $\vec{D}$, sourced by $\vec{\nabla} \times \vec{P}$ and/or a discontinuity in the tangential component of $\vec{P}$, as we saw in for the polarized sphere (Example 5.1) and the spherical cavity in a dielectric with uniform field applied (Example 5.6). To have $\vec{H}$ vanish identically, one needs to have $\vec{\nabla} \cdot \vec{M} = 0$ and also trivial boundary conditions on $\vec{M}$ (no change in $\hat{n} \cdot \vec{M}$).

This all makes sense given the Helmholtz theorem: since $\vec{\nabla} \cdot \vec{H}$ does not vanish, $\vec{H}$ is not just the curl of a vector potential, but must be the sum of the gradient of a scalar potential and the curl of a vector potential. Ampere's Law for $\vec{H}$ tells us that the free current density sources the vector potential, while $-\vec{\nabla} \cdot \vec{M}$ sources the scalar potential. We will see later that the latter point allows us to use our electrostatic boundary value problem techniques.

In particular, in the example of the uniformly magnetized sphere, we see that $\vec{H}$ is identical in form to $\vec{E}$ from the uniformly polarized sphere up to the replacement $\vec{P}/\epsilon_o \to \vec{M}$, so the scalar potential that yields $\vec{M}$ will have the same form, up to this replacement, as the scalar potential that yields $\vec{E}$. We'll pursue this analogy in detail when we discuss boundary value problems for magnetostatic systems.

We can make the same point about ρ_f not being the only source of $\vec{D}$; when $\vec{\nabla} \times \vec{P}$ is nonzero, then $\vec{D}$ receives an additional sourcing term. It was not convenient to make this point when we discussed $\vec{D}$ initially because we had not yet learned about vector potentials and how to discuss sourcing of $\vec{D}$ by a vector field, $\vec{\nabla} \times \vec{P}$. But now we do, and so it should be clear that $\vec{D}$ received a contribution that is sourced by $\vec{\nabla} \times \vec{P}$ in the same way that $\vec{H}$ receives a contribution that is sourced by $\vec{\nabla} \times \vec{H} = \vec{J}_f$.

In particular, in Example 4.5, the capacitor with two side-by-side dielectrics, we saw such a situation, manifested by the discontinuity in the tangential component of $\vec{P}$ at the interface between the two dielectrics.

Who Cares About $\vec{H}$?

Is $\vec{H}$ any more useful than $\vec{D}$ was?

The thing that limits the utility of $\vec{D}$ is that, in practice, one rarely controls free charge, the most obvious source for $\vec{D}$. In practice, one sets potentials using batteries or other voltage sources. Potentials specify $\vec{E}$, not $\vec{D}$. Consider the example of the parallel-plate capacitor with side-by-side dielectrics: σ_f ended being an output of the calculation after calculating $\vec{E}$ rather than an input that yielded $\vec{D}$.

On the other hand, $\vec{H}$ is sourced by the free currents, which is the thing one explicitly controls in the lab. For that reason alone, we expect $\vec{H}$ is of greater utility than $\vec{D}$. We will see this more clearly when we consider specific types of permeable materials.

The other reason we will find $\vec{H}$ more useful is that, in reality, we frequently come across ferromagnets, where $\vec{M}$ is provided and thus we are given the $\vec{\nabla} \cdot \vec{M}$ source for $\vec{H}$, but we rarely encounter ferroelectrics, where $\vec{P}$ and thus the $\vec{\nabla} \times \vec{P}$ source for $\vec{D}$ are provided. We would find $\vec{D}$ useful as a calculation tool if we were given a system in which $\vec{\nabla} \times \vec{P}$ were nonzero or, more likely, $\vec{P}$ were tangent to boundaries between a ferroelectric and vacuum or between different ferroelectrics. Then $\vec{\nabla} \times \vec{P}$ and any discontinuity in $\hat{n} \times \vec{P}$ would source $\vec{D}$ in the same way that $\vec{J}$ and a boundary $\vec{K}$ source $\vec{B}$.

Lecture 18:

Magnetostatics in Matter II:
Magnetic Permeability in Linear Systems
Boundary Value Problems in Magnetostatics

Date Revised: 2024/03/05 06:00
Date Given: 2024/03/05

Magnetic Permeability in Linear Materials

Many magnetic materials we will consider have a linear relationship between the field and the magnetization. The *magnetic susceptibility* of a material is defined to be the constant of proportionality between $\vec{M}$ and $\vec{H}$:

$$\vec{M} = \chi_m \, \vec{H} \tag{6.33}$$

(One can see why $\vec{H}$ is sometimes called the applied field!) Since $\vec{B} = \mu_o \left(\vec{H} + \vec{M} \right)$, we have

$$\vec{B} = \mu_o \left(\vec{H} + \vec{M} \right) = \mu_o \left(1 + \chi_m \right) \vec{H} \equiv \mu \vec{H} \tag{6.34}$$

where we have defined the *magnetic permeability* $\mu = \mu_o(1 + \chi_m)$. The quantity $\mu_r = 1 + \chi_m$ is the *relative permeability*. The definition of χ_m and μ follows a different convention than the definition of χ_e and ϵ. This is for the reason we discussed above: we experimentally control the free current and thus $\vec{H}$, whereas in electrostatics we control the voltages and thus $\vec{E}$. We define the permittivity and the permeability to be the constant of proportionality relating the thing we do control to the thing we do not control.

Paramagnetic materials have $\chi_m > 0$ because the magnetization is in the same direction as the field and so the field due to the free currents is added to by the field from the magnetization.

Diamagnetic materials have $\chi_m < 0$ because the magnetization is in the direction opposite the field and so the field due to the free currents is partially canceled by the field from the magnetization.

For electrostatics in matter, we were concerned entirely with *dielectric* materials: because every atom has some polarizability, every material is dielectric to some extent. In that case, the "di" prefix went with $\chi_e > 0$ (in contrast to $\chi_m < 0$ here) because of the different convention for the relation between $\vec{E}$ and $\vec{D}$.

Diamagnetic materials exist via the same kind of classical argument, now involving the response of currents in materials to applied fields.

The analogous *paraelectric* materials ($\chi_e < 0$) do not exist for the most part — it is hard to understand how one can get an electrically polarizable material to have $\chi_e < 0$. Metals can have negative permittivity at high frequencies (optical), but not DC.

Paramagnetic materials exist only because of quantum mechanics — the existence of magnetic moments not caused by an applied field. There are no such quantum-mechanics-caused electric dipole moments, mainly because such moments violate time-reversal symmetry while magnetic moments do not. They work differently because the "current" sourcing a magnetic dipole moment reverses sign under time reversal while the charges sourcing an electric dipole moment do not.

Example 6.2: Magnetizable Rod with Uniform Current (Griffiths 6.17)

Let's consider a rod of radius R whose axis is in the z direction and which carries a current $\vec{I} = I\,\hat{z}$ distributed uniformly across its cross section. Assume the material is linear with magnetic susceptibility χ_m. Let's find $\vec{H}$, $\vec{M}$, and $\vec{B}$.

Let's first see how far we can get without using χ_m. Ampere's Law for $\vec{H}$ tells us

$$\oint_{\mathcal{C}} \vec{H} \cdot d\vec{\ell} = \int_{\mathcal{S}(\mathcal{C})} da\,\hat{n} \cdot \vec{J_f} \tag{6.35}$$

This system has azimuthal symmetry as well as translational symmetry in z, so we can guess $\vec{H} = \vec{H}(s)$ where s is the radial coordinate in cylindrical coordinates. By the right-hand rule and the z translational symmetry, we expect $\vec{H} = H(s)\,\hat{\phi}$. This eliminates any concern about $\vec{\nabla} \cdot \vec{M}$ or $\hat{n} \cdot \vec{M}$: we know $\vec{M} = \chi_m \vec{H} \propto \vec{H}$, therefore we know, for the assumed form of $\vec{H}$, $\vec{\nabla} \cdot \vec{M} = 0$ inside the cylinder and $\hat{n} \cdot \vec{M} = 0$ at the surface of the cylinder. ($\vec{M} = 0$ outside the cylinder.) Adding in that we know $\vec{J_f} = \hat{z}\,I/\pi R^2$, Ampere's Law in $\vec{J_f}$ and $\vec{H}$ tells us

$$s \leq R: \qquad 2\pi s\,H(s) = \pi s^2\,\frac{I}{\pi R^2} \qquad \Longleftrightarrow \qquad \vec{H}(s) = \frac{I}{2\pi}\,\frac{s}{R^2}\,\hat{\phi} \tag{6.36}$$

$$s \geq R: \qquad 2\pi s\,H(s) = \pi R^2\,\frac{I}{\pi R^2} \qquad \Longleftrightarrow \qquad \vec{H}(s) = \frac{I}{2\pi}\,\frac{1}{s}\,\hat{\phi} \tag{6.37}$$

If we do not know χ_m, we do not know $\vec{M}$ inside the material and so we cannot calculate $\vec{B}$ for $s \leq R$. For $s \geq R$, we have vacuum and so $\vec{M} = 0$ and $\vec{B} = \mu_o \vec{H}$:

$$\vec{B}(s \geq R) = \mu_o \vec{H}(s \geq R) = \frac{\mu_o \, \mathrm{I}}{2 \, \pi \, s} \, \widehat{\phi} \tag{6.38}$$

Note that $\vec{B}(s \geq R)$ is unaffected by the presence of the magnetizable material — this is the same field we would have had with $\chi_m = 0$ inside the rod. We will see why below.

Next, if we use the linearity of the material with susceptibility χ_m, we have

$$\vec{M}(s \leq R) = \chi_m \, \vec{H}(s \leq R) = \chi_m \frac{\mathrm{I}}{2 \, \pi} \frac{s}{R^2} \, \widehat{\phi} = \frac{\mu - \mu_o}{\mu_o} \frac{\mathrm{I}}{2 \, \pi} \frac{s}{R^2} \, \widehat{\phi} \tag{6.39}$$

and therefore

$$\vec{B}(s \leq R) = \mu \, \vec{H}(s \leq R) = \frac{\mu \, \mathrm{I}}{2 \, \pi} \frac{s}{R^2} \, \widehat{\phi} \tag{6.40}$$

All three fields are azimuthal inside and outside R. For paramagnetic materials, $\chi_m \geq 0$ ($\mu \geq \mu_o$), so $\vec{M}$ is parallel to $\vec{H}$ and $|\vec{B}| > \mu_o |\vec{H}|$ inside R. For diamagnetic materials, $\chi_m < 0$ ($\mu \leq \mu_o$), so $\vec{M}$ is antiparallel to $\vec{H}$ and $|\vec{B}| \leq \mu_o |\vec{H}|$ inside R.

Let's check the boundary conditions. All the fields are tangential at the boundary, so continuity of the normal component of $\vec{B}$ trivially satisfied (Equation 5.76), as is the equality of the change in the normal components of $\vec{H}$ and $-\vec{M}$ (Equation 6.26). There is no free surface current density, so we expect the tangential component of $\vec{H}$ to be continuous. We see this indeed holds, with value

$$\widehat{\phi} \cdot \vec{H}(s = R) = \frac{I}{2 \pi R} \tag{6.41}$$

The $\widehat{z}$ tangential component is trivially continuous since it vanishes.

For the sake of completeness, let's calculate the *bound* surface current and check that the tangential boundary condition on $\vec{B}$ is correct. The bound surface current is $\vec{K}_b = \vec{M} \times \hat{n}$ (Equation 6.10). In this case, $\hat{n} = \hat{s}$, the radial unit vector in cylindrical coordinates, so

$$\vec{K}_b(s = R) = M(s = R)\,\hat{\phi} \times \hat{s} = -\chi_m \frac{I}{2\,\pi\,R}\,\hat{z} \qquad (6.42)$$

For a paramagnetic materials ($\chi_m > 0$), the surface current points along $-\hat{z}$ while, for diamagnetic materials ($\chi_m < 0$), it points along $+\hat{z}$. One can see this physically by considering the direction of alignment of the dipoles and which direction the uncancelled current on the boundary flows. From the direction of this surface current, one can then see that the field of this surface current adds to the field of the free current for the paramagnetic case and partially cancels it for the diamagnetic case. (Note that we get this behavior even though the surface current is antiparallel to (parallel to) the direction of the volume current in the paramagnetic (diamagnetic) case. We get this behavior because the region of interest is *inside* the surface current density, not outside it. We'll look at the field outside next.) With the surface bound current in hand, let's check the boundary conditions on $\vec{B}$. It has no normal component in either region, so continuity of the normal component is trivially satisfied. The discontinuity in the tangential component matches Equation 5.84:

$$\hat{n} \times \left[\vec{B}_2 - \vec{B}_1\right] = \hat{s} \times [\mu_o - \mu]\,\frac{I}{2\,\pi\,R}\,\hat{\phi} = -\mu_o \chi_m \frac{I}{2\,\pi\,R}\,\hat{z} = \mu_o\,\vec{K}_b \qquad (6.43)$$

Let's also calculate the bound volume current density, $\vec{J_b} = \vec{\nabla} \times \vec{M}$ from Equation 6.10. It is

$$\vec{J_b}(\vec{r}) = \vec{\nabla} \times \vec{M} = \chi_m \vec{\nabla} \times \vec{H} = \chi_m \vec{J_f} = \chi_m \frac{I}{\pi R^2} \, \widehat{z} \tag{6.44}$$

For paramagnetic materials, $\vec{J_b}$ is parallel to $\vec{J_f}$ and thus its field adds to the field of the free current, while, for diamagnetic materials, it is antiparallel and it partially cancels the free current's field.

Note that the integral of $\vec{J_b}$ over the cross section and the integral of $\vec{K_b}$ over the circumference are equal in magnitude ($\chi_m I$) and opposite in sign, canceling perfectly. This is why the magnetic field outside the wire is only that due to the free current.

A modest extension to this problem would be to include a free surface current in the $\widehat{z}$ direction, which would then cause a discontinuity in the $\widehat{\phi}$ component of $\vec{H}$. You should try this on your own.

Boundary Value Problems in Magnetic Materials

Griffiths does not really consider boundary value problems in magnetostatics, so we follow Jackson §5.9–5.12.

General Conditions for Linear, Homogeneous Magnetic Materials

In linear, homogeneous dielectrics, we showed $\rho_b \propto \rho_f$. We just saw that a similar relation holds for linear, homogeneous magnetic materials, which we can derive generally:

$$\vec{J}_b = \vec{\nabla} \times \vec{M} = \vec{\nabla} \times \left(\frac{\mu - \mu_o}{\mu_o} \vec{H} \right) = \left(\frac{\mu - \mu_o}{\mu_o} \right) \vec{\nabla} \times \vec{H} = \left(\frac{\mu - \mu_o}{\mu_o} \right) \vec{J}_f \qquad (6.45)$$

In particular, if there is no free current in a linear, homogeneous magnetic material, then there is no bound current either. In such situations, the magnetic field is derivable from a scalar potential and Laplace's Equation holds everywhere there is no free current! Boundary conditions, and matching conditions between regions, will determine $\vec{H}$. We'll explore such situations shortly.

The General Technique

In general, it always holds that

$$\vec{B} = \vec{\nabla} \times \vec{A} \qquad \vec{H} = \vec{H}(\vec{B}) \qquad \vec{\nabla} \times \vec{H} = \vec{J_f} \qquad (6.46)$$

Therefore, one can always write the differential equation

$$\vec{\nabla} \times \vec{H}(\vec{\nabla} \times \vec{A}) = \vec{J_f} \qquad (6.47)$$

If the relation between $\vec{H}$ and $\vec{B}$ is not simple, the above equation may be difficult to solve.

For linear magnetic materials, though, the above reduces to

$$\vec{\nabla} \times \left(\frac{1}{\mu} \, \vec{\nabla} \times \vec{A} \right) = \vec{J_f} \tag{6.48}$$

If we further specify that μ is constant over some region, then in that region we have

$$\vec{\nabla} \times \left(\vec{\nabla} \times \vec{A} \right) = \vec{\nabla} \left(\vec{\nabla} \cdot \vec{A} \right) - \nabla^2 \vec{A} = \mu \, \vec{J_f} \tag{6.49}$$

Finally, if we specify $\vec{\nabla} \cdot \vec{A} = 0$, this simplifies to a component-by-component Poisson Equation:

$$\nabla^2 \vec{A} = -\mu \, \vec{J_f} \tag{6.50}$$

In principle, one can apply the same techniques as we used for solving Poisson's Equation in electrostatics to solve this component by component. Boundary conditions must be specified either directly (recall that we proved that if any one of $\vec{A}$, $\vec{B}$, $\hat{n} \times \vec{A}$, or $\hat{n} \times \vec{B}$ is specified at every point on the boundary, then the resulting field (though not necessarily the vector potential) is unique) or by matching using the conditions on the normal and tangential components at boundaries.

Another technical challenge associated with the above equation is that it only separates cleanly into component-by-component Poisson Equations in Cartesian coordinates. If the current distribution it not naturally represented in Cartesian coordinates (*e.g.*, even a simple circular current loop), then separation of variables may not be feasible. Method of images may work, or one may have to resort to other techniques or numerical solution. None of this technical complication takes away from the fact that there will be a unique solution for each component independently. The technical complication just makes it hard to actually obtain that solution.

Hard Ferromagnets ($\vec{M}$ fixed and $\vec{J}_f = 0$): Magnetostatic Scalar Potential

If there are no free currents, then $\vec{\nabla} \times \vec{H} = 0$ and we are assured that $\vec{H}$ can be derived from a magnetostatic scalar potential. Here, we use $\vec{B} = \mu_o \left(\vec{H} + \vec{M} \right)$ with $\vec{M}$ fixed. Then $\vec{\nabla} \cdot \vec{B} = 0$ gives

$$\vec{\nabla} \cdot \mu_o \left(\vec{H} + \vec{M} \right) = 0 \qquad (6.51)$$

$$-\nabla^2 V_M + \vec{\nabla} \cdot \vec{M} = 0 \qquad (6.52)$$

$$\nabla^2 V_M = -\rho_M \qquad \text{with} \quad \rho_M = -\vec{\nabla} \cdot \vec{M} \qquad (6.53)$$

(note the canceling minus signs in the definitions!) where ρ_M is termed the *magnetostatic charge density*. Note the close similarity to the definition of the bound charge density $\rho_b = -\vec{\nabla} \cdot \vec{P}$ for dielectrics. This analogy also implies the existence of surface magnetostatic charge density $\sigma_M = \hat{n} \cdot \vec{M}$, which will be related to the discontinuity in the normal gradient of V_M in a manner timilar to that for V. This equation can be solved by the standard techniques for solving Poisson's Equation.

Example 6.3: Uniformly Magnetized Sphere, Again

Let's apply the above kind of formalism for the uniformly magnetized sphere, which satisfies the hard ferromagnet condition. Again, $\vec{M} = M\,\hat{z}$. This implies $\rho_M = -\vec{\nabla}\cdot\vec{M} = 0$ and $\sigma_M = \hat{n}\cdot\vec{M} = M\cos\theta$. We solved this same problem before for the uniformly polarized dielectric sphere via separation of variables in spherical coordinates, which yielded Equation 4.15. Making the replacement $P/\epsilon_o \to M$ (because ρ/ϵ_o is replaced by ρ_M in Equation 6.53), we obtain

$$V_M(r \leq R) = \frac{M\,z}{3} \qquad V_M(r \geq R) = \frac{\vec{m}\cdot\hat{r}}{4\,\pi\,r^2} \quad \text{with} \quad \vec{m} = \frac{4\,\pi}{3}\,\pi\,R^3\,\vec{M} \qquad (6.54)$$

$$\vec{H} = -\vec{\nabla}V_M = \begin{cases} -\dfrac{\vec{M}}{3} & r \leq R \\ \vec{H} \text{ field of a magnetic dipole } \vec{m} & r \geq R \end{cases} \qquad (6.55)$$

$$\vec{B} = \mu_o\left(\vec{H} + \vec{M}\right) \implies \vec{B}(r \leq R) = \mu_o\left(-\frac{1}{3}\,\vec{M} + \vec{M}\right) = \frac{2}{3}\,\mu_o\,\vec{M} \qquad (6.56)$$

$$\vec{B}(r \geq R) = \mu_o\,\vec{H} = \vec{B} \text{ field of a magnetic dipole } \vec{m} \qquad (6.57)$$

This matches our previous solution for the magnetic field of this system that we obtained by calculating the vector potential of the bound surface current (Example 7.1).

Hard Ferromagnets ($\vec{M}$ fixed and $\vec{J}_f = 0$) via Vector Potential

We have already done this analysis, yielding Equations 6.10 and 6.11:

$$\vec{J}_b(\vec{r}) = \vec{\nabla} \times \vec{M}(\vec{r}) \qquad \vec{K}_b(\vec{r}) = \vec{M}(\vec{r}) \times \hat{n}(\vec{r})$$

$$\vec{A}(\vec{r}) = \frac{\mu_o}{4\pi} \int_{\mathcal{V}} d\tau' \, \frac{\vec{J}_b(\vec{r}\,')}{|\vec{r} - \vec{r}\,'|} + \frac{\mu_o}{4\pi} \oint_{\mathcal{S}(\mathcal{V})} da' \, \frac{\vec{K}_b(\vec{r}\,')}{|\vec{r} - \vec{r}\,'|}$$

We can, in fact, directly calculate the *field* from the bound currents using the Biot-Savart Law. The approach described above of using the magnetostatic scalar potential for such cases will in general be calculationally easier if the problem is amenable to the techniques for solving Poisson's Equation, but the Biot-Savart Law is certainly always guaranteed to work.

Example 6.4: Uniformly Magnetized Sphere, Again

We don't need to do this again: the above vector potential based on the bound current density (in this case, only a bound surface current density) is exactly how we solved this system before in Example 7.1.

No Free Currents, Linear Materials via Scalar Potential

If there are no free currents, then $\vec{\nabla} \times \vec{H} = 0$ and again we are assured that $\vec{H}$ can be derived from a magnetostatic scalar potential

$$\vec{H} = -\vec{\nabla} V_M(\vec{r}) \tag{6.58}$$

Again, if we know the relationship $\vec{B} = \vec{B}(\vec{H})$, then we can use the divergence equation:

$$\vec{\nabla} \cdot \vec{B} \left(-\vec{\nabla} V_M \right) = 0 \tag{6.59}$$

Again, if the relation between $\vec{H}$ and $\vec{B}$ is not simple, the above equation may be difficult to solve.

Again, though, for the case of linear magnetic materials, we have

$$\vec{\nabla} \cdot \left(\mu \, \vec{\nabla} \, V_M \right) = 0 \tag{6.60}$$

In a region where μ is constant, it can be passed through the divergence and we can reduce this to

$$\nabla^2 V_M = 0 \tag{6.61}$$

We now have Laplace's Equation. Again, boundary conditions and/or matching conditions will allow one to solve for V_M. In a region where μ is constant, we could equally well write $\vec{B} = -\vec{\nabla} U_M$ and solve $\nabla^2 U_M = 0$ with appropriate boundary conditions. In general, we will use V_M because its boundary conditions are simpler: the tangential gradient of U_M can have discontinuities due to bound surface currents, while we will see we can write boundary conditions on V_M without referencing the analogue for V_M, the bound magnetostatic surface charge density.

The importance of boundary conditions should be even more clear in such cases: since there is no source term in the equation, the boundary conditions *entirely* determine the solution.

Example 6.5: Magnetically Permeable Sphere in External Field

This is now a "soft," linear material, where we cannot take $\vec{M}$ to be fixed. But it is a situation with no free currents, so Laplace's Equation holds (except at the $r = R$ boundary, but we develop matching conditions there).

Fortunately, we do not need to solve the boundary value problem from scratch because this problem is directly analogous the case of a dielectrically polarizable sphere in an external electric field. We have the following correspondence:

$$\epsilon_o \vec{E} = -\epsilon_o \vec{\nabla} V \qquad\qquad \vec{H} = -\vec{\nabla} V_M \qquad\qquad (6.62)$$

$$\epsilon_o \nabla^2 V = 0 \qquad\qquad \nabla^2 V_M = 0 \qquad\qquad (6.63)$$

$$\vec{P} = \frac{\epsilon - \epsilon_o}{\epsilon_o} \vec{E} \qquad\qquad \vec{M} = \frac{\mu - \mu_o}{\mu_o} \vec{H} \qquad\qquad (6.64)$$

$$\vec{D} = \epsilon_o \vec{E} + \vec{P} \qquad\qquad \vec{B}/\mu_o = \vec{H} + \vec{M} \qquad\qquad (6.65)$$

$$\epsilon_o \vec{E} \stackrel{r\to\infty}{\longrightarrow} \epsilon_o \vec{E}_0 \qquad\qquad \vec{H} \stackrel{r\to\infty}{\longrightarrow} \vec{B}_0/\mu_o \qquad\qquad (6.66)$$

$$\vec{D} \stackrel{r\to\infty}{\longrightarrow} \epsilon_o \vec{E}_0 \qquad\qquad \vec{B}/\mu_o \stackrel{r\to\infty}{\longrightarrow} \vec{B}_0/\mu_o \qquad\qquad (6.67)$$

We have carefully avoided making correspondences in the above between ρ_b and ρ_M and between σ_b and σ_M because, in both cases, these quantities are not specified ahead of time: there is not permanent polarization, there is only polarization in response to applied field.

Let's also compare the matching conditions. We want to use the matching conditions that incorporate only the free charge densities because we do not know the bound charge densities ahead of time. For the electrostatic case, we used

$$\widehat{n} \cdot \left[\vec{D}_>(R) - \vec{D}_<(R) \right] = \sigma_f = 0 \tag{6.68}$$

$$\widehat{s} \cdot \left[\epsilon_o \, \vec{E}_>(R) - \epsilon_o \, \vec{E}_<(R) \right] = 0 \tag{6.69}$$

The corresponding matching conditions for the magnetic case are

$$\widehat{n} \cdot \left[\frac{\vec{B}_>(R)}{\mu_o} - \frac{\vec{B}_<(R)}{\mu_o} \right] = \frac{1}{\mu_o} \, \widehat{n} \cdot \left[\vec{B}_>(R) - \vec{B}_<(R) \right] = 0 \tag{6.70}$$

$$\widehat{s} \cdot \left[\vec{H}_>(R) - \vec{H}_<(R) \right] = \widehat{s} \cdot \left[\vec{K}_f \times \widehat{n} \right] = 0 \tag{6.71}$$

Thus, not only is there a perfect correspondence between fields, potentials, and $r \to \infty$ boundary conditions in the two problems, there is also a correspondence between matching conditions at $r = R$. Thus, we can just apply the solution to the electrostatic problem with the substitutions $\epsilon_o \, \vec{E} \to \vec{H}$, $\epsilon_o \, V \to V_M$, $\vec{P} \to \vec{M}$, and $\vec{D} \to \vec{B}/\mu_o$.

Applying this correpondence to Equations 4.70 and 4.71 gives us

$$V_M(r < R) = -\frac{3\,\mu_o}{2\,\mu_o + \mu}\,H_0\,z = -\frac{3\,\mu_o}{2\,\mu_o + \mu}\,\frac{B_0}{\mu_o}\,z \tag{6.72}$$

$$V_M(r > R) = -H_0\,z + \frac{\vec{m}\cdot\hat{r}}{4\,\pi\,r^2} = -\frac{B_0}{\mu_o}\,z + \frac{\vec{m}\cdot\hat{r}}{4\,\pi\,r^2} \tag{6.73}$$

$$\vec{m} \equiv \frac{4\,\pi}{3}\,R^3\,\vec{M}(r < R) = \frac{4\,\pi}{3}\,R^3\,H_0\,\frac{3\,(\mu - \mu_o)}{2\,\mu_o + \mu}\,\hat{z} = \frac{4\,\pi}{3}\,R^3\,\frac{B_0}{\mu_o}\,\frac{3\,(\mu - \mu_o)}{2\,\mu_o + \mu}\,\hat{z} \tag{6.74}$$

From the above, we calculate the fields and the magnetostatic surface charge density ($\rho_M = 0$ because $\vec{M}$ is uniform):

$$\vec{H}(r < R) = \frac{3\,\mu_o}{2\,\mu_o + \mu}\,\frac{\vec{B}_0}{\mu_o} = \frac{\vec{B}_0}{\mu_o} - \frac{\vec{M}(r < R)}{3} \tag{6.75}$$

$$\vec{M}(r < R) = 3\,\frac{\mu - \mu_o}{2\,\mu_o + \mu}\,\frac{\vec{B}_0}{\mu_o} \qquad \sigma_M = 3\,\frac{\mu - \mu_o}{2\,\mu_o + \mu}\,\frac{B_0}{\mu_o}\,\cos\theta \tag{6.76}$$

$$\vec{B}(r < R) = \mu_o\left[\vec{H}(r < R) + \vec{M}(r < R)\right] = \mu_o\left[\frac{\vec{B}_0}{\mu_o} - \frac{\vec{M}(r < R)}{3} + \vec{M}(r < R)\right]$$

$$= \vec{B}_0 + \frac{2}{3}\,\mu_o\,\vec{M}(r < R) = \left(\frac{3\,\mu}{2\,\mu_o + \mu}\right)\vec{B}_0 \tag{6.77}$$

Explicitly, we find that:

- ▶ Like $\vec{E}$, $\vec{H}$ is uniform inside the sphere and points in the direction of the uniform field. For $\chi_m > 0$, like for $\chi_e > 0$, it is smaller in magnitude than the uniform field at infinity.

- ▶ The magnetization density is in the direction of the uniform field for $\chi_m > 0$ as it was for $\vec{P}$ and $\chi_e > 0$.

- ▶ The magnetostatic surface charge density has a $\cos\theta$ dependence and is positive at the north pole for $\chi_m > 0$, as it was for the electrostatic surface charge density and $\chi_e > 0$.

- ▶ $\vec{B}$ is enhanced relative to the uniform field for $\chi_m > 0$. We did not calculate $\vec{D}$ in the electrostatic case, but we would have found that it, too, was enhanced relative to the uniform field.

We again see the fact that $\vec{H}$ corresponds to $\vec{E}$ and $\vec{B}$ to $\vec{D}$. In the electrostatic case, we noted how the field of the polarization counters the uniform field so that the total field inside the sphere is smaller in magnitude than the uniform field. That is true here too, but for $\vec{H}$, not for $\vec{B}$. $\vec{B}$ itself is enhanced inside the sphere! This difference in the behavior of the "true" fields arises directly from the above somewhat unexpected correspondence of $\vec{H}$ rather than $\vec{B}$ to $\vec{E}$.

There is a shortcut method that is much faster, so good to know from the point of view of technique. It makes the *ansatz* that the sphere magnetizes uniformly so then the total field is the superposition of a uniform field and a uniformly magnetized sphere (Equation 6.16). This assumption is made initially without relating $\vec{M}$ and $\vec{H}$. It then uses the relation $\vec{M} = \chi_m \vec{H}$ (equivalently, $\vec{B} = \mu \vec{H}$) to relate the two and solve for the fields.

The ansatz based on superposition gives

$$\vec{B}(r < R) = \vec{B}_{uniform} + \vec{B}_{sphere} = \vec{B}_0 + \frac{2}{3}\mu_o \vec{M} \tag{6.78}$$

$$\vec{H}(r < R) = \vec{H}_{uniform} + \vec{H}_{sphere} = \vec{H}_{uniform} + \left(\frac{\vec{B}_{sphere}}{\mu_o} - \vec{M}_{sphere} \right)$$

$$= \frac{\vec{B}_0}{\mu_o} - \frac{1}{3}\vec{M} \tag{6.79}$$

Then we apply $\vec{B}(r < R) = \mu \vec{H}(r < R)$ to relate the above two equations and solve for $\vec{M}$. One finds one gets the same result. One can then calculate the field at $r \geq R$ from superposition. Admittedly, this technique is somewhat backhanded; when trying to understand things for the first time, reapplying the scalar potential to the full problem is more straightforward.

Lecture 19:

Magnetostatics in Matter III:
Boundary Value Problems in Magnetostatics (cont.)
Nonlinear Magnetic Materials

Date Revised: 2024/03/07 06:30
Date Given: 2024/03/07

Example 6.6: Magnetically Permeable Spherical Shell

Consider a spherical shell of inner radius a and outer radius b consisting of a highly permeable ($\mu/\mu_o \gg 1$) material placed in a uniform external field $\vec{B}_0$. We shall see that this shell shields its inner volume from the external field by a factor μ/μ_o. This technique is of great importance for magnetically sensitive experiments and equipment.

There are no free currents, so we may use the magnetostatic scalar potential technique. Furthemore, $\vec{\nabla} \cdot \vec{H} = 0$ in each region since μ is constant in each region. So the scalar potential V_M satisfies Laplace's Equation, allowing us to apply our techniques for the solution of Laplace's Equation from electrostatics.

In particular, given the azimuthal symmetry, we may assume the solution in each of the three regions is of the form given in Equation 3.115:

$$V_M(r < a, \theta) \equiv V_1(r, \theta) = \sum_{\ell=0}^{\infty} A_\ell \, r^\ell P_\ell(\cos\theta) \qquad (6.80)$$

$$V_M(a < r < b, \theta) \equiv V_2(r, \theta) = \sum_{\ell=0}^{\infty} \left(C_\ell \, r^\ell + \frac{D_\ell}{r^{\ell+1}} \right) P_\ell(\cos\theta) \qquad (6.81)$$

$$V_M(r > b, \theta) \equiv V_3(r, \theta) = -H_0 \, r \cos\theta + \sum_{\ell=0}^{\infty} \frac{E_\ell}{r^{\ell+1}} P_\ell(\cos\theta) \qquad (6.82)$$

where we have already applied the requirements that V_M be finite as $r \to 0$ and that it yield the uniform field as $r \to \infty$ with $H_0 = B_0/\mu_o$. We have also assumed that V_M has no constant offset as $r \to \infty$.

There are no free currents, so our matching conditions are (as for the magnetically permeable sphere, Equations 6.70 and 6.71) that the normal component of $\vec{B}$ and the tangential component of $\vec{H}$ be continuous. Using $\vec{H} = -\vec{\nabla} V_M$, we thus have the four conditions

$$\mu_o \left. \frac{\partial V_1}{\partial r} \right|_a = \mu \left. \frac{\partial V_2}{\partial r} \right|_a \qquad \qquad \mu \left. \frac{\partial V_2}{\partial r} \right|_b = \mu_o \left. \frac{\partial V_3}{\partial r} \right|_b \qquad (6.83)$$

$$\left. \frac{\partial V_1}{\partial \theta} \right|_a = \left. \frac{\partial V_2}{\partial \theta} \right|_a \qquad \qquad \left. \frac{\partial V_2}{\partial \theta} \right|_b = \left. \frac{\partial V_3}{\partial \theta} \right|_b \qquad (6.84)$$

Note that we do not impose continuity on V_M. In the electrostatic case, we imposed continuity of V and the boundary condition on the normal derivative, ignoring continuity of the tangential derivative. In electrostatics, continuity of V comes from constructing it as the line integral of the electric field, which we in turn were motivated to write down in order to calculate the work done by the electric field on a point charge. Since $\vec{H}$ does not do such work, writing down the line integral is not physically motivated, though it is mathematically reasonable to do so because $\vec{H} = -\vec{\nabla} V_M$. So, here, we instead use continuity of the radial and tangential derivatives. This is an arbitrary choice driven by our physical intuition. We will see below that continuity of V_M would yield information redundant with tangential derivative continuity.

Before we dive into a lot of calculation, let's see what we can figure out without doing much work. The radial derivative equations only connect terms on the two sides of the equations with the same ℓ because they do not modify the orthonormal $P_\ell(\cos\theta)$. What about the angular derivative equations? Recall Equation 3.156:

$$P_\ell^m(x) = (-1)^m (1-x^2)^{m/2} \frac{d^m}{dx^m} P_\ell(x) \tag{6.85}$$

Let's write $\frac{\partial P_\ell(\cos\theta)}{\partial\theta}$ using this:

$$\frac{\partial P_\ell(\cos\theta)}{\partial\theta} = \frac{dP_\ell(\cos\theta)}{d\cos\theta} \frac{d\cos\theta}{d\theta} = \frac{P_\ell^1(\cos\theta)}{(-1)^1(1-\cos^2\theta)^{1/2}}(-\sin\theta) \tag{6.86}$$

$$= P_\ell^1(\cos\theta) \tag{6.87}$$

where we note that, since $0 < \theta < \pi$, there is no sign ambiguity and thus $\sin\theta = (1-\cos^2\theta)^{1/2}$. The $P_\ell^1(\cos\theta)$ are also orthonormal polynomials (the P_ℓ^m over all ℓ at fixed m form an orthonormal set in order for the $Y_{\ell m}$ to form an orthonormal set), so the same point we made above about the equations connecting terms at the same ℓ holds for these equations also. Note however that, for $\ell = 0$, the $\partial/\partial\theta$ matching condition yields zero.

Note also that, for $\ell \geq 1$, these equations are the same as one would have obtained by requiring continuity of V_M since $\partial/\partial\theta$ doesn't modify the radial factor of each term.

Taking the necessary derivatives for the radial derivative equations and then equating the two sides of all six equations (four for $\ell > 0$, only two for $\ell = 0$) term-by-term gives us:

$$\ell > 0 : \qquad \mu_o\, \ell\, A_\ell a^{\ell-1} = \mu\, \ell\, C_\ell\, a^{\ell-1} - \mu\, (\ell+1)\, \frac{D_\ell}{a^{\ell+2}} \qquad (6.88)$$

$$\mu\, \ell\, C_\ell\, b^{\ell-1} - \mu\, (\ell+1)\, \frac{D_\ell}{b^{\ell+2}} = -\mu_o\, H_0\, \delta_{\ell 1} - \mu_o\, (\ell+1)\, \frac{E_\ell}{b^{\ell+2}} \qquad (6.89)$$

$$A_\ell\, a^\ell = C_\ell\, a^\ell + \frac{D_\ell}{a^{\ell+1}} \qquad (6.90)$$

$$C_\ell\, b^\ell + \frac{D_\ell}{b^{\ell+1}} = -H_0 b\, \delta_{\ell 1} + \frac{E_\ell}{b^{\ell+1}} \qquad (6.91)$$

$$\ell = 0 : \qquad 0 = -\mu\, \frac{D_0}{a^2} \qquad\qquad -\mu\, \frac{D_0}{b^2} = -\mu_o\, \frac{E_0}{b^2} \qquad (6.92)$$

We explicitly write out the $\ell = 0$ equations because they yield qualitatively different conditions than the $\ell > 0$ terms.

For $\ell > 1$, solving for C_ℓ and D_ℓ results in both vanishing, so then A_ℓ and E_ℓ vanish for $\ell > 1$.

For $\ell = 0$, the radial derivative matching equations imply $D_0 = E_0 = 0$. We expect $E_0 = 0$ because it would yield a magnetic monopole potential for $r > b$, which we know is physically disallowed.

There are no equations that explicitly determine A_0 and C_0, which correspond to offsets of V_M for $r \leq a$ and $a < r < b$. We actually don't need to find them, since they do not affect $\vec{H}$ when the gradient is taken. (Recall, there is no issue of this potential being related to work or a potential energy, so we do not need to worry about discontinuities due to offsets.) But we can specify them by applying a restricted version of continuity of V_M, which is that we require V_M have the same offset in all regions. The lack of an offset for $r > b$ then implies $A_0 = 0$ and $C_0 = 0$.

For $\ell = 1$, we can do a lot of algebra to find explicit formulae for all the coefficients (you can find these in Jackson §5.12). These formulae are not particularly illuminating, but they become more intuitive when we take the limit $\mu/\mu_o \gg 1$. Inserting those coefficients into the solutions, we obtain

$$V_1(r,\theta) \overset{\frac{\mu}{\mu_o} \gg 1}{=} A_1 \, r \, \cos\theta = -\frac{9}{2\,\frac{\mu}{\mu_o}\left(1 - \frac{a^3}{b^3}\right)} \, H_0 \, r \, \cos\theta \tag{6.93}$$

$$V_2(r,\theta) \overset{\frac{\mu}{\mu_o} \gg 1}{=} \left(C_1 \, r + \frac{D_1}{r^2}\right)\cos\theta = -\frac{3}{\frac{\mu}{\mu_o}\left(1 - \frac{a^3}{b^3}\right)} \, H_0 \left(r + \frac{1}{2}\frac{a^3}{r^2}\right)\cos\theta \tag{6.94}$$

$$V_3(r,\theta) \overset{\frac{\mu}{\mu_o} \gg 1}{=} \left(-H_0 \, r + \frac{E_1}{r^2}\right)\cos\theta = H_0 \left(-r + \frac{b^3}{r^2}\left(1 - \frac{3\left(1 + \frac{1}{2}\frac{a^3}{b^3}\right)}{\frac{\mu}{\mu_o}\left(1 - \frac{a^3}{b^3}\right)}\right)\right)\cos\theta \tag{6.95}$$

Note that we include the term of order μ_o/μ in the $r > b$ solution so we can see that the matching condition on the tangential derivative at $r = b$ (equivalent to matching of V_M itself) is explicitly satisfied even in the limit $\mu/\mu_o \gg 1$.

Here are the resulting fields in the three regions:

$$\vec{H}_1(r, \theta) \overset{\frac{\mu}{\mu_o} \gg 1}{=} -\frac{9}{2} \frac{H_0 \, \hat{z}}{\frac{\mu}{\mu_o} \left(1 - \frac{a^3}{b^3}\right)} \qquad\qquad \vec{B}_1(r, \theta) = \mu_o \, \vec{H}_1(r, \theta) \quad (6.96)$$

$$\vec{H}_2(r, \theta) \overset{\frac{\mu}{\mu_o} \gg 1}{=} \frac{3 \, H_0 \, \hat{z}}{\frac{\mu}{\mu_o} \left(1 - \frac{a^3}{b^3}\right)} + \frac{3 \, (\vec{m}_a \cdot \hat{r}) \, \hat{r} - \vec{m}_a}{4 \, \pi \, r^3} \quad \vec{B}_2(r, \theta) = \mu \, \vec{H}_2(r, \theta) \quad (6.97)$$

$$\vec{H}_3(r, \theta) \overset{\frac{\mu}{\mu_o} \gg 1}{=} H_0 \, \hat{z} + \frac{3 \, (\vec{m}_b \cdot \hat{r}) \, \hat{r} - \vec{m}_b}{4 \, \pi \, r^3} \qquad \vec{B}_3(r, \theta) = \mu_o \, \vec{H}_3(r, \theta) \quad (6.98)$$

$$\text{with} \quad \vec{m}_a = -\frac{9}{2} \frac{H_0}{\frac{\mu}{\mu_o} \left(1 - \frac{a^3}{b^3}\right)} \left(\frac{4 \, \pi}{3} \, a^3\right) \hat{z} \tag{6.99}$$

$$\vec{m}_b = 3 \left[1 - \frac{3 \left(1 + \frac{1}{2} \frac{a^3}{b^3}\right)}{\frac{\mu}{\mu_o} \left(1 - \frac{a^3}{b^3}\right)}\right] H_0 \left(\frac{4 \, \pi}{3} \, b^3\right) \hat{z} \tag{6.100}$$

It is not obvious but it is true that $\vec{m}_b$ incorporates $\vec{m}_a$, which is why there is no explicit contribution from $\vec{m}_a$ to the field at $r > b$. We will see this more clearly below.

The following features can be pointed out:

- ▶ Inside $r < a$, we have a uniform field weakened by a factor of μ/μ_o (for both B and H).

- ▶ In the permeable material, we have a uniform H field as well as a dipole field, but both are of order $(\mu_o/\mu) H_0$ (i.e., attenuated) with the dipole moment pointed to $-\hat{z}$. The dipole field cancels the uniform field at the poles at $r = a$ and adds to it at the equator.

- ▶ In the permeable material, the B field receives a factor of μ, so the B field receives uniform field and dipole contributions of order B_0 in the permeable material, though the vanishing at the poles at $r = a$ remains.

- ▶ One caveat to the above two statements is due to the $(1 - a^3/b^3)$ factor in the denominator of both terms (explicitly in the first term, hiding in $\vec{m}_a$ in the second term). If the shell is quite thin, then a/b is close to unity and this factor is much smaller than unity, resulting in an enhancement in both H_2 and B_2 by the geometric factor $(1 - a^3/b^3)^{-1}$. This factor accounts for the fact that magnetic field lines cannot be broken, and so the vast majority of the field lines that would have threaded through the $r < b$ region (a fraction $1 - \mu_o/\mu$ of them) now must flow entirely through the $a < r < b$ region: the factor is the ratio of the volume of the sphere of radius b to the volume of the shell.

- ▶ Finally, the field outside is the uniform field (for H and B) plus that of a dipole in the $+\hat{z}$ direction. The dominant part of the dipole field cancels the uniform field at the equator at $r = b$, leaving a small residual field of order μ_o/μ smaller. At the poles, the dipole field adds to the uniform field, increasing the fields to $3 H_0$ and $3 B_0$ there.

Here is a picture from Jackson of $\vec{B}$. Note the concentration of field lines in the permeable material and their absence in the empty central region.
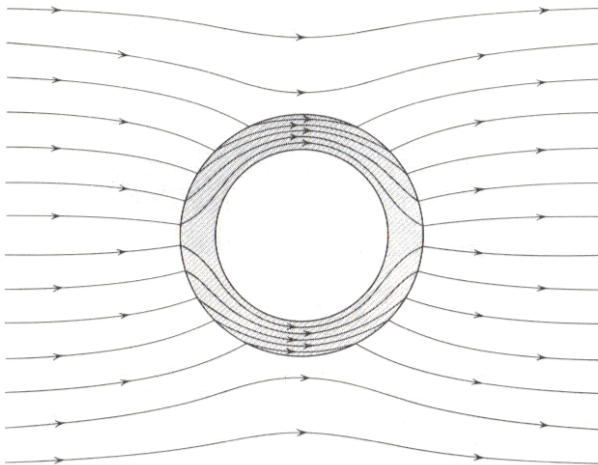
© 1999 Jackson, *Classical Electrodynamics*

Let's now consider the analogy to electrostatic shielding. Electrostatic shielding is easily provided by conductors, and perfect conductors ($\epsilon/\epsilon_o \to \infty$) provide perfect electrostatic shielding. They do this by setting up surface charge that perfectly cancels the externally applied field *both in the conductor and in the cavity*.

The net effect for magnetostatic shielding effect is very similar — zero field in the cavity as $\mu/\mu_o \to \infty$ — but the physical cause is quite different, resulting in a *nonzero (and potentially high)* $\vec{B}$ field in the permeable material rather than zero field as in the electrostatic case. The fundamental reason is that magnetic field lines do not terminate: there are no magnetic monopoles. (The one exception is the field of the magnetic moment of a fundamental particle.) Therefore, the shielding effect is obtained by redirecting the field lines away from the cavity and into the permeable material, not by terminating the field lines. The high magnetic permeability of the materials causes magnetic dipoles to orient themselves such that they generate surface currents that cancel the externally applied field in the cavity (both $\vec{H}$ and $\vec{B}$) and enhance $\vec{B}$ in the material by the volume factor needed to accommodate all the field lines that would have gone through the cavity. In the limit of the equivalent of an electric conductor, with $\mu/\mu_o \to \infty$, the shielding would be perfect, as for an electric conductor.

All that said, we get a perfect analogy to an electric conductor if we instead look at the auxiliary field $\vec{H}$ and the magnetostatic surface charge density σ_M. $\vec{H}$ does get very small in the permeable material and in fact vanishes in the $\mu/\mu_o \to \infty$ limit. If we calculate the magnetostatic surface charge density σ_M from the discontinuity in $\vec{H}$ or $\vec{M}$, one would see that σ_M would look very much like σ_b for the case of a spherical shell of high dielectric susceptibility ($\epsilon/\epsilon_o \gg 1$) and, in the limit $\mu/\mu_o \to \infty$, σ_M would mimic the surface charge density of the electrical conductor limit (which is the same as $\epsilon/\epsilon_o \to \infty$). In particular, σ_M would only be nonzero at $r = b$: $\vec{H}$ would vanish for $r < a$ and $a < r < b$, so there could be no discontinuity in the radial derivative of $\vec{H}$ at $r = a$, which would occur if $\sigma_M \neq 0$ there. That is, in the magnetostatic scalar potential picture, the $\vec{H}$ field lines *do terminate* on the magnetostatic surface charge at $r = b$, yielding vanishing $\vec{H}$ for $r < b$. Of course, $\vec{H}$ is not a physical field, so the analogy is only mathematical, not physical.

Either way one does it, this calculation has important practical implications: such highly permeable materials are in widespread use for magnetic shielding from, especially, Earth's field in magnetically sensitive experiments and equipment such as SQUIDs (very sensitive magnetometers) and photomultiplier tubes (where the electrons' paths can be substantially bent and thus the gain modified by magnetic fields).

While we have benefited from our boundary-value problem techniques to get directly to the fields without having to calculate the bound surface currents, it would be nice to see how the bound surface currents give the observed fields. Recall Equation 5.84, which gives the bound surface current density from the change in the tangential component of the magnetic field:

$$\vec{K}(\vec{r}) = \frac{1}{\mu_o} \widehat{n}(\vec{r}) \times \left[\vec{B}_>(\vec{r}) - \vec{B}_<(\vec{r}) \right] \tag{6.101}$$

where $\widehat{n}$ points from the $<$ region to the $>$ region. In our case, $\vec{K} = \vec{K}_b$ because there are no free currents. Since $\widehat{n} = \widehat{r}$ for our spherical surfaces and $\vec{B}$ only has components in the $\widehat{r}$ and $\widehat{\theta}$ directions, this reduces to

$$K_b(r) \widehat{\phi} = \frac{1}{\mu_o} \left[B_{>,\theta} - B_{<,\theta} \right] \widehat{\phi} = \frac{1}{\mu_o} \left[-\frac{\mu_>}{r} \left. \frac{\partial V_M}{\partial \theta} \right|_{r_>} + \frac{\mu_<}{r} \left. \frac{\partial V_M}{\partial \theta} \right|_{r_<} \right] \widehat{\phi} \tag{6.102}$$

$$= \frac{\mu_< - \mu_>}{\mu_o} \frac{1}{r} \left. \frac{\partial V_M}{\partial \theta} \right|_r \widehat{\phi} \tag{6.103}$$

where $<$ and $>$ indicate the two sides of the particular boundary at r and we use the fact that the tangential component of $\vec{H}$, which is given by $-(1/r) \partial V_M / \partial \theta$ here, is continuous and thus has the same value on both sides of the interface at r. So it is straightforward to calculate the surface currents given V_M.

We also know how to calculate $\vec{B}$ given surface currents derived from a uniform magnetization: we did it in our first calculation of the magnetic field of the permanently magnetized sphere (Equation 6.16) and saw (valid only for $\vec{K}_b \propto \hat{\phi} \sin \theta$!)

$$\vec{B}_{\vec{M}}(r < R) = \frac{2}{3} \mu_o \vec{M} = \frac{2}{3} \mu_o \frac{\vec{K}_b \cdot \hat{\phi}}{\sin \theta} \hat{z}$$

$$\vec{B}_{\vec{M}}(r > R) = \frac{\mu_o}{4\pi} \frac{3(\vec{m} \cdot \hat{r}) \hat{r} - \vec{m}}{r^3} \qquad \vec{m} = \frac{4}{3} \pi R^3 \vec{M} = \frac{4}{3} \pi R^3 \frac{\vec{K}_b \cdot \hat{\phi}}{\sin \theta} \hat{z}$$

where the relation between $\vec{M}$ and $\vec{K}_b$ comes from the definition of the bound surface current, $\vec{K}_b = \vec{M} \times \hat{n} = M \hat{z} \times \hat{r} = \hat{\phi} M \sin \theta$. Here, the $\vec{H}$ and $\vec{B}$ fields are uniform for $r < a$, and the fields are the sum of a uniform field and the field of a dipole for $a < r < b$. It thus looks like we have the superposition of a uniform magnetization in one direction for $r < a$ and a uniform magnetization in the opposite direction for $a < r < b$, with the two magnetizations having a fractional difference of magnitude of μ_o/μ. (The fact that we have permeable materials present is irrelevant for the calculation of $\vec{B}$: once one has all the bound currents, one can calculate $\vec{B}$ directly from them.) So, we expect that, in this case, we can just add the field of the above form due to the bound currents to the uniform applied field to get the total field in the three regions.

That is, we expect (again, valid only for $\vec{K}_a, \vec{K}_b \propto \widehat{\phi} \sin \theta$!)

$$\vec{B}_1(r, \theta) = \vec{B}_0 + \frac{2}{3} \mu_o \frac{\vec{K}_b(a) \cdot \widehat{\phi} + \vec{K}_b(b) \cdot \widehat{\phi}}{\sin \theta} \widehat{z} \tag{6.104}$$

$$\vec{B}_2(r, \theta) = \vec{B}_0 + \frac{2}{3} \mu_o \frac{\vec{K}_b(b) \cdot \widehat{\phi}}{\sin \theta} \widehat{z} + \frac{\mu_o}{4\pi} \frac{3(\vec{m}_a \cdot \widehat{r})\widehat{r} - \vec{m}_a}{r^3} \tag{6.105}$$

$$\vec{B}_3(r, \theta) = \vec{B}_0 + \frac{\mu_o}{4\pi} \frac{3(\vec{m}_b \cdot \widehat{r})\widehat{r} - \vec{m}_b}{r^3} \tag{6.106}$$

$$\vec{m}_a = \frac{4}{3} \pi a^3 \frac{\vec{K}_b(a) \cdot \widehat{\phi}}{\sin \theta} \widehat{z} \qquad \vec{m}_b = \vec{m}_a + \frac{4}{3} \pi b^3 \frac{\vec{K}_b(b) \cdot \widehat{\phi}}{\sin \theta} \widehat{z} \tag{6.107}$$

and then we can obtain $\vec{H}$ from the usual relation $\vec{H}(\vec{r}) = \vec{B}(\vec{r})/\mu(\vec{r})$. Note that we now see explicitly that $\vec{m}_b$ incorporates $\vec{m}_a$ as we had stated without proof above.

The explicit results for the bound surface currents are

$$\vec{K}_b(a, \theta) = \frac{\mu_< - \mu_>}{\mu_o} \frac{1}{a} \frac{\partial V_M}{\partial \theta}\bigg|_a \widehat{\phi} = -\widehat{\phi}\left(\frac{\mu}{\mu_o} - 1\right) \frac{9}{2\frac{\mu}{\mu_o}\left(1 - \frac{a^3}{b^3}\right)} \frac{B_0}{\mu_o} \sin\theta$$

$$\overset{\mathcal{O}(\mu_o/\mu)^0}{\approx} -\widehat{\phi} \frac{9}{2\left(1 - \frac{a^3}{b^3}\right)} \frac{B_0}{\mu_o} \sin\theta \qquad (6.108)$$

$$\vec{K}_b(b, \theta) = \frac{\mu_< - \mu_>}{\mu_o} \frac{1}{b} \frac{\partial V_M}{\partial \theta}\bigg|_b \widehat{\phi} = \widehat{\phi}\left(\frac{\mu}{\mu_o} - 1\right) \frac{3\left(1 + \frac{1}{2}\frac{a^3}{b^3}\right)}{\frac{\mu}{\mu_o}\left(1 - \frac{a^3}{b^3}\right)} \frac{B_0}{\mu_o} \sin\theta$$

$$\overset{\mathcal{O}(\mu_o/\mu)^0}{\approx} \widehat{\phi} \frac{3\left(1 + \frac{1}{2}\frac{a^3}{b^3}\right)}{\left(1 - \frac{a^3}{b^3}\right)} \frac{B_0}{\mu_o} \sin\theta \qquad (6.109)$$

We will explain the choice for the order of the approximation below.

There is an important subtlety in trying to do this calculation of surface currents using the approximate forms for the fields we have written down (valid for $\mu_o/\mu \ll 1$). We expect the magnetic field for $r < a$ to be of order $(\mu_o/\mu) B_0$. But $\vec{B}_0$ is in the expression for $\vec{B}_1$, so that implies the second term in that expression due to the surface currents will carry one term of order B_0 to cancel B_0 and then a second term of order $(\mu_o/\mu) B_0$ to give the residual field. As we explained above, our expressions for the contribution of the surface current to the field are of the following form for $r \lesssim b$:

$$B_K \sim \mu_o \, K \sim \pm \mu_o \frac{\mu - \mu_o}{\mu_o} \frac{1}{r} \frac{\partial V_M}{\partial \theta} \sim \mu_o \left(\frac{\mu}{\mu_o} - 1 \right) H_\theta \tag{6.110}$$

$$\sim \mu_o \left(\mathcal{O}\left(\frac{\mu}{\mu_o} \right)^1 + \mathcal{O}\left(\frac{\mu_o}{\mu} \right)^0 \right) \mathcal{O}\left(\frac{\mu_o}{\mu} \right)^1 H_0 \tag{6.111}$$

$$\sim \left[\mathcal{O}\left(\frac{\mu_o}{\mu} \right)^0 + \mathcal{O}\left(\frac{\mu_o}{\mu} \right)^1 \right] B_0 \tag{6.112}$$

(In the second line, we used $H \sim \mathcal{O}(\mu_o/\mu)^1 H_0$, which one can see from the expressions for $\vec{H}_1$ and $\vec{H}_2$. It is not so obvious that this is true for $\vec{H}_3$ at $r \sim b$, but it must be true because H_θ is continuous. It turns out to be true because the dipole field cancels the applied field to first order in H_0 (*i.e.*, zeroth order in μ_o/μ) at the equator, leaving a residual field of order $\mathcal{O}(\mu_o/\mu)^1 H_0$. The cancellation does not happen at the poles, but $H_\theta = 0$ at the poles.)

We now see the problem. The term that is $\mathcal{O}(\mu_o/\mu)^0$ will cancel the $\vec{B}_0$ term. So then the $\mathcal{O}(\mu_o/\mu)^1 B_0$ term is all that is left and is our full field, as expected. But we have not done the approximation self-consistently. We would have obtained a term of the same order by including terms $\mathcal{O}(\mu_o/\mu)^2$ in the expression for H because they would yield $\mathcal{O}(\mu_o/\mu)^1$ terms when multiplied by the $\mathcal{O}(\mu/\mu_o)^1$ term from the $(\mu/\mu_o - 1)$ prefactor. Without including that term, we will get the incorrect coefficient for the residual field.

We could have included that higher order term, but then we would run into the same problem at the next order: our calculation of the field using the surface currents would be correct to $\mathcal{O}(\mu_o/\mu)^1$, but our expression for the fields would have terms of order $\mathcal{O}(\mu_o/\mu)^2$ that we would not be able to fully reproduce. Given that it would be algebraically challenging to do this even to $\mathcal{O}(\mu_o/\mu)^1$ correctly, we punt on trying to calculate the residual field.

However, we can self-consistently check our results (Equations 6.96-6.100) to $\mathcal{O}(\mu_o/\mu)^0$, so let's do that because it will show us that the zeroth order field does vanish at $r < a$ and it will tell us interesting things for other regions.

First, for $r < a$, we have, *to zeroth order in μ_o/μ,*

$$\hat{z} \cdot \vec{B}_1(r,\theta) \overset{\mathcal{O}(\mu_o/\mu)^0}{\approx} B_0 + \frac{2}{3} \frac{B_0}{1 - \frac{a^3}{b^3}} \left[-\frac{9}{2} + 3 \left(1 + \frac{1}{2} \frac{a^3}{b^3} \right) \right] = 0 \qquad (6.113)$$

$$\hat{z} \cdot \vec{H}_1(r,\theta) = \frac{\vec{B}_1(r,\theta)}{\mu_o} \overset{\mathcal{O}(\mu_o/\mu)^0}{\approx} 0 \qquad (6.114)$$

As expected, both the magnetic and auxiliary fields vanish to zeroth order in μ_o/μ inside the cavity.

For $a < r < b$, we have

$$\widehat{z} \cdot \vec{B}_2(r, \theta) \overset{\mathcal{O}(\mu_o/\mu)^0}{\approx} B_0 + \frac{2}{3} \frac{B_0}{1 - \frac{a^3}{b^3}} (3) \left(1 + \frac{1}{2} \frac{a^3}{b^3}\right) + \frac{\mu_o}{4\pi} \frac{3(\vec{m}_a \cdot \widehat{r}) \cos\theta - m_a}{r^3}$$

$$= \frac{3 B_0}{1 - \frac{a^3}{b^3}} + \frac{\mu_o}{4\pi} m_a \frac{3\cos^2\theta - 1}{r^3} \tag{6.115}$$

$$\widehat{z} \cdot \vec{H}_2(r, \theta) = \frac{\vec{B}_2(r, \theta)}{\mu} \overset{\mathcal{O}(\mu_o/\mu)^0}{\approx} 0 \tag{6.116}$$

$$\text{with} \quad \vec{m}_a \overset{\mathcal{O}(\mu_o/\mu)^0}{\approx} -\frac{4\pi}{3} a^3 \left(\frac{9}{2}\right) \frac{1}{1 - \frac{a^3}{b^3}} \frac{B_0}{\mu_o} \widehat{z} \tag{6.117}$$

The total magnetic field in the permeable material is of order B_0 because both terms shown are of order B_0. In the limit $a \ll b$, one recovers $3 B_0$ as we expect from the case of the permeable sphere (Equation 6.77 with $\mu_o/\mu \to 0$). The auxiliary field vanishes in the permeable material to order $(\mu_o/\mu)^0$ because one must divide the entire expression by μ to get H from B, which combines with the μ_o in the expression for B to give a prefactor of μ_o/μ that vanishes at the level of approximation we are considering. The vanishing of $\vec{H}$ in the permeable material is exact in the limit $\mu/\mu_o \to \infty$ because the above approximations become exact.

Finally, let's look at $r > b$, for which we obtain

$$\widehat{z} \cdot \vec{B}_3(r, \theta) \overset{\mathcal{O}(\mu_o/\mu)^0}{\approx} B_0 + \frac{\mu_o}{4\pi} \frac{3(\vec{m}_b \cdot \widehat{r})\cos\theta - m_b}{r^3} \tag{6.118}$$

$$\text{with} \quad \vec{m}_b \overset{\mathcal{O}(\mu_o/\mu)^0}{\approx} \frac{4\pi}{3} \frac{1}{1 - \frac{a^3}{b^3}} \frac{B_0}{\mu_o} \left[-\frac{9}{2} a^3 + b^3 (3) \left(1 + \frac{1}{2} \frac{a^3}{b^3} \right) \right] \widehat{z}$$

$$= 3 \left(\frac{4\pi}{3} b^3 \right) \frac{B_0}{\mu_o} \widehat{z} \tag{6.119}$$

$$\widehat{z} \cdot \vec{H}_3(r, \theta) = \frac{\vec{B}_3(r, \theta)}{\mu_o} \tag{6.120}$$

One can see that the expressions for $\vec{B}_0$ and $\vec{m}_b$ match *to zeroth order in* μ_o/μ the results we obtained via the boundary value problem technique, Equations 6.96-6.100. The expression for H has the same form with B_0 replaced by H_0 and it also matches the expressions we obtained earlier, again to zeroth order in μ_o/μ.

So, in the end, we see that, to the level of approximation for which we can self-consistently do calculations, the fields we calculate from the surface currents match the fields that we used to calculate those surface currents.

Nonlinear Materials and Ferromagnetism

There are materials whose magnetic response is *nonlinear*. In such materials, in addition to the tendency of magnetic dipoles due to unpaired electrons to align with the applied magnetic field, these dipoles interact with each other in such a way as to prefer aligning with each other, too. This extra preference for magnetization causes the magnetization to depend nonlinearly on $\vec{H}$.

Beyond nonlinearity, there is the phenomenon of *ferromagnetism*, in which there are additional interactions that cause the magnetization to be preserved even after the applied field is reduced.

Both phenomena are caused by unpaired electrons as paramagnetism is; one might have guessed this by the fact that all three phenomena involve the alignment of magnetic dipoles with the applied field. The additional dipole-dipole interaction that causes nonlinearity is due to the *exchange effect* in quantum mechanics. We will explain this in the following.

The Exchange Effect in a Single Atom: Hund's Rules

As you know, electrons in atoms occupy *shells* corresponding to different energies for the electron-nucleus Coulomb interaction. Only a certain number of states are allowed for each shell (n^2 for shell n), and electrons can be put in a shell with spin "up" or spin "down" (multiplying by 2 the number of allowed states).

When a shell is partially filled, the electrons prefer to be *unpaired*, meaning that they have different orbital wavefunctions (probability distributions) and *the same spin direction (*i.e., *aligned spins*) rather than the same orbital wavefunction and different spin directions. This behavior, where electrons prefer to be in different orbitals but have the same spin, is frequently termed *Hund's Rules*.

The reason for this preference against having the same orbital wavefunction is that the electrostatic repulsive energy of two electrons in the same orbital state is high: in quantum mechanics, that energy is determined by the integral of the product of their wavefunctions weighted by $1/|\vec{r} - \vec{r}'|$ where $\vec{r}$ and $\vec{r}'$ are their positions, so the less similar their wavefunctions are, the lower the (positive) electrostatic repulsive energy is.

Next, considering the Pauli exclusion principle, the overall state (spatial wavefunction and spin) must be antisymmetric under exchange of the two electrons. One could achieve this by taking the spatial wavefunction to be antisymmetric and the spin state to be symmetric under exchange or vice versa. It turns out that, when one calculates the Coulomb repulsion energy integral, there is a second term that arises due to the extra terms created by requiring the overall state to be symmetric or antisymmetric under exchange. Moreover, because of the symmetry constraints on the overall state, this *exchange* term is negative when the spatial state is antisymmetric and positive when it is symmetric. Thus, the exchange term ensures that the antisymmetry of the overall state is preferred to be in the spatial wavefunction, not the spin state.

With the above, it would still be possible for the electrons to either have the same spin projection or for them to have opposite spin projections and the spin state to be the symmetric combination of the two possible anti-aligned states. It turns out that spin-orbit coupling causes the latter state to be higher energy, so the case in which the two electrons are aligned is preferred. Thus, we are able to explain Hund's Rules.

The Exchange Effect Among Atoms: Nonlinearity

In addition, though, one needs a mechanism for unpaired electrons in nearby atoms to align with each other; alignment of the unpaired electrons in a single atom is not enough. A similar exchange interaction is required, of which there are many types that depend on the details of the material and how the electrons in nearby atoms interact. The key requirement for such exchange effects to occur, though, is delocalization — the electron wavefunctions must be large enough that they spread to nearby atoms — so that there can be exchange interactions between electrons in adjacent atoms. This explains why nonlinearity occurs only in atoms with d- and f-shell electrons — the electrons in these orbitals are more weakly bound than s- and p-shell electrons, providing the necessary delocalization.

The exchange interaction leads to a nonlinear magnetic permeability, where these interactions cause the magnetic dipoles to prefer to align with each other macroscopically when they have been nudged into alignment by an applied field. This would yield a relationship of the form $\vec{B} = \vec{F}(\vec{H})$, which cannot be summarized by a simple constant of proportionality, but the relation is at least well-defined.

Interactions with the Crystal: Ferromagnetism

Ferromagnets have domains, which are regions of coaligned magnetic dipoles, caused by the aforementioned nonlinearity: it is energetically favorable for the magnetic moments of unpaired electrons in nearby atoms to align. By default, these domains are macroscopic in size (fractions of a mm, perhaps), but they do not align with each other because alignment would create a large magnetic field outside the material, which is not a low-energy state (which we will see when we talk about magnetic energy). When a large field is applied, though, the energy cost of not aligning with the magnetic field ($\vec{m} \cdot \vec{B}_{applied}$) is larger than the energy savings of not having a large field energy ($|\vec{B}_{material}|$), and so the domains align with the applied field.

We then must consider the phenomenon of *saturation*, whereby, at large fields, one gets to the point where *all* the unpaired electrons' dipole moments are aligned with the field and there are no more magnetic dipoles left to align. The magnetization density stops increasing and saturates. The applied field $\vec{H}$ may continue to be increased, but $\vec{B} = \mu_o(\vec{H} + \vec{M}_{sat})$, where $\vec{M}_{sat}$ is the saturated magnetization density. Thus, $\vec{B}$ increases due to the first term only. (At lower fields, $\vec{M}$ increases with $\vec{H}$, leading to a large amplification of $\vec{H}$ to yield $\vec{B}$.)

After bringing a nonlinear paramagnetic material into saturation, what happens when one turns off the applied field? The large exchange interaction energy makes it favorable for the moments to remain aligned with the direction of the applied magnetic field that has been removed. It is not that full alignment is the lowest-energy state, but that there is an energy barrier between the fully aligned state and the lower-energy state with randomly aligned domains. In fact, to reduce $\vec{M}$ and $\vec{B}$ to zero requires a significant $\vec{H}$ in the direction opposite to $\vec{B}$. After $\vec{B}$ goes through zero, it can then begin to align with $\vec{H}$ again and one can reach saturation in the other direction. And so on.

This phenomenon of the magnetization (and thus $\vec{B}$) being dependent on past history is called *hysteresis*: not only is $\vec{B}$ a nonlinear function of $\vec{H}$, but, in addition, $\vec{B}$ depends on the history of $\vec{H}$. Hysteresis curves are shown in Griffiths Figures 6.28 and 6.29.

The exchange phenomenon explains why ferromagnetics becomes less magnetized if they are dropped. The mechanical shock of dropping provides enough vibrational energy to exceed the barrier between the fully aligned state and the random domain state, allowing the domains to randomize in direction again.

We note that ferromagnets have a *Curie* or *transition temperature*, T_c. This temperature corresponds to roughly the exchange energy of nearby dipoles. When the temperature is larger than T_c, the thermal energy available overcomes the exchange energy, causing magnetic ordering to go away. If a saturated ferromagnetic is raised above T_c, the ordering will dissipate. Then, when recooled in zero applied field, randomly oriented domains will appear but there will be no overall ordering of the magnetic dipoles. Cooling in a high enough applied field, by contrast, will result in magnetic ordering and a permanent $\vec{M}$.

That is, if you heat the refrigerator magnet you have dropped many times past its T_c and then cool it slowly in a field large enough to obtain $\vec{M}_{sat}$, it will work well again!

There is not much more we can say about ferromagnetism without considering specific cases.

Section 7
Electrodynamics

Lecture 20:

Electrodynamics I:

Currents and Ohm's Law
Electromotive Forces

Date Revised: 2024/04/02 05:20
Date Given: 2024/04/02

Currents and Ohm's Law

Ohm's Law and Joule Heating: Differential Version

We state the very nonobvious point that the current due to an ensemble of flowing charges is proportional to the *force on a single charge* $\vec{f}$ acting on them:

$$\vec{J} \propto \vec{f} \tag{7.1}$$

Since current is proportional to velocity, and force is proportional to acceleration, why is this true? In an ideal conductor, it would not be true, we would expect current to be proportional to the integral of the force over time. But in all real conductors, there are two important effects that change this picture:

▶ The first is the random thermal motion of the charge carriers. The forces we can apply yield velocities that are small perturbations to this random thermal motion. So the mean *speed* of the carriers is dominated by the thermal speed $v_{thermal}$. This thermal motion is even larger than would be obtained from thermal equipartition, $m v_{thermal}^2 / 2 = 3 k_B T / 2$, because Fermi exclusion causes the electrons active in electrical conduction to have quite high energies (the Fermi energy). In Cu, for example, they have $T_F = 80,000$ K and $v_{thermal} = v_F = 1600$ km/s in Cu.

▶ The second is scattering, which is the cause of the randomness of the thermal motion. The charge carriers scatter off of impurities, defects, and thermal vibrations present in the material. This scattering is elastic in general, resulting in no loss of energy but in a redirection of velocity.

In the presence of such effects, our picture should not be of a charge carrier smoothly accelerating under the influence of an external force, but rather of a carrier with a large randomly directed velocity, scattering frequently, and with acceleration by the force between scatters resulting in a net motion in the direction of the electric force. The scattering randomly redirects the velocity, so the velocity due to the externally applied force is, on average, reset to zero after each collision. If the thermal speed is $v_{thermal}$ and the typical distance traveled between scatters is λ, then the time available for the externally applied force to accelerate a carrier between scatters is

$$t = \frac{\lambda}{v_{thermal}} \qquad (7.2)$$

The *average* velocity acquired from the applied force during this time is

$$\vec{v}_{ave} = \frac{1}{2}\,\vec{a}\,t = \frac{1}{2}\,\frac{\vec{f}}{m}\,\frac{\lambda}{v_{thermal}} \qquad (7.3)$$

This velocity is the average overall velocity because of the zeroing of the instantaneous velocity after each collision.

If we then use $\vec{J} = n\,q\,\vec{v}_{ave}$ where n is the number density of charge carriers and q is the charge per carrier, and we use $\vec{f} = q\,\vec{E}$, we then can write

$$\vec{J} = \left(\frac{n\,q^2\,\lambda}{2\,m\,v_{thermal}} \right) \vec{E} \qquad \implies \qquad \boxed{\vec{J} = \sigma\,\vec{E} \qquad \sigma = \frac{n\,q^2\,\lambda}{2\,m\,v_{thermal}}} \qquad (7.4)$$

Thus, we see our earlier expression is justified. This is *Ohm's Law*.

There is power dissipated in this process — the work done on the charge carriers by the electric field is lost to random motion when they scatter. The infinitesimal amount of energy lost per unit time dP in an infinitesimal volume $d\tau$ is equal to the work done by the electric field on the charge carriers:

$$dP = \text{number density} \cdot \text{velocity} \cdot \frac{\text{force}}{\text{carrier}}\,d\tau = n\,\vec{v}_{ave} \cdot \vec{f}\,d\tau = n\,\frac{\vec{J}}{n\,q} \cdot q\,\vec{E}\,d\tau = \vec{J} \cdot \vec{E}\,d\tau \tag{7.5}$$

This is known as *Joule Dissipation* or *Joule Heating*.

We note that the possibility of $\vec{E} \neq 0$ does not contradict our earlier discussions of conductors in electrostatics; here, we have non-stationary charges, where in that case we considered the final static situation after any currents had flowed.

Integral Version of Ohm's Law and Joule Heating

We integrate the above to obtain a more familiar version of Ohm's Law. We start with:

$$I = \int_{\mathcal{S}} da\, \hat{n} \cdot \vec{J} = \int_{\mathcal{S}} da\, \sigma\, \hat{n} \cdot \vec{E} \tag{7.6}$$

Let's assume the cross-sectional area of the conductor is constant and the conductor is uniform. This lets us do the area integral trivially, yielding $I = \sigma\, A\, \hat{n} \cdot \vec{E}$. If we then do a line integral directed along the wire, such that $d\vec{\ell} \propto \hat{n}$, we have

$$I\ell = \int d\ell\, I = \sigma A \int d\ell\, \hat{n} \cdot \vec{E} = \sigma A \int d\vec{\ell} \cdot \vec{E} = \sigma A\, V \tag{7.7}$$

$$\implies \quad V = I\, R \quad \text{with} \quad R = \frac{\ell}{A}\frac{1}{\sigma} \equiv \frac{\ell}{A}\rho \quad \text{and} \quad \rho = \frac{1}{\sigma} \tag{7.8}$$

which is the familiar version of Ohm's law in terms of current, voltage, and resistance. This is the *integral version of Ohm's Law* while $\vec{J} = \sigma\, \vec{E}$ is the differential (or local) version. We also define the resistivity ρ as the reciprocal of the conductivity σ. We can also integrate the Joule heating expression to get the usual integral expression for Joule heating:

$$P = \int_{\mathcal{V}} dP = \int_{\mathcal{V}} d\tau\, \vec{J} \cdot \vec{E} = \int_{\mathcal{S}} da \int d\ell\, \frac{I}{A}\, \hat{n} \cdot \vec{E} = I\, V = I^2 R = \frac{V^2}{R} \tag{7.9}$$

Steady-State Assumption and Uniform Conductivity $\Longrightarrow$ Zero Charge Density

Do we need to worry about charge accumulation in conductors? Let's calculate the divergence of $\vec{E}$ to find the charge density, assuming uniform conductivity:

$$\vec{\nabla} \cdot \vec{E} = \frac{1}{\sigma} \vec{\nabla} \cdot \vec{J} = 0 \qquad (7.10)$$

where the first step was possible by Ohm's Law and the assumed uniformity of the conductivity and the second step by the steady-state assumption on macroscopic scales. So the answer is no, as long as the conductivity is uniform and the system is steady-state, no charge density accumulates. *Note that this is not a circular argument: the steady-state assumption corresponded to $\partial \rho / \partial t = 0$, not $\rho = 0$. Now, with the combination of $\partial \rho / \partial t = 0$ and Ohm's Law, we conclude $\rho = 0$.*

Later, we will see how it is possible for charge to accumulate when we consider non-steady-state systems (in particular, with sinusoidal currents).

Note that our microscopic picture is not consistent with the steady-state assumption, but, averaged over time, our macroscopic picture is. On the microscopic scale, there are nonzero positive and negative charge density fluctuations away from the mean density associated with the nonuniform motion of the carriers (even on scales over which we can treat the carrier density as continuous), and then, of course, on the most microscopic scales, the smooth charge density consists of individual electrons that have some finite spatial extent.

Uniformity of Electric Field in a Uniform Wire

We implicitly assumed in proving the integral version of Ohm's Law above that the uniformity of the conductor implied that the field and thus the current were uniform over the cross-sectional area. We can prove this. We did not explicitly require that the electric field also be uniform with position along the wire, but we can prove that, too.

We define a uniform conductor to be one with uniform conductivity and uniform cross-sectional area.

We proved above that the charge density vanishes in a uniform conductor with steady currents. Therefore, the conductor satisfies Laplace's Equation. Dirichlet boundary conditions are set at the two ends of the conductor by the potential difference ΔV. We assume these equipotentials are (by connections to a battery) transverse to the wire axis at $z = 0$ and $z = \ell$. On the outer surface of the wire, $\vec{J} \cdot \widehat{n} = 0$ because no current flows out of the wire, which implies that $\vec{E} \cdot \widehat{n} = 0$, which provides a boundary condition on the normal gradient of the potential (a Neumann boundary condition). (Equivalently, this implies the charge density vanishes at the surface.)

We guess a solution that satisfies these boundary conditions,

$$V(\vec{r}) = \frac{\Delta V}{\ell}\, z \qquad \Longrightarrow \qquad \vec{E} = -\vec{\nabla} V = -\frac{\Delta V}{\ell}\, \hat{z} \qquad (7.11)$$

Note that we do not need the sinusoidal solutions from separation of variables here — we only need the linear solution (which we ignored in our discussion of separation of variables in linear coordinates). This will be of relevance for the homework, too!

This linear solution satisfies the boundary conditions — equipotential surfaces at $z = 0$ and $z = \ell$ and vanishing normal derivative at the outer surface (whose normal is always perpendicular to $\hat{z}$) — and therefore it must be the solution.

Therefore, it is valid to assume that the field is uniform over the cross-sectional area of the wire *and* along the length of the wire if the wire is of fixed cross-sectional area, the conductivity is uniform, and the currents are steady-state. The latter two conditions told us Laplace's Equation is satisfied, while the former one provided the *z*-translation symmetry needed to guess the solution.

What happens to this argument if the wire changes in some way along its length; *e.g.*, the conductivity changes, or the wire diameter changes?

Motional Electromotive Force

We deviate from Griffiths somewhat in the introduction of electromotive forces; his §7.1.2 just seems confusing.

Moving Rectangular Loop

Consider a rectangular loop with a resistor in it with part of the loop's area intersecting a region of uniform magnetic field perpendicular to the loop *into the page*, as shown in the figure.
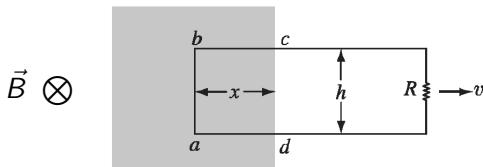
© 2013 Griffiths, *Introduction to Electrodynamics*

Let's consider the force on a charge carrier in the portion of the wire that intersects the field. If $\vec{v} = v\,\hat{x}$ and $\vec{B} = -B\,\hat{z}$ (into the page), then these charges feel a Lorentz force

$$\vec{F}_{mag} = (q\,v\,\hat{x}) \times (-B\,\hat{z}) = q\,v\,B\,\hat{y} \qquad (7.12)$$

Since this force is aligned with the vertical portion of the wire, the carriers in that section can move. Assuming for the moment the charge carriers are positive (the argument can be reversed if they are negative), they would start to collect at b at the top end of the vertical portion and a deficit would appear at a at the bottom end. The local electrostatic repulsion between like charges would cause the charge carriers to start flowing through the rest of the circuit and would prevent this clumping of carriers. In this way, a current is generated around the loop *without the influence of a large-scale electric field in the circuit*. If the loop is pulled at constant speed, one satisfies the steady-state assumption, with no charge buildup.

There is work done on a charge carrier during its movement up the vertical portion of the wire

$$W_{ab} = \int_a^b d\vec{\ell} \cdot \vec{F}_{mag} = q\, v\, B\, h \tag{7.13}$$

We will see below that this work is not done by the Lorentz force as suggested above (recall, the Lorentz force can do no work because $\vec{F}_{mag} \perp \vec{v}$), but it is nevertheless done. The energy gained by the charge carriers via this work is dissipated as Joule heating in the resistor because the carriers quickly reach some steady-state velocity and a steady-state current flows.

We define the work done per unit charge on the charges as they move from a to b as the *motional electromotive force* or *motional emf*:

$$\mathcal{E} = \frac{W_{ab}}{q} = v\, B\, h \tag{7.14}$$

Let's think about how we can interpret $\mathcal{E}$. Let I be the current that flows. I is q times the number of charges that flow per unit time past a given point. Therefore $I\mathcal{E}$ is the power being supplied to the ensemble of charges, the work done on them per unit time. By conservation of energy, it is also the power being dissipated in Joule heating in R. But we know that latter quantity is also $I^2 R$. Equating the two, we see

$$\mathcal{E} = I R \qquad (7.15)$$

That is, $\mathcal{E}$ plays the role of voltage in Ohm's Law for the resistor. $\mathcal{E}$ has the right units for this purpose. In fact, if one attaches a voltmeter across the resistor R, it will report a voltage $V = \mathcal{E}$: a voltmeter works essentially by measuring the current in a very large resistor $R' \gg R$ placed in parallel with R, and the current that will flow through R' is identical to what would flow if a battery $\mathcal{E}$ were placed across R with R' in parallel. So, what appeared to just be a work done on a unit charge now can be interpreted as equivalent to a voltage! But be sure to remember that the current is generated by movement of the circuit in a magnetic field; it is not due to an electric field! We will return to the distinction between $\mathcal{E}$ and a voltage later when we consider electromagnetic induction and Faraday's Law.

Let's now think about what force is doing the work. As we discussed some time ago, the Lorentz force does no work because $\vec{F} \perp \vec{v}$. However, a force must pull the loop. There is a force counteracting this force that the pulling force must match to keep the loop at constant speed: the Lorentz force due to the velocity the carriers have acquired in the $\hat{y}$ direction, which we will denote by $\vec{u} = u\,\hat{y}$. This force is

$$-\vec{F}_{pull} = \vec{F}'_{mag} = q\,u\,\hat{y} \times -B\,\hat{z} = -q\,u\,B\,\hat{x} \tag{7.16}$$

The total velocity of the charge carriers is

$$\vec{w} = \vec{v} + \vec{u} = v\,\hat{x} + u\,\hat{y} \tag{7.17}$$

The pulling force must cancel $\vec{F}'_{mag}$, so the work done per unit time by the pulling force is

$$\frac{dW_{pull}}{dt} = \vec{F}_{pull} \cdot \vec{w} = q\,u\,B\,\hat{x} \cdot (v\,\hat{x} + u\,\hat{y}) = q\,u\,B\,v \tag{7.18}$$

Note that the charge carriers move on a diagonal line relative to the lab frame as they move from point a to point b on the wire, with this line partly in the direction of $\vec{F}_{pull}$. It takes the charge carriers a time $t = h/u$ to move on this trajectory since their $\widehat{y}$ direction speed is u. Therefore, the work done by $\vec{F}_{pull}$ during the movement of a charge from a to b is:

$$W_{pull} = \frac{dW_{pull}}{dt} \frac{h}{u} = q\,B\,v\,h \tag{7.19}$$

$$\implies \quad \frac{W_{pull}}{q} = B\,v\,h = \mathcal{E} \tag{7.20}$$

That is, the work done by the pulling force, per unit charge, matches the motional emf. The pulling force provides the energy that is eventually dissipated as heat as the carriers flow through the resistor.

Mechanically, how does this work? A magnetic field does no work, so it should only change the direction of the velocity of the charge carriers. So, initially, when the pulling force begins to act and the carriers start to move in the x direction and feel a Lorentz force in the y direction, their x velocity starts to be transformed into y velocity. But the loop is being pulled at constant speed $v\,\widehat{x}$, so the walls of the wire exert a force so the carriers' x velocity remains equal to $v\,\widehat{x}$ as the magnetic force acts. Similarly, as the carriers acquire a velocity in the y direction, they feel $\vec{F}'_{mag}$ in the $-\widehat{x}$ direction, and the walls of the wire must exert a force to keep them moving at $v\,\widehat{x}$ in the x direction. By Newton's third law, the charge carriers exert a reaction force on the walls of the wire, which would slow down the loop if there were not a force pulling it. Thus, we see it is the force pulling the loop that ultimately provides the work to drive the current.

And note: All this motion is accomplished without a large-scale electric field. Of course, it relies on the microscopic Coulomb repulsion between like charge carriers and the Coulomb binding to the wire that keeps the charge carriers from flying out of the wire.

Returning to the emf itself, we can rewrite it in a useful form. We define the *magnetic flux* to be the integral of $\vec{B}$ dotted into the normal to a surface over the surface:

$$\Phi = \int_{\mathcal{S}} da\,\hat{n} \cdot \vec{B}(\vec{r}) \tag{7.21}$$

Using the definition of x in the figure, we have in this case

$$\Phi = B\,h\,x \tag{7.22}$$

The time derivative is

$$\frac{d\Phi}{dt} = B\,h\,\frac{dx}{dt} = -B\,h\,v \tag{7.23}$$

(x decreases with time for $v > 0$) which is just the negative of the motional emf. That is, we have

$$\mathcal{E} = -\frac{d\Phi}{dt} \tag{7.24}$$

Moving Arbitrary Loop

Let's prove rigorously that this rule holds more generally for any shape of loop with any type of motion through an arbitrary magnetic field. Consider the motion of a closed loop of arbitrary shape over a time dt. The loop is defined by a contour $\mathcal{C}(t)$ that depends on t. Each point on the loop has a velocity $\vec{v}$ that may depend on the position on the loop. Regardless, each piece of the loop moves by the vector $\vec{v}\,dt$ during this time where $\vec{v}$ is position-dependent. The charges in that piece of the loop acquire a velocity $\vec{u}$ along the direction of the loop due to the action of the Lorentz force during that time.

Surface $\mathcal{S}$

Ribbon

P dl
θ
vdt
P'

P
P'

$d\mathbf{a}$

Loop at   Loop at
time t time $(t + dt)$

Enlargement of $d\mathbf{a}$
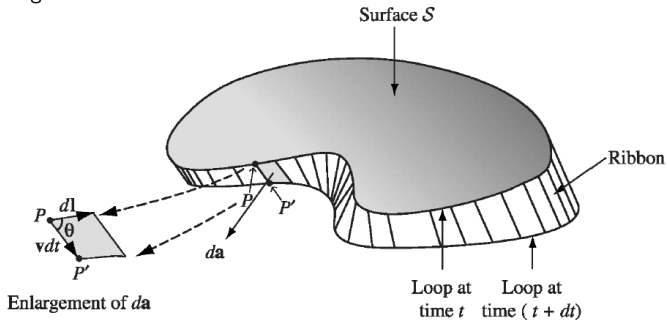
© 2013 Griffiths, *Introduction to Electrodynamics*

We can see, through some work, that the motion $\vec{v}$ also describes the change in the area of and flux through the loop. The flux changes by

$$d\Phi = \int_{\mathcal{S}(\mathcal{C}(t+dt))} da\, \widehat{n}_{\mathcal{C}} \cdot \vec{B} - \int_{\mathcal{S}(\mathcal{C}(t))} da\, \widehat{n}_{\mathcal{C}} \cdot \vec{B} \tag{7.25}$$

We subscript $\widehat{n}$ with $\mathcal{C}$ to distinguish it from a different $\widehat{n}$ we define below. Let's rewrite this expression in a more usable form. Consider a closed surface that consists of the surfaces defined by $\mathcal{C}(t)$ and $\mathcal{C}(t + dt)$ as well as the ribbon-like surface connecting the two contours. (If the two contours were circular loops, the ribbon-like surface would be the wall of the cylinder formed by the two contours.) $\vec{\nabla} \cdot \vec{B} = 0$ tells us the surface integral of $\widehat{n}_{\mathcal{S}} \cdot \vec{B}$ (where $\widehat{n}_{\mathcal{S}}$ is the outward surface normal, identical to $\widehat{n}_{\mathcal{C}}$ for only some parts of the surface) through this surface vanishes. That surface integral is related to the above integrals by

$$0 = \oint_{closed\ \mathcal{S}} da\, \widehat{n}_{\mathcal{S}} \cdot \vec{B} = -\int_{\mathcal{S}(\mathcal{C}(t+dt))} da\, \widehat{n}_{\mathcal{C}} \cdot \vec{B} + \int_{\mathcal{S}(\mathcal{C}(t))} da\, \widehat{n}_{\mathcal{C}} \cdot \vec{B} + \int_{ribbon} da\, \widehat{n}_{\mathcal{S}} \cdot \vec{B} \tag{7.26}$$

where we have used the fact that $\widehat{n}_{\mathcal{S}} = -\widehat{n}_{\mathcal{C}}$ on the $\mathcal{S}(\mathcal{C}(t + dt))$ surface and $\widehat{n}_{\mathcal{S}} = \widehat{n}_{\mathcal{C}}$ on the $\mathcal{S}(\mathcal{C}(t))$ surface. (The direction of $\widehat{n}_{\mathcal{C}}$ is set by the direction of $d\vec{\ell}$ in the figure and the right-hand rule.) The negative sign is present in the former because the orientation of $\widehat{n}_{\mathcal{C}}$ that maintains its direction as the contour moves has $\widehat{n}_{\mathcal{C}}(t + dt)$ on this surface pointing into the enclosed volume rather than outward.

The first two terms give $-d\Phi$, so

$$d\Phi = \int_{ribbon} da \, \widehat{n}_S \cdot \vec{B} \tag{7.27}$$

The area element on the ribbon (with outwardly directed normal as in the above integral) is given by

$$\widehat{n}_S \, da = d\vec{r} \times d\vec{\ell} \tag{7.28}$$

where: $d\vec{\ell}$ is the line element along $\mathcal{C}(t)$ with orientation set by consistency with $\widehat{n}_\mathcal{C}$ for $\mathcal{S}(\mathcal{C}(t))$ and the right-hand rule; and $d\vec{r}$ is the change in the vector position of that line element between t and $t + dt$. The difference between these two positions is related to $\vec{v}$, $d\vec{r} = \vec{v} \, dt$, so:

$$\widehat{n}_S \, da = \vec{v} \, dt \times d\vec{\ell} \tag{7.29}$$

$$\text{Therefore,} \quad d\Phi = \oint_{\mathcal{C}(t)} \left(\vec{v} \, dt \times d\vec{\ell} \right) \cdot \vec{B} \tag{7.30}$$

We turned an area integral into a line integral, but it still calculates magnetic flux.

Since $\vec{u} \parallel d\vec{\ell}$, we can add $\vec{u}$ to $\vec{v}$ to obtain $\vec{w}$ without affecting the integral:

$$d\Phi = \oint_{\mathcal{C}(t)} \left(\vec{w}\, dt \times d\vec{\ell} \right) \cdot \vec{B} \tag{7.31}$$

Using the cyclic property of the triple vector product, reversing the resulting cross product $\vec{B} \times \vec{w}$, and moving dt to the left side, we obtain

$$\frac{d\Phi}{dt} = -\oint_{\mathcal{C}(t)} d\vec{\ell} \cdot \left(\vec{w} \times \vec{B} \right) \tag{7.32}$$

The quantity $\vec{w} \times \vec{B}$ is just the Lorentz force per unit charge:

$$\frac{d\Phi}{dt} = -\oint_{\mathcal{C}(t)} d\vec{\ell} \cdot \frac{\vec{F}_{mag}}{q} \tag{7.33}$$

The integral of the Lorentz force per unit charge integrated around the loop is the generalization for arbitrary loops of our earlier expression for the motional emf for the rectangular loop (earlier, we integrated over only the section of length h from a to b of the rectangular loop for which $\vec{F}_{mag}$ was nonzero), so

$$\boxed{\; \mathcal{E} = \oint_{\mathcal{C}(t)} d\vec{\ell} \cdot \frac{\vec{F}_{mag}}{q} = -\frac{d\Phi}{dt} \qquad \text{moving circuit} \;}$$ (7.34)

The motional emf, as defined by the line integral of the Lorentz force per unit charge around the loop, is given by the negative of the rate of change of the magnetic flux through the loop. The signs of the line integral and the flux are set by requiring that the orientation of the line integral (via $d\vec{\ell}$) be consistent via the right-hand rule with the orientation of the surface normal $\hat{n}_{\mathcal{C}}$ used for the flux calculation.

Example 7.1: Alternating Current Generator (Griffiths 7.10)

The classic and pervasive use of the above relationship is the alternating current generator. Consider a square loop placed in a uniform magnetic field and rotated about a midline at constant angular speed ω. That is, the rotation is such that, at one point of the motion, the magnetic field is normal to the loop while, one fourth of the period before or after this time, the magnetic field is in the plane of the loop. What is the motional emf around the loop generated by this motion?
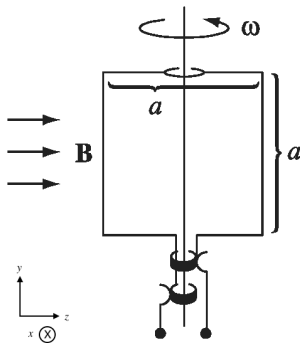
© 2013 Griffiths, *Introduction to Elec-trodynamics*

The magnetic field is constant, so the flux is just given by B times the area of the loop projected onto the direction of $\vec{B}$:

$$\Phi(t) = a^2\,\vec{B} \cdot \widehat{n}(t) = A\,B\,\cos\omega t \qquad (7.35)$$

where we have chosen $\widehat{n} \parallel \vec{B}$ at $t = 0$ and written $a^2 = A$. Thus, the motional emf is

$$\mathcal{E}(t) = -\frac{d\Phi}{dt} = A\,B\,\omega\,\sin\omega t \qquad (7.36)$$

This is of course how 60-Hz AC voltage is generated.

It is instructive to think again about the Lorentz force experienced by the charge carriers in the loop and see how it generates the motional emf. Let the magnetic field be $\vec{B} = B\,\hat{z}$ and let the axis of rotation be $+\hat{y}$. Suppose the loop is just moving past having $\hat{n} = -\hat{x}$, as shown in the figure. Then the carriers all have a velocity parallel to $\pm\hat{x}$ due to the motion of the loop (this is $\vec{v}$). (They also have motion in the $\hat{z}$ direction, but this is parallel to $\vec{B}$ and thus no Lorentz force is generated.) The carriers in the sections of the loop parallel to $\hat{z}$ (perpendicular to the axis of rotation, parallel to the field) cannot move in response to this force because they feel a force in the $\hat{y}$ direction, transverse to the section of wire they are in. Those in the parts of the loop parallel to $\pm\hat{y}$ (parallel to axis of rotation) also feel a force along $\hat{y}$, and they can move along $\hat{y}$. As the loop turns away from this orientation, the arm at $+\hat{z}\,a/2$ has velocity in the $+\hat{x}$ direction and vice versa for the arm at $-\hat{z}\,a/2$. Positive charge carriers in these arms feel forces in the $-\hat{y}$ and $+\hat{y}$ directions, respectively. This forces a current to flow in direction defined by the $-\hat{n} = +\hat{x}$ orientation by right-hand rule, generating a field through the loop in the $-\hat{n} = +\hat{x}$ direction.

As the loop passes through this orientation, the flux is zero and is changing from negative ($\hat{n} \cdot \vec{B} < 0$) to positive ($\hat{n} \cdot \vec{B} > 0$). One can see that the driven current is in the direction needed for its field to counter the change in magnetic flux. This is a manifestation of *Lenz's Law*, which we will return to later.

If one taps the loop as is typical for such a generator, as shown in the figure, the tap connected to the $+\widehat{z}\,a/2$ arm will have positive voltage and the tap connected to the $-\widehat{z}\,a/2$ arm will have negative voltage because they need to drive a current in an external circuit that carries current in the direction consistent with that argued above, from the $+\widehat{z}\,a/2$ arm to the $-\widehat{z}\,a/2$ arm.

Note the polarity of the above statement: we decide the sign of the voltage at the taps not by what is needed to drive the current in the loop (which is driven by the Lorentz force, not by this voltage) but rather by the sign needed to drive the current in the external load (the resistor) so that current exits the loop, goes through the load, and returns to the loop, where it is needed to conserve charge.

Lecture 21:

Electrodynamics II:

Electromagnetic Induction
Faraday's Law
Inductance

Date Revised: 2024/04/09 07:00
Revised lecture break
Date Given: 2024/04/04

Electromagnetic Induction

Faraday's Law

We are going to consider three different physical situations:

- ▶ *Moving loops*: As we considered above, the magnetic field is stationary but the loop is moving.
- ▶ *Moving magnetic fields*: The loop is held fixed but the magnetic field is changing because the currents sourcing the field are being translated.
- ▶ *Changing magnetic fields:* Both the loop and the sources of the field are stationary, but the currents sourcing the field are changing.

We just proved using the Lorentz Force Law that the first situation results in a *motional emf*: a force that causes the flow of a current around the loop, given by Equation 7.34:

$$\mathcal{E} = \oint_{\mathcal{C}(t)} d\vec{\ell} \cdot \frac{\vec{F}_{mag}}{q} = -\frac{d\Phi}{dt} \qquad \text{moving circuit} \qquad (7.37)$$

Faraday's Law consists of the *empirical observation* that the same rule applies for the second and third situations. The subtlety is this: this law could not have been derived using the Lorentz Force applied to the situation described above of a fixed loop and a moving and/or changing magnetic field: there is no magnetic force if the charge carriers are not being forced to move in the magnetic field by the loop being pulled. A natural and important corollary is that the emf that appears for a moving or changing magnetic field is not due to a magnetic force. Rather, since the loop is at rest in the second and third situations, the force that appears arises from a true *electric field*. Mathematically, we write Faraday's Law as

$$\mathcal{E} = \oint_{\mathcal{C}(t)} d\vec{\ell} \cdot \frac{\vec{F}_{elec}}{q} = -\frac{d\Phi}{dt} \qquad \text{moving or changing magnetic field} \qquad (7.38)$$

We see that it is identical in form to the Lorentz Force law applied to a moving loop with the replacement of $\vec{F}_{mag}$ by $\vec{F}_{elec}$.

Combining the two forms, and defining $\vec{E}_{ind} = \vec{F}_{elec}/q$ where $_{ind}$ indicates that the electric field here is not an electrostatic one due to Coulomb's Law but rather an "induced" field due to the changing magnetic flux, we then may write a common law that applies in any situation:

$$\mathcal{E} = \oint_{\mathcal{C}(t)} d\vec{\ell} \cdot \left[\vec{E}_{ind} + \frac{\vec{F}_{mag}}{q} \right] = -\frac{d\Phi}{dt} = -\frac{d}{dt} \int_{\mathcal{S}(\mathcal{C}(t))} da \, \hat{n}(\vec{r}, t) \cdot \vec{B}(\vec{r}, t) \quad (7.39)$$

If there is any ambiguity in the sign, one should apply *Lenz's Law*: the emf has a sign such that the polarity of the current it would drive produces a magnetic field that counters the change in magnetic field. We will prove Lenz's Law explicitly later.

Quasistatic Assumption

Note that we have implicitly assumed in our derivations that the current everywhere in the loop responds instantaneously to the total emf on the left side, that there is no time delay between a buildup of charge at one point in the circuit and the driving of a current around the loop. We made the same assumption in deriving Ohm's Law. This is the "quasistatic assumption," that all fields and currents everywhere change instantaneously and that information is propagated infinitely quickly. Formally, this assumption consists of saying that, given a typical physical scale for a system ℓ and a typical timescale for variation t, we have

$$t \gg \ell/c \tag{7.40}$$

where c is the *speed of light* that will be defined later.

We will release this assumption when we discuss electromagnetic waves and radiation.

Motional EMF, Faraday's Law, Galilean Relativity, and Galilean Field Transformations

When first proposed, Faraday's Law was an empirical observation. However, it could have been justified using the principle of Galilean relativity: physics is the same in all inertial reference frames, those moving at constant velocity.

Consider the problem of the magnetic field moving at fixed velocity. One could go to the rest frame of the magnetic field and consider the loop to be moving at fixed velocity as in our moving loop cases. The magnetic force implied by the motional emf appears. In Galilean relativity, forces are invariant upon change of inertial (fixed velocity) frame. This would imply that the magnetic force in the field-fixed frame is still present in the loop-fixed frame, but now we interpret it as an electric force because the loop is not moving.

In the case of changing magnetic fields, we simply have to invoke the expectation that the loop has no way of knowing whether it experiences a changing field because the current sourcing the field is moving or because it is changing: it only knows about the field that results, not the source of the field.

This Galilean relativity argument was, however, not recognized until after Faraday's observation.

We can make use of this argument to understand how electric and magnetic fields mix with each other under such Galilean (nonrelativistic) transformations. Let's assume we have written down our law, Equation 7.39, in both the rest frame of the loop and in the lab frame in which the loop is moving. The fields and position vectors in the loop rest frame are given $'$ symbols, the ones in the lab frame have no primes. The total emf can be determined explicitly using a voltmeter to measure the voltage across the resistor in the loop, and it is a scalar that is independent of frame (the reading on the voltmeter doesn't change if you see the voltmeter moving with the loop!). So we can equate the lab and rest frame expressions through $\mathcal{E}$:

$$\oint_{\mathcal{C}'} d\vec{\ell}' \cdot \vec{E}'_{ind} = \oint_{\mathcal{C}(t)} d\vec{\ell} \cdot \left[\vec{E}_{ind} + \frac{\vec{F}_{mag}}{q} \right] \tag{7.41}$$

($\mathcal{C}' = \mathcal{C}(t = 0)$ can be assumed by appropriate choice of when the lab and loop rest frame coordinate systems coincide). Now, let's use our expression for the magnetic force term from our derivation of Equation 7.34, dropping the $\vec{u}$ contribution that we had added in:

$$\oint_{\mathcal{C}'} d\vec{\ell}' \cdot \vec{E}'_{ind} = \oint_{\mathcal{C}(t)} d\vec{\ell} \cdot \left[\vec{E}_{ind} + \vec{v} \times \vec{B} \right] \tag{7.42}$$

Since the circuit is arbitrary, we may thus conclude

$$\vec{E}'_{ind} = \vec{E}_{ind} + \vec{v} \times \vec{B} \tag{7.43}$$

The equation can be taken to be completely general because adding a standard electrostatic field to both sides would leave the statement true while accounting for such electrostatic fields:

$$\boxed{\vec{E}' = \vec{E} + \vec{v} \times \vec{B} \qquad \vec{E} = \vec{E}_{Coul} + \vec{E}_{ind}} \tag{7.44}$$

Therefore, this is a rule for how electric fields transform from one frame to another under Galilean relativity, regardless of the source of the field. Electric fields are not the same in a fixed and a moving frame if magnetic fields are present, even before special relativity is considered! Special relativity then *only* adds correction coefficients to the above equation.

It is important to note that the expectation that the electrostatic fields do not depend on frame has been an assertion so far, based on the assumption that Coulomb's Law is unaffected by whether the charges are moving or not. We will return to this point later in connection to Maxwell's Equations, as it will lead to a symmetrization of the above equation between $\vec{E}$ and $\vec{B}$.

Galilean relativity is consistent with the quasistatic assumption. We need only consider special relativity when the nonzero travel time of light becomes important because special relativity says the speed of light is the same in all frames.

Example 7.2: A Stationary Alternating Current Generator

Recall the previous example of an AC generator that used a rotating square loop in a constant magnetic field. Instead, hold the loop fixed but assume that the magnetic field is being varied sinusoidally, $\vec{B}(t) = \vec{B_0} \cos \omega t$. Then the flux is

$$\Phi(t) = A \, \vec{B}(t) \cdot \widehat{n} = A \, B_0 \, \cos \omega t \qquad (7.45)$$

Therefore, the emf generated is

$$\mathcal{E}(t) = -\frac{d\Phi}{dt} = A \, B_0 \, \omega \, \sin \omega t \qquad (7.46)$$

just as before.

Note, again, the polarity of the emf! As before, the emf's polarity is such that it causes current to flow in an external resistor attached to the two ends of the circuit in a direction consistent with the current that flows in the loop.

Something one has to be careful about is incorrectly believing that, because of the emf's sign, it should also drive a current in the zero-resistance *loop* in the direction implied by the emf. That erroneous belief arises because one is assuming the electric field is conservative, that the integral of $\vec{E}$ around a loop vanishes. No: that sign of emf would drive current in the wrong direction! For the current flowing in the loop, the emf measures the work per unit charge done by $\vec{E}_{ind} + \vec{F}_{mag}/q$ as they push the current around the loop, but they are not pushing the charges down an electrostatic potential! The *effect* of having this current flow is that the same current flows through the resistor, creating an apparent potential drop across the resistor that we can measure with a voltmeter. But the voltmeter is just measuring the current flowing through a known resistance, which, by Ohm's Law, is proportional to the line integral of the electric field through the resistor. The voltmeter's ability to measure something that looks like a voltage does not imply that an electrostatic potential can be defined everywhere in the loop and resistor!

In thinking about what causes the current to flow, it is better to visualize the electric field: one recognizes that the changing magnetic field generates an electric field that pushes current in the direction it needs to flow to counter the change in magnetic field. *This electric field has nonzero loop integral around the circuit!* Therefore, the existence of the emf $\mathcal{E}$ at the ends of the circuit does not imply the same emf is experienced by the current flowing in the loop itelf; the nonzero loop integral of the electric field invalidates the rule that the total voltage drop around a loop must vanish, which is the source of the misconception that $\mathcal{E}$, appearing at the ends of the circuit, is also the driver of the current in the loop.

Differential Version of Faraday's Law

Consider the special case of an arbitrary closed contour $\mathcal{C}$ fixed in space. Equation 7.39 tells us

$$\oint_{\mathcal{C}} d\vec{\ell} \cdot \vec{E} = -\frac{d}{dt} \int_{\mathcal{S}(\mathcal{C})} da\, \widehat{n}(\vec{r}) \cdot \vec{B}(\vec{r}, t) \tag{7.47}$$

Let's use Stokes' Theorem on the left side, and, since the contour is time-independent, we can move the time derivative inside the integral on the right side. We turn it into a partial derivative to make it clear that we do not need to worry about any possible time-dependence of $\vec{r}$ (of which there is none here). This yields

$$\oint_{\mathcal{S}(\mathcal{C})} da\, \widehat{n}(\vec{r}) \cdot \left[\vec{\nabla} \times \vec{E}(\vec{r}) \right] = -\int_{\mathcal{S}(\mathcal{C})} da\, \widehat{n}(\vec{r}) \cdot \frac{\partial \vec{B}(\vec{r}, t)}{\partial t} \tag{7.48}$$

Since the loop is arbitrary, the integrands must be equal:

$$\boxed{\vec{\nabla} \times \vec{E}(\vec{r}) = -\frac{\partial \vec{B}(\vec{r}, t)}{\partial t}} \tag{7.49}$$

This differential version of Faraday's Law is the generalization of $\vec{\nabla} \times \vec{E} = 0$ for time-dependent situations. We now explicitly see what was said in the previous example: a changing $\vec{B}$ creates a nonconservative electric field!

Biot-Savart and Ampere's Law for the Induced Electric Field in the Absence of Charges

If we consider the special case of no charge density, then we have

$$\vec{\nabla} \cdot \vec{E} = 0 \qquad \vec{\nabla} \times \vec{E} = -\frac{\partial \vec{B}}{\partial t} \qquad (7.50)$$

This is mathematically identical to the equations of magnetostatics,

$$\vec{\nabla} \cdot \vec{B} = 0 \qquad \vec{\nabla} \times \vec{B} = \mu_o \vec{J} \qquad (7.51)$$

In magnetostatics, we saw that the above two equations, combined with the assumption $\vec{\nabla} \cdot \vec{A} = 0$, yielded Poisson's Equation for $\vec{A}$ with $\mu_o \vec{J}$ as the source (Equation 5.56). By correspondence, we can thus state

$$\boxed{\vec{E} = \vec{\nabla} \times \vec{A}_E \qquad \nabla^2 \vec{A}_E = \frac{\partial \vec{B}}{\partial t} \qquad \vec{\nabla} \cdot \vec{A}_E = 0} \qquad (7.52)$$

This is of course very interesting: we see that $\vec{E}$ receives a contribution from a vector potential that satisfies Poisson's Equation with $\partial \vec{B}/\partial t$ as the source!

Now, if we assume appropriate boundary conditions — fields falling off at infinity, no other surfaces on which the vector potential or field are specified — then we know from Equation 5.56 that the solution to the Poisson's Equation for $\vec{A}_E$ is

$$\vec{A}_E(\vec{r}) = -\frac{1}{4\pi} \int_{\mathcal{V}} d\tau' \frac{\frac{\partial \vec{B}(\vec{r}')}{\partial t}}{|\vec{r} - \vec{r}'|} \tag{7.53}$$

Finally, we may take the curl of the above expression to recover the analogue of the Biot-Savart Law. We did this backwards in the case of magnetostatics: we started with the empirical Biot-Savart Law and derived that the field could be written as the curl of the form of the vector potential corresponding to the above. Nevertheless, that proof could be reversed, so we may conclude that the analogous Biot-Savart Law holds (compare to Equation 5.32)

$$\vec{E}(\vec{r}) = -\frac{1}{4\pi} \int_{\mathcal{V}} d\tau' \frac{\frac{\partial \vec{B}(\vec{r}')}{\partial t} \times (\vec{r} - \vec{r}')}{|\vec{r} - \vec{r}'|^3} = -\frac{1}{4\pi} \frac{\partial}{\partial t} \int_{\mathcal{V}} d\tau' \frac{\vec{B}(\vec{r}') \times (\vec{r} - \vec{r}')}{|\vec{r} - \vec{r}'|^3} \tag{7.54}$$

where we pulled the time derivative outside the integral under the assumption that the volume itself is time-independent.

We also note that, because $\vec{E}$ satisfies the analogue of Ampere's Law, one can apply standard Ampere's Law techniques for finding $\vec{E}$ when $\frac{\partial \vec{B}}{\partial t}$ is given.

Caution: We have made the *quasistatic* assumption, that all time derivatives are small enough that the propagation time for disturbances in the magnetic fields is much less than the timescales on which the field vary. This is what allows us to use the magnetostatic formulae in time-varying situations. If the time derivatives become large, then one needs the full formalism of electromagnetic waves, which we will develop later.

Poisson's Equation for Induced Electric Field, Proof of Lenz's Law

Further pursuing the analogy to magnetostatics, let's see what we get if take the curl of the curl of the induced electric field:

$$\vec{\nabla} \times \left(\vec{\nabla} \times \vec{E}_{ind} \right) = \vec{\nabla} \times \left(-\frac{\partial \vec{B}}{\partial t} \right) = -\frac{\partial}{\partial t} \left(\vec{\nabla} \times \vec{B} \right) = -\mu_o \frac{\partial \vec{J}}{\partial t} \qquad (7.55)$$

We may rewrite the left side using the vector identity for the curl of the curl as we did when deriving Poisson's Equation for $\vec{A}$ in terms of $\vec{J}$ (Equation 5.56):

$$\vec{\nabla} \left(\vec{\nabla} \cdot \vec{E}_{ind} \right) - \nabla^2 \vec{E}_{ind} = -\mu_o \frac{\partial \vec{J}}{\partial t} \qquad (7.56)$$

If we again assume no charge density (valid since we are considering only the induced electric field $\vec{E}_{ind}$) and that the currents are localized so the fields fall off appropriately at infinity, we have a Poisson's Equation for $\vec{E}_{ind}$, whose solution we know:

$$\boxed{\nabla^2 \vec{E}_{ind} = \mu_o \frac{\partial \vec{J}}{\partial t} \quad \overset{\text{localized currents}}{\Longleftrightarrow} \quad \vec{E}_{ind}(\vec{r}) = -\frac{\mu_o}{4\pi} \int_{\mathcal{V}} d\tau' \, \frac{\frac{\partial \vec{J}(\vec{r}')}{\partial t}}{|\vec{r} - \vec{r}'|}} \qquad (7.57)$$

Because of the vector alignment of $\vec{E}_{ind}$ and $-\partial \vec{J}/\partial t$, we thus have *Lenz's Law*: the induced electric field is in the direction needed to drive a current to counter the change in the current that is causing the changing magnetic field.

Furthermore, using Equation 5.56, the relation between $\vec{A}$ and $\vec{J}$ obtained by solving the Poisson Equation $\nabla^2 \vec{A} = -\mu_o \vec{J}$, we have

$$\vec{E}_{ind}(\vec{r}) = -\frac{\partial}{\partial t} \frac{\mu_o}{4\pi} \int_{\mathcal{V}} d\tau' \, \frac{\vec{J}(\vec{r}')}{|\vec{r} - \vec{r}'|} = -\frac{\partial \vec{A}}{\partial t} \tag{7.58}$$

which is a rather remarkable statement — this is an almost trivial sourcing equation for $\vec{E}_{ind}$, certainly much simpler than the sourcing equation involving $\partial \vec{J}/\partial t$ or the Biot-Savart Law for $\vec{E}_{ind}$ sourced by $\partial \vec{B}/\partial t$.

We will return to this contribution to the electric field when we reconsider our scalar and vector potentials after writing down Maxwell's Equations and in light of the fact that Faraday's Law implies the electric field is not curl-free and thus cannot be derived from a scalar potential.

Example 7.3: Induced Electric Field for Coaxial Conductors (Griffiths 7.16)

An alternating current $I = I_0 \cos \omega t$ flows down a long straight wire of negligible radius and returns along a coaxial conducting tube of radius a and negligible thickness. Both conductors are assumed to be perfect (infinite conductivity). We want to find the induced electric field as a function of the transverse radius s in cylindrical coordinates.

For reasons that we will be able to explain later when we discuss EM waves in the presence of conductors, the currents flow in sheets at the surfaces of the conductors because they have infinite conductivity.

In the region between the wire and the outer conductor, the field of the wire is the usual $\vec{B}(s, t) = \widehat{\phi} \, \mu_o \, I(t)/2 \, \pi \, s$. The magnetic field of the return-current cylinder is zero inside (consider an Amperian loop in the xy-plane with radius $s < a$: none of the return current flows through the surface enclosed by that loop). Outside the return-current sheet, its magnetic field is that of a wire carrying the total return current, which has the same magnitude but opposite sign of the field of the inner wire. Thus, the total magnetic field is the inner conductor's magnetic field between the conductors and is zero outside the outer conductor.

The system has azimuthal and z-translation symmetry, so the induced electric field must have the form $\vec{E} = E_s(s) \, \widehat{s} + E_\phi(s) \, \widehat{\phi} + E_z(s) \, \widehat{z}$.

If we think about what kind of Amperian loop has a nonzero flux of $\partial \vec{B}/\partial t$ (not $\vec{J}$!), it is a loop in the sz plane with normal in the $\widehat{\phi}$ direction. Let's first consider a loop of this kind with one z leg at infinity and the other at $s > a$. The contributions to the loop integral of the electric field along the two radial legs cancel, and the contribution from the leg at infinity vanishes assuming the fields fall off as $s \to \infty$, so this loop only gets a contribution from the z leg at finite radius, which picks out $E_z(s > a)$. The enclosed flux of $\partial \vec{B}/\partial t$ vanishes, so we can conclude $E_z(s > a) = 0$.

Now, repeat with one z leg at s between 0 and a and one z leg outside the outer conductor. The radial legs cancel and the z leg outside the outer conductor contributes nothing. When calculating the enclosed flux of $\partial \vec{B}/\partial t$, a similar thing holds: there is no magnetic field outside a, so the area integral only goes from s to a. If the loop's z dimension is ℓ, we have

$$E_z(s < a)\, \ell = -\int_0^\ell dz \int_s^a ds' \, \frac{\partial B_\phi(s', t)}{\partial t} = -\frac{\mu_o}{2\pi} \frac{\partial I}{\partial t} \ell \int_s^a \frac{ds'}{s'} \qquad (7.59)$$

$$= \frac{\mu_o}{2\pi} \omega\, I_0\, \ell\, \sin \omega t\, \ln \frac{a}{s} \qquad (7.60)$$

$$\implies E_z(s < a) = \frac{\mu_o}{2\pi} \omega\, I_0\, \sin \omega t\, \ln \frac{a}{s} \qquad (7.61)$$

Note the sign: taking the loop normal to be $\widehat{\phi}$ implies that the z leg with the nonzero contribution yields a positive contribution. Then the usual minus sign enters, which is cancelled by the sign of the derivative of $\cos \omega t$.

We can see $E_\phi(s)$ vanishes by using a loop in the $s\phi$ plane that has radial legs (ϕ constant) and azimuthal legs (s constant). One azimuthal leg can be taken to infinity so it yields no contribution, and the radial legs' contributions cancel, leaving only the contribution from the azimuthal leg at finite radius. But, unlike the E_z case, this loop has no magnetic flux and thus no $\partial \vec{B}/\partial t$ through it, so $E_\phi(s) = 0$.

Finally, consider E_s, which we can show to vanish by both a conceptual and a mechanical argument. As we argued above, E_s can be a function of s only and must be independent of z. Suppose E_s points outward along $\hat{s}$ at a particular s and consider $E_s(s, z = 0)$. If we rotate the system about this direction by 180°, then the current changes direction. But $E_s(s, z = 0)$ cannot change direction (sign) — it is tied to the current distribution. Yet the reversal of the direction of the current changes the sign of $\vec{B}$ and thus $\partial \vec{B}/\partial t$. Then, by the Biot-Savart Law for $\vec{E}$, $\vec{E}$ should change sign. We have a contradiction unless $E_s(s, z = 0) = 0$. Because of z-translation symmetry, the same must hold at any z.

More mechanically, consider the Biot-Savart integral for $\vec{E}$. Given that $\vec{B}$ and $\partial \vec{B}/\partial t$ are both proportional to $\hat{\phi}$, the vector $\vec{r} - \vec{r}'$ must have a piece proportional to $\hat{z}$ to yield a contribution to the $\hat{s}$ component of $\vec{E}$. But $\vec{B}$ is independent of z, while the $\hat{z}$ component of $\vec{r} - \vec{r}'$ is odd about $z = z'$. So the integrand is odd about $z = z'$, causing the integral to vanish.

Thus, $$\vec{E}(s < a, t) = \hat{z}\, \frac{\mu_o}{2\pi}\, \omega\, I_0\, \sin\omega t\, \ln\frac{a}{s} \qquad \vec{E}(s > a, t) = 0 \qquad (7.62)$$

There are many other approaches to this problem.

▶ We could have derived the same $\vec{E}$ by first determining $\vec{A}$ and taking its time derivative. We calculated $\vec{A}$ of a wire in Ph106b Problem Set 7:

$$\vec{A}(s) = -\frac{\mu_o I}{2\pi} \ln \frac{s}{a} \qquad (7.63)$$

The vector potential vanishes for $s > a$ because the vector potential of the return current cancels the above vector potential of the wire. If we take $I = I_0 \cos \omega t$ and calculate $\vec{E} = -\partial \vec{A}/\partial t$, we get the same result as we obtained already.

▶ We could have also used the same Poisson's Equation solution form for the Poisson's Equation in which $\partial \vec{J}/\partial t$ sources $\vec{E}$.

▶ Lastly, instead of using the Ph106b PS7 solution for $\vec{A}$, we could have obtained $\vec{A}$ from $\vec{B}$ using the same types of loops we used to obtain $\vec{E}$ from $\partial \vec{B}/\partial t$.

No matter how one finds $\vec{E}$, one can see the sign makes sense. This $\vec{E}$ tries to drive a current parallel or antiparallel to the current already flowing in the wire. When the current is decreasing, the electric field is increasing to try to drive a current in the same direction in which current is being lost by the decreasing current. It tries to generates a magnetic field that would compensate for the magnetic field that is being removed by the decreasing central conductor current. And vice versa for an increasing current.

Note how, when we can calculate the induced electric field directly, there is no ambiguity about the direction the driven current would flow, unlike for $\mathcal{E}$.

Mutual Inductance

Except for the example we just did, We have so far considered magnetic fields and fluxes in the abstract, without any concern about where they come from. But they are generated by currents, so it is natural to want to connect the Faraday's Law emf to changing currents. We do that through mutual inductance.

Consider two circuits $\mathcal{C}_1$ and $\mathcal{C}_2$. Suppose a current I_1 is flowing in $\mathcal{C}_1$. The magnetic flux at $\mathcal{C}_2$ is

$$\Phi_{21} = \int_{\mathcal{S}(\mathcal{C}_2)} da_2\, \widehat{n}_2 \cdot \vec{B}_1(\vec{r}_2) = \int_{\mathcal{S}(\mathcal{C}_2)} da_2\, \widehat{n} \cdot \left[\vec{\nabla} \times \vec{A}_1(\vec{r}_2) \right] = \oint_{\mathcal{C}_2} d\vec{\ell}_2 \cdot \vec{A}_1(\vec{r}_2) \quad (7.64)$$

where we used the fact that $\vec{B}$ is derived from a vector potential followed by Stokes' Theorem. Now, let's use the relation between $\vec{A}_1$ and the current in $\mathcal{C}_1$ using the usual solution of the Poisson's Equation for $\vec{A}_1$ (assuming appropriate boundary conditions):

$$\Phi_{21} = \frac{\mu_o}{4\pi} \oint_{\mathcal{C}_2} d\vec{\ell}_2 \cdot \oint_{\mathcal{C}_1} \frac{I_1 d\vec{\ell}_1}{|\vec{r}_2 - \vec{r}_1|} = \frac{\mu_o}{4\pi} I_1 \oint_{\mathcal{C}_2} \oint_{\mathcal{C}_1} \frac{d\vec{\ell}_2 \cdot d\vec{\ell}_1}{|\vec{r}_2 - \vec{r}_1|} \quad (7.65)$$

We rewrite this as follows:

$$\Phi_{21} = M_{21}\,I_1 \qquad M_{21} = \frac{\mu_o}{4\,\pi} \oint_{\mathcal{C}_2} \oint_{\mathcal{C}_1} \frac{d\vec{\ell}_2 \cdot d\vec{\ell}_1}{|\vec{r}_2 - \vec{r}_1|}$$ Neumann Formula (7.66)

where M_{21} is the mutual inductance between $\mathcal{C}_1$ and $\mathcal{C}_2$ and has units of Henries (volt-second/amp). Two important characteristics:

- ▶ $M_{21} = M_{12}$ because the definition is symmetric.
- ▶ M_{21} is a completely geometric quantity: it does not care about the amount of current flowing, just on the relative positions of the two contours. It is like the capacitance matrix in this respect.

We may now take the time derivative to calculate the emf at $\mathcal{C}_2$ due to a change in I_1:

$$\mathcal{E}_2 = -\frac{d\Phi_{21}}{dt} = -M_{21}\,\frac{dI_1}{dt}$$ (7.67)

If unclear, the sign should be chosen to satisfy Lenz's Law.

Self-Inductance

The above derivation works even when $\mathcal{C}_1$ and $\mathcal{C}_2$ are identical: a current loop induces an emf on itself. In practice, calculating the integral can be difficult because of the singularity at $\vec{r}_1 = \vec{r}_2$, but one can be assured that self-inductance exists and is not infinite. The symbol used is L and the corresponding equations are

$$\Phi = L\,\mathrm{I} \qquad L = \frac{\mu_o}{4\,\pi} \oint_{\mathcal{C}} \oint_{\mathcal{C}} \frac{d\vec{\ell}_2 \cdot d\vec{\ell}_1}{|\vec{r}_2 - \vec{r}_1|} \qquad \mathcal{E} = -L\,\frac{d\mathrm{I}}{dt} \qquad (7.68)$$

In both the cases of mutual inductance and self-inductance, one rarely does the integral directly. Instead, one tries to find the field using Ampere's Law, then calculate the flux, and finally get M or L from Φ/I. This eliminates the need to deal directly with the singularity in the above integral.

Generalization to Volume Currents

We may generalize the above to volume currents by using the usual relation between the vector potential and the volume current density (assuming appropriate boundary conditions):

$$\vec{A}(\vec{r}) = \frac{1}{4\pi} \int_{\mathcal{V}} d\tau' \, \frac{\vec{J}(\vec{r}')}{|\vec{r} - \vec{r}'|} \tag{7.69}$$

$$\implies \quad M_{ij} = \frac{\mu_o}{4\pi} \frac{1}{I_i \, I_j} \int_{\mathcal{V}_i} d\tau \int_{\mathcal{V}_j} d\tau' \, \frac{\vec{J}(\vec{r}) \cdot \vec{J}(\vec{r}')}{|\vec{r} - \vec{r}'|} \tag{7.70}$$

$$L = \frac{\mu_o}{4\pi} \frac{1}{I^2} \int_{\mathcal{V}} d\tau \int_{\mathcal{V}} d\tau' \, \frac{\vec{J}(\vec{r}) \cdot \vec{J}(\vec{r}')}{|\vec{r} - \vec{r}'|} \tag{7.71}$$

where $\mathcal{V}_i$ is the volume of the ith inductor. We notice that the currents do not drop out as cleanly, but, assuming linear behavior of the current flow (the current does not flow differently as the overall magnitude of the current is changed), we expect $\vec{J} \propto I$ and indeed, once the functional dependence of the current density on position has been established, the inductances are purely geometrical quantities as for the case of line currents.

This is not a rigorous proof. We will return to this when we discuss magnetic energy.

Example 7.4: Self and Mutual Inductances of Solenoids

Let's first calculate the self-inductance of a solenoid of radius a. Recall that the field of a solenoid is only nonzero inside it and has value

$$\vec{B} = \mu_o \, n \, I \, \hat{z} \qquad (7.72)$$

where n is the number of turns per unit length, the solenoid axis is along $\hat{z}$, and the current flows along the $\hat{\phi}$ direction. The magnetic flux threading the solenoid and the self-inductance are therefore

$$\Phi = n \, \ell \, \pi \, a^2 \, B = \mu_o \, n^2 \, \ell \, \pi \, a^2 \, I \qquad \Longrightarrow \qquad L = \mu_o \, n^2 \, \ell \, \pi \, a^2 \qquad (7.73)$$

If we have two interpenetrating solenoids with turn densities n_1 and n_2, radii $a_1 < a_2$, and lengths $\ell_1 < \ell_2$, then the flux into solenoid 1 of the field from solenoid 2 and the mutual inductance are

$$\Phi_{12} = n_1 \, \ell_1 \, \pi \, a_1^2 \, B_2 = \mu_o \, n_1 \, n_2 \, \ell_1 \, \pi \, a_1^2 \, I_2 \qquad \Longrightarrow \qquad M = \mu_o \, n_1 \, n_2 \, \ell_1 \, \pi \, a_1^2 \quad (7.74)$$

It is interesting and useful to note that we may also calculate Φ_{21} using M given the symmetry of M. This is very convenient, as calculating the contribution to Φ_{21} from the portion of solenoid 1's field past its ends would be nontrivial. It is also interesting to see that the mutual inductance is not manifestly symmetric under index exchange $1 \leftrightarrow 2$. This reflects the asymmetry of the setup between solenoids 1 and 2.

Lecture 22:

Electrodynamics III:

Magnetic Energy of Currents and Fields
Magnetic Energy in the Presence of Magnetizable Materials
Magnetic Forces

Date Revised: 2024/04/10 17:00
Revised lecture break
Date Given: 2024/04/09

Magnetic Energy and Forces

Magnetic Energy in Terms of Currents

Let's consider the work that has be done to drive current against the emf in an inductive object (*e.g.*, a simple loop or a solenoid). The emf is sometimes called the "back emf" because it is the line integral of a force that tries to drive a current that is intended to counter the changing field due to the current one is varying and so the current one is varying must be driven against the emf.

That is, when a varying current is driven through an inductive object, it has to be driven against a force per unit charge whose line integral (note that we did not say potential!) along the current's path is $\mathcal{E}$. (The force is due to an induced electric field for this case of a stationary loop that is experiencing a $d\Phi/dt$ due to its own current varying.) The force that must be exerted, and the work that must be done, is above and beyond the force needed to overcome the inertia of the charge carriers (*i.e.*, the Newton's Law force $F = m\,a$).

The rate at which this work is being done is given by the same expression we derived before for the work done by the pulling force in the case that the field was fixed but the loop was moving: it is the work done per unit charge by the battery to push the current against the back emf, $-\mathcal{E}$, times the charge flowing past a given point per unit time, I:

$$\frac{d\mathrm{W}}{dt} = Power = -\mathrm{I}\,\mathcal{E} = L\,\mathrm{I}\,\frac{d\mathrm{I}}{dt} \tag{7.75}$$

We can integrate this over time to get the total work done and the *magnetic energy* created:

$$W = \frac{1}{2} L I^2 \qquad (7.76)$$

It is natural to ask why, when we considered the situation with the loop being pulled through a magnetic field, we did not worry about this magnetic energy: we said that the work done by the pulling force was completely dissipated in the Joule heating of the resistor. Or, put another way, why did we not need to include a resistor in the calculation here? When we include the resistor, some of the work done by the pulling force *as the loop was accelerated from rest to $\vec{v}$* goes into this magnetic energy, the $L I^2/2$ energy. Once at fixed velocity, however, the current and thus this energy stay constant. The pulling force continues to do work, however. Since we specifically made the steady-state assumption — that the loop had been moving at fixed $\vec{v}$ for all time and would stay moving for all time — this transient process of creating $L I^2/2$ was not relevant, and our conservation of energy argument was valid; we just neglected noting the path that the energy took through the magnetic energy in the steady state. Now, without the resistor, we are focused entirely on the transient portion of the process, hence the importance of the magnetic energy.

Let's generalize the above result. Consider a system of N inductive elements with inductance matrix M_{ij} ($M_{ii} \equiv L_i$, $M_{ij} = M_{ji}$). We turn on the currents in the order $i = 1, 2, \cdots, N$. We first have to maintain the current I_i against the emf on inductor i felt due to its changing current. Once it has reached its final value, we also have to maintain it as the currents in the inductors $j > i$ are increased from 0 to their final values (note: $j > i$, not $j < i$ as we had in electrostatics). The required power is:

$$\frac{dW}{dt} = \sum_{i=1}^{N} \frac{dW_i}{dt} = \sum_{i=1}^{N} (-I_i \, \mathcal{E}_i) = \sum_{i=1}^{N} \left[I_i M_{ii} \frac{dI_i}{dt} + I_i \sum_{j>i}^{N} M_{ij} \frac{dI_j}{dt} \right]$$

$$\implies \quad W = \sum_{i=1}^{N} \left[\frac{1}{2} M_{ii} I_i^2 + I_i \sum_{j>i}^{N} M_{ij} I_j \right] = \frac{1}{2} \sum_{i,j=1}^{N} M_{ij} I_i I_j \tag{7.77}$$

where we have symmetrized the sum over j by including a factor of $1/2$, and then we combined the cross-terms with the self-terms. If we rewrite all our relations using matrix notation, with $\underline{I}$ being a column vector of currents, $\underline{\Phi}$ being a column vector of fluxes, and $\underline{\underline{M}}$ being the matrix of mutual inductances, we have

$$\boxed{\underline{\Phi} = \underline{\underline{M}} \, \underline{I} \qquad W = \frac{1}{2} \underline{\Phi}^T \underline{I} = \frac{1}{2} \underline{I}^T \underline{\underline{M}} \, \underline{I}} \tag{7.78}$$

Note: we could have calculated the above somewhat differently, considering the work done to maintain the loops $j < i$ at their final current values, plus the work done in loop i itself, while loop i is being ramped to its final value. The result would be the same.

Magnetic Energy in Terms of Magnetic Field

Let's manipulate our circuit equations above to try to get the energy in terms of the magnetic field. First, we can rewrite the circuit expressions using the vector potential:

$$L\,\mathrm{I} = \Phi = \int_{\mathcal{S}(\mathcal{C})} da\,\hat{n} \cdot \vec{B} = \oint_{\mathcal{C}} d\vec{\ell} \cdot \vec{A} \tag{7.79}$$

$$\implies \quad W = \frac{1}{2}\,L\,\mathrm{I}^2 = \frac{\mathrm{I}}{2} \oint_{\mathcal{C}} d\vec{\ell} \cdot \vec{A} = \frac{1}{2} \oint_{\mathcal{C}} d\ell\,\vec{\mathrm{I}} \cdot \vec{A} \tag{7.80}$$

We can obviously generalize this for volume currents to

$$\boxed{W = \frac{1}{2} \int_{\mathcal{V}} d\tau\,\vec{J} \cdot \vec{A}} \tag{7.81}$$

Aside: The above equation now justifies Equation 7.71: if one uses Equation 5.56 to write $\vec{A}$ in terms of $\vec{J}$ and then calculates $L = 2\,W/I^2$, one recovers Equation 7.71. By considering two separate volume current distributions, one can recover Equation 7.70 also.

We can use Ampere's Law to obtain

$$W = \frac{1}{2\,\mu_o} \int_{\mathcal{V}} d\tau \, \vec{A} \cdot \left(\vec{\nabla} \times \vec{B}\right) \tag{7.82}$$

We use the product rule for the divergence of a cross-product,
$\vec{\nabla} \cdot (\vec{a} \times \vec{b}) = \vec{b} \cdot (\vec{\nabla} \times \vec{a}) - \vec{a} \cdot (\vec{\nabla} \times \vec{b})$ to rewrite this as

$$W = \frac{1}{2\,\mu_o} \int_{\mathcal{V}} d\tau \left[\vec{B} \cdot \left(\vec{\nabla} \times \vec{A}\right) - \vec{\nabla} \cdot \left(\vec{A} \times \vec{B}\right)\right] \tag{7.83}$$

$$= \frac{1}{2\,\mu_o} \int_{\mathcal{V}} d\tau \, |\vec{B}|^2 - \frac{1}{2\,\mu_o} \oint_{\mathcal{S}(\mathcal{V})} da \, \hat{n} \cdot \left(\vec{A} \times \vec{B}\right) \tag{7.84}$$

Now, the original volume integral was over only the region containing the current, but the volume integral could be extended to a larger region since there would be no additional contribution. So we do the usual thing and expand the volume to include all of space and take the bounding surface to infinity.

We assume that $\vec{A} \times \vec{B}$ falls off more quickly than $1/r^2$ (true for finite current distributions) so that the surface term goes to zero, or that the particulars of the configuration ensure the integral vanishes even if the current distribution is not finite and we expect a finite energy. Therefore,

$$W = \frac{1}{2\,\mu_o} \int d\tau \, |\vec{B}|^2 \tag{7.85}$$

Thus we see that the magnetic energy is just given by the integral of the square of the field. We now see that the magnetic energy created as the currents are ramped from zero to their final values is *stored in the field*. (We specifically avoided saying earlier where it was stored!)

It is interesting to think about how it is possible to store energy in a magnetic field given that the field can do no work. One has to think of this as the work done to drive against the induced electric field as the field was increased from zero to its final value. As usual, this work is supplied by a battery, not the magnetic field.

On the point about the surface term in the case of configurations for which the fields do not fall of at infinity:

▶ For an infinite solenoid, the surface term only includes the endcaps of the solenoid, since $\vec{B}$ vanishes outside the solenoid. The contributions of the two endcaps vanish because $\vec{A} \times \vec{B}$ points along $\hat{s}$ in cylindrical coordinates, but the endcap's normal is along $\hat{z}$. While the surface term vanishes, the energy is still infinite because the volume integral is over an infinite volume with a constant energy density.

▶ For an infinite wire, even when calculated per unit length, all the terms are logarithmically infinite. This is because the current and the fields do not die off quickly enough at infinity. The calculation fails even if one does the calculation using $\vec{J} \cdot \vec{A}$, and even for a finite diameter wire (if one accounts for the fact that $\vec{J}$ becomes a surface current in the perfect conductor case).

Example 7.5: Magnetic Energy in a Solenoid

A solenoid of radius a with n turns per unit length and current I has a field $B = \mu_o n I$. Therefore, the magnetic energy in such a solenoid of length ℓ is

$$W = \pi a^2 \ell \frac{1}{2 \mu_o} B^2 = \frac{1}{2} \mu_o n^2 I^2 \pi a^2 \ell \qquad (7.86)$$

Note that we can extract from this the self-inductance using $W = L I^2 / 2$, yielding $L = \mu_o n^2 \pi a^2 \ell$ as we obtained by calculating the flux. To put some numbers on this, the LHC CMS experiment (http://home.web.cern.ch/about/experiments/cms) has a solenoid with a field of 4 T with radius $a = 3$ m and length 13 m. The stored energy is therefore about 2.5 gigajoules, an enormous number.

Example 7.6: Magnetic Energy in a Coaxial Cable

This is Griffiths Example 7.13. For a coaxial cable of length ℓ with inner and outer conductor radii a and b, the energy and resulting self-inductance are

$$W = \frac{\mu_o}{4 \pi} I^2 \ell \ln \frac{b}{a} \qquad L = \frac{\mu_o}{2 \pi} \ell \ln \frac{b}{a} \qquad (7.87)$$

An Alternate Logical Path

We followed Griffiths in first developing concepts for inductance and energy using current loops and then generalizing both to volume current distributions in a fairly obvious way. That said, there is a more rigorous way to do all of this by first considering the work done in maintaining current densities in the presence of the electric field generated by changes in those current densities. This then leads to the idea of $\vec{J} \cdot \vec{A}$ being the energy density in the magnetic field. Then one can define inductances by writing the field energy in terms of the total currents that normalize the current distributions. It can then be shown that these inductances relate currents to fluxes and thus rates of change of currents to emfs. This alternative logical path is followed in Jackson §5.16–5.17.

Magnetic Energy of an Assembly of Free Currents in the Presence of Magnetizable Materials

This analysis is done similarly to the electrostatic case and follows Jackson §5.16. Recall that we discussed the distinction between the total energy needed to assemble the final configuration, including the construction of the bound dipoles, and the energy needed to bring the free charges in assuming the bound dipoles already exist and neglecting the potential energy of creating them. In this case, we assume the bound magnetic dipoles are created and maintained by someone else — someone else has built them and raised their currents to their full values for us and also maintains those currents in the presence of back emf generated when the free currents change — and we need only consider the work that has to be done to turn on some free currents in the presence of these bound magnetic dipoles.

This separation is not academic: all naturally occurring magnetic materials rely on the magnetic dipole moments of fundamental particles. Those magnetic dipoles are unchangeable, and thus the energy stored in them is effectively a constant offset that we have no experimental access to. It therefore makes sense to ignore it in calculations of magnetic energy.

Consider the differential of work the battery must do to maintain the *free currents* $\vec{J}_f$ during a change in the magnetic field $\delta\vec{B}$. (We need not specify whether this change is due to a change in $\vec{J}_f$ and/or the location of the magnetizable materials — all that matters is $\delta\vec{B}$.) Equation 7.80 implies that the change in energy would be (holding the geometry fixed)

$$\delta W = -I\,\mathcal{E}\,\delta t = I\,\frac{d\Phi}{dt}\,\delta t = I\,\delta\Phi = I\,\delta\left[\int_{\mathcal{S}(\mathcal{C})} da\,\widehat{n}\cdot\vec{B}\right] = I\,\delta\left[\oint_{\mathcal{C}} d\vec{\ell}\cdot\vec{A}\right] = I\oint_{\mathcal{C}} d\vec{\ell}\cdot\delta\vec{A} \tag{7.88}$$

for which the volume generalization would be

$$\delta W = \int_{\mathcal{V}} d\tau\,\vec{J}_f\cdot\delta\vec{A} = \int_{\mathcal{V}} d\tau\left(\vec{\nabla}\times\vec{H}\right)\cdot\delta\vec{A} \tag{7.89}$$

Apply the same algebra and the same discarding of the surface term as in free space:

$$\delta W = \int_{\mathcal{V}} d\tau\,\vec{H}\cdot\left(\vec{\nabla}\times\delta\vec{A}\right) = \int_{\mathcal{V}} d\tau\,\vec{H}\cdot\delta\vec{B} \tag{7.90}$$

For nonlinear materials, we would need to apply the specific $\vec{B}(\vec{H})$ function go further. For linear materials, we use $\delta\vec{B} = \mu\delta\vec{H}$ to do the integral and obtain the expected analogue to the free-space result:

$$\boxed{W = \frac{1}{2\mu}\int_{\mathcal{V}} d\tau\,|\vec{B}|^2 = \frac{\mu}{2}\int_{\mathcal{V}} d\tau\,|\vec{H}|^2 = \frac{1}{2}\int_{\mathcal{V}} d\tau\,\vec{H}\cdot\vec{B}} \tag{7.91}$$

Magnetic Energy of a Magnetizable Material in an External Field

We can do this derivation along the lines of what we did for polarizable materials, following Jackson §5.16. Let's assume that we have a configuration of currents that generates fields $\vec{B}_1$ and $\vec{H}_1$ in a volume containing a permeable material μ_1. Now, bring in a material of permeability μ_2 such that it occupies a volume $\mathcal{V}_2$ contained in $\mathcal{V}$ while holding the free source currents fixed. The fields (everywhere) change to $\vec{B}_2$ and $\vec{H}_2$.

The energy difference we want to calculate is

$$U_2 - U_1 = \frac{1}{2} \int d\tau \left[\vec{B}_2 \cdot \vec{H}_2 - \vec{B}_1 \cdot \vec{H}_1 \right] \tag{7.92}$$

We can apply similar manipulations as we did for the electrostatic case. First, we rewrite the above as

$$U_2 - U_1 = \frac{1}{2} \int d\tau \left[\vec{B}_2 \cdot \vec{H}_1 - \vec{B}_1 \cdot \vec{H}_2 \right] + \frac{1}{2} \int d\tau \left[\vec{B}_1 + \vec{B}_2 \right] \cdot \left[\vec{H}_2 - \vec{H}_1 \right] \quad (7.93)$$

Since $\vec{\nabla} \cdot \left[\vec{B}_1 + \vec{B}_2 \right] = 0$, it can be derived from a vector potential $\vec{A}$, allowing us to rewrite the second term as

$$\frac{1}{2} \int d\tau \left[\vec{H}_2 - \vec{H}_1 \right] \cdot \left(\vec{\nabla} \times \vec{A} \right) \quad (7.94)$$

We use again the vector identity $\vec{\nabla} \cdot (\vec{a} \times \vec{b}) = \vec{b} \cdot (\vec{\nabla} \times \vec{a}) - \vec{a} \cdot (\vec{\nabla} \times \vec{b})$ to integrate by parts, and we turn the divergence into a surface term that we can discard because $\vec{H}_2 - \vec{H}_1$ should vanish as we go far from the permeable material, yielding for the second term

$$\frac{1}{2} \int d\tau \, \vec{A} \cdot \vec{\nabla} \times \left(\vec{H}_2 - \vec{H}_1 \right) \quad (7.95)$$

The curl in the integrand vanishes because $\vec{H}_2$ and $\vec{H}_1$ are sourced by the same free currents.

We are thus left with the first term from the equation we started with

$$U_2 - U_1 = \frac{1}{2} \int d\tau \left[\vec{B}_2 \cdot \vec{H}_1 - \vec{B}_1 \cdot \vec{H}_2 \right] \qquad (7.96)$$

Applying linearity, $\vec{B} = \mu \vec{H}$, we then obtain

$$U_2 - U_1 = \frac{1}{2} \int d\tau \left(\mu_2 - \mu_1 \right) \vec{H}_2 \cdot \vec{H}_1 \qquad (7.97)$$

Finally, we recognize $\mu_2 - \mu_1 = 0$ except in $\mathcal{V}_2$, so

$$U_2 - U_1 = \frac{1}{2} \int_{\mathcal{V}_2} d\tau \left(\mu_2 - \mu_1 \right) \vec{H}_2 \cdot \vec{H}_1 = \frac{1}{2} \int_{\mathcal{V}_2} d\tau \left(\frac{1}{\mu_1} - \frac{1}{\mu_2} \right) \vec{B}_2 \cdot \vec{B}_1 \qquad (7.98)$$

This is the analogue of Equation 4.86 aside from a sign flip, which mechanically is due to the fact that $\vec{B} = \mu \vec{H}$ (rather than $\vec{H} = \mu \vec{B}$). If we take $\mu_1 = \mu_0$ and $\mu_2 = \mu$, we can use $\vec{M}_2 = (\mu_2/\mu_0 - 1)\vec{H}_2 = (\mu/\mu_0 - 1)\vec{H}_2$ to rewrite this as

$$W = U_2 - U_1 = \frac{1}{2} \int_{\mathcal{V}_2} d\tau \, \vec{M} \cdot \vec{B}_1 \qquad \Longleftrightarrow \qquad w = \frac{1}{2} \vec{M} \cdot \vec{B}_1 \qquad (7.99)$$

where now we replace $\vec{M}_2$ by $\vec{M}$ since $\vec{M}_1 = \vec{0}$ if $\mu_1 = \mu_o$. So $\vec{M}$ is the magnetization density of the volume occupied by μ and $\vec{B}_1$ is the *magnetic field in the absence of the permeable material*. There is a sign flip relative to the electrostatic case (Equation 4.87) that, mechanically, came from the sign flip in Equation 7.98.

How do we understand this sign flip conceptually? Trying to track the sign through the derivation is not illuminating. But it can be understood by comparing to our calculation of the energy of magnetic dipole in an external field, Equation 5.153, where we assumed that the magnetic dipole moment and field were given and held fixed without our having to account for how this was done. In that case, the potential energy of the configuration was $U = -\vec{m} \cdot \vec{B}$. (The factor of $1/2$ here comes from the linear relationship between $\vec{m}$ and $\vec{B}$ and the integration from zero field to $\vec{B}$, which is not important for this discussion.) We see that we have a sign flip relative to that situation. It is sensible, then, to attribute the sign flip to the fact that, in deriving the expression $w = |\vec{B}|^2/2\mu$ that was the starting point for this derivation, we accounted for the work done by the batteries to maintain the free currents as the permeable material was brought in. No such work was required in the previously considered case of a fixed dipole moment $\vec{m}$ and fixed field $\vec{B}$.

Note that, importantly, we do not account for how the magnetization density $\vec{M}$ is maintained. This is to be distinguished from not considering how $\vec{M}$ is created, which we argued was just an unchangeable offset. We may ignore this additional consideration here because, again, the magnetization density is, in naturally occurring systems, due to fundamental magnetic dipoles that require no batteries to maintain their magnetic moments.

When we compare to the electrostatic analogy, Equation 4.87, we recognize a sign flip, too. The rationale is the same: in the electrostatic case, we do not have to do any work to maintain the free charges sourcing the applied field $\vec{E}$ at their nominal positions, while here we do have to do work with a battery to maintain currents at the nominal values and positions due to the back emf from the changing $\vec{M}$.

Magnetic Forces from Magnetic Energy with Fluxes Fixed

To evaluate magnetic forces, we need to consider what happens if we have an infinitesimal generalized displacement of one of our inductors. Because it is more straightforward, let's first consider the fixed fluxes case, which is analogous to holding charges fixed in electrostatics. If $d\Phi/dt = 0$, then there are no emfs and there is no need for a battery to do work to drive currents against those emfs. So we only need to consider $dW_{field}|_\Phi$. We can directly calculate the generalized force from the energy holding the fluxes fixed:

$$F_\xi \bigg|_\Phi = -\left(\frac{\partial W_{field}}{\partial \xi}\right)_\Phi = -\frac{1}{2}\sum_{i,j=1}^N \Phi_i\, \Phi_j \frac{\partial M_{ij}^{-1}}{\partial \xi}\bigg|_\Phi = -\frac{1}{2}\underline{\Phi}^T \left[\frac{\partial}{\partial \xi}\underline{\underline{M}}^{-1}\right]\underline{\Phi}\bigg|_\Phi \quad (7.100)$$

which is the analogue of Equation 4.90.

It's not clear at a microscopic level (*i.e.*, what has to happen to the currents) how one maintains fixed fluxes as inductors are moved around. But, certainly, one is assured that, if one sets up a system of inductors with currents and then disconnects them from their batteries, any movement of the loops must keep the fluxes fixed and change the currents accordingly since there are no batteries to work against the emfs and maintain the currents. This issue will be revisited in homework and is discussed in Griffiths Section 8.3 (4th edition).

Magnetic Forces from Magnetic Energy with Currents Fixed

We follow Jackson §5.16. Let's approach this case like we did in electrostatics, first fixing the fluxes (charges), allowing the currents (voltages) to change, and then returning the currents (voltages) to their original values. The contribution to the change in energy from the fluxes-fixed portion is

$$dW_{field}\Big|_\Phi = \frac{1}{2} \sum_{i,j=1}^{N} \Phi_i \, \Phi_j \, d\left[\underline{\underline{M}}^{-1}\right]_{ij} \tag{7.101}$$

This causes changes in the currents (at fixed flux)

$$dI_i\Big|_\Phi = \sum_{j=1}^{N} d\left[\underline{\underline{M}}^{-1}\right]_{ij} \Phi_j \tag{7.102}$$

If we add back current to return to a fixed-current situation, then changes in fluxes result:

$$d\Phi_k\Big|_I = \sum_{i=1}^{N} M_{ki} \, (-dI_i)_\Phi = - \sum_{i,j=1}^{N} M_{ki} d\left[\underline{\underline{M}}^{-1}\right]_{ij} \Phi_j \tag{7.103}$$

The above infinitesimal flux changes cause emfs. The currents I_i must be maintained by batteries in the presence of these emfs. The work done by the batteries over the infinitesimal time dt needed to make the flux changes is

$$
dW_{field}^{bat}\Big|_{I} = dt \left.\frac{dW_{field}^{bat}}{dt}\right|_{I} = dt \sum_{k=1}^{n}(-I_k\,\mathcal{E}_k) = dt \sum_{k=1}^{n} I_k \left.\frac{d\Phi_k}{dt}\right|_{I} \tag{7.104}
$$

$$
= \sum_{k=1}^{n} I_k\, d\Phi_k\Big|_{I} = -\sum_{i,j,k=1}^{n} I_k M_{ki}\, d\left[\underline{\underline{M}}^{-1}\right]_{ij}\Phi_j \tag{7.105}
$$

$$
= -\sum_{i,j=1}^{n} \Phi_i\,\Phi_j\, d\left[\underline{\underline{M}}^{-1}\right]_{ij} = -2\, dW_{field}\Big|_{\Phi} \tag{7.106}
$$

Note that we did not need to worry about the work done by the battery to make the canceling change in current dI_k because this current change would be multiplied against $dt\,\mathcal{E}_k = -d\Phi_k$, which is already infinitesimal. We need only consider the above term consisting of the nominal currents I_k multiplied against $dt\,\mathcal{E}_k$. We had the same situation in electrostatics, where we did not consider the $dV_k\,dQ_k$ terms, only the $V_k\,dQ_k$ terms.

The total change in the field energy is then obtained by adding the two contributions to the field energy: the field energy change at fixed flux followed by the energy added to the field by the batteries as they return the currents to their initial values:

$$dW_{field}\Big|_{\mathrm{I}} = dW_{field}\Big|_{\Phi} + dW_{field}^{bat}\Big|_{\mathrm{I}} = dW_{field}\Big|_{\Phi} - 2\,dW_{field}\Big|_{\Phi} = -dW_{field}\Big|_{\Phi} \qquad (7.107)$$

We thus find a perfect analogy to the electrostatic case, where we found $dW_{field}|_V = -dW_{field}|_Q$. We may thus use the same guidance: the force cannot depend on whether the situation used is fixed flux or fixed current, and so the forces calculated at fixed flux and fixed current must be the same. Thus, we must conclude

$$F_\xi\Big|_{\mathrm{I}} = \left(\frac{\partial W_{field}}{\partial \xi}\right)_{\mathrm{I}} = -\left(\frac{\partial W_{field}}{\partial \xi}\right)_{\Phi} = F_\xi\Big|_{\Phi} \qquad (7.108)$$

That is, just like in the electrostatic case, when the battery is involved and we consider the energy of the entire system, we see we must take the positive gradient of the field energy at fixed current, rather than considering only the energy of the field and taking the negative gradient of the field energy at fixed current. The reason these two gradients are different, with a sign between them, is because the derivative is calculationally different depending on whether I or Φ is held fixed.

We can see this works mathematically by trying it:

$$\left(\frac{\partial W_{field}}{\partial \xi}\right)_{\mathrm{I}} = \frac{\partial}{\partial \xi}\left[\frac{1}{2}\sum_{i,j=1}^{N} \mathrm{I}_i\,\mathrm{I}_j\,M_{ij}\right]_{\mathrm{I}} = \frac{1}{2}\sum_{i,j=1}^{N} \mathrm{I}_i\,\mathrm{I}_j\,\frac{\partial M_{ij}}{\partial \xi}$$

$$= \frac{1}{2}\,\mathrm{I}^T\left[\frac{\partial}{\partial \xi}\underline{\underline{M}}\right]\mathrm{I} \qquad (7.109)$$

Since $\partial \underline{\underline{M}}^{-1}/\partial \xi = -\underline{\underline{M}}^{-1}[\partial \underline{\underline{M}}/\partial \xi]\underline{\underline{M}}^{-1}$ (one can see this in the same way we proved the analogous relationship for $\underline{\underline{C}}$), this form yields Equation 7.100 for $F_\xi|_\Phi$. Thus,

$$F_\xi\big|_{\mathrm{I}} = \left(\frac{\partial W_{field}}{\partial \xi}\right)_{\mathrm{I}} = -\left(\frac{\partial W_{field}}{\partial \xi}\right)_\Phi = F_\xi\big|_\Phi \qquad (7.110)$$

Lecture 23:

Electrodynamics IV:
Displacement Current
Maxwell's Equations

Conservation Laws:
Conservation of Charge and Energy

Electromagnetic Waves I:
Electromagnetic Waves in Vacuum:

Date Revised: 2024/04/10 17:00
Date Given: 2024/04/11

Maxwell's Equations

The Inconsistency in our Equations

Let's write the full set of equations we have come to:

$$\vec{\nabla} \cdot \vec{E} = \frac{\rho}{\epsilon_o} \qquad\qquad \vec{\nabla} \cdot \vec{B} = 0 \qquad (7.111)$$

$$\vec{\nabla} \times \vec{E} = -\frac{\partial \vec{B}}{\partial t} \qquad\qquad \vec{\nabla} \times \vec{B} = \mu_o \, \vec{J} \qquad (7.112)$$

Now, we know that the divergence of a curl vanishes: it's a vector identity. We should check that it holds! For the electric field, we have

$$\vec{\nabla} \cdot \vec{\nabla} \times \vec{E} = \vec{\nabla} \cdot -\frac{\partial \vec{B}}{\partial t} = -\frac{\partial}{\partial t} \vec{\nabla} \cdot \vec{B} = 0 \qquad (7.113)$$

If we repeat with $\vec{B}$, we obtain

$$\vec{\nabla} \cdot \vec{\nabla} \times \vec{B} = \mu_o \, \vec{\nabla} \cdot \vec{J} = -\mu_o \, \frac{\partial \rho}{\partial t} \neq 0 \quad \text{in general} \qquad (7.114)$$

There is a more physical way to see this by applying Ampere's Law to a circuit containing a parallel-plate capacitor. Construct an Ampere's Law loop around the wire carrying the current. Ampere's Law is satisfied because there is magnetic field in an azimuthal direction around the wire (giving a nonzero line integral of $\vec{B}$) and there is current passing through the disk-like surface whose boundary is the contour.

Now pick another surface that passes between the capacitor plates. This is an equally valid surface; nothing about our proof of Ampere's Law from the Biot-Savart Law assumed a particular choice of surface for the Ampere's Law surface integral. But this surface has no current intersecting it because it passes through the capacitor!

The reason this problem happens and we never noticed it before is because we have a non-steady-state case here: charge piles up on the capacitor plates giving $\partial \rho / \partial t \neq 0$; we had assumed all along during magnetostatics and during our discussion of induction that $\partial \rho / \partial t = 0$.

While charge cannot jump across the capacitor plates so that there can be a current to keep Ampere's Law satisfied, we do recognize that, as charge enters one plate of the capacitor, an equal amount of charge leaves the other plate, ensuring that $dQ/dt = 0$ for the capacitor as a whole. This is suggestive of the idea that perhaps there is some sort of current flowing across the capacitor gap, just not the physical movement of charges we are used to. This new current will be called the *displacement current*.

The Displacement Current

In order to solve the above problem, we need something that will cancel

$$\mu_o \, \vec{\nabla} \cdot \vec{J} = -\mu_o \, \frac{\partial \rho}{\partial t} = -\mu_o \, \frac{\partial}{\partial t} \left(\epsilon_o \, \vec{\nabla} \cdot \vec{E} \right) = -\mu_o \, \vec{\nabla} \cdot \left(\epsilon_o \, \frac{\partial \vec{E}}{\partial t} \right) \qquad (7.115)$$

Let's just add the necessary term to Ampere's Law:

$$\vec{\nabla} \times \vec{B} = \mu_o \, \vec{J} + \mu_o \, \epsilon_o \, \frac{\partial \vec{E}}{\partial t} \qquad (7.116)$$

Physically, what we have done is defined a second current density so that the divergence of the *total* current density $\vec{J} + \epsilon_o \, \partial \vec{E}/\partial t$ vanishes. This vanishing is equivalent to the vanishing of the flux of the total current through any surface, which is what is needed to solve the problem we pointed out: now the surface integral of the enclosed current does not depend on the surface chosen.

Was it ok to do this? Does it violate any of our previous conclusions? The only equation we have modified is the $\vec{\nabla} \times \vec{B}$ equation, so we only need to consider our study of magnetostatics, where we applied this equation. The addition preserves the usual behavior of $\vec{\nabla} \times \vec{B}$ for magnetostatics because $\partial \vec{E}/\partial t = 0$ in magnetostatics. Why? Two things can result in time dependence of $\vec{E}$. The first is time dependence in ρ. But in magnetostatics, we make the steady-state assumption, explicitly requiring no buildup of charge and hence $\partial \rho/\partial t = 0$. The second is time dependence of $\vec{B}$, which can yield time dependence of $\vec{E}$ via Faraday's Law. But magnetostatics assumes $\vec{B}$ is constant in time, so there is no worry there.

The added term is called the *displacement current density*,

$$\boxed{\vec{J_d} \equiv \epsilon_o \, \frac{\partial \vec{E}}{\partial t}} \tag{7.117}$$

This is the "current" we foresaw we needed. It is not a physical current carried by charges, but it represents the fact that, when charge builds up in some point in a circuit because of a gap that prevents physical current from flowing, it causes a changing electric field that then pushes charge away from that point, causing current to flow. One needs a *changing* electric field because otherwise one would quickly reach a steady state in which no new charge would move and thus there would be no source for current. (Maintaining fixed charge on a capacitor does not require current to flow.) Effectively, the displacement current carries the current across physical gaps in the circuit. It is therefore justified, both on the basis of units and on physical intuition, to call it a current. One could even argue that the name is suitable: the "displacement" current causes the displacement of charges on the two sides of a gap across which true current cannot flow. (This argument is in disagreement with Griffiths' statement that the displacement current has nothing to do with current.)

More importantly, we also now see for the first time that a changing electric field sources a magnetic field. Unlike with Faraday's Law, however, there is no negative sign and the induced magnetic field does not act in such a way as to try to cancel the changing electric field.

By construction, $\vec{J}_d$ solves the problem with $\vec{\nabla} \cdot \vec{\nabla} \times \vec{B}$, and we already intuitively expect it will sove the problem with the integral version of Ampere's Law, but let's see that explicitly. The electric field in the capacitor is

$$\vec{E} = \frac{\sigma}{\epsilon_o} \, \widehat{n} = \frac{1}{\epsilon_o} \, \frac{Q}{A} \, \widehat{n} \qquad (7.118)$$

where $\widehat{n}$ is the normal from the positive plate to the negative plate. Therefore, the displacement current is

$$\vec{J}_d = \epsilon_o \frac{\partial \vec{E}}{\partial t} = \frac{1}{A} \frac{dQ}{dt} \, \widehat{n} = \frac{\mathrm{I}}{A} \, \widehat{n} \qquad (7.119)$$

The integral form of Ampere's Law with the displacement current is therefore

$$\oint_{\mathcal{C}} d\vec{\ell} \cdot \vec{B} = \mu_o \, \mathrm{I}_{encl} + \mu_o \int_{\mathcal{S}(\mathcal{C})} da \, \widehat{n} \cdot \vec{J}_d \qquad (7.120)$$

If we choose the first surface we discussed earlier, the flat surface in the plane of the contour $\mathcal{C}$, we get the first term but the second term vanishes, yielding $\mu_o \, \mathrm{I}$. If we choose the second surface, the one between the capacitor plates, the first term vanishes but the second term gives $\mu_o \, \mathrm{I}$. Thus, the inconsistency seen earlier has been eliminated.

Example 7.7: Displacement Current for Coaxial Conductors (Griffiths 7.36)

This is a continuation of the example from earlier. We want to calculate the displacement current density and the total displacement current and to compare quantitatively I and I_d.

The displacement current density is $\vec{J}_d = \epsilon_o \, \partial \vec{E}/\partial t$:

$$\vec{J}_d(s < a) = \epsilon_o \, \frac{\partial}{\partial t} \, \widehat{z} \, \frac{\mu_o}{2 \, \pi} \, \omega \, I_0 \, \sin \omega t \, \ln \frac{a}{s} = \widehat{z} \, \mu_o \, \epsilon_o \, \frac{\omega^2 \, I_0}{2 \, \pi} \, \cos \omega t \, \ln \frac{a}{s} \qquad (7.121)$$

Let's integrate over the (s, ϕ) plane to get the total displacement current:

$$\begin{aligned} I_d &= \int_{\mathcal{S}(\mathcal{C})} \widehat{n} \cdot \vec{J}_D = \int_0^{2 \, \pi} d\phi \, \mu_o \, \epsilon_o \, \frac{\omega^2 \, I_0}{2 \, \pi} \, \cos \omega t \int_0^a s \, ds \, \ln \frac{a}{s} \\ &= \mu_o \, \epsilon_o \, \omega^2 \, I_0 \cos \omega t \left(a^2 \left[\frac{x^2}{2} \left(\ln x - \frac{1}{2} \right) \right] \Big|_{x=1}^0 \right) \\ &= \mu_o \, \epsilon_o \, \omega^2 \, I_0 \frac{a^2}{4} \, \cos \omega t \end{aligned} \qquad (7.122)$$

where the indeterminate form $x^2 \ln x$ as $x \to 0$ must be evaluated by L'Hopital's rule (write it as $(\ln x)/(1/x^2)$) to be seen to vanish.

We did not include the displacement current in the calculation of the magnetic field in the system. Is that a problem?

Well, the problem is, in principle, even worse: we ought to include the displacement current in the calculation of $\vec{B}$, but then our calculation of $\vec{E}$ via Faraday's Law needs to also be corrected for the $\vec{B}$ due to the displacement current, yielding a correction to $\vec{E}$, which itself will yield a correction to $\vec{J}_d$, and so on. The proper way to handle this is to develop the formalism for electromagnetic waves, where we self-consistently solve all of Maxwell's Equations.

For now, it is instructive to look at the relative size of $\vec{J}_d$ and $\vec{J}$ so we can understand why these corrections are small and thus why our previous results, while not precisely correct, are an excellent approximation.

The ratio of the amplitudes of the displacement current and the true current, up to factors of order unity, is

$$\frac{I_d(t)}{I(t)} = \frac{\mu_o\,\epsilon_o\,\omega^2\,a^2\,I_0\,\cos\omega t}{I_0\,\cos\omega t} = \frac{\omega^2\,a^2}{c^2} = \frac{(a/c)^2}{1/\omega^2} \tag{7.123}$$

The numerator of the final expression is the square of the light travel time over the length scale of the problem, a. The denominator is, up to a factor $(2\,\pi)^2$, the square of the oscillation period. Thus, this quantity is a measure of how quasistatic the system is. We have mentioned before that, if $a/c \ll 1/\omega$ is not satisfied, then our quasistatic approximation is invalid. This calculation corroborates that point: if the oscillation period becomes comparable to the light travel time so the system is no longer quasistatic, then the displacement current will approach the real current in magnitude and our prior calculation of $\vec{B}$ ignoring the displacement current will be a bad approximation.

The ratio of the displacement current to the true current scales as ω^2, so one must go to high frequency to notice it. Quantitatively, if we ask how high in frequency one must go to obtain $I_d/I = 0.01$ if we take $a = 2$ mm as the dimension of the coaxial conductor, we obtain

$$\nu = \frac{\omega}{2\,\pi} = \frac{1}{2\,\pi}\,\frac{c}{a}\,\sqrt{\frac{I_d}{I}} = \frac{1}{2\,\pi}\,\frac{3\times 10^{11}\ \text{mm/s}}{2\ \text{mm}}\,\sqrt{0.01} \approx 2\ \text{GHz} \tag{7.124}$$

GHz oscillators were not available in Faraday's time, so the fact that he did not observe the effects of the displacement current is not surprising.

Maxwell's Equations in Vacuum

Putting it all together, we obtain *Maxwell's Equations*:

$$\vec{\nabla} \cdot \vec{E} = \frac{\rho}{\epsilon_o} \quad \vec{\nabla} \cdot \vec{B} = 0 \quad \vec{\nabla} \times \vec{E} = -\frac{\partial \vec{B}}{\partial t} \quad \vec{\nabla} \times \vec{B} = \mu_o \vec{J} + \epsilon_o \mu_o \frac{\partial \vec{E}}{\partial t}$$

(7.125)

These, combined with the force law and continuity:

$$\vec{F} = q\left(\vec{E} + \vec{v} \times \vec{B}\right) \qquad \vec{\nabla} \cdot \vec{J} = -\frac{\partial \rho}{\partial t}$$

(7.126)

summarize classical electrodynamics in vacuum. (The above explains why the $\vec{v} \times \vec{B}$ term is not needed explicitly in the differential version of Faraday's Law: it is a consequence of the force law, not of Faraday's Law.) We may rewrite the first set of equations in a way that emphasizes better the source terms:

$$\vec{\nabla} \cdot \vec{E} = \frac{\rho}{\epsilon_o} \quad \vec{\nabla} \cdot \vec{B} = 0 \quad \vec{\nabla} \times \vec{E} + \frac{\partial \vec{B}}{\partial t} = 0 \quad \vec{\nabla} \times \vec{B} - \epsilon_o \mu_o \frac{\partial \vec{E}}{\partial t} = \mu_o \vec{J}$$

(7.127)

Maxwell's Equations in Matter

Just as we found it convenient to rewrite the individual equations of electrostatics and
magnetostatics using only the free charges and currents, it makes sense to do the
same for Maxwell's Equations. The new twists we must take into account are the
time-dependence of $\vec{P}$ and $\vec{M}$. (We have already considered time dependence of ρ, $\vec{J}$,
$\vec{E}$, and $\vec{B}$ for quasistatic situations, corresponding to all length scales in the system
small compared to c/ν.)

How to treat $\vec{P}$ is motivated by the expression

$$\rho_b = -\vec{\nabla} \cdot \vec{P} \tag{7.128}$$

If $\vec{P}$ is time-varying, we expect there to be a current $\vec{J}_p$ associated with the resulting
changes in ρ_b. In fact, the above expression suggests a good definition of $\vec{J}_p$:

$$\vec{J}_p = \frac{\partial \vec{P}}{\partial t} \qquad \Longleftrightarrow \qquad \vec{\nabla} \cdot \vec{J}_p = -\vec{\nabla} \cdot \frac{\partial \vec{P}}{\partial t} = -\frac{\partial}{\partial t} \vec{\nabla} \cdot \vec{P} = -\frac{\partial \rho_b}{\partial t} \tag{7.129}$$

That is, the definition on the left naturally gives the continuity relation between $\vec{J}_p$
and ρ_b one would like.

Intuitively, think of $\vec{J}_p$ as follows. Suppose one has a cylinder of polarizable material of length dz and cross-sectional area da and with polarization vector $\vec{P} = P\,\hat{z}$. The definition $\rho_b = -\vec{\nabla} \cdot \vec{P}$ implies that there is a bound surface charge $Q = \sigma\,da = \hat{n} \cdot \vec{P}\,da = \pm P\,da$ at each end. If, for example, we allow $\vec{P}$ to vary sinusoidally, $\vec{P} = \vec{P}_0 \sin \omega t$, which corresponds to the surface charge obeying $Q(t) = P_0\,da \sin \omega t$, then the current is

$$I_p = \vec{J}_p \cdot \hat{n}\,da = P_0\,da\,\omega \cos \omega t = \frac{dQ}{dt} \qquad (7.130)$$

as would be necessary to transfer charge back and forth between the two ends of the cylinder to yield the corresponding time-dependent surface charge. This current is, literally, the motion of the charges that make up the dipoles as they flip back and forth sinusodally.

Do we have to worry about time dependence of $\vec{M}$? Recall that $\vec{M}$ yields a bound current density

$$\vec{J}_b = \vec{\nabla} \times \vec{M} \qquad (7.131)$$

Time dependence of $\vec{M}$ yields time dependence of $\vec{J}_b$, which produces time dependence of $\vec{B}$ and $\vec{H}$. These time dependences are now fully accounted for by Maxwell's Equations.

Let's use all this to rewrite Maxwell's Equations in terms of free charges and currents. The charge and current densities have the following parts:

$$\rho = \rho_f + \rho_b = \rho_f - \vec{\nabla} \cdot \vec{P} \qquad \vec{J} = \vec{J}_f + \vec{J}_b + \vec{J}_p = \vec{J}_f + \vec{\nabla} \times \vec{M} + \frac{\partial \vec{P}}{\partial t} \qquad (7.132)$$

Using Gauss's law, $\epsilon_o \vec{\nabla} \cdot \vec{E} = \rho_f - \vec{\nabla} \cdot \vec{P}$, and the definition of the displacement field, $\vec{D} = \epsilon_o \vec{E} + \vec{P}$, we obtain

$$\vec{\nabla} \cdot \vec{D} = \rho_f \qquad (7.133)$$

Ampere's Law with the displacement current term is

$$\vec{\nabla} \times \vec{B} = \mu_o \left(\vec{J}_f + \vec{\nabla} \times \vec{M} + \frac{\partial \vec{P}}{\partial t} \right) + \epsilon_o \, \mu_o \, \frac{\partial \vec{E}}{\partial t} \qquad (7.134)$$

We use $\vec{B} = \mu_o \left(\vec{H} + \vec{M} \right)$ as well as $\vec{D} = \epsilon_o \vec{E} + \vec{P}$ to obtain

$$\vec{\nabla} \times \vec{H} = \vec{J}_f + \frac{\partial \vec{D}}{\partial t} \qquad (7.135)$$

Now it is clear why the last term is called the *displacement current* — it is the apparent current due to the time variation of the displacement vector $\vec{D}$!

Faraday's Law and $\vec{\nabla} \cdot \vec{B} = 0$ are not affected since they do not depend on the free and bound currents. Thus, Maxwell's Equations in matter are (again, putting all the fields on the left sides and the sources on the right):

$$\boxed{\vec{\nabla} \cdot \vec{D} = \rho_f} \qquad\qquad \boxed{\vec{\nabla} \cdot \vec{B} = 0} \qquad (7.136)$$

$$\boxed{\vec{\nabla} \times \vec{E} + \frac{\partial \vec{B}}{\partial t} = 0} \qquad\qquad \boxed{\vec{\nabla} \times \vec{H} - \frac{\partial \vec{D}}{\partial t} = \vec{J}_f} \qquad (7.137)$$

Note: the vacuum Maxwell's Equations (Equations 7.125 and 7.127) remain valid, but the above are more practically useful. Similarly, Equations 7.126, the Lorentz Force Law and the continuity equation, remain valid but it is more practically useful to have a version of the continuity equation involving only free currents (note that $\vec{\nabla} \cdot \vec{J}_b = \vec{\nabla} \cdot \vec{\nabla} \times \vec{M} = 0$ by mathematical identity):

$$\boxed{\vec{F} = q\left(\vec{E} + \vec{v} \times \vec{B}\right)} \qquad\qquad \boxed{\vec{\nabla} \cdot \vec{J}_f = -\frac{\partial \rho_f}{\partial t}} \qquad (7.138)$$

These equations must be supplemented by specific *constitutive relations* between $\vec{E}$ and $\vec{D}$ and between $\vec{B}$ and $\vec{H}$ to completely specify the behavior (and, of course, boundary conditions must be provided). For linear media, these relations are:

$$\boxed{\textbf{linear media:} \qquad \vec{P} = \chi_e \epsilon_o \vec{E} \qquad \vec{M} = \chi_m \vec{H} \qquad \vec{D} = \epsilon \vec{E} \qquad \vec{B} = \mu \vec{H}} \qquad (7.139)$$

Boundary Conditions for Maxwell's Equations

For both Maxwell's Equations in vacuum and in matter, we need to review how the fields change between regions that may be separated by surface charge and current densities and which may have different polarization and magnetization.

We recall from our prior calculations of this type that the discontinuity in the normal component of a field is determined by its divergence and the discontinuity in the tangential component by its curl. We also recall charge and current densities can become δ-function singular on a boundary but fields cannot. Thus:

▶ An integral of charge density over a volume containing a boundary reduces, as the height of the volume normal to the boundary is shrunk to zero, to the surface charge density integrated over the intersection of the volume with the boundary. The volume component of the charge density yields zero contribution.

▶ An integral of a current density through an area reduces, as the width of the area normal to the boundary shrinks to zero, to the surface current density passing through the area. The area component of the current density yields zero contribution.

Fields themselves never have singularities like this, so any integral of a field vanishes as the volume or area is shrunk to zero. Hence, the addition of the displacement current does not modify the boundary conditions we have calculated in the past! For Maxwell's Equations in matter, $\vec{J}_p$ has been introduced but it also cannot have δ-function singularity because it is based on $\vec{P}$, which we have defined to be a volume density of dipoles, and so its contribution to $\partial\vec{D}/\partial t$ is consistent with the above.

Individual electric or magnetic dipoles, or a line or sheet of them, which could in principle yield δ-function contributions to $\vec{P}$ or $\vec{M}$ and thus to $\vec{D}$ or $\vec{H}$, cannot be handled using our macroscopic picture of polarizable and magnetizable materials because the fields are δ-function like on the dipoles only (*e.g.*, Griffiths Problems 3.48 and 5.61) and our macroscopic picture assumes that it is valid to average over infinitesimally small volumes. They affect how the field changes at a boundary over an infinitesimally small area, and so we should not even consider their effect. If we want to consider their impact on the fields in such regions, they would need to be treated as free charges and currents and would again not affect the above statements about fields.

Section 8
Conservation Laws

Motivation and Analogy: Conservation of Charge

Back when we first discussed the Lorentz Force, we discussed conservation of charge and the continuity equation:

$$\vec{\nabla} \cdot \vec{J}(\vec{r}) = -\frac{\partial \rho(\vec{r})}{\partial t} \tag{8.1}$$

This is an interesting equation because it enforces *local* conservation of charge: not only is there no creation or destruction of charge over the whole universe, there is also no creation or destruction of charge at a given point. Charge cannot jump from one place to another without a current flowing to move that charge.

In electrodynamics, we want to ask the same question for energy and momentum because we want to understand whether the fields we have constructed have true physical meaning or are just mathematical constructs. Determining whether they carry energy and momentum is one way to answer that question, and such a consideration leads to the question of conservation of these quantities.

We will do all this in vacuum. It of course applies to polarizable and magnetizable materials, too, since our study of them is just a rewriting of our vacuum equations in a more convenient form.

Poynting's Theorem: Conservation of Energy

We have shown that the work required to set up distributions of charge or current is

$$W_e = \frac{\epsilon_o}{2} \int d\tau \, |\vec{E}|^2 \qquad W_m = \frac{1}{2\mu_o} \int d\tau \, |\vec{B}|^2 \qquad (8.2)$$

Recall that this is the work needed to move new charge in from infinity due to the repulsion from the charge already there or the work that needs to be done to raise a current from zero to its final value against a back emf (induced electric field). It is thus natural to expect that the total energy density in the electric and magnetic fields is

$$u_{field} = \frac{1}{2} \left(\epsilon_o |\vec{E}|^2 + \frac{1}{\mu_o} |\vec{B}|^2 \right) \qquad (8.3)$$

We will show this is valid by considering the exchange of energy between the fields and charges/currents. We use the term *electromagnetic field* to reflect the fact that the fields influence each other and their energies are on the same footing.

Given a single particle of charge q acted on by the electromagnetic field, the work done on it as it moves by $d\vec{\ell}$ is

$$dW = \vec{F} \cdot d\vec{\ell} = q \left(\vec{E} + \vec{v} \times \vec{B} \right) dt = q\, \vec{E} \cdot \vec{v}\, dt \qquad (8.4)$$

Now, if we consider a continuous distribution of charge and current, we may replace $q = \rho\, d\tau$ and $\rho\, \vec{v} = \vec{J}$, giving that the power is (as we saw from Ohm's Law)

$$\frac{dW}{dt} = \int d\tau\, \vec{E} \cdot \vec{J} \qquad (8.5)$$

Let's manipulate the integrand using Ampere's Law:

$$\vec{E} \cdot \vec{J} = \frac{1}{\mu_o}\, \vec{E} \cdot \left(\vec{\nabla} \times \vec{B} \right) - \epsilon_o\, \vec{E} \cdot \frac{\partial \vec{E}}{\partial t} \qquad (8.6)$$

One subtlety here: we started off talking about $\vec{J}$ being acted upon by an electromagnetic field, and now it seems like we are treating $\vec{J}$ as the source of that field. It is not the sole source of the field because, now with the displacement current term combined with Faraday's Law, there can be electric and magnetic fields that are sourced by each other's time variation rather than by physical currents. The above substitution is nevertheless valid because the second term subtracts off the displacement current term that is due to changing fields rather than physical current: one should not be able to do work on the displacement current!

Another subtlety is the issue of whether $\vec{J}$ can create fields that do work on itself. This is entirely possible, as we saw in the example of the time-varying current in the coaxial conductor: a time-varying current generated a time-varying magnetic field that generated a time-varying electric field aligned with the original current. If there were no battery driving the current, then the work being done by the field on the current should reduce the energy in the current in exactly the way that would be needed to conserve energy. Of course, if a battery is involved, then it can supply energy and we do not expect the energy of the currents and fields alone to be conserved.

Returning to our expression for $\vec{E} \cdot \vec{J}$, we can use the product rule
$\vec{\nabla} \cdot (\vec{a} \times \vec{b}) = \vec{a} \cdot (\vec{\nabla} \times \vec{b}) - \vec{b} \cdot (\vec{\nabla} \times \vec{a})$ to rewrite it and then use Faraday's Law:

$$\vec{E} \cdot \vec{J} = \frac{1}{\mu_o} \vec{B} \cdot \left(\vec{\nabla} \times \vec{E} \right) - \frac{1}{\mu_o} \vec{\nabla} \cdot \left(\vec{E} \times \vec{B} \right) - \epsilon_o \vec{E} \cdot \frac{\partial \vec{E}}{\partial t} \tag{8.7}$$

$$= -\frac{1}{\mu_o} \vec{B} \cdot \frac{\partial \vec{B}}{\partial t} - \epsilon_o \vec{E} \cdot \frac{\partial \vec{E}}{\partial t} - \frac{1}{\mu_o} \vec{\nabla} \cdot \left(\vec{E} \times \vec{B} \right) \tag{8.8}$$

$$= -\frac{\partial}{\partial t} \frac{1}{2} \left(\epsilon_o |\vec{E}|^2 + \frac{1}{\mu_o} |\vec{B}|^2 \right) - \frac{1}{\mu_o} \vec{\nabla} \cdot \left(\vec{E} \times \vec{B} \right) \tag{8.9}$$

Incorporating the above and applying the divergence theorem to the last term, we thus obtain *Poynting's Theorem*:

$$\frac{dW}{dt} = -\left[\frac{d}{dt} \int_{\mathcal{V}} d\tau \frac{1}{2} \left(\epsilon_o |\vec{E}|^2 + \frac{1}{\mu_o} |\vec{B}|^2 \right) + \oint_{\mathcal{S}(\mathcal{V})} da \, \hat{n} \cdot \vec{S} \right] \tag{8.10}$$

with the *Poynting vector* defined to be $\quad \boxed{\vec{S} = \frac{1}{\mu_o} \left(\vec{E} \times \vec{B} \right)} \tag{8.11}$

Poynting's Theorem says that *the work per unit time done on the charges and currents in a volume $\mathcal{V}$ by electromagnetic forces is equal to negative of the sum of the change per unit time of the energy in the fields and the energy flowing outward through the surface of $\mathcal{V}$.* $\vec{S}$ has units of energy per unit time per unit area and is considered the *energy flux density*.

Another useful form is given by putting the field energy density term on the left side:

$$\frac{dW}{dt} + \frac{d}{dt}\int_{\mathcal{V}} d\tau \, \frac{1}{2}\left(\epsilon_o|\vec{E}|^2 + \frac{1}{\mu_o}|\vec{B}|^2\right) = -\frac{1}{\mu_o}\oint_{\mathcal{S}(\mathcal{V})} da\,\hat{n}\cdot\vec{S} \qquad (8.12)$$

$$\boxed{\frac{d}{dt}\left(E_{mech} + E_{field}\right) = -\oint_{\mathcal{S}(\mathcal{V})} da\,\hat{n}\cdot\vec{S}} \qquad (8.13)$$

The rate of change of the total energy in the volume is given by the flux of the Poynting vector through the boundary of the volume — this much more explicitly puts mechanical and field energy on the same footing and shows that both can be transported by the Poynting flux.

Note that this allays our fears about a current doing work on itself: while it may do so, energy remains conserved as long as one takes into account the field energy.

We may write both versions in a local form by recognizing that the volume being integrated over is arbitrary. If we define $u_{mech}(\vec{r})$ to be the density of mechanical energy ($W = \int_{\mathcal{V}} d\tau\, u_{mech}$) as a function of position and

$$u_{field} = \frac{1}{2} \left(\epsilon_o |\vec{E}(\vec{r})|^2 + \frac{1}{\mu_o} |\vec{B}(\vec{r})|^2 \right) \tag{8.14}$$

to be the energy density of the electromagnetic field, then our two versions of Poynting's theorem yield the local relations (after converting the surface integral of $\vec{S}$ back to a volume integral using the divergence theorem):

$$\boxed{\frac{\partial u_{mech}(\vec{r})}{\partial t} = -\frac{\partial u_{field}(\vec{r})}{\partial t} - \vec{\nabla} \cdot \vec{S}(\vec{r}) \quad \Longleftrightarrow \quad \frac{\partial}{\partial t} \left[u_{mech}(\vec{r}) + u_{field}(\vec{r}) \right] = -\vec{\nabla} \cdot \vec{S}(\vec{r})} \tag{8.15}$$

This is the kind of local conservation theorem we wanted, relating *the rate of change of a density* (here the energy density) *to the divergence of a current density* (here the Poynting vector).

When there is no change in mechanical energy — *e.g.*, in empty space — then we can specialize the above to obtain the *continuity equation for the energy of the electromagnetic field*:

$$\boxed{u_{mech} = 0: \quad \frac{dE_{field}}{dt} = -\oint_{\mathcal{S}(\mathcal{V})} da\, \hat{n} \cdot \vec{S} \quad \Longleftrightarrow \quad \vec{\nabla} \cdot \vec{S}(\vec{r}) = -\frac{\partial u_{field}(\vec{r})}{\partial t}} \tag{8.16}$$

Example 8.1: Power Transported Down a Coaxial Cable (Griffiths Problem 8.1)

Consider a coaxial cable with central conductor of diameter a and outer conductor of radius b and zero thickness. A *static* current flows along $+\hat{z}$ on the central conductor and back along $-\hat{z}$ on the outer shell. We used a similar geometry in a previous example, but with a time-varying current in that case. The inner conductor is held at voltage V and the outer conductor at $V = 0$ (ground) at one end of the cylinder, and there is a resistive sheet of *sheet* conductivity $\sigma_\square$ (definition to be provided) capping the other end.

Because the inner conductor is assumed to have infinite conductivity, there can be no electric field inside and thus all the current must flow on the surface (consequence of Ohm's Law: $\vec{J} = 0$ because $\vec{E} = 0$). The calculation of the magnetic field is thus the same as the prior example in the same geometry with time-varying current. The magnetic field between the conductors is

$$\vec{B}(s) = \frac{\mu_o \, \mathrm{I}}{2\,\pi\,s} \, \hat{\phi} \tag{8.17}$$

where I is the current due to V (value to be determined). In the prior example, we did not explicitly have a voltage on the inner conductor (effectively, the conductivity of the sheet at the end was infinite).

Here, since we have such a voltage, there is a line charge density on the inner conductor and a radial electric field. You are no doubt familiar with the Gauss' Law calculation of this configuration, which yields

$$\vec{E}(s) = \frac{\lambda}{2\pi\epsilon_o s}\,\hat{s} \tag{8.18}$$

Let's find λ by matching to the applied voltage. The potential and field are

$$V(s) \propto \ln s \qquad V = V(a) - V(b) = \frac{\lambda}{2\pi\epsilon_o}\ln\frac{b}{a} \implies \vec{E}(s) = \frac{V}{s\ln\frac{b}{a}}\,\hat{s} \tag{8.19}$$

The Poynting vector is

$$\vec{S} = \frac{1}{\mu_o}\vec{E}\times\vec{B} = \frac{IV}{2\pi s^2\ln\frac{b}{a}}\,\hat{z} \tag{8.20}$$

The energy and energy current are between the conductors, not in them! The power flowing down the cable is found by integrating the Poynting vector over the cross-sectional area where the fields are:

$$P = \int_{\mathcal{S}} da\,\hat{n}\cdot\vec{S}(s) = \int_a^b s\,ds\int_0^{2\pi} d\phi\,\frac{IV}{2\pi s^2\ln\frac{b}{a}} = \frac{IV}{\ln\frac{b}{a}}\int_a^b\frac{ds}{s} = IV \tag{8.21}$$

Let's calculate the power dissipated in the resistive sheet at the end. The sheet does not disturb the potential because the sheet continues to satisfy Laplace's equation with the same boundary conditions in s: $V(a) = V$, $V(b) = 0$. Therefore, our electric field above is valid in the conducting sheet, and the surface current density and total current are

$$\vec{K}(s) = \sigma_\square \, \vec{E}(s) = \frac{\sigma_\square \, V}{s \, \ln \frac{b}{a}} \, \hat{s} \tag{8.22}$$

where $\sigma_\square$ is the *conductivity per square*, which can be thought of as $\sigma_\square = \lim_{t \to 0} \sigma/t$ where σ is the usual conductivity and t is the thickness of the sheet. The total current is

$$\mathrm{I} = \int_0^{2\pi} s \, d\phi \, K(s) = 2\pi s \, K(s) = \frac{2\pi \sigma_\square \, V}{\ln \frac{b}{a}} \tag{8.23}$$

We do not need it, but it is interesting to note that the resistance is

$$R = \frac{V}{I} = \frac{\ln \frac{b}{a}}{2\,\pi\,\sigma_\square} \tag{8.24}$$

The power dissipated in the resistor is

$$P = \int_S da\,\vec{K}(s) \cdot \vec{E}(s) = \int_a^b s\,ds \int_0^{2\pi} d\phi\,\sigma_\square \left(\frac{V}{s \ln \frac{b}{a}}\right)^2 \tag{8.25}$$

$$= 2\,\pi\,\sigma_\square \left(\frac{V}{\ln \frac{b}{a}}\right)^2 \int_a^b \frac{ds}{s} = \frac{2\,\pi\,\sigma_\square V^2}{\ln \frac{b}{a}} = I\,V = I^2 R = \frac{V^2}{R} \tag{8.26}$$

as expected since this is the power coming down the central conductor and it cannot go beyond the resistive sheet since the fields go to zero out there (no current or charge density beyond the sheet).

Lecture 39:

Conservation Laws (Optional Material):
Stress Tensor/Linear Momentum
Torque Tensor/Angular Momentum

Relativistic Conservation Laws (Optional Material):
Maxwell Energy-Momentum Tensor and Conservation of
Energy-Momentum
Relativistic Angular Momentum Tensor

Applications of Radiation: Diffraction (Optional Material)

Date Revised: 2023/06/07 07:00
Date Given: Not given in 2024

These three sections,
consisting of material colored green and orange,
is optional!

The Maxwell Stress Tensor: Conservation of Linear Momentum (Skip)

We showed in the previous section that fields carry energy and that one must account for that energy in order for conservation of energy to hold. The natural next question to ask is whether the electromagnetic fields carry momentum. The matter and fields are related by the fields exerting forces on the matter, so let's use these forces to connect the momentum of the matter and fields. The Lorentz Force Law is

$$\frac{d\vec{P}_{mech}}{dt} = \vec{F} = q\left(\vec{E} + \vec{v} \times \vec{B}\right) \tag{8.27}$$

Integrating this over a charge and current density gives

$$\frac{d\vec{P}_{mech}}{dt} = \int_{\mathcal{V}} d\tau \left(\rho\,\vec{E} + \vec{J} \times \vec{B}\right) \tag{8.28}$$

Using Maxwell's Equations, we can write this purely in terms of the fields:

$$\frac{d\vec{P}_{mech}}{dt} = \int_{\mathcal{V}} d\tau \left(\epsilon_o \left[\vec{\nabla} \cdot \vec{E}\right]\vec{E} + \left[\frac{1}{\mu_o}\,\vec{\nabla} \times \vec{B} - \epsilon_o\,\frac{\partial \vec{E}}{\partial t}\right] \times \vec{B}\right) \tag{8.29}$$

We recall the same subtleties as for energy: ρ and $\vec{J}$ now being taken as source of fields, and the last term subtracts off the displacement current since the magnetic field exerts no force on it.

After a remarkable amount of manipulation that we will not reproduce here—see Griffiths §8.2.2 or Jackson §6.8—one arrives at

$$\frac{d\vec{P}_{mech}}{dt} = \int_{\mathcal{V}} d\tau \left[\vec{\nabla} \cdot \underline{\underline{T}} - \epsilon_o\, \mu_o\, \frac{\partial \vec{S}}{\partial t} \right] \tag{8.30}$$

where $\underline{\underline{T}}$ is the *Maxwell Stress Tensor*

$$\underline{\underline{T}}(\vec{r}) = \sum_{i,j=1}^{3} T_{ij}(\vec{r})\, \hat{r}_i\, \hat{r}_j \qquad T_{ij}(\vec{r}) = \epsilon_o \left[E_i\, E_j - \frac{1}{2}\, \delta_{ij}\, E^2 \right] + \frac{1}{\mu_o} \left[B_i\, B_j - \frac{1}{2}\, \delta_{ij}\, B^2 \right] \tag{8.31}$$

(we do not show the fields' dependence on position for brevity) and where the vector dot products and divergence of $\underline{\underline{T}}$ are given by

$$\vec{a} \cdot \underline{\underline{T}} = \sum_{i=1}^{3} a_i\, T_{ij}\, \hat{r}_j \qquad \underline{\underline{T}} \cdot \vec{a} = \sum_{j=1}^{3} \hat{r}_i\, T_{ij}\, a_j \qquad \vec{\nabla} \cdot \underline{\underline{T}} = \sum_{i=1}^{3} \hat{r}_j\, \frac{\partial}{\partial r_i}\, T_{ij} \tag{8.32}$$

Note that T_{ij} is symmetric in its indices. We are not terribly concerned in this course with the transformation properties of scalars, vectors, and tensors under coordinate system rotations, so we will not comment further on what a tensor is. Recall we encountered the quadrupole moment tensor earlier.

Using the divergence theorem on the first term and moving the time derivative in the second term outside the integral, we obtain

$$\boxed{\frac{d\vec{P}_{mech}}{dt} = \oint_{\mathcal{S}(\mathcal{V})} da\,\hat{n}(\vec{r}) \cdot \underline{\underline{T}}(\vec{r}) - \epsilon_o\,\mu_o\,\frac{d}{dt}\int_{\mathcal{V}} d\tau\,\vec{S}(\vec{r})} \qquad (8.33)$$

This equation states that *the rate of change of the mechanical momentum in a volume $\mathcal{V}$ is equal to the integral over the surface of the volume of the stress tensor's flux through that surface minus the rate of change of the volume integral of the Poynting vector.*

Let us consider a situation in which the second term vanishes and we are left with the flux of $\underline{\underline{T}}$ over the surface. This justifies the naming of $\underline{\underline{T}}$: it gives the force per unit area due to the electromagnetic fields, or the *stress*. T_{ij} is the force per unit area acting in the ith direction on an area element who normal is in the jth direction. The diagonal elements are *pressures* and the off-diagonal forces are *shears*. More generally, the force per unit area in the $\hat{n}_1$ direction on an area element whose normal is in the $\hat{n}_2$ direction (not necessarily parallel or perpendicular to $\hat{n}_1$), or vice versa, is

$$\frac{F(\vec{r}, \hat{n}_1, \hat{n}_2)}{A} = \hat{n}_1 \cdot \underline{\underline{T}}(\vec{r}) \cdot \hat{n}_2 \qquad (8.34)$$

We may abstract out mechanical momentum and force densities $\vec{p}_{mech}(\vec{r})$ and $\vec{f}(\vec{r})$; *i.e.*, per unit volume expressions:

$$\boxed{\vec{p}_{mech}(\vec{r}) \equiv \rho_m(\vec{r})\vec{v}(\vec{r}) \qquad\qquad \vec{f}(\vec{r}) \equiv \vec{\nabla} \cdot \underline{\underline{T}}(\vec{r}) - \epsilon_o\,\mu_o\,\frac{\partial \vec{S}(\vec{r})}{\partial t}} \qquad (8.35)$$

where $\rho_m(\vec{r})$ is the *mass density*
$\vec{v}(\vec{r})$ is the *velocity field* of the mass density

We may conclude that these quantities are related locally because of the arbitrariness of the volume over which we are integrating:

$$\int_{\mathcal{V}} d\tau\, \frac{\partial \vec{p}_{mech}}{\partial t} = \frac{d}{dt}\int_{\mathcal{V}} d\tau\, \vec{p}_{mech} = \frac{d\vec{P}_{mech}}{dt} = \int_{\mathcal{V}} d\tau\, \vec{f} \qquad (8.36)$$

$$\implies \boxed{\frac{\partial \vec{p}_{mech}(\vec{r})}{\partial t} = \vec{f}(\vec{r}) = \vec{\nabla} \cdot \underline{\underline{T}}(\vec{r}) - \epsilon_o\,\mu_o\,\frac{\partial \vec{S}(\vec{r})}{\partial t}} \qquad (8.37)$$

This is the kind of conservation law we wanted to get to, a local one that relates *the rate of change of the local momentum density to the divergence of the local stress tensor and the rate of change of the Poynting vector*. It can also be viewed as a local force law, the generalization of Newton's Second Law.

As we did with the energy, and motivated by the appearance of a time derivative on the right side, we may rewrite the above as

$$\frac{d}{dt}\left(\vec{P}_{mech} + \epsilon_o\,\mu_o \int_{\mathcal{V}} d\tau\,\vec{S}\right) = \oint_{\mathcal{S}(\mathcal{V})} da\,\hat{n}\cdot\underline{\underline{T}} \qquad (8.38)$$

We are thus motivated to define the *linear momentum density* and *linear momentum* of the electromagnetic field as

$$\boxed{\vec{p}_{field}(\vec{r}) \equiv \vec{g}(\vec{r}) \equiv \epsilon_o\,\mu_o\,\vec{S}(\vec{r}) = \epsilon_o\,\vec{E}(\vec{r})\times\vec{B}(\vec{r}) \qquad \vec{P}_{field} = \int_{\mathcal{V}} d\tau\,\vec{g}}\qquad (8.39)$$

With that definition, we obtain

$$\frac{d}{dt}\left(\vec{P}_{mech} + \vec{P}_{field}\right) = \oint_{\mathcal{S}(V)} da\,\hat{n}\cdot\underline{\underline{T}} \quad\Longleftrightarrow\quad \frac{\partial}{\partial t}\left[\vec{p}_{mech}(\vec{r}) + \vec{g}(\vec{r})\right] = \vec{\nabla}\cdot\underline{\underline{T}}(\vec{r})$$

(8.40)

We thus see that *the rate of change of the total (mechanical + field) linear momentum in a volume is given by the integral of the stress tensor over the surface, or that the rate of change of the total momentum density at a point is given by the divergence of the stress tensor at that point.* The stress tensor is thus seen to be the momentum current density in the same way that $\vec{J}$ is the electric current density and $\vec{S}$ is the energy current density (up to a sign): all satisfy local continuity equations.

The second equation can also be considered a generalized force law, where now we consider the rate of change of the momentum of both the particles and the fields, with $\vec{\nabla}\cdot\underline{\underline{T}}$ being the "force" that acts on both.

When there is no change in mechanical momentum—e.g., in empty space—we obtain the *continuity equation for the linear momentum of the electromagnetic field*:

$$\vec{p}_{mech} = 0 : \quad \frac{d\vec{P}_{field}}{dt} = \oint_{\mathcal{S}(V)} da\,\hat{n}\cdot\underline{\underline{T}} \quad\Longleftrightarrow\quad \vec{\nabla}\cdot\underline{\underline{T}}(\vec{r}) = \frac{\partial\vec{g}(\vec{r})}{\partial t}$$

(8.41)

It is interesting to note that both $\vec{S}$ and $\underline{\underline{T}}$ play two roles:

- $\vec{S}$ is the power per unit area transported by the electromagnetic field, while $\vec{g} = \epsilon_o\,\mu_o\,\vec{S}$ is the linear momentum per unit volume stored in the field. This intimate connection between energy and momentum for the electromagnetic field reflects the photon's masslessness in quantum field theory.

- Similarly, $\underline{\underline{T}}$ plays two roles; both as a force per unit area (the stress) applied by the electromagnetic field as well as the momentum current density carried by the electromagnetic field (with a minus sign; units of momentum per unit area per unit time). This makes sense: for the electromagnetic field to exert a force, it must provide momentum.

Note this issue of the sign. If we wanted $\underline{\underline{T}}$ to have a continuity equation like current and energy, where the rate of change of the conserved quantity is equal to the negative of the divergence of the current (loss of conserved quantity corresponds to outflow of current), we would have had to define $\underline{\underline{T}}$ with the opposite sign. But the sign given ensures that $\underline{\underline{T}}$ can be used to calculate forces without a sign flip. This makes sense: $\underline{\underline{T}}$ pointing into a volume should have a positive surface integral so that it indicates it is adding momentum to the volume. The only way out of this choice would be if we wanted to flip the sign and interpret $\underline{\underline{T}}$ as the force that the mechanical system exerts on the field (and then the continuity equation would behave the way we want), but that would be nonintuitive since we generally want to calculate the forces the field exerts on the mechanical system.

Example 8.2: Magnetic Force Between Two Spinning Charged Hemispheres (Griffiths Problem 8.3)

Given two hemispherical shells of radius R and uniform surface charge density σ spinning at angular frequency ω about the z axis, what is the magnetic force between the north and south hemispheres? (Griffiths Example 8.2 calculates the electrostatic force for a similar situation.)

We have calculated the magnetic field for a similar configuration when we calculated the field of the uniformly magnetized sphere, which was

$$\vec{B}(r \leq R) = \frac{2}{3}\,\mu_o\,M\,\hat{z} \tag{8.42}$$

$$\vec{B}(r \geq R) = \frac{\mu_o}{4\,\pi}\,\frac{3\,(\vec{m}\cdot\hat{r})\hat{r} - \vec{m}}{r^3} \quad \text{with} \quad \vec{m} = \frac{4}{3}\,\pi\,R^3 M\,\hat{z} \tag{8.43}$$

The surface current was $\vec{K} = \hat{\phi}\,M\sin\theta$. In the new problem, the surface current is $\vec{K} = \hat{\phi}\,\sigma\,\omega\,R\sin\theta$, so we just need to replace M with $\sigma\,\omega\,R$, giving

$$\vec{B}(r \leq R) = \frac{2}{3}\,\mu_o\,\sigma\,\omega\,R\,\hat{z} \tag{8.44}$$

$$\vec{B}(r \geq R) = \frac{\mu_o}{4\,\pi}\,\frac{3\,(\vec{m}\cdot\hat{r})\hat{r} - \vec{m}}{r^3} \quad \text{with} \quad \vec{m} = \frac{4}{3}\,\pi\,R^4\sigma\,\omega\,\hat{z} \tag{8.45}$$

To calculate the force, we would nominally expect to calculate the flux of the stress tensor over the hemisphere (the plane at $z = 0$ for $r < R$ and the hemispherical shell $r = R$ at $z > 0$). However, the derivation implies that any volume containing the matter on which we would like to calculate the force suffices for the calculation. So let's do the calculation more easily by setting the surface to be the $z = 0$ plane. The force will only be in the z direction by symmetry, so we need only the T_{3i} components. Moreover, because the plane we want to do the calculation for has a surface normal only in the z direction, we can restrict to the T_{33} component:

$$T_{33} = \frac{1}{2\,\mu_o}\,B_z^2 \quad\Longrightarrow\quad T_{33}(r < R, z = 0) = \frac{2}{9}\,\mu_o\,\sigma^2\omega^2\,R^2 \tag{8.46}$$

$$T_{33}(r > R, z = 0) = \frac{\mu_o\,\sigma^2\,\omega^2\,R^8}{18\,r^6} \tag{8.47}$$

We can do the area integral easily ($\hat{n} = -\hat{z}$ because we want the force on the upper half space and $-\hat{z}$ is the outward surface normal):

$$F_z = -\int_0^{2\pi} d\phi \left[\int_0^R r\,dr\,T_{33}(r < R, z = 0) + \int_R^\infty r\,dr\,T_{33}(r > R, z = 0) \right] \tag{8.48}$$

$$= -2\pi \left[\frac{R^2}{2}\,\frac{2}{9}\,\mu_o\,\sigma^2\omega^2\,R^2 + \frac{\mu_o\,\sigma^2\,\omega^2\,R^8}{72\,R^4} \right] = -\frac{\pi}{4}\,\mu_o\,\sigma^2\,R^4\,\omega^2 \tag{8.49}$$

Conservation of Angular Momentum (Skip)

One can go back and write analogues of everything we did for linear momentum for the case of angular momentum. The key point is that the manipulations that led us to Equation 8.30 did not rely on any transformations of integrals; we just needed to manipulate the integrand. Those manipulations remain valid, but now with a $\vec{r} \times$ in front inside the integral. That is, we start with

$$\frac{d\vec{L}_{mech}}{dt} = \vec{N} = \vec{r} \times \vec{F} = \vec{r} \times q \left(\vec{E} + \vec{v} \times \vec{B} \right) \qquad (8.50)$$

Again, we integrate over the charge and current density to obtain

$$\frac{d\vec{L}_{mech}}{dt} = \int_{\mathcal{V}} d\tau \, \vec{r} \times \left(\rho \, \vec{E} + \vec{J} \times \vec{B} \right) \qquad (8.51)$$

Then we perform the same manipulations of the expression in parentheses as before, obtaining

$$\frac{d\vec{L}_{mech}}{dt} = \int_{\mathcal{V}} d\tau \left[\vec{r} \times \left(\vec{\nabla} \cdot \underline{\underline{T}} \right) - \epsilon_o \, \mu_o \, \frac{\partial}{\partial t} \left(\vec{r} \times \vec{S} \right) \right] \qquad (8.52)$$

Let's manipulate the expression $\vec{r} \times \left(\vec{\nabla} \cdot \underline{\underline{T}} \right)$: we would obviously like to turn it into a pure divergence. Using Equation 8.32,

$$\vec{r} \times \left(\vec{\nabla} \cdot \underline{\underline{T}}(\vec{r}) \right) = \vec{r} \times \sum_{i,j=1}^{3} \widehat{r}_j \frac{\partial T_{ij}}{\partial r_i} = \sum_{i,j,k=1}^{3} \widehat{r}_k \times \widehat{r}_j \, r_k \frac{\partial T_{ij}}{\partial r_i} \tag{8.53}$$

$$= \sum_{i,j,k=1}^{3} \widehat{r}_k \times \widehat{r}_j \left(\frac{\partial}{\partial r_i} \, r_k \, T_{ij} \right) - \sum_{i,j,k=1}^{3} \widehat{r}_k \times \widehat{r}_j \, T_{ij} \frac{\partial r_k}{\partial r_i} \tag{8.54}$$

$$= \sum_{i,j,k=1}^{3} \frac{\partial}{\partial r_i} \left(r_k \, \widehat{r}_k \times T_{ij} \, \widehat{r}_j \right) - \sum_{i,j,k=1}^{3} \widehat{r}_k \times \widehat{r}_j \, T_{ij} \, \delta_{ik} \tag{8.55}$$

$$= - \sum_{i,j,k=1}^{3} \frac{\partial}{\partial r_i} \left(T_{ij} \, \widehat{r}_j \times r_k \, \widehat{r}_k \right) - \sum_{i,j,k=1}^{3} \widehat{r}_k \times \widehat{r}_j \, T_{kj} \tag{8.56}$$

$$= \vec{\nabla} \cdot \left(-\underline{\underline{T}}(\vec{r}) \times \vec{r} \right) = \vec{\nabla} \cdot \underline{\underline{M}}(\vec{r}) \qquad \text{with} \qquad \underline{\underline{M}}(\vec{r}) = -\underline{\underline{T}}(\vec{r}) \times \vec{r} \tag{8.57}$$

(we did not show the explicit $\vec{r}$ dependence for the intermediate steps for brevity) where the second term in the penultimate line vanishes because it is the product of quantities that are antisymmetric in j and k ($\widehat{r}_k \times \widehat{r}_j$) and symmetric in j and k (T_{kj}). $\underline{\underline{M}}$ is the analogue of the stress tensor, but now for torque, which we will call the *torque tensor*. We reordered $\vec{r}$ and $\underline{\underline{T}}$ to obtain $-\underline{\underline{T}} \times \vec{r}$ rather than $\vec{r} \times \underline{\underline{T}}$ so the coordinate index of the divergence matches up with the first coordinate index of $\underline{\underline{M}}$.

Explicitly (undoing the reordering of $\underline{\underline{\mathcal{T}}}$ and $\vec{r}$ so we get $\vec{r}\times$, not $\times\vec{r}$, below, and not showing explicitly the position dependence for brevity):

$$\underline{\mathcal{M}} = -\underline{\underline{\mathcal{T}}} \times \vec{r} = -\sum_{i,j,k=1}^{3} T_{ij}\,\hat{r}_i\,\hat{r}_j \times r_k\,\hat{r}_k = -\sum_{i,j,k,m=1}^{3} \hat{r}_i\,\hat{r}_m\,\epsilon_{mjk}\,T_{ij}\,r_k \qquad (8.58)$$

$$= \sum_{i,j,k,m=1}^{3} \hat{r}_m\,\hat{r}_i\,r_k\,\epsilon_{mkj}\left(\epsilon_o\left[E_j\,E_i - \frac{1}{2}\,\delta_{ji}\,E^2\right] + \frac{1}{\mu_o}\left[B_j\,B_i - \frac{1}{2}\,\delta_{ji}\,B^2\right]\right) \qquad (8.59)$$

$$= \epsilon_o\left[\vec{r}\times\vec{E}\right]\vec{E} + \frac{1}{\mu_o}\left[\vec{r}\times\vec{B}\right]\vec{B} - \frac{1}{2}\sum_{i,k,m=1}^{3} \hat{r}_m\,\hat{r}_i\,r_k\,\epsilon_{mki}\left[\epsilon_o E^2 + \frac{1}{\mu_o}B^2\right] \qquad (8.60)$$

$$= \epsilon_o\left[\vec{r}\times\vec{E}\right]\vec{E} + \frac{1}{\mu_o}\left[\vec{r}\times\vec{B}\right]\vec{B} + \frac{1}{2}\left[\epsilon_o E^2 + \frac{1}{\mu_o}B^2\right]\begin{bmatrix} 0 & z & -y \\ -z & 0 & x \\ y & -x & 0 \end{bmatrix}$$
$$(8.61)$$

Aside: In Ph106a, one shows that angular momentum is more rigorously written as an antisymmetric second-rank (pseudo)tensor, but, because such an object has only 3 independent quantities, it can be reduced to a (pseudo)vector (first-rank tensor) using cross-product notation. That applies here to both $\vec{L}_{mech}$ and to $\vec{r}\times\vec{S}$. By extrapolation, $\underline{\mathcal{M}}$ may be written as an a completely antisymmetric *third-rank* (pseudo)tensor. Since we do not use any of the transformation properties of these objects under rotations in this course, there is no need to use these higher-rank objects and so we stick with the less sophisticated vector notation for cross products. But this concept will return when we consider the relativistic generalization of $\underline{\mathcal{M}}$ because the reduction to a second-rank tensor is only possible in three dimensions.

With our expression in terms of the divergence of $\underline{\underline{\mathcal{M}}}$, we may write the analogue of Equation 8.30 for torque:

$$\frac{d\vec{L}_{mech}}{dt} = \int_{\mathcal{V}} d\tau \left[\vec{\nabla} \cdot \underline{\underline{\mathcal{M}}} - \epsilon_o\,\mu_o\,\frac{\partial}{\partial t}\left(\vec{r}\times\vec{S}\right) \right] \tag{8.62}$$

Using the divergence theorem, we may rewrite as we did the force equation

$$\frac{d\vec{L}_{mech}}{dt} = \oint_{S(\mathcal{V})} da\,\hat{n}(\vec{r}) \cdot \underline{\underline{\mathcal{M}}}(\vec{r}) - \epsilon_o\,\mu_o\,\frac{d}{dt}\int_{\mathcal{V}} d\tau\,\vec{r}\times\vec{S}(\vec{r}) \tag{8.63}$$

We thus have a relation between *the rate of change of mechanical angular momentum and the flux of the torque tensor* $\underline{\underline{\mathcal{M}}}$ *into/out of the volume and the rate of the change of integral of the funny quantity containing the Poynting vector.*

Let's turn this into a differential version. We need to define the mechanical momentum density and the torque density:

$$\vec{\ell}_{mech}(\vec{r}) \equiv \vec{r} \times \vec{p}_{mech}(\vec{r}) = \vec{r} \times \rho_m(\vec{r})\vec{v}(\vec{r}) \qquad \vec{n}_{torque}(\vec{r}) \equiv \vec{\nabla} \cdot \underline{\underline{\mathcal{M}}}(\vec{r}) - \epsilon_o\,\mu_o\,\frac{\partial}{\partial t}\left(\vec{r} \times \vec{S}(\vec{r})\right)$$

$$(8.64)$$

Then we have

$$\int_{\mathcal{V}} d\tau\,\frac{\partial \vec{\ell}_{mech}}{\partial t} = \frac{d}{dt}\int_{\mathcal{V}} d\tau\,\vec{\ell}_{mech} = \frac{d\vec{L}_{mech}}{dt} = \int_{\mathcal{V}} d\tau\,\vec{n}_{torque} \qquad (8.65)$$

$$\implies \quad \boxed{\frac{\partial \vec{\ell}_{mech}(\vec{r})}{\partial t} = \vec{n}_{torque}(\vec{r}) = \vec{\nabla} \cdot \underline{\underline{\mathcal{M}}}(\vec{r}) - \epsilon_o\,\mu_o\,\frac{\partial}{\partial t}\left(\vec{r} \times \vec{S}(\vec{r})\right)} \qquad (8.66)$$

We thus obtain a local conservation law that relates *the rate of change of the local angular momentum density to the divergence of the local torque tensor and the rate of change of the rate of the change of the funny quantity containing the Poynting vector.* It can also be viewed as a local torque version of Newton's Second Law.

As before, it is natural to define a *field angular momentum density* and move it to the left side of the above equations:

$$\vec{\ell}_{field}(\vec{r}) = \vec{r} \times \vec{g}(\vec{r}) = \epsilon_o\,\mu_o\,\vec{r} \times \vec{S}(\vec{r}) = \epsilon_o\,\vec{r} \times (\vec{E}(\vec{r}) \times \vec{B}(\vec{r})) \qquad \vec{L}_{field} = \int_{\mathcal{V}} d\tau\,\vec{\ell}_{field}$$

(8.67)

With that definition, we obtain

$$\frac{d}{dt}\left(\vec{L}_{mech} + \vec{L}_{field}\right) = \oint_{\mathcal{S}(\mathcal{V})} da\,\hat{n} \cdot \underline{\underline{\mathcal{M}}} \iff \frac{\partial}{\partial t}\left(\vec{\ell}_{mech}(\vec{r}) + \vec{\ell}_{field}(\vec{r})\right) = \vec{\nabla} \cdot \underline{\underline{\mathcal{M}}}(\vec{r})$$

(8.68)

Again, we obtain an integral conservation equation relating *the rate of change of the total angular momentum in a volume to the integral of the torque tensor over the surface* and a local conservation equation relating *the rate of change of the total angular momentum density to the divergence of a current density, here now the angular momentum current density* (which has units of angular momentum per unit area per unit time). The second equation is a generalized local "torque" equation.

Note the choice of sign for $\underline{\underline{\mathcal{M}}}$ follows the same convention as for the stress tensor: it gives a continuity equation with a sign flip but is the correct sign for torque. Be aware that this sign convention is the opposite of Jackson's (his Problem 6.9).

Note that field angular momentum is not the same as photon spin or circular polarization; we will come back to this later when we discuss polarization of EM waves.

Section 9
Electromagnetic Waves

Introduction and Study Guidelines

Maxwell's Equations have in them the seeds of self-propagating disturbances in the electromagnetic field: though time-varying charges and currents must generate the waves, they can propagate on their own once initiated. So, in this section, we will develop the theory of such waves propagating in either free space or linear dielectric media, without any free charges. Later on, we will discuss radiation, the process by which time-varying charges and currents generate electromagnetic waves.

We deviate from Griffiths' ordering of topics because you have seen the wave equation three times before, in Ph1c, Ph2/12a, and Ph106a, so we do not need to reintroduce it from scratch. Let's just launch into it and bring the formalism of waves in as we go.

Electromagnetic Waves in Vacuum

From Maxwell's Equations to the Wave Equation

As noted earlier, we will consider Maxwell's Equations in free space with no sources:

$$\vec{\nabla} \cdot \vec{E} = 0 \qquad \vec{\nabla} \cdot \vec{B} = 0 \qquad \vec{\nabla} \times \vec{E} + \frac{\partial \vec{B}}{\partial t} = 0 \qquad \vec{\nabla} \times \vec{B} - \epsilon_o \, \mu_o \, \frac{\partial \vec{E}}{\partial t} = 0 \quad (9.1)$$

These equations couple $\vec{E}$ and $\vec{B}$, so let's try to find uncoupled equations by eliminating $\vec{B}$ from the $\vec{\nabla} \times \vec{E}$ equation and $\vec{E}$ from the $\vec{\nabla} \times \vec{B}$ equation by taking the curl again and using one of our standard vector identities:

$$\vec{\nabla} \times \left(\vec{\nabla} \times \vec{E} \right) = \vec{\nabla} \times \left(-\frac{\partial \vec{B}}{\partial t} \right) \qquad\qquad \vec{\nabla} \times \left(\vec{\nabla} \times \vec{B} \right) = \epsilon_o \, \mu_o \, \vec{\nabla} \times \left(-\frac{\partial \vec{E}}{\partial t} \right)$$

$$\vec{\nabla} \left(\vec{\nabla} \cdot \vec{E} \right) - \nabla^2 \vec{E} = -\epsilon_o \, \mu_o \, \frac{\partial}{\partial t} \, \vec{\nabla} \times \vec{B} \qquad \vec{\nabla}(\vec{\nabla} \cdot \vec{B}) - \nabla^2 \vec{B} = -\epsilon_o \, \mu_o \, \frac{\partial}{\partial t} \, \vec{\nabla} \times \vec{E}$$

$$\nabla^2 \vec{E} = \epsilon_o \, \mu_o \, \frac{\partial^2 \vec{E}}{\partial t^2} \qquad\qquad\qquad \nabla^2 \vec{B} = \epsilon_o \, \mu_o \, \frac{\partial^2 \vec{B}}{\partial t^2}$$

where $\vec{\nabla} \cdot \vec{E} = 0$ because there is no charge density. These are copies of the *wave equation*.

Specifically, these are component-by-component versions of the equation

$$\nabla^2 f(\vec{r}, t) = \frac{1}{v^2} \frac{\partial^2}{\partial t^2} f(\vec{r}, t) \tag{9.2}$$

One can see by substitution that any function of the form

$$f(\vec{r}, t) = g(w) \qquad \text{with} \qquad w = \vec{k} \cdot \vec{r} - \omega t \tag{9.3}$$

satisfies the wave equation. To prove this, we first need to calculate the derivatives:

$$\nabla^2 f(\vec{r}, t) = \sum_{i=1}^{3} \frac{\partial^2 f}{\partial r_i^2} = \sum_{i=1}^{3} \frac{\partial}{\partial r_i} \frac{dg}{dw} \frac{\partial w}{\partial r_i} = \sum_{i=1}^{3} \frac{\partial}{\partial r_i} \frac{dg}{dw} k_i \tag{9.4}$$

$$= \sum_{i=1}^{3} k_i \frac{d^2 g}{dw^2} \frac{\partial w}{\partial r_i} = \sum_{i=1}^{3} k_i^2 \frac{d^2 g}{dw^2} = |\vec{k}|^2 \frac{d^2 g}{dw^2} \tag{9.5}$$

$$\frac{\partial^2}{\partial t^2} f(\vec{r}, t) = \frac{\partial}{\partial t} \frac{dg}{dw} \frac{\partial w}{\partial t} = \frac{\partial}{\partial t} \frac{dg}{dw} (-\omega) = -\omega \frac{d^2 g}{dw^2} \frac{\partial w}{\partial t} = \omega^2 \frac{d^2 g}{dw^2} \tag{9.6}$$

The wave equation is satisfied by the assumed form if

$$|\vec{k}|^2 = \frac{\omega^2}{v^2} \tag{9.7}$$

The above condition says we can rewrite the argument, eliminating either ω or $|\vec{k}|$, as

$$w = \pm\vec{k} \cdot \vec{r} - |\vec{k}| \, v \, t = \frac{\omega}{v} \left(\pm\widehat{k} \cdot \vec{r} - v \, t \right) \tag{9.8}$$

where we have chosen ω and v to be always nonnegative while $\vec{k}$ is allowed to take on any sign and direction. We can see that surfaces of constant w are given by

$$\delta w = 0 \quad \Longrightarrow \quad \omega \left(\pm\widehat{k} \cdot \delta\vec{r} - v \, \delta t \right) = 0 \quad \Longrightarrow \quad \pm\widehat{k} \cdot \frac{\delta\vec{r}}{\delta t} = v \tag{9.9}$$

That is, the surfaces of constant w propagate in space along the direction $\pm\widehat{k}$ at speed v. This implies that the "shape function" $g(w)$ propagates at this speed. Returning to our electromagnetic wave equations, we thus see that these waves in the electric and magnetic fields propagate at speed $v = 1/\sqrt{\epsilon_o \, \mu_o}$ which is now, by definition, the *speed of light*, denoted by c.

The interpretation of ω and $k = |\vec{k}|$ are not clear yet, and in fact they are no longer strictly necessary (the factor $\omega/v = k$ could be absorbed into $g(w)$ now that we know ω and k are not independent), but they will become so below when we consider sinusoidal waves.

Lecture 24:

Electromagnetic Waves II:

Electromagnetic Waves in Vacuum (cont.):
Generic Properties,
Time-Averaging, Complex Notation,
Types of Polarization

Electromagnetic Waves in Perfectly Nonconducting, Linear,
Isotropic Matter:
Generic Properties,
Reflection and Refraction

Date Revised: 2024/04/16 07:00
Date Given: 2024/04/16

General Properties of Solutions to the EM Wave Equations

We can use Maxwell's Equations to derive some general properties about electromagnetic waves. Many of these connect to the fact that $\vec{E}$ and $\vec{B}$ are vector quantities. We will begin by assuming the waves are sinusoidal solutions of the most general form allowed so far

$$\vec{E}(\vec{r}, t) = \vec{E}_0 \cos\left(\vec{k}_E \cdot \vec{r} - \omega_E\, t + \delta_E\right) \tag{9.10}$$

$$\vec{B}(\vec{r}, t) = \vec{B}_0 \cos\left(\vec{k}_B \cdot \vec{r} - \omega_B\, t + \delta_B\right) \tag{9.11}$$

where we have allowed different ω, propagation directions $\widehat{k}$, and phase shifts δ because nothing has restricted that freedom yet. (The sign freedom on $\vec{k}$ has been absorbed into $\widehat{k}$.) We have assumed sinusoidal solutions because they form a complete basis for solution of the wave equation, so any solution can be decomposed in terms of them.

This sinusoidal assumption now enables an interpretation of ω and k. The time dependence at a given point in space has *angular frequency* ω, *frequency* $\nu = \omega/2\pi$, and *period* $T = 1/\nu$. The quantity k is the *propagation constant*, and the spatial dependence implies a *wavelength* $\lambda = 2\pi/k = v/\nu$. When we consider more general solutions that are the sums of sinusoids, these interpretations fail again because the sum does not correspond to single ω and k values.

With the sinusoidal assumption, we can demonstrate the following:

▶ *Transversality*
We can rewrite the divergence equations:

$$0 = \vec{\nabla} \cdot \vec{E} = \frac{d\vec{E}}{dw} \cdot \vec{\nabla} w = -\frac{\omega_E}{c} \, \widehat{k}_E \cdot \vec{E}_0 \, \sin\left(\vec{k}_E \cdot \vec{r} - \omega_E \, t + \delta_E\right) \quad (9.12)$$

$$0 = \vec{\nabla} \cdot \vec{B} = \frac{d\vec{B}}{dw} \cdot \vec{\nabla} w = -\frac{\omega_B}{c} \, \widehat{k}_B \cdot \vec{B}_0 \, \sin\left(\vec{k}_B \cdot \vec{r} - \omega_B \, t + \delta_B\right) \quad (9.13)$$

For the above equations to hold at all points in space, it is necessary for $\vec{E}_0$ and $\vec{B}_0$ to be perpendicular to their respective propagation directions. EM waves are thus *transverse* waves: the field disturbance is in the direction perpendicular to propagation.

It is somewhat remarkable that the same equations that relate fields to charges (and, in particular, imply there are no magnetic point charges) also show that, in the absence of charges, the fields are transverse. This turns out to be a very profound fact. You may recall from Ph106b PS1, when we considered a Coulomb law in which the potential included an exponential decay. In this case, $\vec{\nabla} \cdot \vec{E} \neq 0$ even when charges are not present. The non-vanishing of $\vec{\nabla} \cdot \vec{E}$ would permit a longitudinal component to the wave. In quantum field theory, a non-vanishing transverse component implies the particle corresponding to the field has mass and the static potential for the field is short range with decay length $\lambda \propto 1/M$.

▶ *Orthogonality and Equality of $\vec{k}$, ω, and δ*
Let's write the curl equations. First, we take the necessary derivatives:

$$
\vec{\nabla} \times \vec{E} = \sum_{i,j,k=1}^{3} \epsilon_{ijk}\, \widehat{r}_i\, \frac{\partial E_k}{\partial r_j} = \sum_{i,j,k=1}^{3} \epsilon_{ijk}\, \widehat{r}_i\, \frac{dE_k}{dw}\, \frac{\partial w}{\partial r_j}
$$

$$
= \sum_{i,j,k=1}^{3} \epsilon_{ijk}\, \widehat{r}_i\, \frac{dE_k}{dw}\, k_{E,j} = \vec{k}_E \times \frac{d\vec{E}}{dw} = \frac{\omega_E}{c}\, \widehat{k}_E \times \frac{d\vec{E}}{dw}
$$

$$
= -\frac{\omega_E}{c}\, \widehat{k}_E \times \vec{E}_0 \sin\left(\vec{k}_E \cdot \vec{r} - \omega_E\, t + \delta_E\right) \tag{9.14}
$$

$$
\frac{\partial \vec{E}}{\partial t} = \frac{d\vec{E}}{dw}\, \frac{\partial w}{\partial t} = -\omega_E\, \frac{d\vec{E}}{dw} = \omega_E\, \vec{E}_0 \sin\left(\vec{k}_E \cdot \vec{r} - \omega_E\, t + \delta_E\right) \tag{9.15}
$$

$\vec{B}$ has similar derivatives. Plugging the above into Faraday's Law and Ampere's Law:

$$
-\frac{\omega_E}{c}\, \widehat{k}_E \times \vec{E}_0 \sin\left(\vec{k}_E \cdot \vec{r} - \omega_E\, t + \delta_E\right) = -\omega_B\, \vec{B}_0 \sin\left(\vec{k}_B \cdot \vec{r} - \omega_B\, t + \delta_B\right) \tag{9.16}
$$

$$
\frac{\omega_B}{c}\, \widehat{k}_B \times \vec{B}_0 \sin\left(\vec{k}_B \cdot \vec{r} - \omega_B\, t + \delta_B\right) = -\frac{\omega_E}{c^2}\, \vec{E}_0 \sin\left(\vec{k}_E \cdot \vec{r} - \omega_E\, t + \delta_E\right) \tag{9.17}
$$

We can conclude two things from these equations:

▶ In order for the equality to hold at all points in space and at all times, the arguments of the sin functions on the two sides must be the same, $\vec{k}_E = \vec{k}_B$, $\omega_E = \omega_B$, and $\delta_E = \delta_B$

▶ $\vec{k}$, $\vec{E}_0$, and $\vec{B}_0$ form a mutually orthogonal set of vectors

In the end, we therefore have the following relation between $\vec{E}$ and $\vec{B}$:

$$\vec{E}(\vec{r}, t) = \vec{E}_0 \cos\left(\vec{k} \cdot \vec{r} - \omega\, t + \delta\right) \tag{9.18}$$

$$\vec{B}(\vec{r}, t) = \vec{B}_0 \cos\left(\vec{k} \cdot \vec{r} - \omega\, t + \delta\right) \tag{9.19}$$

$$\vec{B}_0 = \frac{1}{c}\,\widehat{k} \times \vec{E}_0 \quad \Longleftrightarrow \quad \vec{E}_0 = -c\,\widehat{k} \times \vec{B}_0 \tag{9.20}$$

One note on the behavior of independent polarizations. Given a propagation direction $\widehat{k}$, we can pick two directions, $\widehat{n}_1$ and $\widehat{n}_2$, to form a basis for $\vec{E}$. A natural choice is to require $\widehat{n}_1 \times \widehat{n}_2 = \widehat{k}$, which also implies $\widehat{k} \times \widehat{n}_1 = \widehat{n}_2$. Then $\vec{E}_0 = E_1 \widehat{n}_1 + E_2 \widehat{n}_2$ and $\vec{B}_0 = B_1 \widehat{n}_1 + B_2 \widehat{n}_2$. E_1 and E_2 are the two possible *polarizations* of the electric field. (Of course, we can pick any $\widehat{n}_1$ we want; once $\widehat{n}_1$ has been picked, then the polarization directions are set.) The curl of $\vec{E}$ relation then implies

$$B_2 = \vec{B}_0 \cdot \widehat{n}_2 = \frac{1}{c} \vec{E}_0 \cdot \widehat{n}_1 = \frac{E_1}{c} \qquad B_1 = \vec{B}_0 \cdot \widehat{n}_1 = -\frac{1}{c} \vec{E}_0 \cdot \widehat{n}_2 = -\frac{E_2}{c} \qquad (9.21)$$

We thus see that, aside from picking consistent $\widehat{k}$, $\widehat{n}_1$, and $\widehat{n}_2$, there is no connection between the (E_1, B_2) pair and the (E_2, B_1) pair. The waves in the two complementary polarizations can have different ω and thus different k. They are two completely independent waves. There is no fixed relationship between the waves in the two polarizations, and they can get out of phase with each other as they propagate if they have different ω.

If we consider two waves that have the same ω, then the two waves propagate together — their relative phase does not change with time or position. But there remains no condition connecting E_1 and E_2, or B_1 and B_2, so there is no requirement that the complementary polarizations have matching amplitude or phase (δ). We will see later that this independence of the two polarization amplitudes and phases can be used to generate a diverse set of possible polarizations: linear, circular, and elliptical.

Rewriting EM Waves using the Auxiliary Field

It is interesting to note at this point that $\vec{B}$ is not the most natural field quantity to work with: it is smaller than $\vec{E}$ by a factor c, which is large. If we instead use $\vec{H} = \vec{B}/\mu_o$, then we obtain

$$\vec{H}_0 = \frac{1}{Z_0} \, \widehat{k} \times \vec{E}_0 \qquad Z_0 = \sqrt{\frac{\mu_o}{\epsilon_o}} \qquad (9.22)$$

The quantity $Z_0 \approx 377 \ \Omega$ is known as the *impedance of free space* and has units of resistance (impedance). We see that $\vec{H}$ is now only a factor of 377 smaller than $\vec{E}$. We also recall that $\vec{H}$ has units of surface current density. This foreshadows the way $\vec{H}$ will be related to the surface currents that the electric field drives in polarizable/magnetizable media and in conductors.

Energy and Momentum in Electromagnetic Waves

The energy density in an electromagnetic wave, now using $|\vec{B}| = |\vec{E}|/c = |\vec{E}|\sqrt{\epsilon_o\,\mu_o}$, is

$$u = \frac{1}{2}\left(\epsilon_o\,E^2 + \frac{1}{\mu_o}B^2\right) = \frac{1}{2}\left(\epsilon_o\,E^2 + \frac{\epsilon_o\,\mu_o}{\mu_o}E^2\right) = \epsilon_o\,E^2 \qquad (9.23)$$

The energy flux per unit area is the Poynting vector:

$$\vec{S} = \frac{1}{\mu_o}\,\vec{E}\times\vec{B} = \frac{E^2}{c\,\mu_o}\,\widehat{k} = c\,\epsilon_o\,E^2\,\widehat{k} = c\,u\,\widehat{k} \qquad (9.24)$$

Thus, we see that the energy transported by the electromagnetic wave travels at the speed of light, just as the wave does. The momentum density vector is

$$\vec{g} = \epsilon_o\,\mu_o\,\vec{S} = \frac{u}{c}\,\widehat{k} = \frac{\epsilon_o\,E^2}{c}\,\widehat{k} \qquad (9.25)$$

Note that the energy flux and momentum density differ by a factor of c^2, not just c (as we would expect for a relation between energy and momentum) because one is a *flux* (energy/area/time) and the other is a *density* (momentum/volume); the difference in spatial and temporal units introduces another factor of velocity between the two.

We can also write down the stress tensor. First, consider the special case $\widehat{k} = k\,\widehat{z}$ so $\widehat{n}_1 = \widehat{x}$ and $\widehat{n}_2 = \widehat{y}$. Consider a wave polarized along $\widehat{x}$. Then the fields are

$$\vec{E} = E\,\widehat{x} \qquad \vec{B} = B\,\widehat{y} = \frac{E}{c}\,\widehat{y} = \sqrt{\mu_o\,\epsilon_o}\,E\,\widehat{y} \tag{9.26}$$

The stress tensor is

$$T_{11} = \epsilon_o\left(E^2 - \frac{1}{2}\,E^2\right) + \frac{1}{\mu_o}\left(-\frac{1}{2}\,B^2\right) = 0 \tag{9.27}$$

$$T_{22} = \epsilon_o\left(-\frac{1}{2}\,E^2\right) + \frac{1}{\mu_o}\left(B^2 - \frac{1}{2}\,B^2\right) = 0 \tag{9.28}$$

$$T_{33} = \epsilon_o\left(-\frac{1}{2}\,E^2\right) + \frac{1}{\mu_o}\left(-\frac{1}{2}\,B^2\right) = -u \tag{9.29}$$

$$T_{12} = T_{13} = T_{23} = 0 \tag{9.30}$$

$$\Longrightarrow \quad \underline{\underline{T}} = -u\,\widehat{z}\,\widehat{z} \tag{9.31}$$

The stress tensor for the complementary polarization is the same.

It is reasonable to extrapolate from the above that the generic stress tensor (now making the time dependence explicit) is

$$\underline{\underline{T}} = -\,\widehat{k}\,\widehat{k}\,\epsilon_o\,E_0^2\,\cos^2\left(\vec{k}\cdot\vec{r} - \omega\,t + \delta\right) \tag{9.32}$$

One explanation of the reason for the negative sign is that $\underline{\underline{T}}$ is the *negative* of the momentum current density.

From the stress tensor, we can calculate the *radiation pressure*, the force per unit area that would be applied to an object that absorbs the electromagnetic wave. Recall that $\widehat{n}_1 \cdot \underline{\underline{T}} \cdot \widehat{n}_2$ gives the force acting in the $\widehat{n}_1$ direction on a surface element whose normal is in the $\widehat{n}_2$ direction. Since $\underline{\underline{T}} \propto -\widehat{k}\,\widehat{k}$, the force is only nonzero (and positive) in the $\widehat{k}$ direction on an area element whose *outward* normal is in the $-\widehat{k}$ direction. (Recall how, in our example of using the stress tensor to calculate the force between the two spinning charged hemispheres, the surface normal was in the $-\widehat{z}$ direction for calculating the force on the hemisphere in the upper half-space.) The radiation pressure in the $\widehat{k}$ direction is then

$$Pressure = \widehat{k}\cdot\underline{\underline{T}}\cdot-\widehat{k} = E_0^2\,\cos^2\left(\vec{k}\cdot\vec{r} - \omega\,t + \delta\right) = u \tag{9.33}$$

We will see later that, if the wave is not absorbed but reflected, the wave maintains its amplitude $|\vec{E}|$ but its $\widehat{k}$ reverses sign, implying that the momentum transfer and thus the pressure are increased by a factor of 2.

It is not particularly useful to write down the angular momentum density and the angular momentum tensor for a plane wave. They are position dependent, reflecting the fact that, if a charged particle absorbs energy and momentum from a plane wave at some nonzero distance from the origin, it acquires a linear momentum $\vec{p}$ and thus an angular momentum $\vec{r} \times \vec{p}$. The latter carries no information beyond that of the former. Only if the wave has a nontrivial dependence of $\vec{E}$ and $\vec{B}$ on position — for example, $\vec{E} \times \vec{B} \propto \hat{\phi}$ — is the angular momentum of the wave interesting. Such waves are beyond the scope of our current discussion of plane waves.

Time Averaging for Plane Waves

For sinusoidally oscillating plane waves, it is standard to take time averages of quantities. Obviously, the fields themselves time-average to zero. But energy and momentum do not:

$$\langle u(\vec{r}_0)\rangle = \left\langle \epsilon_o E_0^2 \cos^2\left(\vec{k}\cdot\vec{r}_0 - \omega\,t + \delta\right)\right\rangle = \frac{1}{2}\,\epsilon_o E_0^2 \tag{9.34}$$

$$\left\langle \vec{S}(\vec{r}_0)\right\rangle = \frac{1}{2}\,c\,\epsilon_o E_0^2\,\widehat{k} \qquad \langle \vec{g}(\vec{r}_0)\rangle = \frac{1}{2}\,\frac{\epsilon_o E_0^2}{c}\,\widehat{k} \qquad \left\langle \underline{\mathcal{T}}(\vec{r}_0)\right\rangle = -\frac{1}{2}\,\epsilon_o E_0^2\,\widehat{k}\,\widehat{k} \tag{9.35}$$

The average power per unit area transported by the wave is the *intensity*

$$\mathrm{I} = \left\langle |\vec{S}|\right\rangle = \frac{1}{2}\,c\,\epsilon_o E_0^2 \tag{9.36}$$

Note that the magnitude refers to the *vector magnitude*. Starting with the next slide, when we consider complex notation for fields, we will always define $\vec{S}$ in such a way as to be real, so the magnitude will continue to refer to vector magnitude even for complex fields.

Complex Notation for Plane Waves

For the sake of convenience in manipulation, we will from now on use *complex notation* for plane waves,

$$\vec{E}(\vec{r}, t) = \hat{n}\,\mathcal{R}\left[\widetilde{E}_0\,e^{i(\vec{k}\cdot\vec{r} - \omega t)}\right] \qquad\qquad \vec{B}(\vec{r}, t) = \frac{\hat{k} \times \hat{n}}{c}\,\mathcal{R}\left[\widetilde{E}_0\,e^{i(\vec{k}\cdot\vec{r} - \omega t)}\right] \quad (9.37)$$

$$\widetilde{E}_0 = E_0\,e^{i\delta} \tag{9.38}$$

where $\widetilde{E}_0$ is now a complex number into which we have absorbed the phase factor $e^{i\delta}$ and $\mathcal{R}$ means "take the real part." We will not carry along tildes on the vectors $\vec{E}$ and $\vec{E}_0$ because it would be too cumbersome for the notation. It will be clear from context whether we mean the complex or real fields.

To calculate quadratic quantities like u, $\vec{S}$, $\vec{g}$, and $\underline{\underline{\mathcal{T}}}$ with full space- and time-dependence requires that one first take the real part and then apply the previously provided formulae for these quadratic quantities.

Fortunately, if one is only interested in time averages, there is a simple extension to the prescription for time averages involving complex conjugation and multiplication by a factor of $1/2$:

$$\langle u(\vec{r}_0) \rangle = \frac{1}{4} \left(\epsilon_o \vec{E}^* \cdot \vec{E} + \frac{1}{\mu_o} \vec{B}^* \cdot \vec{B} \right) \overset{pw\ vac}{=} \frac{1}{2} \epsilon_o E_0^2 \tag{9.39}$$

$$\left\langle \vec{S}(\vec{r}_0) \right\rangle = c^2 \left\langle \vec{g}(\vec{r}_0) \right\rangle = \frac{1}{2\mu_o} \mathcal{R} \left(\vec{E}^* \times \vec{B} \right) \overset{pw\ vac}{=} \frac{1}{2} c \epsilon_o E_0^2 \, \hat{k} \tag{9.40}$$

$$\left\langle \left(\underline{\underline{T}} \right)_{ij} (\vec{r}_0) \right\rangle = \frac{1}{2} \left[\epsilon_o \left(\mathcal{R} \left[E_i^* E_j \right] - \frac{1}{2} \delta_{ij} \vec{E}^* \cdot \vec{E} \right) + \frac{1}{\mu_o} \left(\mathcal{R} \left[B_i^* B_j \right] - \frac{1}{2} \delta_{ij} \vec{B}^* \cdot \vec{B} \right) \right] \tag{9.41}$$

$$= -\frac{1}{2} \epsilon_o E_0^2 \, \hat{k} \cdot \hat{r}_i \, \hat{k} \cdot \hat{r}_j \tag{9.42}$$

where the first expression for each quantity is always valid (even later for conductors when $\vec{E}$ and $\vec{B}$ can be out of phase) while the final evaluation is only valid for the plane waves in vacuum we have been considering (indicated by the $\overset{pw\ vac}{=}$ notation). The factors of $1/2$ in front come from time-averaging. Note that there may be spatial dependence remaining in the result if the wave amplitude has a dependence on position outside of the sinusoidal wave-propagation factor (not possible for plane EM waves, but it will happen for radiation). It is not necessary to take the real part for u and certain pieces of $\underline{\underline{T}}$ because they are manifestly real.

Types of Polarization

So far, we have only discussed linear polarization, wherein the direction of $\vec{E}$ at a particular point in space stays constant over time (aside from sign flips). However, by combining two orthogonal linear polarizations with appropriate complex coefficients, one can obtain more "complex" behavior.

The simplest extension is to consider what happens when you add two orthogonal polarizations of the same amplitude but with a possible phase shift:

$$\vec{E}(\vec{r}, t) = \frac{\widetilde{E}_0}{\sqrt{2}} \left(\widehat{n}_1 + \widehat{n}_2 \, e^{i\delta} \right) e^{i\left(\vec{k} \cdot \vec{r} - \omega t \right)} \tag{9.43}$$

If $\delta = 0$ or $\delta = \pi$, then one just obtains a linear polarization in the direction $(\widehat{n}_1 \pm \widehat{n}_2) / \sqrt{2}$.

But, if $\delta = \pm \pi/2$, then the wave polarized along $\widehat{n}_2$ is $\pi/2$ out of phase with the wave polarized along $\widehat{n}_1$: when the $\widehat{n}_1$ mode has zero amplitude, the $\widehat{n}_2$ mode has maximum amplitude and vice versa. If we take the real part, it is clear what is going on:

$$\mathcal{R}\left[\vec{E}(\vec{r}, t)\right] = \frac{E_0}{\sqrt{2}}\left[\widehat{n}_1 \cos\left(\vec{k}\cdot\vec{r} - \omega t + \delta_0\right) + \widehat{n}_2 \cos\left(\vec{k}\cdot\vec{r} - \omega t + \delta_0 \pm \frac{\pi}{2}\right)\right]$$
(9.44)

$$= \frac{E_0}{\sqrt{2}}\left[\widehat{n}_1 \cos\left(\vec{k}\cdot\vec{r} - \omega t + \delta_0\right) \mp \widehat{n}_2 \sin\left(\vec{k}\cdot\vec{r} - \omega t + \delta_0\right)\right] \quad (9.45)$$

The polarization vector maintains an amplitude $E_0/\sqrt{2}$ but it sweeps around in a circle with period $T = 2\pi/\omega$: this is *circular polarization*. To understand which direction the polarization vector rotates, let's look *into* the wave (toward $-\widehat{k}$) while sitting at a fixed point in space (fixed $\vec{r}$). The time-varying component of the arguments of the sinusoids is $-\omega t$, and thus, as time evolves positively, the arguments of the sinusoids evolve negatively. For $\delta = +\pi/2$, the sign on the second term is negative, the rotation is counterclockwise, and the wave is called *left circularly polarized*. Conversely, $\delta = -\pi/2$ yields clockwise rotation and is called *right circularly polarized*. One also speaks in terms of helicity, in which case one considers the rotation of the polarization *relative to the direction of motion* using the right-hand rule. The left circularly polarized wave has *positive helicity* because the polarization vector rotates around $+\widehat{k}$ according to the right-hand rule (thumb along $+\widehat{k}$). The right circularly polarized wave has *negative helicity* because it obeys the left-hand rule.

The next possibility is to allow unequal coefficients:

$$\vec{E}(\vec{r}, t) = \widetilde{E}_0 \left(\alpha \, \widehat{n}_1 + \beta \, \widehat{n}_2 \, e^{i\delta} \right) e^{i(\vec{k}\cdot\vec{r} - \omega\, t)} \qquad \alpha^2 + \beta^2 = 1 \qquad (9.46)$$

When $\delta = 0$ or $\delta = \pi$, we again obtain linear polarization, but now making an angle $\theta = \tan^{-1}(\pm\beta/\alpha)$ with the $\widehat{n}_1$ axis (sign is the same as sign of $e^{i\delta} = \pm 1$).

If we now consider $\delta = \pm\pi/2$ and $\alpha \neq \beta$, we obtain an *elliptically polarized* wave: at a fixed point, the polarization vector sweeps out an ellipse whose semimajor and semiminor axes are along $\widehat{n}_1$ and $\widehat{n}_2$. If δ is an arbitrary value, then the semimajor and semiminor axes are rotated from the $\widehat{n}_1$–$\widehat{n}_2$ system by an angle related to δ. We also obtain an elliptically polarized wave if we consider $\delta \neq 0$, $\pm\pi/2$, or π but $\alpha = \beta$,

It turns out that elliptically polarized waves are easier to analyze if they are rewritten in terms of the two helicities (or circular polarizations). That is, if we take as our polarization basis and field decomposition

$$\widehat{n}_\pm = \frac{1}{\sqrt{2}} \left(\widehat{n}_1 \pm e^{i\,\pi/2} \, \widehat{n}_2 \right) \qquad \vec{E}(\vec{r}, t) = \left(\widetilde{E}_+ \, \widehat{n}_+ + \widetilde{E}_- \, \widehat{n}_- \right) e^{i(\vec{k}\cdot\vec{r} - \omega\, t)} \qquad (9.47)$$

then the parameters of the ellipse traced out by the polarization vector are:

$$r\, e^{i\,\theta} = \frac{\widetilde{E}_-}{\widetilde{E}_+} \qquad \frac{\text{semiminor axis}}{\text{semimajor axis}} = \left| \frac{1 - r}{1 + r} \right| \qquad \text{angle wrt } \widehat{n}_1 = \frac{\theta}{2} \qquad (9.48)$$

where the angle is measured *looking into the wave* (*i.e.*, looking in the $-\widehat{k}$ direction).

It is interesting to note that a circularly polarized plane wave is, in terms of angular momentum, no different from a linearly polarized plane wave according to the definition of the angular momentum density, Equation 8.67:

$$\vec{\ell}_{field} = \epsilon_o \, \vec{r} \times \left(\vec{E} \times \vec{B} \right) = \frac{\epsilon_o \, E_0^2}{c} \, \vec{r} \times \hat{k} \tag{9.49}$$

One can see that the angular momentum has to do with the relative orientation of the propagation direction and the position vector, not with the nature of the polarization. This reflects the fact that, in quantum mechanics, the helicity of the wave becomes the intrinsic *spin angular momentum* of the photon, while the quantity calculated above is the *orbital angular momentum* of the photon and has to do with the *spatial distribution* of the EM wave, in much the same way that orbital angular momentum in quantum mechanics is determined by the spatial distribution of the wavefunction and is unassociated with the particle's spin.

Electromagnetic Waves in Perfectly Nonconducting, Linear, Isotropic Matter

Propagation in Linear, Isotropic Media

Maxwell's Equations in matter in the absence of free charges and currents are

$$\vec{\nabla} \cdot \vec{D} = 0 \qquad \vec{\nabla} \cdot \vec{B} = 0 \qquad \vec{\nabla} \times \vec{E} + \frac{\partial \vec{B}}{\partial t} = 0 \qquad \vec{\nabla} \times \vec{H} - \frac{\partial \vec{D}}{\partial t} = 0 \qquad (9.50)$$

As noted earlier, we need relations between $\vec{D}$ and $\vec{E}$ and between $\vec{H}$ and $\vec{B}$ to make use of these. If we assume linear, isotropic media (ϵ and μ scalars, not tensors)

$$\vec{D} = \epsilon \vec{E} \qquad \vec{B} = \mu \vec{H} \qquad (9.51)$$

then the equations reduce to

$$\vec{\nabla} \cdot \vec{E} = 0 \qquad \vec{\nabla} \cdot \vec{B} = 0 \qquad \vec{\nabla} \times \vec{E} + \frac{\partial \vec{B}}{\partial t} = 0 \qquad \vec{\nabla} \times \vec{B} - \epsilon \mu \frac{\partial \vec{E}}{\partial t} = 0 \qquad (9.52)$$

These are the same as our equations in vacuum, leading to the same kinds of waves, but with the modification

$$v = \frac{1}{\sqrt{\epsilon \mu}} = \frac{c}{n} \qquad \text{with} \qquad n = \sqrt{\frac{\epsilon \mu}{\epsilon_o \mu_o}} = \textit{index of refraction} \qquad (9.53)$$

This mathematical transformation reflects a remarkable fact: the complicated polarization and magnetization of the medium occurring as the wave passes through it do nothing except change its speed and, we shall see, affect the wave amplitude. This is a consequence of the linearity of the medium we assume.

Most materials in which waves can propagate (as we will see, this means materials that do not have high conductivities) have $\mu \approx \mu_o$, so $n \approx \sqrt{\epsilon/\epsilon_o} = \sqrt{\epsilon_r}$. Since $\epsilon_r > 1$ in general (there are very few paraelectric materials that enhance the field rather than act to decrease it), light generally goes more slowly in dielectrics. Though, creation of metamaterials in which the effective index of refraction is less than unity (over a limited frequency range via use of resonant structures) is an area of active research!

The relation between $\vec{B}$ and $\vec{E}$, Equation 9.20, is modified in the obvious manner:

$$\vec{B}(\vec{r}, t) = \frac{1}{v} \, \widehat{k} \times \vec{E}(\vec{r}, t) \tag{9.54}$$

As we did for free space, we also have that $\vec{H}$ and $\vec{E}$ are related by an impedance, $Z = \sqrt{\mu/\epsilon}$, which we now call the *wave impedance*. With it, we have

$$\vec{H}(\vec{r}, t) = \frac{\vec{B}(\vec{r}, t)}{\mu} = \frac{1}{Z} \, \widehat{k} \times \vec{E}(\vec{r}, t) \tag{9.55}$$

Recall that $\vec{H}$ carries units of surface current density.

Energy density, the Poynting vector, the momentum vector, and the stress tensor take on unsurprising forms given the above modification:

$$u = \frac{1}{2}\left(\epsilon E^2 + \frac{1}{\mu}B^2\right) = \epsilon E^2 \qquad \vec{S} = \frac{1}{\mu}\vec{E}^* \times \vec{B} = v\,\epsilon\,E^2\,\widehat{k} = v\,u\,\widehat{k} \qquad (9.56)$$

$$\vec{g} = \epsilon\,\mu\,\vec{S} = \frac{\epsilon\,E^2}{v}\,\widehat{k} = \frac{u}{v}\,\widehat{k} \qquad \underline{\underline{T}} = -\epsilon\,E^2\,\widehat{k}\,\widehat{k} = -u\,\widehat{k}\,\widehat{k} \qquad (9.57)$$

The time averages for a sinusoidal wave of (real) amplitude E_0 are (now including intensity):

$$\langle u \rangle = \frac{1}{2}\,\epsilon\,E_0^2 \qquad \left\langle \vec{S} \right\rangle = \frac{1}{2\,\mu}\left\langle \vec{E}^* \times \vec{B} \right\rangle = \frac{1}{2}\,v\,\epsilon\,E_0^2\,\widehat{k} = v\,\langle u \rangle\,\widehat{k} \qquad (9.58)$$

$$\mathrm{I} = \left\langle |\vec{S}| \right\rangle = \frac{1}{2}\,v\,\epsilon\,E_0^2 = v\,\langle u \rangle \qquad (9.59)$$

$$\langle \vec{g} \rangle = \epsilon\,\mu\,\left\langle \vec{S} \right\rangle = \frac{1}{2}\,\frac{\epsilon\,E_0^2}{v}\,\widehat{k} = \frac{\langle u \rangle}{v}\,\widehat{k} \qquad \left\langle \underline{\underline{T}} \right\rangle = -\frac{1}{2}\,\epsilon\,E_0^2\,\widehat{k}\,\widehat{k} = -\langle u \rangle\,\widehat{k}\,\widehat{k} \qquad (9.60)$$

Since $v > c$ is possible, we cannot so easily interpret the above equations as implying that energy propagates at speed v. You know from Ph2/12 that energy in a wave propagates with the group velocity $v_g = d\omega/dk$ and so it becomes important to know $v(\omega)$. We will build an approximate physical model for $v(\omega)$ in §5.

Boundary Conditions

Recalling our boundary conditions for linear media (Equations 2.61, 4.23, ??, ??) and applying our assumption of no free currents, we have ($\hat{n}$ = normal from 1 to 2, $\hat{s}$ = any tangential vector at interface):

$$\hat{n} \cdot \epsilon_1 \vec{E}_1 = \hat{n} \cdot \epsilon_2 \vec{E}_2 \qquad \hat{n} \cdot \vec{B}_1 = \hat{n} \cdot \vec{B}_2 \qquad \hat{s} \cdot \vec{E}_1 = \hat{s} \cdot \vec{E}_2 \qquad \hat{s} \cdot \frac{\vec{B}_1}{\mu_1} = \hat{s} \cdot \frac{\vec{B}_2}{\mu_2} \quad (9.61)$$

We will apply these to calculate the reflection and refraction of EM waves at the interface between different linear media. We will write the magnetic field boundary condition in terms of $\vec{H}$ because it makes them look like the electric field boundary conditions, which will be convenient during our discussion of reflection and refraction:

$$\hat{n} \cdot \mu_1 \vec{H}_1 = \hat{n} \cdot \mu_2 \vec{H}_2 \qquad \hat{s} \cdot \vec{H}_1 = \hat{s} \cdot \vec{H}_2 \qquad (9.62)$$

Reflection and Refraction: General Considerations

We will skip past the case of normal incidence, which you studied in Ph1c, but we will consider it as a special case of our generic results.

Assume we have a wave with propagation vector $\vec{k}_i$ propagating in medium 1 and incident on an interface with medium 2, with normal $\widehat{n}$ pointing from 1 into 2. We expect there to be reflected and transmitted waves. We write these all as

$$\vec{E}_i(\vec{r}, t) = \vec{E}_{0,i} \, e^{i(\vec{k}_i \cdot \vec{r} - \omega t)} \qquad\qquad \vec{H}_i(\vec{r}, t) = \frac{1}{Z_1} \, \widehat{k}_i \times \vec{E}_i \qquad (9.63)$$

$$\vec{E}_r(\vec{r}, t) = \vec{E}_{0,r} \, e^{i(\vec{k}_r \cdot \vec{r} - \omega t)} \qquad\qquad \vec{H}_r(\vec{r}, t) = \frac{1}{Z_1} \, \widehat{k}_r \times \vec{E}_r \qquad (9.64)$$

$$\vec{E}_t(\vec{r}, t) = \vec{E}_{0,t} \, e^{i(\vec{k}_t \cdot \vec{r} - \omega t)} \qquad\qquad \vec{H}_t(\vec{r}, t) = \frac{1}{Z_2} \, \widehat{k}_t \times \vec{E}_t \qquad (9.65)$$

We use $\vec{H}$ instead of $\vec{B}$ because the boundary conditions are more easily written in terms of $\vec{H}$. We have already applied the condition that the frequencies of the three waves are identical. This is necessary for any boundary conditions connecting them to be applicable at all time. Then

$$k_i \, v_1 = k_r \, v_1 = k_t \, v_2 = \omega \qquad \Longrightarrow \qquad k_r = k_i \qquad k_t = \frac{v_1}{v_2} \, k_i = \frac{n_2}{n_1} \, k_i \qquad (9.66)$$

Since $k = 2\pi/\lambda$, this implies the wavelength differs by n_2/n_1 in the two media!

Now, consider the kinds of matching conditions we will write down. They must always hold over all $\vec{r}$ in the interface. In particular, if we take the interface to be the $z = 0$ plane, with $\widehat{n} = \widehat{z}$, then the matching conditions that must hold at all x and y are of the form

$$( )\, e^{i\left(\vec{k}_i \cdot (x\widehat{x} + y\widehat{y}) - \omega t\right)} + ( )\, e^{i\left(\vec{k}_r \cdot (x\widehat{x} + y\widehat{y}) - \omega t\right)} = ( )\, e^{i\left(\vec{k}_t \cdot (x\widehat{x} + y\widehat{y}) - \omega t\right)} \qquad (9.67)$$

In order for these to hold at arbitrary x and y, it must be that

$$\widehat{x} \cdot \vec{k}_i = \widehat{x} \cdot \vec{k}_r = \widehat{x} \cdot \vec{k}_t \qquad (9.68)$$
$$\widehat{y} \cdot \vec{k}_i = \widehat{y} \cdot \vec{k}_r = \widehat{y} \cdot \vec{k}_t \qquad (9.69)$$
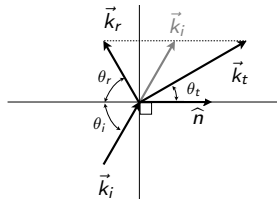
These conditions imply that if you project all three propagation vectors into the plane of the interface (whose normal is $\widehat{n}$), then their projections in that plane are equal. Furthermore, there is a plane formed by this common xy projection of the propagation vectors and the normal $\widehat{n}$ (which is normal to the projection plane), and all three vectors lie in this plane.

Each propagation vector makes an angle with the interface normal, $\widehat{n}$. We label them θ_i, θ_r, and θ_t. The projection of a given wavevector parallel to the normal is therefore $|\vec{k}| \cos \theta$ while the projection perpendicular to the normal (in the plane of the interface) is $|\vec{k}| \sin \theta$. Since we have argued that these projections into the plane of the interface are equal, we thus have

$$k_i \sin \theta_i = k_r \sin \theta_r = k_t \sin \theta_t \qquad (9.70)$$

Now, using our relations $k_r = k_i$ and $k_t = \frac{n_2}{n_1} k_i$, we may conclude:

$$\textit{law of reflection:} \qquad \theta_i = \theta_r \qquad (9.71)$$
$$\textit{law of refraction (Snell's Law):} \qquad n_1 \sin \theta_i = n_2 \sin \theta_t \qquad (9.72)$$

Snell's Law tells us that the light ray bends toward the normal if $n_2 > n_1$, and it bends away from the normal if $n_2 < n_1$.

Total internal reflection occurs when $n_2 < n_1$ and $\sin \theta_i > n_2/n_1$: in this case, $\sin \theta_t > 1$ and there is no solution for k_t. This happens because, when $n_2 < n_1$, then the magnitude $k_t < k_i$, so then the projection $k_i \sin \theta_i$ can be (but doesn't have to be, depending on the value of $\sin \theta_i$) too large for k_t to match. We will study this case in more detail in a homework problem.

Note that none of these results depending on knowing anything about Maxwell's Equations or boundary conditions: these are generic properties of any waves, which is why they hold for sound waves, phonons, etc. They result only from matching time and space dependences at the boundary.

Lecture 25:

Electromagnetic Waves III:
Electromagnetic Waves in Perfectly Nonconducting, Linear,
Isotropic Matter (cont.):
Reflection and Refraction (cont.)

Date Revised: 2024/04/18 06:30
Date Given: 2024/04/18

Reflected and Transmitted Field Relations

Now, let's apply our electromagnetic boundary conditions to figure out how the amplitudes and energies are related. We may drop all the exponential factors because we have established that they are identical at the interface. We need to define two new unit vectors:

$$\widehat{u} \propto \widehat{n} \times \widehat{k}_i = \widehat{n} \times \widehat{k}_r = \widehat{n} \times \widehat{k}_t$$
$$\widehat{w} = \widehat{u} \times \widehat{n} \qquad (9.73)$$
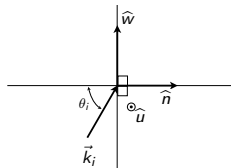
It doesn't matter which $\widehat{k}$ we use to define $\widehat{u}$ because we argued earlier that all of them lie in the same plane with $\widehat{n}$, so the direction normal to the plane containing them and $\widehat{n}$ is independent of which one is used. (The equality of the three cross-products follows from Equations 9.68 and 9.69.) Clearly, then, $\widehat{u}$ is perpendicular to this plane in which $\widehat{n}$ and the propagation vectors lie. $\widehat{w}$ is then the obvious third direction, and it and $\widehat{n}$ define the plane that the $\widehat{k}$ live in.

With these definitions, our boundary conditions can be written as

$$\widehat{n} \cdot \epsilon_1 \left(\vec{E}_{0,i} + \vec{E}_{0,r} \right) = \widehat{n} \cdot \epsilon_2 \, \vec{E}_{0,t} \qquad \left[\begin{array}{c} \widehat{u} \\ \widehat{w} \end{array} \right] \cdot \left(\vec{E}_{0,i} + \vec{E}_{0,r} \right) = \left[\begin{array}{c} \widehat{u} \\ \widehat{w} \end{array} \right] \cdot \vec{E}_{0,t} \quad (9.74)$$

$$\widehat{n} \cdot \mu_1 \left(\vec{H}_{0,i} + \vec{H}_{0,r} \right) = \widehat{n} \cdot \mu_2 \vec{H}_{0,t} \qquad \left[\begin{array}{c} \widehat{u} \\ \widehat{w} \end{array} \right] \cdot \left(\vec{H}_{0,i} + \vec{H}_{0,r} \right) = \left[\begin{array}{c} \widehat{u} \\ \widehat{w} \end{array} \right] \cdot \vec{H}_{0,t} \quad (9.75)$$

where the stacking of $\widehat{u}$ and $\widehat{w}$ is just meant to indicate that those equations apply
with either $\widehat{u}$ on both sides or $\widehat{w}$ on both sides.

Now, we must consider two cases for the polarization of the incoming wave: *in
(parallel to) the plane of incidence* or *perpendicular to the plane of incidence*, also
termed *transverse magnetic (TM)* and *transverse electric (TE)* for obvious reasons:

$$
\begin{array}{ccc}
\begin{array}{c} \textit{in (parallel to) the} \\ \textit{plane of incidence} \\ \textit{or transverse magnetic (TM)} \end{array} & \vec{E}_{0,i} \cdot \widehat{u} = 0 \quad \vec{H}_{0,i} \cdot \left[\begin{array}{c} \widehat{n} \\ \widehat{w} \end{array} \right] = 0 & (9.76)
\end{array}
$$

$$
\begin{array}{ccc}
\begin{array}{c} \textit{perpendicular to the} \\ \textit{plane of incidence} \\ \textit{or transverse electric (TE)} \end{array} & \vec{E}_{0,i} \cdot \left[\begin{array}{c} \widehat{n} \\ \widehat{w} \end{array} \right] = 0 \qquad \vec{H}_{0,i} \cdot \widehat{u} = 0 & (9.77)
\end{array}
$$

In (parallel to) the plane of incidence (TM):
When $\vec{E}_{0,i}$ is in the plane of incidence, we can decompose it into pieces along $\widehat{w}$ and along $\widehat{n}$. There is freedom on the sign convention, and we choose

$$\vec{E}_{0,i} = \widetilde{E}_{0,i} \left(\widehat{w} \cos \theta_i - \widehat{n} \sin \theta_i \right) \tag{9.78}$$

We pick the conventions for $\vec{E}_{0,r}$ and $\vec{E}_{0,t}$ so that all three electric field vectors align for normal incidence. (The $\vec{H}$ orientations are then defined by this choice and the direction of the corresponding $\widehat{k}$ vectors.) With these choices, we then have (remember, $\vec{H} = \widehat{k} \times \vec{E}/Z$):

$$\vec{E}_{0,r} = \widetilde{E}_{0,r} \left(\widehat{w} \cos \theta_r + \widehat{n} \sin \theta_r \right) \tag{9.79}$$

$$\vec{E}_{0,t} = \widetilde{E}_{0,t} \left(\widehat{w} \cos \theta_t - \widehat{n} \sin \theta_t \right) \tag{9.80}$$

$$\vec{H}_{0,i} = \frac{\widetilde{E}_{0,i}}{Z_1} \widehat{u} \tag{9.81}$$

$$\vec{H}_{0,r} = -\frac{\widetilde{E}_{0,r}}{Z_1} \widehat{u} \tag{9.82}$$

$$\vec{H}_{0,t} = \frac{\widetilde{E}_{0,t}}{Z_2} \widehat{u} \tag{9.83}$$
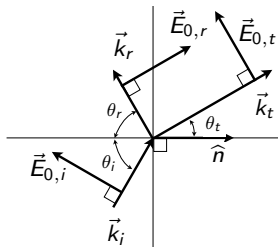
So, restricting to the boundary conditions with information, we obtain

$$\widehat{n}: \qquad \epsilon_1 \left(-\widetilde{E}_{0,i}\,\sin\theta_i + \widetilde{E}_{0,r}\,\sin\theta_r\right) = -\epsilon_2 \widetilde{E}_{0,t}\,\sin\theta_t \qquad (9.84)$$

$$\widehat{w}: \qquad \widetilde{E}_{0,i}\,\cos\theta_i + \widetilde{E}_{0,r}\,\cos\theta_r = \widetilde{E}_{0,t}\,\cos\theta_t \qquad (9.85)$$

$$\widehat{u}: \qquad \frac{1}{Z_1}\left(\widetilde{E}_{0,i} - \widetilde{E}_{0,r}\right) = \frac{1}{Z_2}\,\widetilde{E}_{0,t} \qquad (9.86)$$

Since the incident wave amplitude must be allowed to be arbitrary, we expect to only be able to determine the ratios $\widetilde{E}_{r,0}/\widetilde{E}_{i,0}$ and $\widetilde{E}_{t,0}/\widetilde{E}_{i,0}$. Only two of the above equations can therefore be independent. (One can easily see that the $\widehat{n}$ and $\widehat{u}$ equations are equivalent via Snell's Law.) Picking the last two because they are easiest to work with, we obtain *Fresnel's Equations in (parallel to) the plane of incidence (for TM waves)*:

$$\alpha = \frac{\cos\theta_t}{\cos\theta_i} \qquad \beta = \frac{Z_1}{Z_2} \qquad (9.87)$$

$$\boxed{\frac{\widetilde{E}_{0,r}}{\widetilde{E}_{0,i}} = \left(\frac{\alpha - \beta}{\alpha + \beta}\right) \qquad \frac{\widetilde{E}_{0,t}}{\widetilde{E}_{0,i}} = \left(\frac{2}{\alpha + \beta}\right)} \qquad \begin{array}{l} \textit{Fresnel's} \\ \textit{Equations} \\ \textit{in (parallel to)} \\ \textit{the plane of} \\ \textit{incidence (TM)} \end{array} \qquad (9.88)$$

Perpendicular to the plane of incidence (TE):
When $\vec{E}_{0,i}$ is perpendicular to the plane of incidence, it must be parallel to $\widehat{u}$ because $\widehat{u}$ defines the normal to that plane (again, up to a sign choice.) Now $\vec{H}_{0,i}$ is in the plane of incidence. We have

$$\vec{E}_{0,i} = \widetilde{E}_{0,i}\widehat{u} \tag{9.89}$$

We again pick the convention so that all three electric field vectors align for normal incidence. With these choices, we then have (again, $\vec{H} = \widehat{k} \times \vec{E}/Z$):

$$\vec{E}_{0,r} = \widetilde{E}_{0,r}\widehat{u} \tag{9.90}$$

$$\vec{E}_{0,t} = \widetilde{E}_{0,t}\widehat{u} \tag{9.91}$$

$$\vec{H}_{0,i} = \frac{\widetilde{E}_{0,i}}{Z_1}\left(-\widehat{w}\cos\theta_i + \widehat{n}\sin\theta_i\right) \tag{9.92}$$

$$\vec{H}_{0,r} = \frac{\widetilde{E}_{0,r}}{Z_1}\left(\widehat{w}\cos\theta_r + \widehat{n}\sin\theta_r\right) \tag{9.93}$$

$$\vec{H}_{0,t} = \frac{\widetilde{E}_{0,t}}{Z_2}\left(-\widehat{w}\cos\theta_t + \widehat{n}\sin\theta_t\right) \tag{9.94}$$
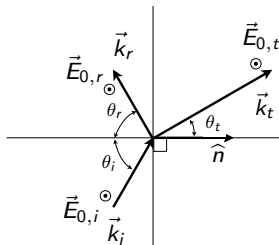
Restricting to the boundary conditions with information, we obtain

$$\widehat{n}: \qquad \frac{\mu_1}{Z_1}\left(\widetilde{E}_{0,i}\,\sin\theta_i + \widetilde{E}_{0,r}\,\sin\theta_r\right) = \frac{\mu_2}{Z_2}\,\widetilde{E}_{0,t}\,\sin\theta_t \qquad (9.95)$$

$$\widehat{w}: \qquad \frac{1}{Z_1}\left(\widetilde{E}_{0,i}\,\cos\theta_i - \widetilde{E}_{0,r}\,\cos\theta_r\right) = \frac{1}{Z_2}\,\widetilde{E}_{0,t}\,\cos\theta_t \qquad (9.96)$$

$$\widehat{u}: \qquad\qquad\qquad \widetilde{E}_{0,i} + \widetilde{E}_{0,r} = \widetilde{E}_{0,t} \qquad (9.97)$$

Again, we only need two of the equations, so we use the latter two because the $\widehat{n}$ equation can be reduced to the $\widehat{u}$ equation via Snell's Law. We solve to obtain Fresnel's Equations perpendicular to the plane of incidence (TE):

$$\boxed{\frac{\widetilde{E}_{0,r}}{\widetilde{E}_{0,i}} = \left(\frac{1-\alpha\,\beta}{1+\alpha\,\beta}\right) \qquad \frac{\widetilde{E}_{0,t}}{\widetilde{E}_{0,i}} = \left(\frac{2}{1+\alpha\,\beta}\right) \qquad \begin{array}{l}\textit{Fresnel's}\\\textit{Equations}\\\textit{perpendicular}\\\textit{to the plane of}\\\textit{incidence (TE)}\end{array}} \qquad (9.98)$$

with α and β as defined earlier.

Reflected and Transmitted Energy

The energy flux (intensity) at a particular point is

$$I_j = \left\langle |\vec{S}_j| \right\rangle = \frac{1}{2}\, \epsilon_j\, v_j\, E_j^2 \cos\theta_j = \frac{1}{2}\, \frac{c}{Z_j}\, E_j^2 \cos\theta_j \tag{9.99}$$

We can calculate from this the reflected and transmitted energy or power ratios:

$$\mathcal{R} = \frac{I_r}{I_i} = \left(\frac{\widetilde{E}_{0,r}}{\widetilde{E}_{0,i}}\right)^2 = \left(\frac{\alpha - \beta}{\alpha + \beta}\right)^2 \qquad \text{parallel (TM)} \tag{9.100}$$

$$= \left(\frac{1 - \alpha\,\beta}{1 + \alpha\,\beta}\right)^2 \qquad \text{perpendicular (TE)} \tag{9.101}$$

$$\mathcal{T} = \frac{I_t}{I_i} = \frac{Z_1}{Z_2}\left(\frac{\widetilde{E}_{0,t}}{\widetilde{E}_{0,i}}\right)^2 \frac{\cos\theta_t}{\cos\theta_i} = \alpha\,\beta \left(\frac{2}{\alpha + \beta}\right)^2 \qquad \text{parallel (TM)} \tag{9.102}$$

$$= \alpha\,\beta \left(\frac{2}{1 + \alpha\,\beta}\right)^2 \qquad \text{perpendicular (TE)} \tag{9.103}$$

By calculating $\mathcal{R} + \mathcal{T}$ explicitly, one can see that $\mathcal{R} + \mathcal{T} = 1$ always in both cases. Notice the $\alpha\,\beta$ prefactor for $\mathcal{T}$.

Normal Incidence

Let's summarize the results for normal incidence, $\theta_i = \theta_r = \theta_t = 0$, for which things simplify substantially: $\alpha = 1$ and thus the TM and TE cases are equivalent, yielding:

$$\frac{\widetilde{E}_{0,r}}{\widetilde{E}_{0,i}} = \frac{1-\beta}{1+\beta} = \frac{1-\frac{Z_1}{Z_2}}{1+\frac{Z_1}{Z_2}} \qquad \mathcal{R} = \left(\frac{1-\beta}{1+\beta}\right)^2 = \left(\frac{1-\frac{Z_1}{Z_2}}{1+\frac{Z_1}{Z_2}}\right)^2 \qquad (9.104)$$

$$\frac{\widetilde{E}_{0,t}}{\widetilde{E}_{0,i}} = \frac{2}{1+\beta} = \frac{2}{1+\frac{Z_1}{Z_2}} \qquad \mathcal{T} = \beta\left(\frac{2}{1+\beta}\right)^2 = \frac{Z_1}{Z_2}\left(\frac{2}{1+\frac{Z_1}{Z_2}}\right)^2 \qquad (9.105)$$

We will see that we get similar equations for transmission lines and waveguides. The concept of "impedance matching" becomes apparent: when the wave impedances of the two media match, then $\mathcal{R} = 0$.

For the special but frequently encountered case $\mu_1 = \mu_2 = \mu_0$, $\beta = Z_1/Z_2$ reduces to $n_2/n_1 = v_1/v_2$ and we recover the results for a wave on a string from Ph2/12:

$$\frac{\widetilde{E}_{0,r}}{\widetilde{E}_{0,i}} = \frac{n_1 - n_2}{n_1 + n_2} = \frac{v_2 - v_1}{v_2 + v_1} \qquad \mathcal{R} = \left(\frac{n_2 - n_1}{n_2 + n_1}\right)^2 = \left(\frac{v_2 - v_1}{v_2 + v_1}\right)^2 \qquad (9.106)$$

$$\frac{\widetilde{E}_{0,t}}{\widetilde{E}_{0,i}} = \frac{2n_1}{n_1 + n_2} = \frac{2v_2}{v_2 + v_1} \qquad \mathcal{T} = \frac{n_2}{n_1}\left(\frac{2n_1}{n_1 + n_2}\right)^2 = \frac{v_1}{v_2}\left(\frac{2v_2}{v_2 + v_1}\right)^2 \qquad (9.107)$$

Typical Behavior

We first show the simplest case $\mu_1/\mu_2 = 1$, which implies $n_2/n_1 = (Z_2/Z_1)^{-1} = \beta$.

Notice how, for the $n_2/n_1 < 1$ case, the transmitted field amplitude exceeds unity but the transmitted power $\mathcal{T}$ does not and, in particular, how the transmitted field amplitude appears to diverge but the transmitted energy vanishes at $\theta_{TIR} = \sin^{-1} n_2/n_1$. These behaviors are simultaneously possible because of the $\alpha\beta$ pre-factor in $\mathcal{T}$. For both cases of n_2/n_1, notice also the nonmonotonic behavior in $\mathcal{R}$ and $\mathcal{T}$ near the Brewster Angle

(θ_B, to be defined later) for the parallel incidence case: this behavior occurs because the reflected field, while monotonic, passes through zero, requiring the squared quantity $\mathcal{R}$ to be positive, zero, and then positive again. Note also that the transmitted field does not pass through unity magnitude even though $\mathcal{T}$ does. Both this behavior, and $\mathcal{T}$'s non-monoticity, occur because of the α prefactor, which starts at 1 at normal incidence and then grows with θ_i.
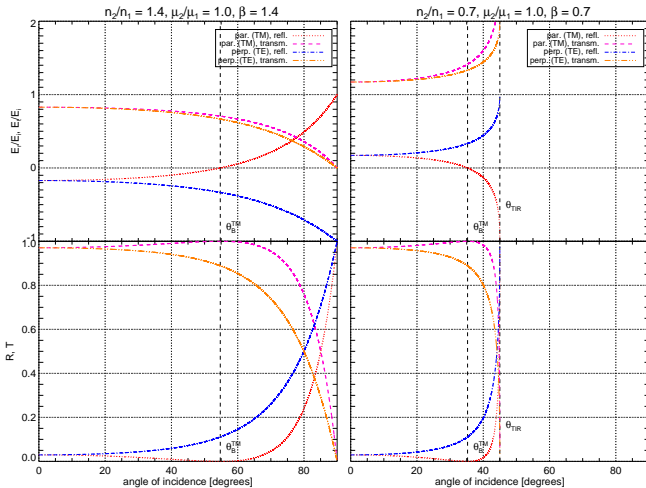
Now we allow $\mu_2/\mu_1 \neq 1$. In general, $\beta = Z_1/Z_2 = (n_2/n_1)/(\mu_2/\mu_1)$, so we first take a value for $n_2/n_1 > 1$ and consider different cases for μ_2/μ_1 such that all three cases for $\beta - 1$ are explored, and then we do the same for $n_2/n_1 < 1$. In the $\beta = 1.0$ case $(\mu_2/\mu_1 = n_2/n_1)$, we have impedance matching and thus unity transmission at normal incidence, but the deviation of μ_2/μ_1 from unity results in nontrivial behavior away from normal incidence $(\alpha \neq 1)$. Notice how $\theta_B^{TM} = \theta_B^{TE} = 0$ for $\beta = 1.0$ and how the TE and TM behavior in energy are identical (though not in field). There is no θ_{TIR} because $n_2/n_1 > 1$.

$n_2/n_1 = 1.4$, $\mu_2/\mu_1 = 1.4$, $\beta = 1.0$
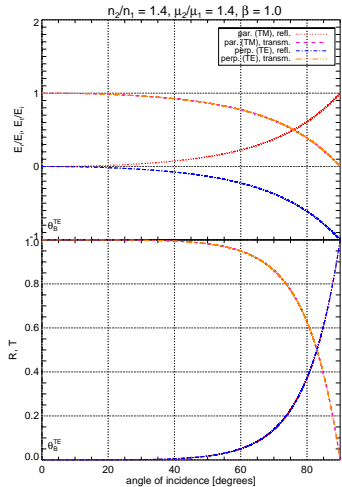
We now consider values of μ_2/μ_1 that change the sign of $\beta - 1$ for the chosen value of n_2/n_1. θ_B^{TE} appears for $\beta = Z_1/Z_2 < 1$ and θ_B^{TM} for $\beta > 1$. Again, we have no θ_{TIR} because $n_2/n_1 > 1$.
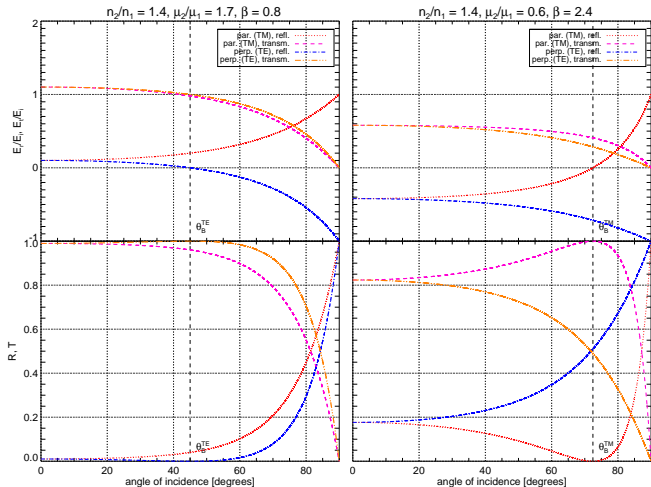
In this case, we hold the sign of $\beta - 1$ fixed but allow the sign of $\mu_2/\mu_1 - 1$ to change. We see it is not μ_2/μ_1 that matters but rather it is $\beta = Z_1/Z_2$.
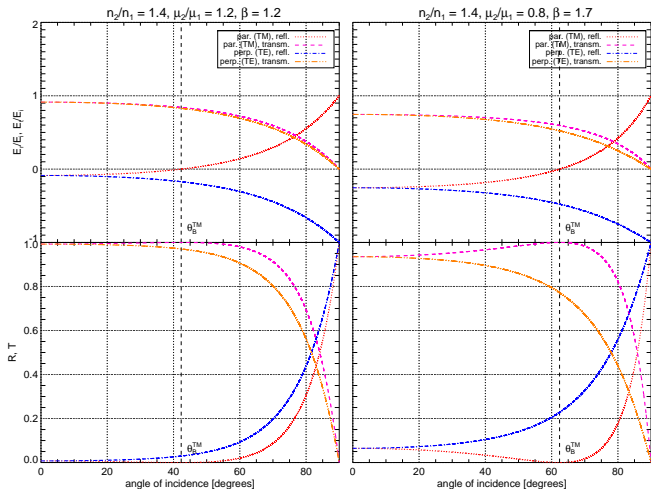
We now take $n_2/n_1 < 1$, which yields a θ_{TIR} for these and the remaining cases. Here, we again consider the $\beta = 1$ special case, for which we again see $\theta_B^{TM} = \theta_B^{TE} = 0$.

$n_2/n_1 = 0.7$, $\mu_2/\mu_1 = 0.7$, $\beta = 1.0$
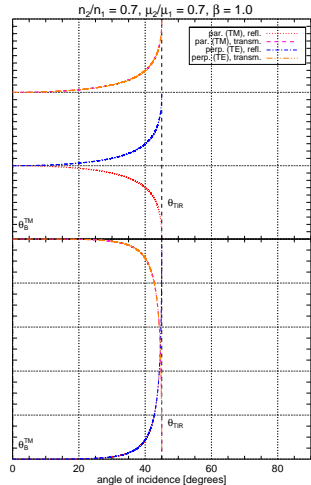
We now change the sign of $\mu_2/\mu_1 - 1$ to change the sign of $\beta - 1$, yielding θ_B^{TM} for $\beta < 1$ and θ_B^{TE} for $\beta > 1$.
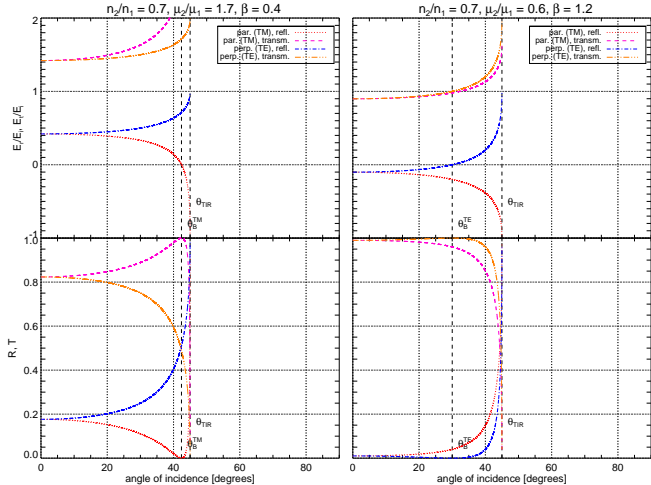
Here we again see how β is more important than μ_2/μ_1: we change the sign of $\mu_2/\mu_1 - 1$ while holding the sign of $\beta - 1$ fixed and we see the behavior is qualitatively unchanged.
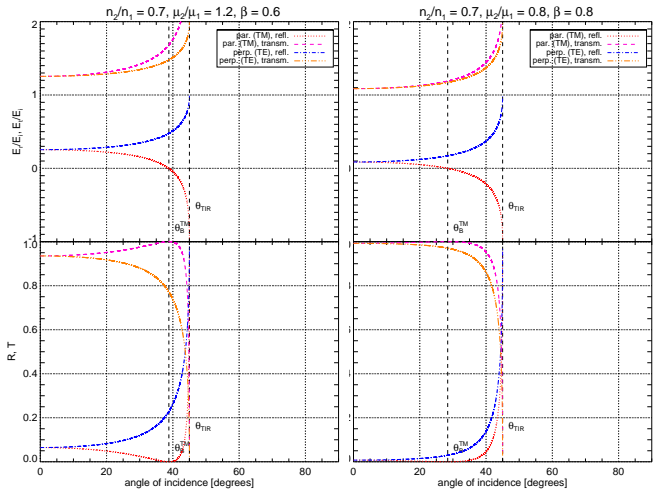
Understanding the Signs and Magnitudes of the Transmitted and Reflected Fields

▶ *Sign of transmitted wave*
To understand the sign of the transmitted wave, we just need to notice that α and β are always positive numbers. α is always positive because θ_i and θ_r are restricted to the first quadrant. Therefore, all the quantities in the expressions for $\widetilde{E}_{0,t}$ are positive, and thus the transmitted wave always has the same sign electric field as the incident wave.

Since $\widehat{k}_i \cdot \widehat{n}$ and $\widehat{k}_t \cdot \widehat{n}$ have the same sign and $\vec{H} \propto \widehat{k} \times \vec{E}$, we may also conclude that the sign of the magnetic field of the transmitted wave is unchanged.

▶ *Sign of electric field of reflected wave (general case) and Brewster's Angle*
The sign of the reflected wave depends on the sizes of α and β. What general statements can we make?

We know that, for a given pair of media, either $\sin\theta_t / \sin\theta_i < 1$ or $\sin\theta_t / \sin\theta_i > 1$ is true for all angles because this ratio is set by Snell's Law, $\sin\theta_t / \sin\theta_i = n_1/n_2$. Since sin and cos are monotonic over the first quadrant, we can conclude that $\alpha = \cos\theta_t / \cos\theta_i$ is also either smaller than or greater than 1 for all angles, with the case being determined by n_1/n_2:

$$\frac{n_1}{n_2} > 1 \iff \frac{\sin\theta_t}{\sin\theta_i} > 1 \iff \frac{\cos\theta_t}{\cos\theta_i} < 1 \iff \alpha < 1 \quad (9.108)$$

$$\frac{n_1}{n_2} < 1 \iff \frac{\sin\theta_t}{\sin\theta_i} < 1 \iff \frac{\cos\theta_t}{\cos\theta_i} > 1 \iff \alpha > 1 \quad (9.109)$$

However, $\beta = Z_1/Z_2 = (n_2/n_1)/(\mu_2/\mu_1)$, so the size of n_2/n_1 relative to unity does not completely determine the size of β relative to unity. No generic statement can be made about the relative size of α, β, and $1/\beta$.

That said, we can make conditional statements that depend on whether we are considering the parallel (TM) or perpendicular (TE) incidence cases and whether β is smaller or larger than unity. We break this up into two steps: what is the sign of the reflection at normal incidence, and whether that sign changes as the angle of incidence changes:

▶ Sign of reflected wave at normal incidence, $\theta_i = \theta_t = 0$ (parallel (TM) and perpendicular (TE) incidence degenerate at this angle)
   The sign of the denominator of the expression for the reflected amplitude is always positive, and $\alpha = 1$ because $\theta_i = \theta_t = 0$ at normal incidence, so the sign of the reflection at normal incidence is set by the sign of the numerator $1 - \beta$, which is set by the ratio of the wave impedances:

$$ \text{if} \quad \beta = \frac{n_2/n_1}{\mu_2/\mu_1} = \frac{Z_1}{Z_2} > 1 \quad \implies \quad \text{sign}\left(\frac{\widetilde{E}_{0,r}}{\widetilde{E}_{0,i}}\right)_{\theta_i=0} < 0 \quad (9.110) $$

$$ \text{if} \quad \beta = \frac{n_2/n_1}{\mu_2/\mu_1} = \frac{Z_1}{Z_2} < 1 \quad \implies \quad \text{sign}\left(\frac{\widetilde{E}_{0,r}}{\widetilde{E}_{0,i}}\right)_{\theta_i=0} > 0 \quad (9.111) $$

▶ Does the sign of the reflected wave change with angle of incidence?
   We can answer this question by determining whether there is an angle at which zero reflection occurs. If so, the sign of the reflected wave will change from its sign at normal incidence to the opposite sign. Whether there is such an angle depends on the relative sizes of α and β. We will call this angle *Brewster's Angle*, though that nomenclature generally only applies for the case $\mu_1 = \mu_2 = \mu_o$, which we will consider shortly.

▶ *Parallel incidence (TM)*
The condition for zero reflection is $\alpha = \beta$. If we square this
requirement, use a trigonometric identity, and use Snell's law, we
find

$$\frac{1 - \frac{n_1^2}{n_2^2} \sin^2 \theta_B}{1 - \sin^2 \theta_B} = \beta^2 \quad \Longleftrightarrow \quad \sin^2 \theta_B = \frac{\beta^2 - 1}{\beta^2 - \frac{n_1^2}{n_2^2}} \qquad (9.112)$$

It is clear that the condition is driven by $\beta = Z_1/Z_2$, the ratio of the
wave impedances, and n_1/n_2, the ratio of the indices of refraction.
We have to consider multiple cases to determine when there is a
solution for θ_B, which can only happen if the above expression takes
on a value between 0 (numerator and denominator have same sign)
and 1 (numerator smaller in magnitude than denominator):

$$\text{if} \quad \beta = \frac{n_2/n_1}{\mu_2/\mu_1} = \frac{Z_1}{Z_2} > 1 \implies \text{need } \frac{n_2}{n_1} > 1, \ \frac{n_2}{n_1} > \frac{\mu_2}{\mu_1} \quad (9.113)$$

$$\text{if} \quad \beta = \frac{n_2/n_1}{\mu_2/\mu_1} = \frac{Z_1}{Z_2} < 1 \implies \text{need } \frac{n_2}{n_1} < 1, \ \frac{n_2}{n_1} < \frac{\mu_2}{\mu_1} \quad (9.114)$$

► *Perpendicular incidence (TE)*
   The condition for zero reflection is $\alpha = 1/\beta$. Repeating the same
   process to find a solution, we have

$$\frac{1 - \frac{n_1^2}{n_2^2}\sin^2\theta_B}{1 - \sin^2\theta_B} = \frac{1}{\beta^2} \quad \Longleftrightarrow \quad \sin^2\theta_B = \frac{\frac{1}{\beta^2} - 1}{\frac{1}{\beta^2} - \frac{n_1^2}{n_2^2}} \qquad (9.115)$$

This is basically just the converse of the TM case, so the same set
of requirements for a solution for θ_B implies

$$\text{if} \quad \beta = \frac{n_2/n_1}{\mu_2/\mu_1} = \frac{Z_1}{Z_2} > 1 \implies \text{need } \frac{n_2}{n_1} < 1,\ \frac{n_2}{n_1} > \frac{\mu_2}{\mu_1} \quad (9.116)$$

$$\text{if} \quad \beta = \frac{n_2/n_1}{\mu_2/\mu_1} = \frac{Z_1}{Z_2} < 1 \implies \text{need } \frac{n_2}{n_1} > 1,\ \frac{n_2}{n_1} < \frac{\mu_2}{\mu_1} \quad (9.117)$$

Note how the pairing of the two n_2/n_1 conditions changes between the
TM and TE cases. The above behavior was seen in the plots prior to this
section.

▶ *Sign of electric field of reflected wave, and Brewster's Angle, for $\mu_1 = \mu_2$*
  This case is typical for everyday experience; there are few light-transmitting yet
  also $\mu \neq \mu_0$ materials. We can be much more specific in this case because now
  $\beta = Z_1/Z_2 = n_2/n_1$ and there is a clear relationship between α and β, which we
  may now rewrite as

$$\frac{n_1}{n_2} > 1 \quad \Longleftrightarrow \quad \beta = \frac{n_2}{n_1} < 1 \quad \text{and} \quad \alpha = \frac{\cos\theta_t}{\cos\theta_i} < 1 \qquad (9.118)$$

$$\frac{n_1}{n_2} < 1 \quad \Longleftrightarrow \quad \beta = \frac{n_2}{n_1} > 1 \quad \text{and} \quad \alpha = \frac{\cos\theta_t}{\cos\theta_i} > 1 \qquad (9.119)$$

Considering the two cases separately:

  ▶ *Parallel incidence (TM), $\mu_1 = \mu_2$*
    Since α and β are both either < 1 or > 1, it is possible for $\alpha = \beta$ to be
    true and therefore the sign may depend on the angle. At $\theta_i = \theta_t = 0$, we
    have $\alpha = 1$ identically, so the sign of the reflected wave is $1 - \beta$, which is
    positive if $\beta < 1$ and negative if $\beta > 1$. This sets the polarity of the
    reflected wave for small θ_i.

    To understand the polarity of the reflected wave for larger $\theta_i \neq 0$, we then
    have to ask whether it is possible for the polarity of $\alpha - \beta$ to change as α
    varies with θ_i, which would happen if there is a zero in $\alpha - \beta$.

To solve for this angle where $\alpha = \beta$ (*Brewster's Angle* again, θ_B), we can square the ratio of cosines so it can be written in terms of sines, then use Snell's Law to obtain:

$$\frac{1 - \frac{n_1^2}{n_2^2} \sin^2 \theta_B}{1 - \sin^2 \theta_B} = \frac{n_2^2}{n_1^2} \quad \Longleftrightarrow \quad \sin^2 \theta_B = \frac{n_2^2}{n_1^2 + n_2^2} \tag{9.120}$$

$$\Longleftrightarrow \quad \boxed{\tan \theta_B = \frac{n_2}{n_1} = \beta = \frac{Z_1}{Z_2} \quad \begin{array}{l} \textit{Brewster's} \\ \textit{Angle for} \\ \mu_1 = \mu_2 \end{array}} \tag{9.121}$$

Note that $0 < \theta_B < \pi/4$ for $n_2 < n_1$ and $\pi/4 < \theta_B < \pi/2$ for $n_2 > n_1$. Therefore, we may summarize this case as (using $1 - \beta = (n_1 - n_2)/n_1$):

sign of reflected wave for parallel incidence (TM) and $\mu_1 = \mu_2$:

$$0 < \theta_i < \theta_B : \quad \text{sign} \left(\frac{\widetilde{E}_{0,r}}{\widetilde{E}_{0,i}} \right) = -\text{sign} \left(\frac{n_2}{n_1} - 1 \right) \tag{9.122}$$

$$\theta_B < \theta_i < \theta_{max} : \quad \text{sign} \left(\frac{\widetilde{E}_{0,r}}{\widetilde{E}_{0,i}} \right) = \text{sign} \left(\frac{n_2}{n_1} - 1 \right) \tag{9.123}$$

$$\theta_{max} = \sin^{-1} \left(\min \left(1, \frac{n_2}{n_1} \right) \right) \tag{9.124}$$

▶ *Perpendicular incidence (TE),* $\mu_1 = \mu_2$

The perpendicular incidence case is easier to analyze. At normal incidence, $\alpha = 1$ again and the formulae become identical to the parallel incidence case, giving us the same behavior: the reflected wave is positive if $\beta < 1$ and negative if $\beta > 1$. This common behavior must hold, as the parallel and perpendicular cases are degenerate for normal incidence.

To see if there is an angle at which the sign of the reflected wave can flip, we need to know if there is an angle at which the reflected wave vanishes. It is easy to see there is not: for $1 - \alpha\beta$ to vanish, we require $\alpha = 1/\beta$. For the case $\mu_1 = \mu_2$, we thus require $\alpha = n_1/n_2$. But we saw above that, when $1/\beta = n_1/n_2 > 1$ we have $\alpha < 1$ and when $1/\beta = n_1/n_2 < 1$ we have $\alpha > 1$. Thus, there is never a zero in the reflected wave, and the sign can never flip.

Therefore, we may summarize this case as:

---

sign of reflected wave for perpendicular incidence (TE) and $\mu_1 = \mu_2$:

$$0 < \theta_i < \theta_{max} : \quad \text{sign}\left(\frac{\widetilde{E}_{0,r}}{\widetilde{E}_{0,i}}\right) = -\text{sign}\left(\frac{n_2}{n_1} - 1\right) \quad (9.125)$$

$$\theta_{max} = \sin^{-1}\left(\min\left(1, \frac{n_2}{n_1}\right)\right) \quad (9.126)$$

► *Practical Implications of Brewster's Angle*
  Because the reflected amplitude for parallel incidence goes through a zero
  at θ_B, it is small for angles near θ_B. For the everyday materials water
  ($n = 1.33$) and glass ($n \approx 1.5$), this angle is $53°$ and $56°$, which is a
  typical viewing angle. This explains why polarized sunglasses reduce glare:
  by blocking the reflected polarization that is perpendicular to the plane of
  incidence (TE), they block the only component of the reflected wave that
  has appreciable amplitude. They are designed to pass the parallel
  component because it has no reflection near Brewster's angle. Be sure to
  get the orientation right: if the interface is a horizontal surface, then that
  plane is actually parallel to the surface for light coming from above, so
  the plane of parallel incidence is vertical for the viewer.

▶ *Relation between sign of reflected wave and possibility of total internal reflection*
Both sets of equations imply that total internal reflection occurs only if $n_2 < n_1$, which also implies that the sign of the reflected wave at normal incidence is equal to the sign of the incident wave. Thus, the two conditions are equivalent:

no sign flip at normal incidence $\iff$ total internal reflection possible

▶ *Sign of magnetic field of reflected wave*
For any of these cases, we can obtain the sign of the *magnetic field* of the reflected wave by applying the rule $\vec{H} \propto \hat{k} \times \vec{E}$ to the reflected wave. From this, and from the sign flip of $\hat{k}_r \cdot \hat{n}$ relative to $\hat{k}_i \cdot \hat{n}$, we can conclude that the magnetic field of the reflected wave has the opposite behavior as the electric field in both the parallel and perpendicular cases: if the electric field receives a sign flip, the magnetic field does not, and vice versa.

Lecture 26:

Electromagnetic Waves IV:
Electromagnetic Waves in Conducting Matter

Date Revised: 2024/04/25 06:00
Revised lecture break
Date Given: 2024/04/23

Electromagnetic Waves in Conducting Matter

Maxwell's Equations and the Wave Equation for Conductors; Plane-Wave Solutions

The primary distinction between a conducting medium and a nonconducting medium is that Ohm's Law is now obeyed, $\vec{J}_f = \sigma \vec{E}$. Incorporating it into Maxwell's Equations for a linear medium gives

$$\vec{\nabla} \cdot \vec{E} = \frac{\rho_f}{\epsilon} \qquad \vec{\nabla} \cdot \vec{B} = 0 \qquad \vec{\nabla} \times \vec{E} + \frac{\partial \vec{B}}{\partial t} = 0 \qquad \vec{\nabla} \times \vec{B} - \mu \sigma \vec{E} - \epsilon \mu \frac{\partial \vec{E}}{\partial t} = 0$$
$$(9.127)$$

Do we need to worry about free charge? No. If we combine the continuity equation, $\vec{\nabla} \cdot \vec{J}_f = -\partial \rho_f / \partial t$ with Gauss's Law, we obtain

$$\frac{\partial \rho_f}{\partial t} = -\sigma \left(\vec{\nabla} \cdot \vec{E} \right) = -\frac{\sigma}{\epsilon} \rho_f \qquad (9.128)$$

The solution to this equation (with $\sigma / \epsilon > 0$) is a decaying exponential in time, $\rho_f(t) = \rho_f(t = 0) \exp(-t/\tau)$ with $\tau = \epsilon / \sigma$. Thus, *any free charge decays away*. The time constant for the decay as compared to the relevant field variation timescale (the wave oscillation period $T = 2\pi/\omega$) is a measure of how good the conductor is at the frequencies of interest: for a good conductor, one must have $\tau \ll T$ (or, equivalently, $\omega \tau \ll 1$, neglecting the 2π). Regardless, the appearance of any free charge is transient, while we will see the wave, though decaying in space, is stable in time.

Assuming this condition is satisfied, we then have again a homogeneous set of Maxwell's Equations:

$$\vec{\nabla} \cdot \vec{E} = 0 \qquad \vec{\nabla} \cdot \vec{B} = 0 \qquad \vec{\nabla} \times \vec{E} + \frac{\partial \vec{B}}{\partial t} = 0 \qquad \vec{\nabla} \times \vec{B} - \mu\,\sigma\,\vec{E} - \epsilon\,\mu\,\frac{\partial \vec{E}}{\partial t} = 0 \tag{9.129}$$

We may do as before to obtain wave equations:

$$\nabla^2 \vec{E} = \epsilon\,\mu\,\frac{\partial^2 \vec{E}}{\partial t^2} + \sigma\,\mu\,\frac{\partial \vec{E}}{\partial t} \qquad \nabla^2 \vec{B} = \epsilon\,\mu\,\frac{\partial^2 \vec{B}}{\partial t^2} + \sigma\,\mu\,\frac{\partial \vec{B}}{\partial t} \tag{9.130}$$

As we did for nonconducting matter, we assume plane wave solutions because any solution can be built from them by linearity (now using complex notation instead of real sinusoids for convenience, as we will see):

$$\vec{E}(\vec{r}, t) = \vec{E}_0\,e^{i(\vec{k}\cdot\vec{r} - \omega\,t)} \qquad \vec{B}(\vec{r}, t) = \vec{B}_0\,e^{i(\vec{k}\cdot\vec{r} - \omega\,t)} \tag{9.131}$$

We dispense with repeating the proof that $\vec{k}$ and ω are the same for $\vec{E}$ and $\vec{B}$ (the proof is similar in form) and simply assume it and check it below. We determine later the relation between $\vec{E}_0$ and $\vec{B}_0$ (including any possible δ, which they absorb when using complex notation).

Plugging in, one obtains an equation for $\vec{k} \cdot \vec{k}$ (which we do not write as $|\vec{k}|^2$ because the right-hand side is complex! It was real for non-conducting matter):

$$\vec{k} \cdot \vec{k} = \epsilon \mu \omega^2 + i \sigma \mu \omega \tag{9.132}$$

We take the square root, yielding

$$\sqrt{\vec{k} \cdot \vec{k}} \equiv k + i \kappa \qquad v_{\epsilon\mu} = \frac{1}{\sqrt{\epsilon\mu}} \qquad k_{\epsilon\mu} = \frac{\omega}{v_{\epsilon\mu}} \qquad \lambda_{\epsilon\mu} = \frac{2\pi}{k_{\epsilon\mu}} \qquad \tau = \frac{\epsilon}{\sigma} \tag{9.133}$$

$$k = k_{\epsilon\mu} \left[\frac{\sqrt{1 + \frac{1}{\omega^2 \tau^2}} + 1}{2} \right]^{1/2} \qquad \kappa = k_{\epsilon\mu} \left[\frac{\sqrt{1 + \frac{1}{\omega^2 \tau^2}} - 1}{2} \right]^{1/2} \equiv \frac{1}{\delta} \tag{9.134}$$

The wave thus takes the form

$$\vec{E}(\vec{r}, t) = \vec{E}_0 \, e^{-(\widehat{k} \cdot \vec{r}/\delta)} e^{i(k\,\widehat{k} \cdot \vec{r} - \omega t)} \qquad \vec{B}(\vec{r}, t) = \vec{B}_0 \, e^{-(\widehat{k} \cdot \vec{r}/\delta)} e^{i(k\,\widehat{k} \cdot \vec{r} - \omega t)} \tag{9.135}$$

which is a plane wave propagating in the direction $\widehat{k}$ with propagation constant k, wavelength $\lambda = 2\pi/k$, speed $v = \omega/k = \lambda\nu$, and index of refraction $n = c/v$, but with decaying amplitude. The decay length is the *skin depth*, $\delta = 1/\kappa$. Notice that the time dependence is unchanged and still given by ω and ν while the spatial dependence has been modified. Note also that, while the wave decays in space, it is oscillatory (stable in amplitude) in time.

Let's consider two limits, the *poor conductor limit* and the *good conductor limit*:

▶ **Poor conductor limit**
This is the limit $\omega \tau \gg 1$. (Note that we have to wait many wave periods after the appearance of the wave in order for the free charge to decay away because $\omega \tau \gg 1$ ($\tau \gg T/2\pi$), but we do reach $\rho_f = 0$ if we wait this long.) The second term under the inner square root is small and one can Taylor expand to obtain

$$k \overset{\omega\tau \gg 1}{\longrightarrow} \frac{k_{\epsilon\mu}}{\sqrt{2}} \left[\left(1 + \frac{1}{2}\, \frac{1}{\omega^2\, \tau^2} \right) + 1 \right]^{1/2} \approx k_{\epsilon\mu} \tag{9.136}$$

$$\kappa \overset{\omega\tau \gg 1}{\longrightarrow} \frac{\omega}{v_{\epsilon\mu}\, \sqrt{2}} \left[\left(1 + \frac{1}{2}\, \frac{1}{\omega^2\, \tau^2} \right) - 1 \right]^{1/2} \approx \frac{k_{\epsilon\mu}}{2\, \omega\, \tau} = \frac{\sigma}{2} \sqrt{\frac{\mu}{\epsilon}} = \frac{Z_{\epsilon\mu}\, \sigma}{2} \tag{9.137}$$

$$\delta \overset{\omega\tau \gg 1}{\longrightarrow} \frac{2}{\sigma} \sqrt{\frac{\epsilon}{\mu}} = \frac{2}{Z_{\epsilon\mu}\, \sigma} = \frac{2\,\rho}{Z_{\epsilon\mu}} \tag{9.138}$$

In this case, the wave propagation speed and wavelength are the same as in a nonconducting medium. The form for the skin depth is particularly interesting because it involves the ratio of the resistivity and the wave impedance, which connect to the Ohmic and displacement currents, respectively. We can write the skin depth another way, now using the wavelength in the medium:

$$\delta \overset{\omega\tau \gg 1}{\longrightarrow} \frac{2\,\omega\,\tau}{k_{\epsilon\mu}} = \frac{2\,\tau}{T}\, \lambda_{\epsilon\mu} \gg \lambda_{\epsilon\mu} \tag{9.139}$$

$2\,\tau/T \gg 1$ counts the number of wavelengths over which the amplitude decays: in a poor conductor, the decay of the wave happens over many wavelengths.

▶ **Good Conductor Limit**

Similarly, we can take the *good conductor* limit, $\omega\,\tau \ll 1$, but here without the caveat on waiting for the free charge density to decay. Here, the $\frac{1}{\omega^2\,\tau^2}$ term dominates the square root, so k and κ converge to the same value:

$$k, \kappa \overset{\omega\tau \ll 1}{\longrightarrow} \frac{\omega}{v_{\epsilon\mu}} \left[\frac{1}{2}\sqrt{\frac{1}{\omega^2\,\tau^2}}\right]^{1/2} = \sqrt{\frac{\omega}{2\,v_{\epsilon\mu}^2\,\tau}} = \sqrt{\frac{\mu\,\sigma\,\omega}{2}} \qquad (9.140)$$

$$\delta \overset{\omega\tau \ll 1}{\longrightarrow} \sqrt{\frac{2}{\mu\,\sigma\,\omega}} \qquad (9.141)$$

There is no propagating wave because the decay constant and the propagation constant are the same; put another way, $\delta = 1/\kappa = 1/k = \lambda/2\,\pi$, indicating the wave decays in about $1/6$ of a wavelength!

Notice how ϵ becomes irrelevant for a good conductor. We will see that $Z_{\epsilon\mu}$ is effectively replaced by $Z_\sigma \sim 1/\sigma\,\delta$ so $\delta \sim 1/Z_\sigma\,\sigma$, a form similar to the poor conductor case. However, $Z_\sigma \ll Z_{\epsilon\mu}$ for a good conductor, so $H \sim E/Z$ is much larger than it would be for the same E in a poor conductor. This expresses the fact that, in a good conductor, the Ohmic currents dominate over the displacement current.

Motivated by the above, we can rewrite the important parameter $1/\omega\,\tau$ as $\sigma/\epsilon\,\omega$. Referring back to Equations 9.130 and 9.132, we see that this quantity is the ratio of the Ohmic current to the displacement current. One can consider $\epsilon\,\omega$ to the "effective conductivity" for the displacement current or σ/ω to be the "effective permittivity" for the Ohmic current. The latter will recur when we consider a dispersive matter model for conductors. We will also return to dominance of Ohmic over displacement current in good conductors.

General Results

Coming back to our general plane-wave solution, we write (with $\widehat{n}$ defining the polarization of $\vec{E}$)

$$\vec{E}(\vec{r}, t) = \widetilde{E}_0 \,\widehat{n}\, e^{-\left(\widehat{k}\cdot\vec{r}/\delta\right)} \, e^{i\left(k\,\widehat{k}\cdot\vec{r}-\omega\,t\right)} \tag{9.142}$$

and we use the $\vec{\nabla} \times \vec{E}$ equation to find

$$\vec{B}(\vec{r}, t) = \frac{k + i\,\kappa}{\omega} \,\widetilde{E}_0 \,\widehat{k} \times \widehat{n}\, e^{-\left(\widehat{k}\cdot\vec{r}/\delta\right)} \, e^{i\left(k\,\widehat{k}\cdot\vec{r}-\omega\,t\right)} \tag{9.143}$$

We may write

$$k + i\,\kappa = K\,e^{i\,\phi} \qquad K = \sqrt{k^2 + \kappa^2} = k_{\epsilon\mu}\left[\sqrt{1 + \left(\frac{\sigma}{\epsilon\,\omega}\right)^2}\right]^{1/2} \tag{9.144}$$

$$\tan\phi = \frac{\kappa}{k} = \left[\frac{\sqrt{1 + \left(\frac{\sigma}{\epsilon\,\omega}\right)^2} - 1}{\sqrt{1 + \left(\frac{\sigma}{\epsilon\,\omega}\right)^2} + 1}\right]^{1/2} \tag{9.145}$$

Motivated by the discussion on the last slide, we have rewritten $1/\omega\,\tau$ as $\sigma/\epsilon\,\omega$.

With the above definition, we can more simply write the relation between the phases and amplitudes of $\vec{E}$ and $\vec{B}$:

$$\frac{\widetilde{B}_0}{\widetilde{E}_0} = \frac{k + i\,\kappa}{\omega} = \frac{K\,e^{i\,\phi}}{k_{\epsilon\mu}\,v_{\epsilon\mu}} = \frac{e^{i\,\phi}}{v_{\epsilon\mu}}\left[1 + \left(\frac{\sigma}{\epsilon\,\omega}\right)^2\right]^{1/4} \qquad (9.146)$$

We see that there is now a phase shift between electric and magnetic fields, with the magnetic field lagging behind the electric field. The amplitudes are related in a somewhat complicated manner that depends on the wave velocity as before but now with a modification factor that depends on the conductivity.

It now also makes sense to introduce the relation between $\vec{H}$ and $\vec{E}$ using $\vec{H} = \vec{B}/\mu$:

$$\vec{H}(\vec{r}, t) = \frac{\vec{B}(\vec{r}, t)}{\mu} = \frac{k + i\,\kappa}{\omega\,\mu}\,\widetilde{E}_0\,\widehat{k}\times\widehat{n}\,e^{-(\widehat{k}\cdot\vec{r}/\delta)}\,e^{i(k\,\widehat{k}\cdot\vec{r} - \omega\,t)} \qquad (9.147)$$

$$\frac{\widetilde{H}_0}{\widetilde{E}_0} = \frac{1}{\mu}\frac{\widetilde{B}_0}{\widetilde{E}_0} = \frac{e^{i\,\phi}}{v_{\epsilon\mu}\,\mu}\left[1 + \left(\frac{\sigma}{\epsilon\,\omega}\right)^2\right]^{1/4} = \frac{e^{i\,\phi}}{Z_{\epsilon\mu}}\left[1 + \left(\frac{\sigma}{\epsilon\,\omega}\right)^2\right]^{1/4} \equiv \frac{1}{\widetilde{Z}} \qquad (9.148)$$

where we have defined a *complex wave impedance* $\widetilde{Z}$ for the conductor in the same way as we defined the wave impedance for a nonconductor. The auxiliary field acquires the same phase shift as the magnetic field except that it is now the complex wave impedance — the wave impedance with the same modification factor — that relates the electric and auxiliary fields.

Again, taking some limits:

$$poor\ conductor: \quad \frac{1}{\omega\tau} = \frac{\sigma}{\epsilon\,\omega} \ll 1 \quad \implies \quad \frac{\widetilde{B}_0}{\widetilde{E}_0} \to \frac{1}{v_{\epsilon\mu}}, \quad \frac{\widetilde{H}_0}{\widetilde{E}_0} \to \frac{1}{Z_{\epsilon\mu}} \quad (9.149)$$

$$good\ conductor: \quad \frac{1}{\omega\tau} = \frac{\sigma}{\epsilon\,\omega} \gg 1 \quad \implies \quad \frac{\widetilde{B}_0}{\widetilde{E}_0} \to \frac{e^{i\pi/4}}{v_{\epsilon\mu}}\sqrt{\frac{\sigma}{\epsilon\,\omega}} \gg \frac{1}{v_{\epsilon\mu}} \quad (9.150)$$

$$\frac{\widetilde{H}_0}{\widetilde{E}_0} \to \frac{e^{i\pi/4}}{Z_{\epsilon\mu}}\sqrt{\frac{\sigma}{\epsilon\,\omega}} \gg \frac{1}{Z_{\epsilon\mu}} \quad (9.151)$$

In a poor conductor, the relation approaches the nonconducting case (as we expect).

In contrast, in a good conductor, the magnetic and auxiliary fields are enhanced (because there are free currents to enhance the magnetic field but no free charges to enhance the electric field) and they lag the electric field by exactly $\pi/4$. Why $\pi/4$ and not $\pi/2$? The two current densities are $J_d = i\,\epsilon\,\omega E$ and $J_f = \sigma\,E$. So $|J_d| = \epsilon\,\omega/\sigma|J_f| \ll |J_f|$ in the good conductor limit, and thus the imaginary piece $i\,\sigma\,\omega$ dominates the right side of the wave equations, Equations 9.130, and therefore $\vec{k}\cdot\vec{k} \propto i = e^{i\,\pi/2}$. But it is $(\vec{k}\cdot\vec{k})^{1/2}/\omega = (k + i\,\kappa)/\omega$ that appears as the coefficient relating E to B (from Faraday's Law), so then $(k + i\,\kappa)/\omega \propto e^{i\,\pi/4}$ hence the $\pi/4$ phase shift.

In a poor conductor, the ohmic term is negligible and so we only get the piece due to $J_d = \epsilon\,\partial E/\partial t$. It is in phase in spite of the $\partial/\partial t$ because $\partial E/\partial t$ is related by Ampere's Law to spatial derivatives of B, which introduce a canceling factor of $i = e^{i\,\pi/2}$. This causes $\vec{k}\cdot\vec{k}$ to be real and positive, so no additional phase is introduced.

For the good conductor case, we can rewrite the relation between the electric and auxiliary fields in a more suggestive form if we recognize the skin depth is present:

$$\frac{1}{\widetilde{\overline{Z}}} = \frac{\widetilde{H}_0}{\widetilde{E}_0} = \frac{e^{i\pi/4}}{Z_{\epsilon\mu}} \sqrt{\frac{\sigma}{\epsilon\,\omega}} = e^{i\pi/4} \sqrt{\frac{\epsilon}{\mu}} \sqrt{\frac{\sigma}{\epsilon\,\omega}} \tag{9.152}$$

$$= e^{i\pi/4} \sqrt{\frac{\sigma^2}{2} \frac{2}{\sigma\,\mu\,\omega}} = \frac{1+i}{2}\,\sigma\,\delta \equiv \frac{1+i}{2}\,\frac{1}{Z_\sigma} \tag{9.153}$$

where $Z_\sigma = 1/\sigma\,\delta$ and ϵ has vanished from the expression! Why does the ratio have this form? In a good conductor, the fields, and thus the ohmic current density, decay to zero over the distance δ — the bulk of the current flows in a layer of thickness δ. If we imagine integrating this current over the skin depth to get an effective surface current density, we would multiply $J \approx \sigma\,E$ by δ to get $K \approx \sigma\,\delta E$ and thus $K/E \approx \sigma\delta = 1/Z_\sigma$. Recall that H has the same units as K, so it makes perfect sense that $H \sim K$ and the ratio $H/E \sim K/E \approx 1/Z_\sigma$. The prefactor is only present to take care of the phase issue, and it has (almost) unity magnitude. In fact, we can define the *surface impedance* of the conductor by

$$Z_s = \frac{2}{1+i}\,Z_\sigma = (1-i)\,Z_\sigma \qquad \frac{1}{\widetilde{\overline{Z}}} = \frac{\widetilde{H}_0}{\widetilde{E}_0} = \frac{1}{Z_s} \tag{9.154}$$

We will come back to this quantity in the context of waveguides with good but not perfectly conducting walls.

Energy Density, Poynting Vector, and Intensity

We calculate these using our standard formulae, now taking the real part as needed:

$$\langle u \rangle = \frac{1}{4} \left(\epsilon \left| \widetilde{E}_0 \right|^2 + \frac{1}{\mu} \left| \widetilde{B}_0 \right|^2 \right) e^{-\left(2\,\widehat{k}\cdot\vec{r}/\delta \right)} \tag{9.155}$$

$$= \frac{1}{4} \left(\epsilon \left| \widetilde{E}_0 \right|^2 + \epsilon\, Z_{\epsilon\mu}^2 \left| \widetilde{H}_0 \right|^2 \right) e^{-\left(2\,\widehat{k}\cdot\vec{r}/\delta \right)} \tag{9.156}$$

$$= \frac{1}{4}\, \epsilon \left| \widetilde{E}_0 \right|^2 \left(1 + \left[\left(\frac{\sigma}{\epsilon\,\omega} \right)^2 + 1 \right]^{1/2} \right) e^{-\left(2\,\widehat{k}\cdot\vec{r}/\delta \right)} \tag{9.157}$$

$$= \frac{1}{2}\, \epsilon\, \frac{k^2}{k_{\epsilon\mu}^2} \left| \widetilde{E}_0 \right|^2 e^{-\left(2\,\widehat{k}\cdot\vec{r}/\delta \right)} = \frac{1}{4}\, \epsilon \left(1 + \frac{Z_{\epsilon\mu}^2}{|\widetilde{Z}|^2} \right) \left| \widetilde{E}_0 \right|^2 e^{-\left(2\,\widehat{k}\cdot\vec{r}/\delta \right)} \tag{9.158}$$

$$\left\langle \vec{S} \right\rangle = \frac{1}{2\,\mu}\, \mathcal{R} \left(\left\langle \vec{E}^* \times \vec{B} \right\rangle \right) = \frac{\widehat{k}}{2\,\mu} \left| \widetilde{E}_0 \right|^2 \frac{e^{-\left(2\,\widehat{k}\cdot\vec{r}/\delta \right)}}{v_{\epsilon\mu}} \left[1 + \left(\frac{\sigma}{\epsilon\,\omega} \right)^2 \right]^{1/4} \cos\phi \tag{9.159}$$

$$= \frac{\widehat{k}}{2}\, v_{\epsilon\mu}\, \epsilon \left| \widetilde{E}_0 \right|^2 \left[1 + \left(\frac{\sigma}{\epsilon\,\omega} \right)^2 \right]^{1/4} e^{-\left(2\,\widehat{k}\cdot\vec{r}/\delta \right)} \cos\phi \tag{9.160}$$

$$= \frac{\widehat{k}}{2}\, v_{\epsilon\mu}\, \epsilon\, \frac{K}{k_{\epsilon\mu}} \left| \widetilde{E}_0 \right|^2 e^{-\left(2\,\widehat{k}\cdot\vec{r}/\delta \right)} \cos\phi = \widehat{k}\, \frac{k_{\epsilon\mu}}{k}\, v_{\epsilon\mu} \langle u \rangle \tag{9.161}$$

$$\mathrm{I} = \left\langle \left| \vec{S} \right| \right\rangle \tag{9.162}$$

where we have written $\langle u \rangle$ in two forms, one involving the propagation constant and the other involving the wave impedance.

Let's take the poor conductor and good conductor limits:

$1/\omega\,\tau = \sigma/\epsilon\,\omega \ll 1$, poor conductor:

$$\langle u \rangle \to \frac{1}{2}\,\epsilon\,\left|\widetilde{E}_0\right|^2 e^{-\left(2\,\widehat{k}\cdot\vec{r}/\delta\right)} \tag{9.163}$$

$$\left\langle \vec{S} \right\rangle \to \frac{\widehat{k}}{2}\,v_{\epsilon\mu}\,\epsilon\,\left|\widetilde{E}_0\right|^2 e^{-\left(2\,\widehat{k}\cdot\vec{r}/\delta\right)}\,\cos\phi = v_{\epsilon\mu}\langle u \rangle\,\widehat{k} \qquad I \to v_{\epsilon\mu}\langle u \rangle \tag{9.164}$$

The poor conductor expressions match our expressions for a wave in a nonconducting medium except that they decay with depth as one would expect.

$1/\omega\,\tau = \sigma/\epsilon\,\omega \gg 1$, good conductor:

$$\langle u \rangle \to \frac{1}{4}\,\epsilon\,\left|\widetilde{E}_0\right|^2\,\frac{\sigma}{\epsilon\,\omega}\,e^{-(2\,\widehat{k}\cdot\vec{r}/\delta)} = \frac{1}{4}\,\frac{\left|\widetilde{J}_0\right|\left|\widetilde{E}_0\right|}{\omega}\,e^{-(2\,\widehat{k}\cdot\vec{r}/\delta)} \tag{9.165}$$

$$\left\langle \vec{S} \right\rangle \to \frac{\widehat{k}}{2}\,v_{\epsilon\mu}\,\epsilon\,\left|\widetilde{E}_0\right|^2\,e^{-(2\,\widehat{k}\cdot\vec{r}/\delta)}\,\sqrt{\frac{\sigma}{\epsilon\,\omega}}\,\frac{1}{\sqrt{2}} = \frac{1}{4}\,\left|\widetilde{J}_0\right|\left|\widetilde{E}_0\right|\,e^{-(2\,\widehat{k}\cdot\vec{r}/\delta)}\,\delta\,\widehat{k} \tag{9.166}$$

$$= \omega\,\delta\langle u \rangle\widehat{k} \qquad\qquad\qquad\qquad\qquad \text{I} \to \omega\,\delta\langle u \rangle \tag{9.167}$$

From the above, we can calculate the total power dissipated in the conductive medium per unit area. We just need to know how much power flows in per unit area at the $\widehat{k}\cdot\vec{r} = 0$ plane: the wave is propagating along $\widehat{k}$, and there is no power flowing backward along $\widehat{k}$. This power flow at $\widehat{k}\cdot\vec{r} = 0$ is $\text{I}(\widehat{k}\cdot\vec{r} = 0)$ because the plane we are considering has normal $\widehat{k}$. Therefore, the power dissipated per unit area is

$$\frac{dP_{diss}}{dA} = \text{I}(\widehat{k}\cdot\vec{r} = 0) = \frac{1}{4}\,\left|\widetilde{J}_0\right|\left|\widetilde{E}_0\right|\,\delta = \frac{1}{2}\,\left|\widetilde{J}_0\right|\left|\widetilde{E}_0\right|\,\frac{\delta}{2} \tag{9.168}$$

which is as one would expect: $\left|\widetilde{J}_0\right|\left|\widetilde{E}_0\right|/2$ is the Joule power dissipation per unit volume (including time-averaging, and note $\left|\widetilde{J}_0\right|\left|\widetilde{E}_0\right| = \vec{J}_0\cdot\vec{E}_0$ because the Ohmic current, which dominates $\vec{J}$ for a good conductor, is in phase with the electric field), and $\delta/2$ is the *energy* decay length and so provides the thickness over which that power is dissipated to give a power dissipation per unit area.

Reflection at a Conducting Surface

Because free charges and currents are now possible, our boundary conditions differ from those for an interface between two nonconducting media (using the form that is symmetric between $\vec{E}$ and $\vec{H}$):

$$\hat{n} \cdot \left[\epsilon_1 \vec{E}_1 - \epsilon_2 \vec{E}_2 \right] = \sigma_f \qquad \hat{n} \cdot \left[\mu_1 \vec{H}_1 - \mu_2 \vec{H}_2 \right] = 0 \qquad (9.169)$$

$$\hat{s} \cdot \left[\vec{E}_1 - \vec{E}_2 \right] = 0 \qquad \hat{s} \cdot \left[\vec{H}_1 - \vec{H}_2 \right] = \left(\vec{K}_f \times \hat{n} \right) \cdot \hat{s} \qquad (9.170)$$

If we restrict to conductors that obey Ohm's Law, $\vec{J} = \sigma \vec{E}$, then we may conclude $\vec{K}_f = 0$ because the current singularity of a surface current requires a singularity in the field, which cannot happen.

The general case is very complicated to analyze; a measure of the difficulty is that not even Jackson tries to do it! Let's consider only the case of normal incidence. Since $\hat{n} \cdot \vec{E} = 0$, we have $\sigma_f = 0$ and the situation is similar to analyzing the case of nonconducting media.

Recall our vectors $\widehat{w}$ and $\widehat{u}$ in the plane of the interface. Our three expressions for the fields are

$$\vec{E}_i(\vec{r}, t) = \widetilde{E}_{0,i}\, \widehat{w}\, e^{i(k_i\, \widehat{n}\cdot\vec{r}-\omega\, t)} \qquad \vec{H}_i(\vec{r}, t) = \frac{1}{Z_{\epsilon\mu,1}}\, \widetilde{E}_{0,i}\, \widehat{u}\, e^{i(k_i\, \widehat{n}\cdot\vec{r}-\omega\, t)} \qquad (9.171)$$

$$\vec{E}_r(\vec{r}, t) = \widetilde{E}_{0,r}\, \widehat{w}\, e^{i(-k_i\, \widehat{n}\cdot\vec{r}-\omega\, t)} \qquad \vec{H}_r(\vec{r}, t) = -\frac{1}{Z_{\epsilon\mu,1}}\, \widetilde{E}_{0,r}\, \widehat{u}\, e^{i(-k_i\, \widehat{n}\cdot\vec{r}-\omega\, t)} \quad (9.172)$$

$$\vec{E}_t(\vec{r}, t) = \widetilde{E}_{0,t}\, \widehat{w}\, e^{i(k_t\, \widehat{n}\cdot\vec{r}-\omega\, t)} \qquad \vec{H}_t(\vec{r}, t) = \frac{k_t + i\, \kappa_t}{\mu_2\, \omega}\, \widetilde{E}_{0,t}\, \widehat{u}\, e^{i(k_t\, \widehat{n}\cdot\vec{r}-\omega\, t)} \quad (9.173)$$

$$\times\, e^{-(\kappa_t\, \widehat{n}\cdot\vec{r})} \qquad\qquad\qquad \times\, e^{-(\kappa_t\, \widehat{n}\cdot\vec{r})}$$

where $Z_{\epsilon\mu,1} = \sqrt{\mu_1/\epsilon_1}$. Inserting these expressions into the tangential boundary condition equations (the normal BC vanish because of normal incidence), evaluated with $\widehat{n} = \widehat{z}$ and at $z = 0$, we obtain

$$\widetilde{E}_{0,i} + \widetilde{E}_{0,r} = \widetilde{E}_{0,t} \qquad (9.174)$$

$$\frac{1}{Z_{\epsilon\mu,1}} \left[\widetilde{E}_{0,i} - \widetilde{E}_{0,r} \right] = \frac{k_t + i\, \kappa_t}{\mu_2\, \omega}\, \widetilde{E}_{0,t} \qquad (9.175)$$

Solving, we obtain

$$\frac{\widetilde{E}_{0,r}}{\widetilde{E}_{0,i}} = \left(\frac{1 - \widetilde{\beta}}{1 + \widetilde{\beta}}\right) \qquad \frac{\widetilde{E}_{0,t}}{\widetilde{E}_{0,i}} = \left(\frac{2}{1 + \widetilde{\beta}}\right) \tag{9.176}$$

$$\text{with } \widetilde{\beta} \equiv \frac{Z_{\epsilon\mu,1}}{\mu_2\,\omega}\,(k_t + i\,\kappa_t) = \frac{Z_{\epsilon\mu,1}}{Z_{\epsilon\mu,2}}\,\frac{K_t\,e^{i\,\phi}}{\omega/v_{\epsilon\mu,2}} = \beta\,\frac{K_t\,e^{i\,\phi_t}}{k_{\epsilon\mu,2}} \tag{9.177}$$

where $Z_{\epsilon\mu,1} = \sqrt{\mu_1/\epsilon_1}$, $Z_{\epsilon\mu,2} = \sqrt{\mu_2/\epsilon_2}$, $v_{\epsilon\mu,2} = 1/\sqrt{\epsilon_2\mu_2}$, $k_{\epsilon\mu,2} = \omega/v_{\epsilon\mu,2}$, $K_t\,e^{i\,\phi_t} = k_t + i\,\kappa_t$, and $\beta = \mu_1\,n_2/\mu_2\,n_1 = Z_{\epsilon\mu,1}/Z_{\epsilon\mu,2}$ was defined in Equation 9.87. The reflected power is

$$\mathcal{R} = \left|\frac{\widetilde{E}_{0,r}}{\widetilde{E}_{0,i}}\right|^2 = \left|\frac{1 - \widetilde{\beta}}{1 + \widetilde{\beta}}\right|^2 \tag{9.178}$$

The expression for the transmitted power, if calculated directly from the Poynting vector in the conductor, is rather complicated. Instead, one can just use $\mathcal{T} = 1 - \mathcal{R}$.

These expressions are very similar in spirit to the normal incidence equations for nonconducting media, Equations 9.104 and 9.105, except that β is replaced by $\widetilde{\beta}$. Of course, this causes the reflected and transmitted waves to acquire nontrivial phase shifts relative to the incident wave.

As usual, let's take the poor conductor and good conductor limits:

$$\frac{1}{\omega\,\tau} = \frac{\sigma}{\epsilon\,\omega} \ll 1 \qquad \begin{array}{l} poor \\ conductor \end{array} \qquad \frac{1}{\omega\,\tau} = \frac{\sigma}{\epsilon\,\omega} \gg 1 \qquad \begin{array}{l} good \\ conductor \end{array} \tag{9.179}$$

$$\widetilde{\beta} \to \beta \qquad\qquad \widetilde{\beta} \to \frac{Z_{\epsilon\mu,1}}{Z_\sigma}\frac{1+i}{2} \to \infty \tag{9.180}$$

$$\frac{\widetilde{E}_{0,r}}{\widetilde{E}_{0,i}} \to \left(\frac{1-\beta}{1+\beta}\right) \qquad \frac{\widetilde{E}_{0,r}}{\widetilde{E}_{0,i}} \to -1 \tag{9.181}$$

$$\frac{\widetilde{E}_{0,t}}{\widetilde{E}_{0,i}} \to \left(\frac{2}{1+\beta}\right) \qquad \frac{\widetilde{E}_{0,t}}{\widetilde{E}_{0,i}} \ll 1 \tag{9.182}$$

$$\mathcal{R} \to \left(\frac{1-\beta}{1+\beta}\right)^2 \qquad \mathcal{R} \to 1 \tag{9.183}$$

$$\mathcal{T} \to \beta\left(\frac{2}{1+\beta}\right)^2 \qquad \mathcal{T} \ll 1 \tag{9.184}$$

The poor conductor limit is the normal incidence case for two nonconducting media, Equations 9.104 and 9.105, and the good conductor limit gives perfect reflection with a sign flip for the reflected wave.

The case of a wave not at normal incidence becomes more algebraically challenging because three of the boundary condition equations are required and two of them involve the angles of incidence and transmission, as we saw in the nonconducting case. The complex nature of $k + i\,\kappa$ further complicates the situation. You will calculate an approximate solution for this case in homework.

Lecture 27:

Electromagnetic Waves V:
Electromagnetic Waves in Dispersive Matter
Transmission Lines

Date Revised: 2024/04/30 05:45
Correction to Equation 9.206
Date Given: 2024/04/25

Electromagnetic Waves in Dispersive Matter

Classical Model for Frequency Dependence of Permittivity — Dispersive Matter

All of the parameters we have been working with — σ, μ, and ϵ — are generally frequency-dependent. The one that is usually most obvious is the frequency dependence of the dielectric constant; this is, for example, how a prism works to disperse optical light, via the frequency dependence of ϵ and thus of $n \approx \sqrt{\epsilon/\epsilon_o}$. This dependence of the speed of light on frequency is, therefore, called *dispersion*.

We will build a simple model for dispersion. The key elements of the model are:

▶ Electrons in a nonconducting medium are bound to their locations and can move around the minimum of the potential. Any (reasonable) potential looks quadratic near its minimum because the first nonzero term in a Taylor expansion of the potential near a minimum ($dU/dx = 0$) is the quadratic term. So we assume the electron moves in a quadratic potential and hence feels a binding force

$$F_{binding} = -m\,\omega_0^2\,x \qquad (9.185)$$

where x is the displacement from the equilibrium position and ω_0 is the natural frequency of the oscillation. The assumed quadratic potential is $U(x) = m\,\omega_0^2\,x^2/2$.

▶ There is some damping force by which the electron can lose energy as it oscillates about the minimum of the potential. In most materials, this damping force is from emission and absorption of phonons (quantized crystal acoustic vibrations). This can also be thought of as "scattering with" phonons, which will become relevant when we consider the "conductor" limit later. The damping force is assumed to follow a standard damping force law

$$F_{damping} = -m\gamma \frac{dx}{dt} \tag{9.186}$$

The m is extracted from the damping coefficient for simplicity later.

▶ There is an incident electromagnetic wave with electric field polarization such that it drives the electron in the $\pm x$ direction:

$$F_{driving} = q\,E = q\,E_0 \cos\omega t \tag{9.187}$$

The wave angular frequency ω and the natural frequency of oscillation ω_0 are *not* assumed to be the same.

The resulting equation of motion for the electron is

$$m \frac{d^2x}{dt^2} = F = F_{binding} + F_{damping} + F_{driving} \tag{9.188}$$

$$\implies \quad \frac{d^2x}{dt^2} + \gamma \frac{dx}{dt} + \omega_0^2 x = \frac{q}{m} E_0 \cos \omega t \tag{9.189}$$

You've seen this equation, the *driven, damped simple harmonic oscillator*, in Ph1a and Ph106a. You know that we switch to complex notation so that the right side becomes $\widetilde{E}_0 \, e^{-i \omega t}$ and we assume a harmonic solution $\widetilde{x}(t) = \widetilde{x}_0 \, e^{-i \omega t}$. Plugging in this solution gives an algebraic equation that can be solved for $\widetilde{x}_0$, which gives the amplitude and phase of the solution:

$$\widetilde{x}_0 = \frac{q/m}{\omega_0^2 - \omega^2 - i \, \gamma \, \omega} \, \widetilde{E}_0 \tag{9.190}$$

From this, we can calculate the dipole moment of the electron, which will lead us to the polarization density of the medium:

$$\widetilde{p}(t) = q \, \widetilde{x}(t) = \frac{q^2/m}{\omega_0^2 - \omega^2 - i \, \gamma \, \omega} \, \widetilde{E}_0 \, e^{-i \omega t} \tag{9.191}$$

We can obtain the polarization density by the usual formula of summing over all the electrons. Suppose there are multiple electrons per atom or molecule (which we will term a "site") with different binding and damping forces. If there are f_j electrons per site with natural frequency ω_j and damping γ_j, and there are N sites per unit volume, then the polarization density is

$$\vec{P} = \frac{N\,q^2}{m} \left(\sum_j \frac{f_j}{\omega_j^2 - \omega^2 - i\,\gamma_j\,\omega} \right) \vec{E} \qquad (9.192)$$

where $\vec{E}$ and $\vec{P}$ are complex vectors. Clearly, there is a proportionality between $\vec{E}$ and $\vec{P}$, and we may define a *complex susceptibility*, $\widetilde{\chi}_e$, *complex permittivity*, $\widetilde{\epsilon}$, and *complex dielectric constant*, $\widetilde{\epsilon}_r$,

$$\vec{P} = \widetilde{\chi}_e\,\epsilon_o\,\vec{E} \quad \Longrightarrow \quad \widetilde{\epsilon}_r = \frac{\widetilde{\epsilon}}{\epsilon_o} = 1 + \widetilde{\chi}_e = 1 + \frac{N\,q^2}{m\,\epsilon_o} \sum_j \frac{f_j}{\omega_j^2 - \omega^2 - i\,\gamma_j\,\omega} \qquad (9.193)$$

We will use this complex permittivity in the wave equation for EM waves.

Wave Solutions for Dispersive Matter

With the complex permittivity, the wave equation becomes

$$\nabla^2 \vec{E} = \widetilde{\epsilon}\, \mu_o \frac{\partial^2 \vec{E}}{\partial t^2} \tag{9.194}$$

with $\widetilde{\epsilon}$ being a complex number. If we assume a plane wave solution, this becomes very much like the wave equation in conducting matter:

$$\text{complex permittivity:} \quad \vec{k}\cdot\vec{k}\,\vec{E}_0\,e^{i\left(\vec{k}\cdot\vec{r}-\omega\,t\right)} = \widetilde{\epsilon}\,\mu_o\,\omega^2\,\vec{E}_0\,e^{i\left(\vec{k}\cdot\vec{r}-\omega\,t\right)} \tag{9.195}$$

$$\text{conducting matter:} \quad \vec{k}\cdot\vec{k}\,\vec{E}_0\,e^{i\left(\vec{k}\cdot\vec{r}-\omega\,t\right)} = \left(\epsilon\,\mu\,\omega^2 + i\,\sigma\,\mu\,\omega\right)\vec{E}_0\,e^{i\left(\vec{k}\cdot\vec{r}-\omega\,t\right)} \tag{9.196}$$

So, as with conducting matter, we define a complex propagation vector

$$\vec{k}\cdot\vec{k} = \widetilde{\epsilon}\,\mu_o\,\omega^2 \qquad \Longleftrightarrow \qquad \sqrt{\vec{k}\cdot\vec{k}} = k + i\,\kappa \tag{9.197}$$

and the wave obeys

$$\vec{E}(\vec{r},t) = \vec{E}_0\,e^{-\left(\kappa\,\widehat{k}\cdot\vec{r}\right)}\,e^{i\left(k\,\widehat{k}\cdot\vec{r}-\omega\,t\right)} \tag{9.198}$$

The wave speed is $v = \omega/k$, the index of refraction is $n = c\,k/\omega$, and the intensity (energy, power) attenuation constant is $\alpha = 2\,\kappa$.

Now, let's figure out what k and κ are. Let's specialize to gases where the density and thus the N-proportional term in $\widetilde{\epsilon}_r$ is small so we can Taylor expand the square root:

$$k + i\,\kappa = \sqrt{\widetilde{\epsilon}\,\mu_o\,\omega^2} = \frac{\omega}{c}\,\sqrt{\widetilde{\epsilon}_r} \approx \frac{\omega}{c}\left[1 + \frac{N\,q^2}{2\,m\,\epsilon_o}\sum_j \frac{f_j}{\omega_j^2 - \omega^2 - i\,\gamma_j\,\omega}\right] \qquad (9.199)$$

We separate the real and imaginary parts by making the denominator real to obtain

$$n = \frac{c\,k}{\omega} \approx 1 + \frac{N\,q^2}{2\,m\,\epsilon_o}\sum_j \frac{f_j\left(\omega_j^2 - \omega^2\right)}{\left(\omega_j^2 - \omega^2\right)^2 + \gamma_j^2\,\omega^2} \qquad (9.200)$$

$$\frac{\alpha}{2} = \kappa = \frac{\omega^2}{c}\,\frac{N\,q^2}{2\,m\,\epsilon_o}\sum_j \frac{\gamma_j\,f_j}{\left(\omega_j^2 - \omega^2\right)^2 + \gamma_j^2\,\omega^2} \qquad (9.201)$$

Our assumption of α being small is very similar to the poor conductor case: the wave propagates into the medium, with $\widetilde{B}_0 = \widetilde{E}_0/v = \widetilde{E}_0\,n/c$ and with the magnetic and electric fields in phase, and with the wave decaying over a large number of wavelengths given by the ratio $\delta/\lambda = k/(2\,\pi\,\kappa)$. (There are no free charges in the medium because the oscillating electrons are a bound charge density whose effect is incorporated into $\widetilde{\epsilon}$, so the caveat about the validity of this limit is lifted.)

We see how the damping terms γ_j are the equivalent of Ohmic losses in a conductor: they lead to loss of energy from the wave. *Because the current is a bound current, Ohmic loss is absorbed into $\widetilde{\epsilon}$ rather than provided by an explicit σ.*

Low-Frequency Behavior, $\omega \ll \omega_j$

If the frequency is much lower than any of the resonant frequencies, $\omega \ll \omega_j$, and the damping term is also small, $\gamma_j \ll \omega$ (weak damping), then we may drop the imaginary piece in the denominator of Equation 9.199 and Taylor expand the remainder, yielding:

$$n \overset{\omega \ll \omega_j}{\longrightarrow} 1 + \frac{N q^2}{2 m \epsilon_o} \sum_j \frac{f_j}{\omega_j^2 - \omega^2} = 1 + \frac{N q^2}{2 m \epsilon_o} \sum_j \frac{f_j}{\omega_j^2 \left(1 - \frac{\omega^2}{\omega_j^2}\right)} \tag{9.202}$$

$$\approx 1 + \frac{N q^2}{2 m \epsilon_o} \sum_j \frac{f_j}{\omega_j^2} \left(1 + \frac{\omega^2}{\omega_j^2}\right) = 1 + \frac{N q^2}{2 m \epsilon_o} \sum_j \frac{f_j}{\omega_j^2} + \omega^2 \frac{N q^2}{2 m \epsilon_o} \sum_j \frac{f_j}{\omega_j^4} \tag{9.203}$$

We see the behavior consists of a frequency-independent offset of n from unity combined with a quadratic dependence on frequency. This can be rewritten as

$$n = 1 + A \left(1 + \frac{B}{\lambda^2}\right) \tag{9.204}$$

This is *Cauchy's Formula*, with A being the *coefficient of refraction* (the offset from unity) and B being the *coefficient of dispersion* (the normalization of the frequency-dependent, dispersive term). The rising index of refraction with frequency reflects the increasing ability of the individual electrons to be displaced by the field and thus the sites to be polarized as the frequency approaches the resonant frequencies. This formula explains the dispersion of light by a glass prism at visible wavelengths (see, *e.g.*, https://en.wikipedia.org/wiki/Cauchy%27s_equation).

The reason that the expression asymptotes to a constant at low frequency is that, in this limit, the index of refraction is just the square root of the static dielectric constant, which is determined by the polarizability per atom, α, which itself is set by the effective spring constant for static distortions of the electron distribution about the sites:

$$q\,\vec{E} = \vec{F} = k\,\vec{x} \implies \vec{p} = q\,\vec{x} = \frac{q^2}{k}\,\vec{E} \implies \alpha = \frac{|\vec{p}|}{|\vec{E}|} = \frac{q^2}{k} \qquad (9.205)$$

$$\implies \chi_e = \frac{|\vec{P}|}{\epsilon_o|\vec{E}|} = \frac{N|\vec{p}|}{\epsilon_o|\vec{E}|} = \frac{N\,q^2}{\epsilon_o k} \implies \epsilon_r = 1 + \chi_e = 1 + \frac{N\,q^2}{\epsilon_o k} \qquad (9.206)$$

We can see this behavior by applying the above Taylor expansion to the dielectric constant (Equation 9.193, so no factor of $1/2$), dropping the quadratic term, and rewriting in terms of the spring constants $k_j = m\omega_j^2$:

$$\widetilde{\epsilon}_r = \frac{\widetilde{\epsilon}}{\epsilon_o} = 1 + \frac{N\,q^2}{\epsilon_o} \sum_j \frac{f_j}{k_j} \qquad (9.207)$$

The two expressions are identical, with the second one performing the necessary sum over the number of electrons per state (spring constant) f_j and the different electron states j.

Conductor Behavior, $\omega_0 = 0, \omega \ll \omega_j$

If we neglect some of our assumptions and set $\omega_0 = 0$ but assume $\omega \ll \omega_j$ for $j > 0$, we obtain (going back to Equation 9.199 and looking at $\widetilde{\epsilon}_r$ not n, so no factor of $1/2$):

$$\widetilde{\epsilon}_r = \frac{\widetilde{\epsilon}}{\epsilon_o} = 1 + \frac{N q^2}{\epsilon_o} \sum_j \frac{f_j}{k_j} + i \frac{N_e q^2}{\epsilon_o m} \frac{1}{\gamma_0 - i \omega} \frac{1}{\omega} + \mathcal{O}(\omega^2) \tag{9.208}$$

where $N_e = N f_0$ is the number density of free conduction electrons (no restoring force, hence $\omega_j = \omega_0 = 0$ for them). Recalling our analogy between dielectrics with complex permittivity and conductors,

$$\vec{k} \cdot \vec{k} = [\mathcal{R}(\widetilde{\epsilon}) + i \mathcal{I}(\widetilde{\epsilon})] \mu_o \omega^2 \quad \longleftrightarrow \quad \vec{k} \cdot \vec{k} = \left(\epsilon \mu \omega^2 + i \sigma \mu \omega\right) \tag{9.209}$$

$$= \left(\epsilon + i \frac{\sigma}{\omega}\right) \mu \omega^2 \tag{9.210}$$

$$\implies \quad \sigma = \omega \mathcal{I}\left(\widetilde{\epsilon}|_{\omega \ll \omega_j}\right) = \frac{N_e q^2}{m} \mathcal{R}\left(\frac{1}{\gamma_0 - i\omega}\right) \overset{\omega/\gamma_0 \ll 1}{\longrightarrow} \frac{N_e q^2}{m} \frac{1}{\gamma_0} \tag{9.211}$$

($\omega/\gamma_0 \ll 1$ implies the damping time $1/\gamma_0$ is much less than the wave period T.) That is, a conductor can be treated like a dispersive medium with one "resonant frequency," $\omega_0 = 0$, which corresponds to totally unbound electrons with no restoring force! Nonconductors have their first resonance at $\omega_1 > 0$. Thus, we can view conductors and nonconductors in a unified way depending on whether there is a "resonant frequency" at $\omega = 0$. As we noted in discussing conductors, σ/ω can be considered to be the "conductive permittivity."

The reason our model can accommodate conductors is because it contains the resistive damping that results in a terminal velocity for the electrons in the same way that the Drude model results in a linear relationship between field and electron speed. Recall our definition of the damping force from Equation 9.186, $|F_{damping}| = m\gamma v$, which we can rewrite as

$$\frac{1}{\gamma} = \frac{v}{|F_{damping}/m|} \tag{9.212}$$

$1/\gamma$ has units of time: it is the characteristic damping time. Recall also our Drude model for DC conductivity, with

$$\sigma = \frac{N_e q^2 \lambda}{2\, m\, v_{thermal}} \tag{9.213}$$

Using Equation 9.211, which relates the two, we obtain:

$$\frac{v}{|F_{damping}|/m} = \frac{1}{\gamma_0} = \frac{m}{N_e q^2}\, \sigma = \frac{\lambda}{2\, v_{thermal}} = \frac{\tau}{2} \tag{9.214}$$

where τ is the time between scatters in the Drude model. Recall (Equation 7.3) that $v_{ave} = (\tau/2)\,(q\,E/m)$ is average drift velocity in the Drude model and thus $\tau/2$ is the damping timescale for the motion driven by the DC electric field. That is exactly the definition of γ_0^{-1}, the damping time for the DC electric field.

Plasma Behavior, $\omega \gg \omega_j$

Again, we can ignore the damping behavior because now $\omega \gg \omega_j$ (the $\omega^2 - \omega_j^2$ term dominates and we can neglect its ω_j^2 piece), providing

$$n \overset{\omega \gg \omega_j}{\longrightarrow} 1 - \frac{N q^2}{2 m \epsilon_o} \sum_j \frac{f_j}{\omega^2} \equiv 1 - \frac{1}{2} \frac{\omega_p^2}{\omega^2} \qquad \omega_p^2 \equiv \frac{N q^2}{m \epsilon_o} \sum_j f_j = \frac{N Z q^2}{m \epsilon_o} \qquad (9.215)$$

($Z = \sum_j f_j$ is the number of electrons per site.) This is the same relationship we would have obtained if we had ignored $F_{binding}$ and $F_{damping}$: this behavior arises only from the force being given by the driving electric field. At these high frequencies, the oscillation occurs much faster than any binding ($\omega \gg \omega_j$) or damping ($\omega \gg \gamma_j$) forces can have effect and the electrons can be considered to be free. The asymptotic return of the index of refraction back toward unity reflects the fact that, due to the electrons' inertia, the distance they can move during one cycle of the EM wave, and hence the polarizability of the medium, decreases as $1/m\omega^2$ at high frequency. ω_p is called the *plasma frequency* because, above this frequency, we see the behavior of a plasma of free electrons.

Notice that $n < 1$ is possible in the plasma limit, yielding wave speeds that are larger than c. Griffiths discusses how the group velocity of waves prevents a violation of special relativity.

Strangely enough, the plasma behavior can hold at much lower frequencies in tenuous plasmas where the electrons are free and binding *and* damping forces are negligible, with ω_j^2 and γ_j going to zero. In these cases, one has to return to the full permittivity (Equation 9.193) because the approximation that $n-1$ is small compared to unity fails. In such cases, one has a permittivity

$$\frac{\widetilde{\epsilon}}{\epsilon_o} = 1 - \frac{\omega_p^2}{\omega^2} \tag{9.216}$$

It is possible for this relation to hold even for $\omega < \omega_p$, in which case, $\widetilde{\epsilon}$ becomes negative. Note that this is *not* equivalent to the perfect conductor limit: in that case, $\widetilde{\epsilon}$ acquires a significant imaginary component, as we just saw in the previous section. The implications of the purely negative $\widetilde{\epsilon}$ are that there is reflection at an interface into such a medium (total internal reflection for all angles) and the field falls off exponentially into the medium with skin depth and decay constant

$$\frac{1}{\delta} = \kappa = \frac{\omega_p}{c} \tag{9.217}$$

It turns out that this plasma phenomenon, *not the perfect conductor limit*, explains the optical reflectivity of metals. There are resonances in the ultraviolet in metals that then bring $\widetilde{\epsilon} > 0$ and allow metals to be ultraviolet transparent. This phenomenon also explains how the ionosphere reflects AM (500–1600 kHz) and shortwave (2.3–26 MHz) radio waves. AM radio reflection off the ionosphere is better at night than during day because the ionosphere cools and becomes more dense (ω_p increases).

Near-Resonance Behavior

The behavior of the refractive index and skin depth near a resonance are shown below. The decay constant displays a maximum (the skin depth displays a minimum) on resonance because the system is best able to absorb (and dissipate) energy on resonance.
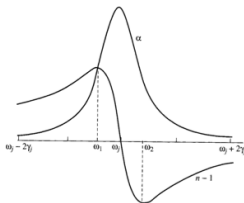
The index of refraction takes on its continuum (ω far from ω_j) value on-resonance but displays a characteristic negative slope behavior through the resonance. This negative slope of the refractive index is termed *anomalous dispersion* because it deviates from the smooth behavior seen well below ($\omega \ll \omega_j$) or above ($\omega \gg \omega_j$) the resonances.

One interesting feature is that, above ω_j, n can drop below unity, yielding $v > c$. Another interesting feature arises from the anomalous dispersion. Let's relate the negative slope in $n(\omega)$ to the group velocity to see its implications:

$$n = c\,\frac{k}{\omega} \quad \implies \quad \frac{dn}{d\omega} = \frac{c}{\omega}\,\frac{dk}{d\omega} - c\,\frac{k}{\omega^2} \quad \implies \quad v_g = \frac{d\omega}{dk} = \frac{v}{1 + \frac{d\ln n}{d\ln \omega}} \quad (9.218)$$

Since $d\ln n/d\ln \omega < 0$ in the anomalous dispersion region, it is possible to have $v_g > v$ and thus even to have $v_g > c$! Causality is not violated because, in a region of anomalous dispersion, the approximation made ($|d\ln n/d\ln \omega| \ll 1$) to model the propagation of a wave packet and define group velocity fail; see Jackson §7.8 for more details. These strange behaviors of v and v_g are used in AMO physics.

Rotational Contribution of Permanently Polarized Molecules

One can show (we will not) that the rotational degrees of freedom of liquid or gas molecules with *permanent* dipole moments give a contribution to the permittivity

$$\widetilde{\epsilon}_{r,rot} = 1 + \frac{\chi_{rot}}{1 - i\,\omega\,\tau_{rot}} \tag{9.219}$$

where χ_{rot} is the zero-frequency contribution to the susceptibility and τ_{rot} is the time constant for relaxation of an imposed polarization due to thermal fluctuations in the absence of an electric field. The real part falls off for $\omega > \tau_{rot}^{-1}$ while the imaginary part has a peak at $\omega = \tau_{rot}^{-1}$. The real part yields standard non-conducting matter behavior while the imaginary part yields conductor-like behavior in terms of the relative phase of B and E and of dissipation (into the thermal modes that cause relaxation of polarization). One never arrives, however, at the good conductor limit because the second term has an upper limit of χ_{rot} while the corresponding term for a conductor (Equation 9.210) is $\sigma/(\epsilon\,\omega)$, which has no upper limit, and because χ_{rot} itself has an upper limit set by full alignment of the intrinsic dipole moments and $\vec{E}$. Another way of saying this is that the electrons that are responding are still bound and thus can never display the almost-free propagation characteristic of a good conductor.

Overall Picture of Dispersive Behavior

The generic behavior of the real and imaginary parts of $\widetilde{\epsilon}/\epsilon_o = \widetilde{\epsilon}_r$ (usually denoted by ϵ' and ϵ''), in the rotational/conduction limits and through the anomalous dispersion features, is shown here.

Notice how, after each resonance is passed, the continuum value of ϵ' decreases a bit: this is because the numerator of each term in the sum for n ($\sim \sqrt{\epsilon'}$) changes sign after ω passes through ω_j. Eventually, ϵ' goes to the plasma limit and then rises back toward unity as $\omega \to \infty$.
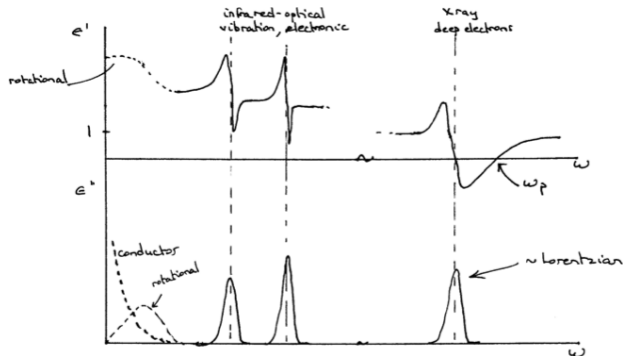
Figure courtesy of M. Cross

Lecture 28:

Electromagnetic Waves VI:
Transmission Lines

Date Revised: 2024/05/21 05:45
Corrected sketched proof in §9.7.3 of
$Z_{\mathcal{LC}} = (\mathcal{L}/\mu)\, Z_{\epsilon\mu} = (\mathcal{L}_0/\mu_o)\, Z_{\epsilon\mu}$
Date Given: 2024/04/30

Introduction and Study Guide: Guided Waves

We are going to undertake a more in-depth treatment of guided waves and transmission lines than Griffiths because it is an interesting and very useful topic in modern physics research. Some areas that use guided waves include superconducting quantum computing platforms and microwave-optical transduction; particle accelerators; astronomical instrumentation at wavelengths longer than 100 μm that preserves the coherent (wave) nature of light, even into the optical when one considers routing of light by fiber optics; and high speed data generation, sampling, and transmission, from photomultipliers (vacuum and silicon) to radio astronomy detection. More broadly, almost any experimental apparatus that uses electrical currents in any way (try to think of one that does not!) can be susceptible to radiofrequency interference, especially because of pervasive cell phones and WiFi but even due to radio, computers, and even fluorescent lights. Knowing how this radiation can get into an apparatus is the first step in stopping it. The systematic trial and error approach to solving experimental problems only works when one uses physics as a guide to where susceptibilities might be present.

We will first consider transmission lines from the circuit theory perspective, then consider Maxwell's Equations in confined regions and the propagating solutions. The material on transmission lines can be found in Heald and Marion §7.1. The material on waveguides is found in Heald and Marion §7.3–7.5 and Jackson §8.1–8.5.

Transmission Lines

Circuit Transmission Line Theory

Here, we consider a specialized arrangement consisting of two electrodes, translation-invariant in the z direction, with a voltage between them and currents flowing on them. We allow for the voltage and current to be position-dependent, which is a necessity for a propagating wave. We shall see that the critical ingredients necessary for propagating solutions are having both inductance and capacitance in much the same way that free-space propagation occurs by having both electric and magnetic fields that re-excite one another via Faraday's and Ampere's Laws.

We start with the figure below, which consists of two wires on which there is a position-dependent voltage difference and a position-dependent current. We treat the two wires completely (anti)symmetrically: if there is a voltage $V(z)$ on one wire, there is a voltage $-V(z)$ on the other wire at the same point. Thus, the voltage difference between the wires is $\Delta V(z) = 2\,V(z)$, which is the quantity we will be primarily concerned with. There is a current flowing on each wire, again antisymmetric, with values $\mathrm{I}(z)$ and $-\mathrm{I}(z)$. There will also be a net charge $Q(z)$ and $-Q(z)$ on the wires. We assume the system has an inductance per unit length *for the pair of wires* of $\mathcal{L}$ and a capacitance per unit length between the wires of $\mathcal{C}$.
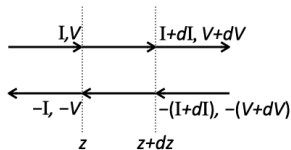
We write down equations connecting the various quantities. An incremental accumulation of charge on each wire is related to the capacitance and the voltage between them

$$dQ = \mathcal{C} \, dz \, d(\Delta V) \qquad (9.220)$$

Now, if there is an accumulation of charge at a point z, it is because, during a time interval dt, the current leaving that point z is less than the current entering. This *difference* in current between $z + dz$ and z is $d\mathrm{I}$, which can be related to dQ by :

$$-d\mathrm{I} \, dt = dQ = d(\Delta V) \, \mathcal{C} \, dz \qquad (9.221)$$

(polarity: if $d\mathrm{I} < 0$, then less current leaves the point z on the top wire than enters, so there is a net gain of positive charge on the top wire at point z). Dividing by both differentials gives

$$\frac{\partial \mathrm{I}}{\partial z} = -\mathcal{C} \, \frac{\partial \Delta V}{\partial t} \qquad (9.222)$$

We use partial derivatives because I and ΔV depend on both t and z.

The voltage drop due to the inductance (we assume zero resistance) is:

$$d(\Delta V) = -\mathcal{L} \, dz \, \frac{\partial \mathrm{I}}{\partial t} \qquad \implies \qquad dV = -\frac{\mathcal{L}}{2} \, dz \, \frac{\partial \mathrm{I}}{\partial t} \qquad (9.223)$$

To understand the sign, think about the loop formed by the four points in the diagram. For $\mathrm{I} > 0$ and $\partial \mathrm{I}/\partial t > 0$, the magnetic flux through the loop is into the page and increasing in magnitude. So an electric field is induced with polarity such that it tries to drive current to counter the increasing flux, which means that the electric field goes counter-clockwise around the loop so the current it would generate adds flux coming *out* of the page. (Remember that, because this electric field does not have to have zero loop integral, the electric field lines can go around in a loop like this.) The field lines go from the top to the bottom wire at z and from the bottom to the top wire at $z + dz$. Voltage is the line integral of electric field, so this direction of the field lines explains the sign above: the added voltage drop due to the induced field is positive at z and negative at $z + dz$ when $\partial \mathrm{I}/\partial t$ is positive. (If it is easier to think about, consider putting a voltmeter (*i.e.*, a large resistor through which the induced electric field drives a current) between the electrodes.)

Be careful about applying the same argument to the induced field along each wire, which might lead you to think the sign is the opposite (field lines go from $z + dz$ to z in the top wire and the opposite direction in the bottom wire). The change in V between z and $z + dz$ is not due to the line integral along field lines in the two wires but rather because the system is not quasistatic — V can be different in different places just due to time delay, without there being field lines. (Note that the points at $z + dz$ on the two electrodes should be considered to not be spatially separated from each other — time delay happens along the z direction, not between the electrodes.)

To understand the factor of $1/2$. we just need to recognize that the full emf affects ΔV, so, when we want to calculate the dV of a single wire, it is $\Delta V/2$.

Moving the dz to the left side gives

$$\frac{\partial \Delta V}{\partial z} = -\mathcal{L}\frac{\partial \mathrm{I}}{\partial t} \tag{9.224}$$

We can take derivatives of our two equations (Equations 9.222 and 9.224) and combine them to obtain

$$\frac{\partial^2 \mathrm{I}}{\partial z^2} = \mathcal{L}\mathcal{C}\frac{\partial^2 \mathrm{I}}{\partial t^2} \qquad \frac{\partial^2 \Delta V}{\partial z^2} = \mathcal{L}\mathcal{C}\frac{\partial^2 \Delta V}{\partial t^2} \tag{9.225}$$

We thus have wave equations describing waves in one dimension moving with speed $v = 1/\sqrt{\mathcal{L}\mathcal{C}}$.

In one dimension and for a fixed frequency ω, there are two solutions with velocity $+v$ and $-v$ and they have the general form (as we derived earlier for EM waves)

$$\mathrm{I}_\pm(z,t) = \mathrm{I}_\pm F\left(\pm\frac{\omega}{v}z - \omega t\right) \qquad \Delta V(z,t) = V_\pm(z,t) = Z_\pm \,\mathrm{I}_\pm \,F\left(\pm\frac{\omega}{v}z - \omega t\right)$$
(9.226)

where $\pm$ indicates the direction of propagation and $Z_\pm$ (to be determined) defines the ratio $(\Delta V)/\mathrm{I}$. As for $\vec{E}$ and $\vec{B}$, we assume the same functional form for the current and voltage because they are tied to each other through derivatives. (In homework, you will add finite wire conductivity, which will make $Z_\pm$ complex, allowing a phase shift.) We can determine $Z_\pm$, which gives their ratio, by plugging into $\frac{\partial \mathrm{I}}{\partial z} = -\mathcal{C}\,\frac{\partial \Delta V}{\partial t}$:

$$\pm \mathrm{I}_\pm \frac{\omega}{v} \frac{\partial F}{\partial u} = \mathcal{C}\,Z_\pm \,\mathrm{I}_\pm \,\omega \frac{\partial F}{\partial u} \quad \Longrightarrow \quad Z_\pm = \pm\frac{1}{v\,\mathcal{C}} = \pm\sqrt{\frac{\mathcal{L}}{\mathcal{C}}} \equiv \pm Z_{\mathcal{LC}} \qquad (9.227)$$

We see that the wave propagating to $+z$ has current flow to the right in the top wire when the top wire has a positive voltage while the $-z$ mode has current flow to the left when the top wire has a positive voltage. Rather than allowing the negative impedance $-Z_{\mathcal{LC}}$, we will consider these two modes to have the same *characteristic impedance*, $Z_{\mathcal{LC}}$, remembering that the currents flow in opposite directions. While $Z_{\mathcal{LC}}$ has units of ohms because it is the ratio of a voltage to a current, realize that it has nothing to do with resistivity in the wires! It is set by the geometry of the electrodes and the materials (ϵ and μ) surrounding them, which set $\mathcal{L}$ and $\mathcal{C}$.

Example 9.1: Broadside-Coupled Trace Transmission Line

This structure consists of two thin metal ribbons of width w and separation $h \ll w$. The capacitance is just that of a parallel plate capacitor. The inductance is easily obtained by calculating the field between two infinite sheets of current flowing in opposite directions ($B = \mu K$) and then calculating from that the flux per unit length ($\Phi/\ell = \mu K h$) and dividing by the current ($I = K w$). Thus, we have

$$\mathcal{C} = \frac{\epsilon\, w}{h} \qquad \mathcal{L} = \frac{\mu\, h}{w} \tag{9.228}$$

$$v = \frac{1}{\sqrt{\mathcal{L}\mathcal{C}}} = \frac{1}{\sqrt{\epsilon\,\mu}} = \frac{c}{n} \qquad Z_{\mathcal{L}\mathcal{C}} = \sqrt{\frac{\mathcal{L}}{\mathcal{C}}} = \frac{h}{w}\sqrt{\frac{\mu}{\epsilon}} \tag{9.229}$$

Example 9.2: Coaxial Cable Transmission Line

The formalism works even though the system is not symmetric. It will become clear in the next section why it works.

For coaxial cable with an inner solid conductor of radius a and outer shell conductor of radius b, we know from prior calculations that the capacitance and inductance per unit length are

$$\mathcal{C} = \frac{2\pi\epsilon}{\ln\frac{b}{a}} \qquad \mathcal{L} = \frac{\mu\ln\frac{b}{a}}{2\pi} \tag{9.230}$$

Therefore, the transmission line properties are:

$$v = \frac{1}{\sqrt{\mathcal{L}\mathcal{C}}} = \frac{1}{\sqrt{\epsilon\mu}} = \frac{c}{n} \qquad Z_{\mathcal{L}\mathcal{C}} = \sqrt{\frac{\mathcal{L}}{\mathcal{C}}} = \frac{\ln\frac{b}{a}}{2\pi}\sqrt{\frac{\mu}{\epsilon}} \tag{9.231}$$

Recall that ϵ, μ, n, v, and $Z_{\epsilon\mu}$ characterize the dielectric between the conductors, not the perfect conductors themselves! For $\mu \approx \mu_o$, this gives $Z_{\mathcal{L}\mathcal{C}} \approx (60\Omega)\ln\frac{b}{a}/\sqrt{\epsilon_r}$. Teflon and polyethylene are frequently used as the dielectric, having $\epsilon_r = 2.0$ for teflon and $\epsilon_r = 2.3$ for high-density polyethylene. Standard coaxial cable impedances are 50 Ω and 75 Ω. 75 Ω is a typical cable impedance for television antennas because it is a good match to the antenna impedance (which we will discuss later). Not that any of you are old enough to have seen a television antenna...

Generalization to Arbitrary Electrode Shapes via Wave Formulation

We derived the above for the case of two wires, but we can extend to two electrodes of arbitrary shape that are *z*-translation independent, as indicated below.
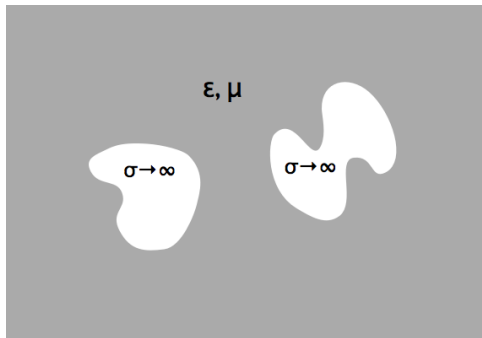
To do this, let's think about the fields and the charge and current densities rather than just voltages and currents. We aim to find a solution to Maxwell's Equations that takes the form of a wave propagating in the z-direction and that has transverse fields (consistent with our transmission line picture involving capacitance and inductance per unit length). We use $\vec{H}$ instead of $\vec{B}$ because the former is more directly related to the surface current density. Recall $k\, v_{\epsilon\mu} = \omega$. We assume:

$$\vec{E}(\vec{r}, t) = [\hat{x}\, E_x(x, y) + \hat{y}\, E_y(x, y)]\, e^{i(k\,z - \omega\, t)} \qquad \rho(\vec{r}, t) = \rho_0(x, y)\, e^{i(k\,z - \omega\, t)} \qquad (9.232)$$

$$\vec{H}(\vec{r}, t) = [\hat{x}\, H_x(x, y) + \hat{y}\, H_y(x, y)]\, e^{i(k\,z - \omega\, t)} \qquad \vec{J}(\vec{r}, t) = \hat{z}\, J_0(x, y)\, e^{i(k\,z - \omega\, t)} \qquad (9.233)$$

ρ_0 and J_0 are assumed to be surface charge and current densities, so they have δ functions in the transverse coordinates in them, but we don't write them out explicitly because it is unnecessary for our discussion here.

We will show that, under the assumptions made above, in which the charge and current densities are the same in form and have a wave dependence on z, then the propagating wave solution is just a static solution in (x, y) multiplied by the $e^{i(k\,z - \omega\, t)}$ dependence, where the static solution is that sourced by ρ_0 and $J_0 = v_{\epsilon\mu}\, \rho_0$.

When we get to waveguides, we will see that this special case yields the TEM (transverse-electric-magnetic) modes, while more general current distributions yield other, more complex modes.

To do this, let's write out Maxwell's Equations assuming the above form. The four equations involving x and y derivatives are

$$\left[\vec{\nabla} \times \vec{E}\right]_z = -\frac{\partial \vec{B}}{\partial t} \cdot \hat{z} \qquad \Longrightarrow \qquad \frac{\partial E_y}{\partial x} - \frac{\partial E_x}{\partial y} = 0 \qquad (9.234)$$

$$\left[\vec{\nabla} \cdot \vec{E}\right] = \frac{\rho}{\epsilon} \qquad \Longrightarrow \qquad \frac{\partial E_x}{\partial x} + \frac{\partial E_y}{\partial y} = \frac{\rho_0(x,y)}{\epsilon} \qquad (9.235)$$

$$\left[\vec{\nabla} \times \vec{H}\right]_z = \vec{J}_f \cdot \hat{z} + \epsilon \frac{\partial \vec{E}}{\partial t} \cdot \hat{z} \qquad \Longrightarrow \qquad \frac{\partial H_y}{\partial x} - \frac{\partial H_x}{\partial y} = J_0(x,y) \qquad (9.236)$$

$$\left[\vec{\nabla} \cdot \vec{H}\right] = 0 \qquad \Longrightarrow \qquad \frac{\partial H_x}{\partial x} + \frac{\partial H_y}{\partial y} = 0 \qquad (9.237)$$

These four equations are satisfied by the form we have chosen; the exponential factor multiplies both sides of every equation and thus cancels out. They will yield a static solution because they are sourced by the charge and current densities ρ_0 and J_0, which are not time-dependent.

The four equations in z and t derivatives are

$$\left[\vec{\nabla} \times \vec{E}\right]_x = -\frac{\partial \vec{B}}{\partial t} \cdot \hat{x} \qquad \Longrightarrow \qquad -i\,k\,E_y = i\,\omega\,\mu\,H_x \qquad (9.238)$$

$$\left[\vec{\nabla} \times \vec{E}\right]_y = -\frac{\partial \vec{B}}{\partial t} \cdot \hat{y} \qquad \Longrightarrow \qquad i\,k\,E_x = i\,\omega\,\mu\,H_y \qquad (9.239)$$

$$\left[\vec{\nabla} \times \vec{H}\right]_x = \vec{J}_f \cdot \hat{x} + \epsilon\,\frac{\partial \vec{E}}{\partial t} \cdot \hat{x} \qquad \Longrightarrow \qquad -i\,k\,H_y = -i\,\omega\,\epsilon\,E_x \qquad (9.240)$$

$$\left[\vec{\nabla} \times \vec{H}\right]_y = \vec{J}_f \cdot \hat{y} + \epsilon\,\frac{\partial \vec{E}}{\partial t} \cdot \hat{y} \qquad \Longrightarrow \qquad i\,k\,H_x = -i\,\omega\,\epsilon\,E_y \qquad (9.241)$$

These equations are satisfied by the exponential factor along with the relation

$$\hat{x}\,H_x + \hat{y}\,H_y = \frac{1}{Z_{\epsilon\mu}}\left(-\hat{x}\,E_y + \hat{y}\,E_x\right) \quad \Longrightarrow \quad \vec{H} = \frac{1}{Z_{\epsilon\mu}}\,\hat{k} \times \vec{E} \qquad (9.242)$$

Let's relate the charge and current densities by applying continuity, $\vec{\nabla} \cdot \vec{J} + \partial \rho / \partial t = 0$:

$$k \, J_0(x, y) = \omega \, \rho_0(x, y) \qquad \Longrightarrow \qquad J_0(x, y) = v_{\epsilon \mu} \, \rho_0(x, y) \qquad (9.243)$$

In addition to relating the two densities, this equation tells us that current in a perfect conductor flows only on the surfaces: since we know there is no charge density inside a perfect conductor (the decay constant for the charge density was $\tau = \epsilon / \sigma \to 0$, in the language of our solutions for EM waves in conductors), the above equation tells us there is no current density inside the conductor either.

Another way of seeing this would be to recognize that the skin depth $\delta = \sqrt{2 / (\sigma \, \mu \, \omega)}$ vanishes as $\sigma \to \infty$: the fields do not penetrate into the conductors, so neither does the current density. This happens because a surface current density is driven that cancels the EM field inside the conductor.

This, along with the fact that $\widehat{x} \, E_x + \widehat{y} \, E_y$ and $\widehat{x} \, H_x + \widehat{y} \, H_y$ are sourced by the constant charge density ρ_0 and current density J_0, shows what we set out to prove, that a transmission line has propagating wave solutions in z whose transverse field configuration is identical to that of the two-dimensional static problem sourced by charge density ρ_0 and current density $J_0 = v_{\epsilon \mu} \, \rho_0$. Note that these are not the only solutions, as we will show later.

To build a picture of the fields, it is very important that they look like static solutions. Most importantly, $\vec{E}$ is normal to the conducting surfaces and $\vec{H}$ is tangential to them, both because of the usual boundary conditions at the surface of a perfect conductor and also because $\vec{H} \propto \widehat{k} \times \vec{E}$, which determines the field contours as one approaches the conductors. We can literally draw the $\vec{E}$ that we would have drawn for the static configuration with the same conducting boundary, and then $\vec{H}$ is determined.

Connection between Circuit and Wave Formulations of Transmission Lines

Now let's recover the circuit formulation from the wave formulation. Given the wave formulation solution, we can calculate the voltage and current via

$$\Delta V(z, t) = \int_{\vec{r}_+}^{\vec{r}_-} d\vec{\ell} \cdot \vec{E} \tag{9.244}$$

$$I(z, t) = \pm \oint_{\mathcal{C}_\pm} d\ell \, K_0 = \pm v_{\epsilon\mu} \oint_{\mathcal{C}_\pm} d\ell \, \sigma_0 = \pm \frac{1}{Z_{\epsilon\mu}} \oint_{\mathcal{C}_\pm} d\ell \, \hat{n} \cdot \vec{E} \tag{9.245}$$

where we used the relation between the surface charge density and the discontinuity in the normal component of $\vec{D}$ in obtaining the last expression, noting $\epsilon \, v_{\epsilon\mu} = \sqrt{\epsilon/\mu} = 1/Z_{\epsilon\mu}$. K_0 is the surface current density, with J_0 and K_0 related by a δ function that defines the conductors' surfaces, and σ_0 is the surface charge density, with σ_0 and ρ_0 also related by this δ function. The voltage integral is from any point on the $+$ electrode to any point on the $-$ electrode (since each is an equipotential in (x, y)) and the current (charge) integral is over the entire $+$ or $-$ electrode.

We can then determine the characteristic impedance from

$$Z_{\mathcal{LC}} = \left| \frac{\Delta V(z,t)}{\mathrm{I}(z,t)} \right| = Z_{\epsilon\mu} \left| \frac{\Delta V(z,t)}{\oint_{\mathcal{C}_{\pm}} d\ell \, \widehat{n} \cdot \vec{E}} \right| = Z_{\epsilon\mu} \left| \frac{\Delta V(z,t)}{\epsilon^{-1} \oint d\ell \, \sigma_0} \right| \tag{9.246}$$

$$= Z_{\epsilon\mu}\epsilon \left| \frac{\Delta V(z,t)}{\lambda} \right| = \frac{\epsilon}{\mathcal{C}} Z_{\epsilon\mu} = \frac{\epsilon_o}{\mathcal{C}_0} Z_{\epsilon\mu} \tag{9.247}$$

where $\mathcal{C}_0 = \mathcal{C}\,\epsilon_o/\epsilon$ is the capacitance per unit length if the ϵ medium is replaced with ϵ_o. We see that the characteristic impedance is proportional to $Z_{\epsilon\mu}$ with the constant of proportionality being related to the capacitance per unit length scaled by ϵ. This is a purely geometric quantity (not even dependent on ϵ or μ). Clearly, the above quantity is determined only by the physical *shape* of the solutions, implying it is set purely by the geometry. This is what we found earlier in our coaxial cable and broadside-coupled trace examples, that the transmission line impedance $Z_{\mathcal{LC}}$ was $Z_{\epsilon\mu}$ times a factor determined by the geometry alone.

If one retains I in the denominator above but instead recognizes $\vec{A} = V \vec{v}/c^2$, one can relate the voltage difference in the numerator to a line integral of $\vec{A}$, which gives the magnetic flux per unit length Φ_ℓ and thus $\Phi_\ell/\mathrm{I} = \mathcal{L}$. Thus, one can show $Z_{\mathcal{LC}} = (\mathcal{L}/\mu) Z_{\epsilon\mu} = (\mathcal{L}_0/\mu_o) Z_{\epsilon\mu}$, where we have analogously defined $\mathcal{L}_0 = \mathcal{L}\,\mu_o/\mu$ as the inductance if the μ medium were replaced by vacuum, again a purely geometric quantity. This proof is less trivial, and, in Problem Set 5, you will effectively obtain the necessary relation in the process of proving

$$\mathcal{L}\,\mathcal{C} = \epsilon\,\mu \qquad \Longrightarrow \qquad v_{\epsilon\mu} = \frac{1}{\sqrt{\epsilon\,\mu}} = \frac{1}{\sqrt{\mathcal{L}\,\mathcal{C}}} = v_{\mathcal{LC}} \tag{9.248}$$

The above relation offers another approach to proving $Z_{\mathcal{LC}} = (\mathcal{L}_0/\mu_o)\, Z_{\epsilon\mu}$:

$$Z_{\mathcal{LC}} = \frac{\epsilon}{\mathcal{C}}\, Z_{\epsilon\mu} = \frac{\epsilon\,\mu}{\mathcal{C}\,\mathcal{L}}\,\frac{\mathcal{L}}{\mu}\, Z_{\epsilon\mu} = \frac{\mathcal{L}}{\mu}\, Z_{\epsilon\mu} = \frac{\mathcal{L}_0}{\mu_0}\, Z_{\epsilon\mu} \tag{9.249}$$

The above could of course be inverted to use $Z_{\mathcal{LC}} = (\epsilon/\mathcal{C})\, Z_{\epsilon\mu}$ and $Z_{\mathcal{LC}} = (\mathcal{L}/\mu)\, Z_{\epsilon\mu}$ to show $\mathcal{LC} = \epsilon\mu$.

Finally, we note that the combination of the above relations tell us that the geometrical factor in $\mathcal{C}$ and $\mathcal{L}$ is the same (up to reciprocation):

$$\frac{\epsilon_o}{\mathcal{C}_0}\, Z_{\epsilon\mu} = Z_{\mathcal{LC}} = \frac{\mathcal{L}_0}{\mu_o}\, Z_{\epsilon\mu} \quad \Longleftrightarrow \quad \frac{\mathcal{L}_0}{\mu_o} = \frac{\epsilon_o}{\mathcal{C}_0} \tag{9.250}$$

$$\text{or} \qquad \mathcal{L}\,\mathcal{C} = \epsilon\,\mu \quad \Longleftrightarrow \quad \frac{\mathcal{L}}{\mu} = \frac{\epsilon}{\mathcal{C}} \quad \Longleftrightarrow \quad \frac{\mathcal{L}_0}{\mu_o} = \frac{\epsilon_o}{\mathcal{C}_0} \tag{9.251}$$

Together, these various results — the circuit and wave formulations both have solutions involving waves propagating in the z direction; the propagation speeds are equal, $v_{\epsilon\mu} = v_{\mathcal{LC}}$; the circuit formulation voltage and current can be written in terms of the wave formulation fields; and the characteristic impedance $Z_{\mathcal{LC}}$ is equal to the medium's wave impedance $Z_{\epsilon\mu}$ times a geometrical factor that can be calculated from the wave solution — imply that the formulations are equivalent. The wave formulation is more generic and also gives us the underlying basis for the circuit formulation, while the circuit formulation abstracts out the key features we need to do calculations in a less cumbersome manner, as we will see.

Energy and Power in Transmission Lines

The time-averaged energy flux of the EM wave is given by the Poynting vector:

$$\left\langle \vec{S} \right\rangle = \frac{1}{2} \, \mathcal{R} \left(\left\langle \vec{E}^* \times \vec{H} \right\rangle \right) = \widehat{k} \, \frac{1}{2 \, Z_{\epsilon\mu}} \left| \vec{E}_0 \right|^2 = \widehat{k} \, \frac{1}{2} \, v_{\epsilon\mu} \, \epsilon \left| \vec{E}_0 \right|^2 = \widehat{k} \, v_{\epsilon\mu} \langle u \rangle \qquad (9.252)$$

where $\vec{E}(\vec{r}, t) = \vec{E}_0(x, y) \, e^{i(k z - \omega t)}$ and $\vec{E}_0(x, y)$ is an electrostatic solution in two dimensions. The Poynting vector has units of power/area. Therefore, if we integrate it over the xy-plane, we get the total time-averaged power flowing past a given point:

$$\langle P \rangle = \int_{\mathcal{S}} da \, \widehat{z} \cdot \left\langle \vec{S} \right\rangle = \frac{1}{2} \, v_{\epsilon\mu} \int_{\mathcal{S}} da \, \epsilon \left| \vec{E}_0 \right|^2 \qquad (9.253)$$

The integral (with the $1/2$) is the time-averaged energy per unit length in the electric field, which can be rewritten using the capacitance per unit length and the voltages because $\vec{E}_0(x, y)$ is an electrostatic solution:

$$\langle P \rangle = \frac{1}{2} \, v_{\epsilon\mu} \, \mathcal{C} \, |V_0|^2 = \frac{1}{2} \, \frac{\mathcal{C}}{\sqrt{\mathcal{L}\mathcal{C}}} \, |V_0|^2 = \frac{1}{2} \, \frac{|V_0|^2}{Z_{\mathcal{LC}}} = \frac{1}{2} \, |\mathrm{I}_0|^2 \, Z_{\mathcal{LC}} = \frac{1}{2} \, \mathcal{R} \left(\mathrm{I}_0^* \, V_0 \right) \quad (9.254)$$

where V_0 is amplitude of the wave giving ΔV, $\mathrm{I}_0 = V_0 / Z_{\mathcal{LC}}$ is the amplitude of the wave giving I, and we take the real part so the result is generalizable (so far, I and ΔV are in phase). We thus have an expression for the (time-averaged) power flowing down the transmission line in terms of the transmission line parameters instead of the microscopic fields.

Reflection and Transmission at a Transmission Line Junction

Now, let's consider a wave incident on a junction between two transmission lines (at $z = 0$) of characteristic impedances Z_1 and Z_2. We assume a solution consisting of an incident and reflected wave at $z < 0$ and a transmitted wave at $z > 0$. The voltage must be continuous as always, and the current must be continuous to avoid a buildup of charge at the boundary. (If we choose to match fields instead: $\vec{E}$ and $\vec{B}$ are transverse and there are no free charges or currents at the boundary, so their normal components vanish and their transverse components are continuous, yielding these conditions on V and I.) We write the waves as (including a reflected wave for $z < 0$):

$$z < 0 : \qquad V(z, t) = e^{i(k_1 z - \omega\, t)} + \widetilde{r}\, e^{i(-k_1 z - \omega\, t)} \tag{9.255}$$

$$\mathrm{I}(z, t) = \frac{1}{Z_1}\, e^{i(k_1 z - \omega\, t)} - \frac{\widetilde{r}}{Z_1}\, e^{i(-k_1 z - \omega\, t)} \tag{9.256}$$

$$z > 0 : \qquad V(z, t) = \widetilde{t}\, e^{i(k_2 z - \omega\, t)} \tag{9.257}$$

$$\mathrm{I}(z, t) = \frac{\widetilde{t}}{Z_2}\, e^{i(k_2 z - \omega\, t)} \tag{9.258}$$

where $\widetilde{r}$ and $\widetilde{t}$ are to be determined. Note that we know $k_2\, v_2 = \omega = k_1\, v_1$. As usual, we have assumed the same time dependence because the matching conditions must be satisfied at all times. Notice the sign on the left-going reflected current term.

Setting $z = 0$ and matching the solutions as we described, we obtain

$$1 + \widetilde{r} = \widetilde{t} \qquad \frac{1 - \widetilde{r}}{Z_1} = \frac{\widetilde{t}}{Z_2} \implies \widetilde{r} = \frac{Z_2 - Z_1}{Z_2 + Z_1} \qquad \widetilde{t} = \frac{2 Z_2}{Z_1 + Z_2} \qquad (9.259)$$

We see that these equations are identical to Equations 9.104 and 9.105 for the fields in the case of normal incidence of an EM wave on the interface between two different perfectly non-conducting materials, except that now the impedances are the characteristic impedances of the transmission lines $Z_{\mathcal{LC}}$ instead of the wave impedances $Z_{\epsilon\mu}$. Of course, the former incorporates the latter, but it also depends on the transmission line geometry, which does not exist for the free space case.

We may calculate the power reflection and transmission coefficients based on Equation 9.254:

$$\mathcal{R} = \frac{\frac{1}{2}|\widetilde{r}|^2/Z_1}{\frac{1}{2}|1|^2/Z_1} = \left(\frac{Z_2 - Z_1}{Z_2 + Z_1}\right)^2 \qquad \mathcal{T} = \frac{\frac{1}{2}|\widetilde{t}|^2/Z_2}{\frac{1}{2}|1|^2/Z_1} = \frac{Z_1}{Z_2}\left(\frac{2 Z_2}{Z_1 + Z_2}\right)^2 \qquad (9.260)$$

One can easily see $\mathcal{R} + \mathcal{T} = 1$ and that these expressions also match what we derived for an EM wave at normal incidence.

We now see the origin of the term impedance matching: $\mathcal{R} = 0$ if $Z_1 = Z_2$.

Reflection and Transmission at a Load Impedance

If, instead of a second transmission line section, we have a junction to a
lumped-element load with complex impedance Z_L (some combination of resistances,
capacitances, and inductances), then we just have

$$z > 0: \qquad V(t) = \widetilde{t}\, e^{-i\,\omega\,t} \qquad I(t) = \frac{\widetilde{t}}{Z_L}\, e^{-i\,\omega\,t} \tag{9.261}$$

This doesn't change the matching conditions, so we have the same relations as
Equations 9.259 and 9.260 with Z_2 replaced by Z_L. In particular, there is full
transmission of the power from the transmission line into the load if $Z_L = Z_1$: this is
another version of impedance matching. Note that $Z_L = 0$ and $Z_L = \infty$ both result in
unity reflection, though with opposite signs for $\widetilde{r}$.

Input Impedance of a Terminated Transmission Line

Another interesting and frequently considered question is what happens at the interface between transmission lines when the second line is terminated in a load impedance after a length ℓ. We can phrase this in terms of an effective impedance "looking into" (from the point of view of a wave incident on the interface from Z_1 into) the loaded transmission line.

Let's assume that there are left- and right-going waves on this length ℓ of transmission line with amplitudes V_+ and V_- to be determined. (We won't write the $e^{-i\omega t}$ time dependence for now; it is common to all of the expression and so will cancel out in any equations.) Define the "complex wave ratio" as the ratio of the left-going and right-going amplitudes:

$$r(z) = \frac{V_-(z)}{V_+(z)} \tag{9.262}$$

The voltage and current at any point on the line are

$$V(z) = V_+(z) + V_-(z) = V_+(z)\left[1 + r(z)\right] \tag{9.263}$$

$$\mathrm{I}(z) = \mathrm{I}_+(z) - \mathrm{I}_-(z) = \frac{V_+(z)}{Z_{\mathcal{LC}}} - \frac{V_-(z)}{Z_{\mathcal{LC}}} = \frac{V_+(z)}{Z_{\mathcal{LC}}}\left[1 - r(z)\right] \tag{9.264}$$

We can then calculate an impedance at any point on the line:

$$Z(z) = \frac{V(z)}{I(z)} = Z_{\mathcal{LC}} \frac{1 + r(z)}{1 - r(z)} \tag{9.265}$$

We know or can now calculate the functions $V_+(z)$, $V_-(z)$, $r(z)$, and $Z(z)$:

$$V_+(z) = V_+(0)\, e^{i\,k\,z} \qquad\qquad r(z) = r(0)\, e^{-2\,i\,k\,z} \tag{9.266}$$

$$V_-(z) = V_-(0)\, e^{-i\,k\,z} \qquad\qquad Z(z) = Z_{\mathcal{LC}}\, \frac{1 + r(0)\, e^{-2\,i\,k\,z}}{1 - r(0)\, e^{-2\,i\,k\,z}} \tag{9.267}$$

Now, to find $Z_i = Z(0)$, the effective input impedance of the transmission line, we simply set $V(\ell) = Z_L\, I(\ell)$, which implies $Z(\ell) = Z_L$, find $r(0)$ from the relation between $Z(\ell)$ and $r(0)$, and then find $Z_i = Z(0)$ from $r(0)$:

$$Z(\ell) = Z_L \qquad \Longrightarrow \qquad r(0) = e^{2\,i\,k\,\ell}\, \frac{\dfrac{Z_L}{Z_{\mathcal{LC}}} - 1}{\dfrac{Z_L}{Z_{\mathcal{LC}}} + 1} \tag{9.268}$$

$$\Longrightarrow \qquad Z_i = Z(0) = Z_{\mathcal{LC}}\, \frac{1 + r(0)}{1 - r(0)} = Z_{\mathcal{LC}}\, \frac{Z_L - i\, Z_{\mathcal{LC}} \tan k\ell}{Z_{\mathcal{LC}} - i\, Z_L \tan k\ell} \tag{9.269}$$

Based on this formula, some examples of the behavior are given below. It is crucial to recognize that the behavior depends on k so is frequency-dependent!

▶ Short-circuit termination

$$Z_L = 0 \qquad \Longrightarrow \qquad Z_i = -i\, Z_{\mathcal{LC}} \tan(k\ell) \qquad (9.270)$$

The line acts purely reactive (no resistance). It is capacitive or inductive depending on the length in numbers of quarter-wavelengths:

$$(2\,n)\,\frac{\lambda}{4} < \ell < (2\,n+1)\,\frac{\lambda}{4} \quad \Longrightarrow \quad \mathcal{I}(Z_i) < 0 \quad \Longrightarrow \quad \text{inductive} \qquad (9.271)$$

$$(2\,n-1)\,\frac{\lambda}{4} < \ell < (2\,n)\,\frac{\lambda}{4} \quad \Longrightarrow \quad \mathcal{I}(Z_i) > 0 \quad \Longrightarrow \quad \text{capacitive} \qquad (9.272)$$

Note that the sign conventions for inductive and capacitive reactances are the opposite of what is used in Ph1c and in engineering because we use here a $e^{-i\,\omega\,t}$ time dependence instead of the usual $e^{i\,\omega\,t}$ engineering time dependence.

▶ Open-circuit termination

$$Z_L = \infty \qquad \Longrightarrow \qquad Z_i = i\, Z_{\mathcal{LC}} \cot(k\ell) \qquad (9.273)$$

Again, capacitive or inductive depending on the length.

▶ Quarter-wavelength

$$\ell = \frac{\lambda}{4} \qquad Z_i = \frac{Z_{\mathcal{LC}}^2}{Z_L} \tag{9.274}$$

Acts like a transformer of the load impedance.

▶ Half-wavelength

$$\ell = \frac{\lambda}{2} \qquad Z_i = Z_L \tag{9.275}$$

No change in impedance.

Lecture 29:

Electromagnetic Waves VII:
Waveguides: Form of Solution, Boundary Conditions,
Propagation Properties, Example Solutions

Date Revised: 2024/05/02 06:30
Date Given: 2024/05/02

Waveguides

What is a Waveguide?

There are two critical ingredients in the concept of a *waveguide*:

- ▶ First, that an EM wave is propagating along the z direction in a (possibly not simply connected in the xy-plane) region of nonconducting material that is bounded by (for now) lossless metal boundaries in the xy-plane. There *need not* be an outer boundary of metal (see also our EM field discussion of transmission lines earlier). A hollow rectangular waveguide is a simply connected example, while a coaxial cable is not. A transmission line consisting of two straight wires in vacuum is an example in which there is no outer metal boundary *and* the dielectric is not simply connected.

- ▶ Second, that voltages appear and currents flow sinusoidally on those walls to support the propagating wave. The sinusoidal flow eliminates the need for an explicit return conductor because the time average of the current at a given point vanishes: at any point, the current goes in the $+z$ direction half the time, then in the $-z$ direction half the time. We assume that the waveguide goes on forever or is terminated in an impedance that matches the waveguide impedance (this will be explained below) so we can neglect the end effects.

Eigenvector-Eigenvalue Equation for Waveguide Solutions

We will assume that the electric and magnetic fields have propagating wave solutions in the z direction with $e^{i(k\,z - \omega\,t)}$ dependences (k positive or negative):

$$\vec{E}(\vec{r}, t) = \vec{E}_0(\vec{r}_\perp)\, e^{i(k\,z - \omega\,t)} \qquad \vec{H}(\vec{r}, t) = \vec{H}_0(\vec{r}_\perp)\, e^{i(k\,z - \omega\,t)} \tag{9.276}$$

where $\vec{E}_0$ and $\vec{H}_0$ are, as usual, complex, and $\vec{r}_\perp = x\,\widehat{x} + y\,\widehat{y}$. We again work with $\vec{H}$ since it is more closely related to the currents and also because the equations become more symmetric between $\vec{E}$ and $\vec{H}$. We calculated earlier what happens when you plug such a solution into Maxwell's Equations, but now we may not assume that the waves are transverse, we must allow for z components. Assuming the medium in the waveguide to has uniform ϵ and μ, plugging into the two curl equations yields

$$\frac{\partial E_y}{\partial x} - \frac{\partial E_x}{\partial y} = i\,\mu\,\omega\,H_z \qquad\qquad \frac{\partial H_y}{\partial x} - \frac{\partial H_x}{\partial y} = -i\,\epsilon\,\omega\,E_z \tag{9.277}$$

$$\frac{\partial E_z}{\partial y} - i\,k\,E_y = i\,\mu\,\omega\,H_x \qquad\qquad \frac{\partial H_z}{\partial y} - i\,k\,H_y = -i\,\epsilon\,\omega\,E_x \tag{9.278}$$

$$i\,k\,E_x - \frac{\partial E_z}{\partial x} = i\,\mu\,\omega\,H_y \qquad\qquad i\,k\,H_x - \frac{\partial H_z}{\partial x} = -i\,\epsilon\,\omega\,E_y \tag{9.279}$$

We see a clear symmetry between E and H in the equations.

We may solve these reduced equations for the transverse components in terms of the z, or *longitudinal* components. (*e.g.*, plug the equation with H_y on the RHS into the equation for E_x on the RHS and solve for E_x and one gets an equation for E_x in terms of E_z and H_z.) Defining any vector with a $\perp$ subscript to be the vector's component in the xy-plane, and recalling $k_{\epsilon\mu} = \omega/v_{\epsilon\mu}$ and $Z_{\epsilon\mu} = \sqrt{\mu/\epsilon}$, we obtain

$$\vec{H}_{0,\perp} = \frac{1}{k_{\epsilon\mu}^2 - k^2}\left[i\,k\,\vec{\nabla}_\perp H_{0,z} + i\,\frac{k_{\epsilon\mu}}{Z_{\epsilon\mu}}\,\widehat{z}\times\vec{\nabla}_\perp E_{0,z}\right] \tag{9.280}$$

$$\vec{E}_{0,\perp} = \frac{1}{k_{\epsilon\mu}^2 - k^2}\left[i\,k\,\vec{\nabla}_\perp E_{0,z} - i\,k_{\epsilon\mu}\,Z_{\epsilon\mu}\,\widehat{z}\times\vec{\nabla}_\perp H_{0,z}\right] \tag{9.281}$$

Note that the transverse components end up out of phase with the longitudinal component by $\pi/2$ due to the factor i. Note that one cannot find a generic, simple relationship between $\vec{E}_{0,\perp}$ and $\vec{H}_{0,\perp}$ like we had before for transverse waves. We will see below that we can obtain such a relationship, but differently for different modes.

For propagation in the $-\widehat{z}$ direction, one can work through the algebra to see that only the sign on k in the first terms changes. The second terms are unchanged because the prefactors in the second terms, as well as the sign of $\widehat{z}$, do not depend on k.

In order for the units on the two sides to match, the factor of $1/k$ cancels the units of the $\vec{\nabla}_\perp$ operator, and there is the usual factor of $Z_{\epsilon\mu}$ connecting E and H.

Next, let's use the Maxwell divergence equations by taking the divergence of the above. We note that

$$\vec{\nabla} \cdot \left(\hat{z} \times \vec{\nabla}_\perp\right) = \left(\hat{x}\,\frac{\partial}{\partial x} + \hat{y}\,\frac{\partial}{\partial y} + \hat{z}\,\frac{\partial}{\partial z}\right) \cdot \left(\hat{y}\,\frac{\partial}{\partial x} - \hat{x}\,\frac{\partial}{\partial y}\right) = 0 \qquad (9.282)$$

Therefore,

$$0 = \vec{\nabla} \cdot \vec{E} = e^{i(k\,z - \omega\,t)} \left(\vec{\nabla}_\perp \cdot \vec{E}_{0,\perp} + i\,k\,E_{0,z}\right) \qquad (9.283)$$

$$= \frac{i\,k}{k_{\epsilon\mu}^2 - k^2}\,\nabla_\perp^2\,E_{0,z} + i\,k\,E_{0,z} \qquad (9.284)$$

Rewriting, and doing the same with $\vec{\nabla} \cdot \vec{B}$, we obtain

$$\boxed{\nabla_\perp^2\,E_{0,z} + \left(k_{\epsilon\mu}^2 - k^2\right) E_{0,z} = 0 \qquad \nabla_\perp^2\,H_{0,z} + \left(k_{\epsilon\mu}^2 - k^2\right) H_{0,z} = 0 \qquad k_{\epsilon\mu} = \frac{\omega}{v_{\epsilon\mu}}}$$
$$(9.285)$$

These are clearly eigenvalue-eigenvector equations, where the action of an operator on the solution is equal to a constant times the solution. Note how there still remains no relationship between $E_{0,z}$ and $H_{0,z}$: they remain completely independent, giving rise to two different kinds of modes in which one or the other vanishes.

If we choose $E_{0,z} = 0$, then the waves are called *transverse electric (TE)* modes, and if we choose $H_{0,z} = 0$, they are called *transverse magnetic (TM)* modes. Note that, because $\vec{E}_{0,\perp}$ depends on $E_{0,z}$ and $H_{0,z}$, setting $E_{0,z} = 0$ still allows for nonzero $\vec{E}_{0,\perp}$ for TE modes, and similarly for $\vec{H}_{0,\perp}$ for TM modes. By linearity, any arbitrary sum of TE and TM modes is still a valid field configuration. Note that the mode wavelength, set by k, is not necessarily the same as it is in unconfined space or transmission lines, which would be given by $k_{\epsilon\mu}$. Once one has specified TE or TM, a simple relationship between the transverse fields is obtained (from Equations 9.280 and 9.281):

$$\text{TE:} \quad \vec{H}_{0,\perp} = \frac{1}{Z_{\epsilon\mu}} \frac{k}{k_{\epsilon\mu}} \hat{z} \times \vec{E}_{0,\perp} \qquad \text{TM:} \quad \vec{H}_{0,\perp} = \frac{1}{Z_{\epsilon\mu}} \frac{k_{\epsilon\mu}}{k} \hat{z} \times \vec{E}_{0,\perp} \qquad (9.286)$$

Note that these relations are k-dependent ($\Leftrightarrow \nu$-dependent) because the ratio $k/k_{\epsilon\mu}$ or $k_{\epsilon\mu}/k$ appears. So, while they are generic relations, there will be no such simple relation if one sums modes with different k ($\Leftrightarrow \nu$) or sums TE and TM modes.

There are also *transverse electric and magnetic (TEM)* modes. These are not obtained by setting $E_{0,z} = H_{0,z} = 0$ in the above wave equations. Rather, one must go back to the curl equations and set the z components to be zero there, which results in the simplified equations for $\vec{E}_{0,\perp}$ and $\vec{H}_{0,\perp}$ we derived in our wave formulation of transmission lines (Equations 9.232 and 9.233): static solutions in two dimensions give the transverse field behavior for the propagating TEM mode (with $\vec{H}_{0,\perp} = \hat{z} \times \vec{E}_{0,\perp}/Z_{\epsilon\mu}$ and $k = k_{\epsilon\mu}$). TEM modes are more generic than transmission lines, as one can have TEM modes with more than two conductors. Nevertheless, it is conventional to equate "TEM mode" with "transmission line mode."

Generic Boundary Conditions in a Waveguide

We have the partial differential equations for the fields, but we need boundary conditions to have a fully defined mathematical problem to solve, so let's consider those now.

There will be free charges and currents in the walls to generate the fields, so we may not assume anything about $\hat{n} \cdot \vec{E}$ and $\hat{t} \cdot \vec{H}$ at the walls.

For a perfect conductor, the skin depth vanishes. Therefore, there is no magnetic or electric field in the conductor. This, along with $\vec{\nabla} \cdot \vec{B} = 0$, implies $\hat{n} \cdot \vec{H} = 0$ at the walls.

The vanishing of the fields in the conductor along with Faraday's Law tells us $\hat{t} \cdot \vec{E} = 0$ at the walls. (Remember, fields cannot have a δ-function spatial dependence, so the right side of Faraday's Law yields no contribution to the surface integral that is related to the usual loop integral we do to determine tangential boundary conditions like this.)

No TEM Mode in Hollow Waveguide

These boundary conditions enable us to immediately prove that *hollow (simply connected in the xy-plane) waveguides have no TEM modes*. Returning to Maxwell's equations for the waveguide and using $E_{0,z} = 0 = H_{0,z}$ for the TEM mode, we obtain

$$\vec{\nabla} \cdot \vec{E} = 0 \quad \Longrightarrow \quad \frac{\partial E_{0,x}}{\partial x} + \frac{\partial E_{0,y}}{\partial y} = 0 \qquad \left[\vec{\nabla} \times \vec{E}\right]_z = 0 \quad \Longrightarrow \quad \frac{\partial E_{0,y}}{\partial x} - \frac{\partial E_{0,x}}{\partial y} = 0$$

The other curl equation terms vanish because $\vec{E}_0(x, y)$ has no z dependence and $E_{0,z}$ vanishes. Thus, $\vec{E}_0$ has no curl, so it is the gradient of a scalar potential $V_0(x, y)$. The vanishing of $\vec{\nabla} \cdot \vec{E}$ tells us that this potential $V_0(x, y)$ satisfies Laplace's (instead of Poisson's) Equation in two dimensions. The boundary condition $\hat{t} \cdot \vec{E} = 0$ implies any connected boundary of the waveguide is an equipotential for V_0 (at fixed z). If there is only one boundary, then we may use the fact that solutions to Laplace's Equation have no local minima or maxima to conclude that the potential must be constant across the (x, y)-plane and thus that $\vec{E}_{0,\perp} = 0$. Given the generic relation for TEM modes, $\vec{H}_{0,\perp} = \hat{z} \times \vec{E}_{0,\perp}/Z_{\epsilon\mu}$, this also implies $\vec{H}_{0,\perp} = 0$. (The lack of fields also implies $\rho_0(x, y)$ and $J_0(x, y)$ vanish too.)

This argument does not apply for a non-hollow waveguide because there are at least two boundaries that may not be at the same potential (*e.g.*, coaxial cable).

This result makes sense: transmission lines need two (and exactly two) non-grounded electrodes with complementary currents and voltages to define $\mathcal{L}$ and $\mathcal{C}$.

TEM Mode in a Coaxial Cable

Before jumping into TE and TM modes, let's consider an example of a TEM mode from the waveguide perspective. A simple TEM example is the coaxial cable, to which our "no TEM mode in hollow waveguide" theorem does not apply. We learned that the TEM modes have fields that are the same as the static fields for the configuration with the relation Equation 9.242 between the electric and magnetic fields:

$$\vec{H} = \frac{1}{Z_{\epsilon\mu}} \, \widehat{k} \times \vec{E} \tag{9.287}$$

We thus know the electrostatic and magnetostatic fields for this configuration have the form

$$
\boxed{
\begin{array}{ccc}
\text{coaxial} & & \\
\text{waveguide} & \vec{E}_{0,\perp} = \dfrac{A}{s} \, \widehat{s} & \qquad \vec{H}_{0,\perp} = \dfrac{A}{Z_{\epsilon\mu} \, s} \, \widehat{\phi} \\
\text{TEM mode} & &
\end{array}
} \tag{9.288}
$$

where the relative normalization has been determined from the aforementioned relation between the fields.

Note that such cables do also support TE and TM modes; see, for example,
`http://www.microwaves101.com/encyclopedias/coax#TE11`.

Boundary Conditions for TE and TM Modes

We now return to the generic case in which either $E_{0,z} \neq 0$ or $H_{0,z} \neq 0$. Since we have reduced the problem to one of finding these z components of the fields, we need boundary conditions on $E_{0,z}$ and $H_{0,z}$ on the walls, which we will define by a contour $\mathcal{C}$ (consisting of multiple disconnected pieces when there is more than one conductor). We may now apply the previously developed generic boundary conditions to these fields and use Equations 9.280 and 9.281.

Starting with Faraday's Law, $\hat{t} \cdot \vec{E} = 0$ at the walls, one allowed direction of $\hat{t}$ is $\hat{t} = \hat{z}$, which gives us one of the boundary conditions we will use, $\hat{z} \cdot \vec{E} = 0$ at the walls, or

$$\boxed{E_{0,z}\big|_{\mathcal{C}} = 0} \tag{9.289}$$

Note that this holds even for (especially for!) TM modes, which have nonzero $E_{0,z}$. Even though $E_{0,z}$ is nonzero in general for TM modes, it does vanish at the walls, and the vanishing at the walls does not imply it vanishes everywhere!

We can also apply Faraday's law with the orthogonal tangent vector, $\widehat{t}_\perp = \widehat{z} \times \widehat{n}$. This vector is tangent to the walls and lies in the xy plane. Using $\widehat{t}_\perp \cdot \vec{E} = 0$ and Equation 9.281:

$$0 = \widehat{t}_\perp \cdot \vec{E} = \widehat{t}_\perp \cdot \vec{E}_{0,\perp} = \widehat{t}_\perp \cdot \frac{1}{k_{\epsilon\mu}^2 - k^2} \left[i\, k\, \vec{\nabla}_\perp E_{0,z} - i\, Z_{\epsilon\mu}\, k_{\epsilon\mu}\, \widehat{z} \times \vec{\nabla}_\perp H_{0,z} \right]$$

The first term on the right side vanishes because it calculates the gradient of $E_{0,z}$ along $\widehat{t}_\perp$, which vanishes because $E_{0,z}$ vanishes everywhere on the walls. The second term can be cleaned up using the vector identity $\vec{a} \cdot (\vec{b} \times \vec{\nabla}) = (\vec{a} \times \vec{b}) \cdot \vec{\nabla}$ to yield another boundary condition on the z components of the fields:

$$0 = \widehat{t}_\perp \cdot \left(\widehat{z} \times \vec{\nabla}_\perp \right) H_{0,z} = \left(\widehat{t}_\perp \times \widehat{z} \right) \cdot \vec{\nabla}_\perp H_{0,z} \quad \Longrightarrow \quad \boxed{\left. \widehat{n} \cdot \vec{\nabla}_\perp H_{0,z} \right|_{\mathcal{C}} = 0} \quad (9.290)$$

We may also try to use $\widehat{n} \cdot \vec{H} = 0$ combined with Equation 9.280:

$$0 = \widehat{n} \cdot \vec{H} = \widehat{n} \cdot \vec{H}_{0,\perp} = \widehat{n} \cdot \frac{1}{k_{\epsilon\mu}^2 - k^2} \left[i\, k\, \vec{\nabla}_\perp H_{0,z} + i\, \frac{k_{\epsilon\mu}}{Z_{\epsilon\mu}}\, \widehat{z} \times \vec{\nabla}_\perp E_{0,z} \right]$$

Both terms vanish already: the first term because $\widehat{n} \cdot \vec{\nabla}_\perp H_{0,z} = 0$, the second because $\widehat{n} \cdot (\widehat{z} \times \vec{\nabla}_\perp) E_{0,z} = (\widehat{n} \times \widehat{z}) \cdot \vec{\nabla}_\perp E_{0,z} = -\widehat{t}_\perp \cdot \vec{\nabla}_\perp E_{0,z}$, which vanishes because $E_{0,z}$ vanishes everywhere on the walls. So no further boundary conditions are available.

Generic Properties of TM Modes

For a TM mode, we set $H_{0,z} = 0$ and drop the $H_{0,z}$ wave equation in Equation 9.285 and focus on the $E_{0,z}$ equation. The relevant boundary condition is $E_{0,z}|_{\mathcal{C}} = 0$ because the $H_{0,z}$ condition is useless. Rewriting the $E_{0,z}$ piece of Equation 9.285 in a generic form gives

$$\nabla_\perp^2 \psi + \gamma^2 \psi = 0 \qquad \gamma^2 = k_{\epsilon\mu}^2 - k^2 \qquad \psi|_{\mathcal{C}} = 0 \qquad (9.291)$$

This is an eigenvalue-eigenvector problem (like the Schrodinger equation in two dimensions) with a *Dirichlet* boundary condition. The boundary condition results in a discrete set of allowed $\{\gamma_n^{TM}\}$ and thus solutions $\{\psi_n^{TM}\}$, just like such a boundary condition imposes quantization in quantum mechanics or solutions to Laplace's Equation. One thus has

$$\boxed{\text{TM: } E_{0,z} = \psi_n^{TM} \quad \vec{E}_{0,\perp} = \frac{i\,k}{\left(\gamma_n^{TM}\right)^2} \vec{\nabla}_\perp \psi_n^{TM} \quad \vec{H}_{0,\perp} = \frac{1}{Z_{\epsilon\mu}} \frac{i\,k_{\epsilon\mu}}{\left(\gamma_n^{TM}\right)^2} \hat{z} \times \vec{\nabla}_\perp \psi_n^{TM}}$$

$$(9.292)$$

The specific solutions depend on the shape of the boundary chosen.

Generic Properties of TE Modes

For a TE mode, we set $E_{0,z} = 0$ and drop the $E_{0,z}$ wave equation in Equation 9.285 and focus on the $H_{0,z}$ equation. The relevant boundary condition is $\hat{n} \cdot \vec{\nabla}_\perp H_{0,z} = 0 \big|_C$ because the $E_{0,z}$ condition is useless. Rewriting the $H_{0,z}$ piece of Equation 9.285 in a generic form gives

$$\nabla_\perp^2 \psi + \gamma^2 \psi = 0 \qquad \gamma^2 = k_{\epsilon\mu}^2 - k^2 \qquad \hat{n} \cdot \vec{\nabla}\psi \big|_C = 0 \qquad (9.293)$$

Again, we have an eigenvalue-eigenvector problem with a discrete set of eigenvalues $\{\gamma_n^{TE}\}$ due to the boundary condition (now *Neumann*) and a corresponding set of eigenvectors $\{\psi_n^{TE}\}$. *These eigenvalues and solutions are different from the TM case because the boundary condition is different!*. The full solution is then

$$\boxed{\text{TE:} \quad H_{0,z} = \psi_n^{TE} \quad \vec{H}_{0,\perp} = \frac{i\,k}{\left(\gamma_n^{TE}\right)^2} \vec{\nabla}_\perp \psi_n^{TE} \quad \vec{E}_{0,\perp} = -Z_{\epsilon\mu} \frac{i\,k_{\epsilon\mu}}{\left(\gamma_n^{TE}\right)^2} \hat{z} \times \vec{\nabla}_\perp \psi_n^{TE}}$$

$$(9.294)$$

Again, the specific solutions depend on the shape of the boundary chosen.

Complementarity and Completeness of TM and TE Modes

It should be clear that the TM and TE solutions are complementary in that they solve the same eigenvalue-eigenvector equation but with complementary boundary conditions, Dirichlet and Neumann. These two types of boundary conditions are the only types of boundary conditions implied by Maxwell's Equations, so the solutions, along with the TEM mode solutions if they exist, form a complete set to describe any propagating field in the waveguide. As noted above, there is no reason for the γ_n's of the TE and TM modes to be the same (though they can be), and certainly the $\{\psi_n^{TM}\}$ and $\{\psi_n^{TE}\}$ will be different due to the different boundary conditions.

Propagation Properties: Propagation Constant, Wave Speed, Dispersion Relations, and Wave Impedance

This all very simple for TEM modes. There is no eigenvalue-eigenvector equation to be solved, so the propagation constant is always $k = k_{\epsilon\mu} = \omega/v_{\epsilon\mu}$, the wavelength is $\lambda = 2\pi/k_{\epsilon\mu}$, the wave speed is $v_{\epsilon\mu}$, the wave impedance is $Z_{\epsilon\mu}$ independent of frequency, and the *full* fields are always related to each other by $\vec{H} = \hat{k} \times \vec{E}/Z_{\epsilon\mu}$. Unlike TE and TM modes, there is no cutoff frequency, which makes sense because the transverse fields solve the static Maxwell's Equations in two dimensions, so the solutions approach static, z-independent solutions as $k_{\epsilon\mu} \to 0$.

On the other hand, the TE and TM modes satisfy the condition (from $\gamma^2 = k_{\epsilon\mu}^2 - k^2$ and $k_{\epsilon\mu} = \omega/v_{\epsilon\mu}$, with γ_n being γ_n^{TM} or γ_n^{TE})

$$\omega^2 = v_{\epsilon\mu}^2 \left(k^2 + \gamma_n^2 \right) \qquad (9.295)$$

Thus, there is a cutoff frequency $\omega_{c,n} = v_{\epsilon\mu}\,\gamma_n$ such that, for $\omega < \omega_{c,n}$, we find $k^2 < 0$, yielding a decaying mode with decay constant

$$\boxed{\begin{array}{l} \omega_{c,n} = v_{\epsilon\mu}\,\gamma_n \\ \omega < \omega_{c,n} \end{array} \;:\; \kappa_n(\omega) = -i\,k_n(\omega) = \sqrt{\gamma_n^2 - \frac{\omega^2}{v_{\epsilon\mu}^2}} = \frac{\omega_{c,n}}{v_{\epsilon\mu}}\sqrt{1 - \frac{\omega^2}{\omega_{c,n}^2}}} \qquad (9.296)$$

Note that the decay constant depends on which type (TM or TE) and number n of mode one considers! We see that the decay length $d_n(\omega) = 1/\kappa_n(\omega) \to \infty$ as $\omega \to \omega_{c,n}$ from below because propagation must become possible for $\omega > \omega_{c,n}$.

For the propagating modes with $\omega > \omega_{c,n}$, the propagation constant, wavelength, and wave speed are

$$
\boxed{
\begin{array}{l}
\omega_{c,n} = v_{\epsilon\mu}\,\gamma_n \\
\omega > \omega_{c,n}
\end{array}
\quad : \quad
k_n(\omega) = \sqrt{\frac{\omega^2}{v_{\epsilon\mu}^2} - \gamma_n^2} = \frac{\omega}{v_{\epsilon\mu}}\sqrt{1 - \frac{\omega_{c,n}^2}{\omega^2}} = k_{\epsilon\mu}\sqrt{1 - \frac{\omega_{c,n}^2}{\omega^2}}
}
\tag{9.297}
$$

$$
v_n(\omega) = \frac{\omega}{k_n(\omega)} = \frac{v_{\epsilon\mu}}{\sqrt{1 - \frac{\omega_{c,n}^2}{\omega^2}}}
\qquad
\lambda_n(\omega) = \frac{2\pi}{k_n(\omega)} = \frac{\lambda_{\epsilon\mu}}{\sqrt{1 - \frac{\omega_{c,n}^2}{\omega^2}}}
\tag{9.298}
$$

We see that the propagation constant, wavelength, and the wave speed go to the unguided values $k_{\epsilon\mu} = \omega/v_{\epsilon\mu}$, $\lambda_{\epsilon\mu} = 2\pi/k_{\epsilon\mu} = v_{\epsilon\mu}/\nu$, and $v_{\epsilon\mu}$, respectively, as $\omega \to \infty$: the system approaches the transmission-line/free-space limit. If one goes back to Equations 9.280 and 9.281, one sees that ratio of the transverse fields to the longitudinal field increases as $1/(k_{\epsilon\mu}^2 - k^2)$ as $\omega \to \infty$ (i.e., $k \to k_{\epsilon\mu}$), making the longitudinal component small in this limit, as one would expect. Essentially, propagation looks like that in free space as the wavelength becomes small compared to the transverse dimensions: the wave no longer notices the boundaries.

As $\omega \to \omega_{c,n}$, the propagation constant vanishes, the wavelength becomes infinite, and the wave speed becomes infinite. (The latter two must become infinite together because their ratio is ν, which remains finite. Similarly, because $k_n(\omega)\,v_n(\omega) = \omega$, $k_n(\omega) \to 0$ as $v_n(\omega) \to \infty$ so ω can remain constant.)

We recall from our discussion of EM waves in matter that the *wave impedance* is implicitly defined via the relation between $\vec{H}$ and $\vec{E}$, Equation 9.55. The generalization in the case of waveguides uses the *transverse* components of $\vec{H}$ and $\vec{E}$ (following the relation between the transverse components of $\vec{H}$ and $\vec{E}$, Equations 9.286):

$$\vec{H}_{0,\perp} = \frac{1}{Z}\,\widehat{k}\times\vec{E}_{0,\perp} \tag{9.299}$$

Using our relations between $\vec{H}_{0,\perp}$ and $\vec{E}_{0,\perp}$ for the TM and TE modes and the expression for $k_n(\omega)$ we just derived, we obtain

$$Z_n^{\mathsf{TM}}(\omega) = \frac{k_n^{\mathsf{TM}}(\omega)}{k_{\epsilon\mu}}\,Z_{\epsilon\mu} = Z_{\epsilon\mu}\sqrt{1-\frac{\omega_{c,n}^2}{\omega^2}} \quad Z_n^{\mathsf{TE}}(\omega) = \frac{k_{\epsilon\mu}}{k_n^{\mathsf{TE}}(\omega)}\,Z_{\epsilon\mu} = \frac{Z_{\epsilon\mu}}{\sqrt{1-\frac{\omega_{c,n}^2}{\omega^2}}} \tag{9.300}$$

The wave impedance behaves differently as $\omega \to \omega_{c,n}$ for the TM and TE modes! This is explained by the current flow. In the TE mode, $\vec{H}_\perp \propto k_n^{\mathsf{TE}}(\omega) \to 0$ as $\omega \to \omega_{c,n}$ (while $\vec{E}_\perp \propto k_{\epsilon\mu}\,Z_{\epsilon\mu}$). Thus, no current flows longitudinally while the voltage remains finite, so we need $Z_n^{\mathsf{TE}}(\omega) \to \infty$ to obtain this. On the other hand, in the TM mode, $\vec{E}_\perp \propto k_n^{\mathsf{TM}}(\omega) \to 0$ as $\omega \to \omega_{c,n}$ (while $\vec{H}_\perp \propto k_{\epsilon\mu}/Z_{\epsilon\mu}$). In this case, current flows longitudinally but no voltage drop appears, so we need $Z_n^{\mathsf{TM}}(\omega) \to 0$ to obtain this.

In the $\omega \to \infty$ limit, we recover $Z_{\epsilon\mu}$ for both impedances, just as we obtained the free-space limits for propagation constant, wave speed, and wavelength in this limit.

Rectangular Waveguide Modes

Let's now specialize to a particular geometry, which will let us find the modes that solve the previously mentioned eigenvector-eigenvalue problem. Our first example will be rectangular waveguide of height a in x and width b in y. One corner of the waveguide is at $(x, y) = 0$ and the diagonal corner is at $(x, y) = (a, b)$.

As one might expect, we try to solve the eigenvalue-eigenvector equation by separation of variables. Let's first try the TE mode, so $H_{0,z}(x, y)$ is what we will solve for. We assume the usual separation of variables form $H_{0,z}(x, y) = X(x) Y(y)$, which yields:

$$\frac{1}{X(x)} \frac{d^2 X}{dx^2} + \frac{1}{Y(y)} \frac{d^2 Y}{dy^2} + \left(k_{\epsilon\mu}^2 - k^2 \right) = 0 \qquad (9.301)$$

As usual, assume that the first two terms are constants:

$$\frac{1}{X(x)} \frac{d^2 X}{dx^2} = -k_x^2 \qquad \frac{1}{Y(y)} \frac{d^2 Y}{dy^2} = -k_y^2 \qquad k^2 = k_{\epsilon\mu}^2 - k_x^2 - k_y^2 \qquad (9.302)$$

The solutions to the differential equations are clearly exponentials, and they have imaginary arguments for $k_x^2 \geq 0$, which we will find it is most convenient to write as sinusoids:

$$X(x) = A_x \sin k_x x + B_x \cos k_x x \qquad (9.303)$$

Our boundary coundition is $\left.\widehat{n} \cdot \vec{\nabla} H_{0,z}\right|_{\mathcal{C}} = 0$. Explicitly,

$$\left.\frac{dX}{dx}\right|_{x=0,a} = 0 \qquad \Longrightarrow \qquad A_x = 0 \qquad k_x = \frac{m\,\pi}{a} \qquad X(x) \propto \cos\frac{m\,\pi\,x}{a} \quad (9.304)$$

$$\left.\frac{dY}{dy}\right|_{y=0,b} = 0 \qquad \Longrightarrow \qquad A_y = 0 \qquad k_y = \frac{n\,\pi}{b} \qquad Y(y) \propto \cos\frac{n\,\pi\,y}{b} \quad (9.305)$$

$$\Longrightarrow \quad \boxed{\begin{array}{l} \text{rectangular} \\ \text{waveguide} \\ \text{TE}_{mn} \text{ mode} \end{array} \quad H_{0,z} \propto \cos\frac{m\,\pi\,x}{a}\,\cos\frac{n\,\pi\,y}{b} \quad \gamma_{mn}^2 = k_x^2 + k_y^2 = \frac{m^2\pi^2}{a^2} + \frac{n^2\pi^2}{b^2}} \quad (9.306)$$

This is called the TE_{mn} mode. From $H_{0,z}$, we can obtain $\vec{H}_{0,\perp}$ and $\vec{E}_{0,\perp}$ via Equations 9.294.

The propagation constant can now be written

$$k = \sqrt{k_{\epsilon\mu}^2 - k_x^2 - k_y^2} = \sqrt{k_{\epsilon\mu}^2 - \gamma_{mn}^2} = \sqrt{k_{\epsilon\mu}^2 - \frac{\pi^2 m^2}{a^2} - \frac{\pi^2 n^2}{b^2}} \tag{9.307}$$

and thus the cutoff frequency for the TE_{mn} mode is

$$\boxed{\begin{array}{l} \text{rectangular} \\ \text{waveguide} \qquad \omega_{c,mn} = v_{\epsilon\mu}\,\pi\sqrt{\dfrac{m^2}{a^2} + \dfrac{n^2}{b^2}} \\ TE_{mn} \text{ mode} \end{array}} \tag{9.308}$$

The rest of our generic discussion of dispersion relations applies.

The TE mode with the lowest cutoff frequency is either TE_{10} or TE_{01}, depending on whether a or b is larger. (TE_{00} vanishes; try plugging it in to Equation 9.277.) There may be multiple TE_{m0} modes below the first TE_{0n} mode, or vice versa, depending on the relative values of a and b.

Griffiths has a nice discussion of interpreting these modes as the result of a free EM wave propagating through the waveguide with an infinite number of reflections off the walls. It can help to build some intuition for how these modes arise. The actual calculation clearly can only be applied in the simple case of rectangular waveguide.

It is straightforward to obtain the TM modes. The same math applies except the boundary condition is $X(x)|_{x=0,a} = 0$ and $Y(y)|_{y=0,b} = 0$, so one obtains sines instead of cosines (like particle in a box in quantum mechanics):

$$
\boxed{
\begin{array}{l}
\text{rectangular} \\
\text{waveguide} \\
\text{TM}_{mn} \text{ mode}
\end{array}
\qquad
E_{0,z} \propto \sin\frac{m\pi x}{a}\,\sin\frac{n\pi y}{b}
\qquad
\omega_{c,mn} = v_{\epsilon\mu}\,\pi\sqrt{\frac{m^2}{a^2} + \frac{n^2}{b^2}}
}
\tag{9.309}
$$

These sine modes are of course necessary to ensure $E_{0,z}$ vanishes at the walls as required by our boundary conditions. From $E_{0,z}$, we can obviously obtain $\vec{H}_{0,\perp}$ and $\vec{E}_{0,\perp}$ via Equations 9.292.

The cutoff frequencies have the same form as for TE, but, because sines instead of cosines are present, the lowest TM mode is TM_{11}, which has a higher cutoff frequency than both TE_{10} and TE_{01}.

In general, the TM modes for any system require less work to determine because the boundary condition is easier to calculate and enforce. Whenever you do a problem that asks for both the TE and TM modes, do the TM modes first (in contrast to what we have done here!).

On the next few slides, we show some visualizations of the fields and surface charge and current densities for the TE_{10} mode, which is particularly easy to visualize because the fields are independent of y.

TE$_{10}$ mode, xz plane. The mode is independent of y.
Lines: $\vec{H}$; Shading: E_y, σ
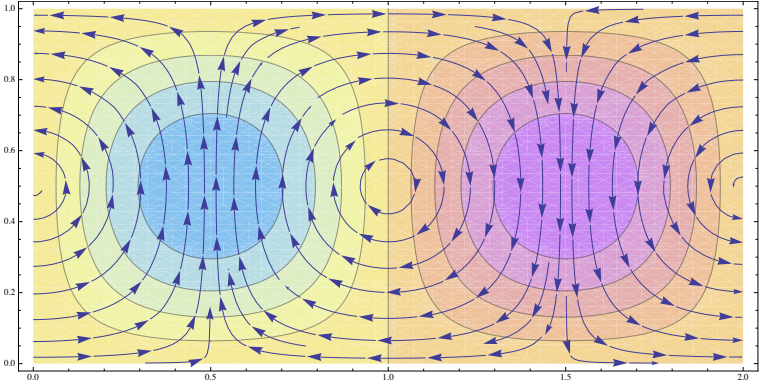
Figure courtesy of M. Cross

TE_{10} mode, xz plane. The mode is independent of y.
Lines: $\vec{K}$; Shading: E_y, σ
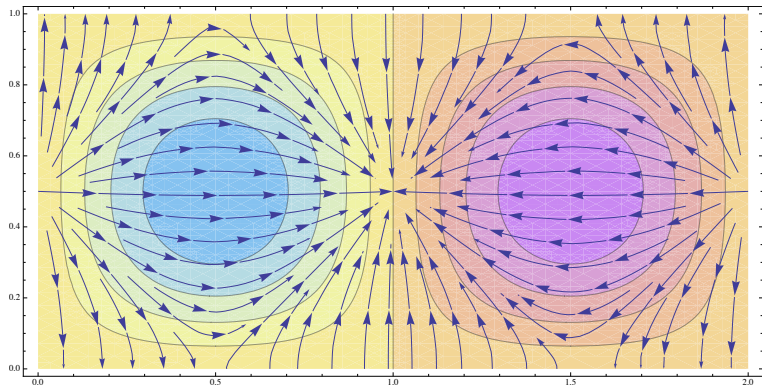
The continuity equation is critical here: The locations of peak charge, where $|\sigma|$ is maximum and $|\partial\sigma/\partial t|$ vanishes, have no net current in or out (i.e., $\vec{\nabla} \cdot \vec{K} = 0$) while the locations of zero charge, where $\sigma = 0$ but $|\partial\sigma/\partial t|$ is maximized, have a net current flow in or out (i.e., $\vec{\nabla} \cdot \vec{K} \neq 0$).

Figure courtesy of M. Cross

TE_{10} mode. The mode is independent of y.
This figure is displaced by $\lambda/4$ to the left relative to the cross sections.
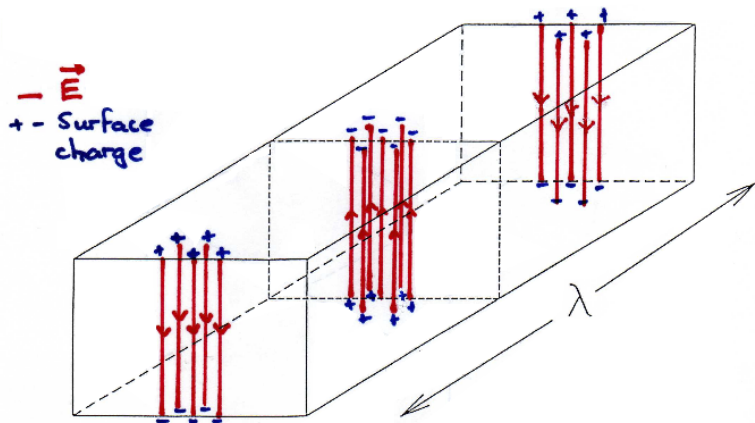
Note again on this and the following slide the $\lambda/4$ displacement between locations of peak $|\sigma|$ and of peak $|\vec{\nabla} \cdot \vec{K}|$.

Figure courtesy of M. Cross

TE_{10} mode. The mode is independent of y.
This figure is displaced by $\lambda/4$ to the left relative to the cross sections.

— $\vec{B}$ (uniform in y)

— Surface Current

λ
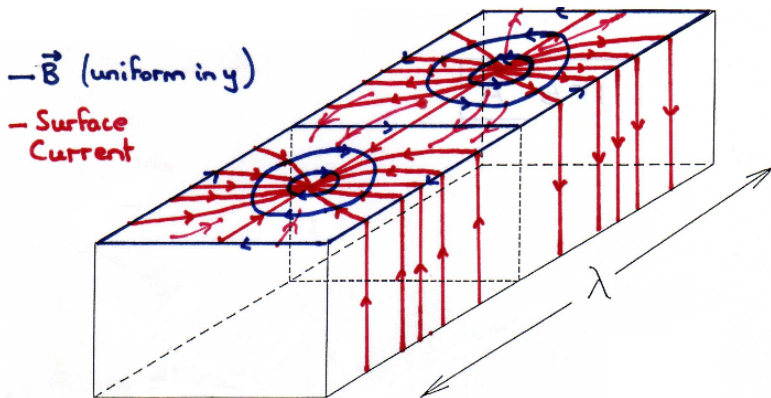
Figure courtesy of M. Cross

Lecture 30:

Electromagnetic Waves VIII:
Energy in Waveguides Waveguides with Finite Conductivity

Date Revised: 2024/05/09 05:00
Correct error in first equation in Equation 9.325
Date Given: 2024/05/07

Energy, Poynting Vector, and Group Velocity

Let's calculate the Poynting Vector for a waveguide:

$$\left\langle \vec{S} \right\rangle = \frac{1}{2} \mathcal{R} \left(\left\langle \vec{E}^* \times \vec{H} \right\rangle \right) \tag{9.310}$$

Let's evaluate $\vec{E}^* \times \vec{H}$ for a particular mode TM mode n (remember, $\hat{z} \times \hat{z} = 0$ and $H_{0,z} = 0$!):

$$\vec{E}^* \times \vec{H} = E_{0,z}^* \, \hat{z} \times \vec{H}_{0,\perp} + \vec{E}_{0,\perp}^* \times H_{0,z} \, \hat{z} + \vec{E}_{0,\perp}^* \times \vec{H}_{0,\perp}$$

$$\overset{\text{TM},n}{=} E_{0,z,n}^* \, \hat{z} \times \left(\frac{1}{Z_{\epsilon\mu}} \frac{i \, k_{\epsilon\mu}}{\gamma_n^2} \, \hat{z} \times \vec{\nabla}_\perp E_{0,z,n} \right)$$

$$+ \left(\frac{i \, k_n^{\text{TM}}(\omega)}{\gamma_n^2} \vec{\nabla}_\perp E_{0,z,n} \right)^* \times \left(\frac{i \, k_{\epsilon\mu}}{Z_{\epsilon\mu} \gamma_n^2} \, \hat{z} \times \vec{\nabla}_\perp E_{0,z,n} \right)$$

The expression is complicated because the fields are not purely transverse and there are phase shifts among the components.

Using the $BAC - CAB$ rule for the triple-vector products, and also $\hat{z} \cdot \vec{\nabla}_\perp = 0$, we obtain (now showing the TE analogue)

$$\vec{E}^* \times \vec{H} \overset{\text{TM},n}{=} -\frac{1}{Z_{\epsilon\mu}} \frac{i\,k_{\epsilon\mu}}{\gamma_n^2} E_{0,z,n}^* \vec{\nabla}_\perp E_{0,z,n} + \hat{z} \frac{1}{Z_{\epsilon\mu}} \frac{k_n^{\text{TM}}(\omega)\,k_{\epsilon\mu}}{\gamma_n^4} \left| \vec{\nabla}_\perp E_{0,z,n} \right|^2 \quad (9.311)$$

$$\vec{E}^* \times \vec{H} \overset{\text{TE},n}{=} Z_{\epsilon\mu} \frac{i\,k_{\epsilon\mu}}{\gamma_n^2} H_{0,z,n} \vec{\nabla}_\perp H_{0,z,n}^* + \hat{z}\, Z_{\epsilon\mu} \frac{k_n^{\text{TE}}(\omega)\,k_{\epsilon\mu}}{\gamma_n^4} \left| \vec{\nabla}_\perp H_{0,z,n} \right|^2 \quad (9.312)$$

The first term points transverse to the direction of propagation. Since the walls are perfectly conducting, there can be no time-averaged, net energy flow in that direction. So, when we time-average, we neglect it. That leaves us with

$$\left\langle \vec{S}_n^{\text{TM}} \right\rangle = \frac{\hat{z}}{2} \frac{\omega\, k_n^{\text{TM}}(\omega)}{\gamma_n^4} \epsilon \left| \vec{\nabla}_\perp E_{0,z,n} \right|^2 \qquad \left\langle \vec{S}_n^{\text{TE}} \right\rangle = \frac{\hat{z}}{2} \frac{\omega\, k_n^{\text{TE}}(\omega)}{\gamma_n^4} \mu \left| \vec{\nabla}_\perp H_{0,z,n} \right|^2 \quad (9.313)$$

Notice how the use of $\vec{H}$ instead of $\vec{B}$ makes the expressions for the two modes very similar in form.

Let's integrate $\left\langle \vec{S} \right\rangle \cdot \hat{z}$ over the waveguide cross-sectional area to get the total power. Using ψ_n to represent $E_{0,z,n}$ or $H_{0,z,n}$ as appropriate, we can manipulate the area integral using the two-dimensional analogue of Green's First Identity (Equation 3.12):

$$\int_{\mathcal{S}} da \left| \vec{\nabla}_\perp \psi_n \right|^2 = \oint_{\mathcal{C}(\mathcal{S})} d\ell \, \psi_n^* \, \hat{n} \cdot \vec{\nabla}_\perp \psi_n - \int_{\mathcal{S}} da \, \psi_n^* \, \nabla_\perp^2 \psi_n \tag{9.314}$$

(We are considering power in a single mode for now. We'll generalize shortly.)

The first term vanishes due to our boundary conditions: either ψ_n or $\hat{n} \cdot \vec{\nabla}_\perp \psi_n$ always vanishes on the boundary. The second term can be transformed using the eigenvector-eigenvalue equation that yields the solutions ψ_n. Thus,

$$\int da \left| \vec{\nabla}_\perp \psi_n \right|^2 = \gamma_n^2 \int_{\mathcal{S}} da \, |\psi_n|^2 \tag{9.315}$$

Thus, the time-averaged power flow in a single mode is

$$\boxed{\left\langle P_n^{\text{TM}} \right\rangle = \frac{1}{2} \frac{\omega\, k_n^{\text{TM}}(\omega)}{\gamma_n^2} \epsilon \int_{\mathcal{S}} da\, |E_{0,z,n}|^2 = \frac{1}{2} v_{\epsilon\mu} \frac{\omega^2}{\omega_{c,n}^2} \sqrt{1 - \frac{\omega_{c,n}^2}{\omega^2}}\, \epsilon \int_{\mathcal{S}} da\, |E_{0,z,n}|^2} \quad (9.316)$$

$$\boxed{\left\langle P_n^{\text{TE}} \right\rangle = \frac{1}{2} \frac{\omega\, k_n^{\text{TE}}(\omega)}{\gamma_n^2} \mu \int_{\mathcal{S}} da\, |H_{0,z,n}|^2 = \frac{1}{2} v_{\epsilon\mu} \frac{\omega^2}{\omega_{c,n}^2} \sqrt{1 - \frac{\omega_{c,n}^2}{\omega^2}}\, \mu \int_{\mathcal{S}} da\, |H_{0,z,n}|^2} \quad (9.317)$$

(Note the symmetry between E and H in the above; the expressions would look more different if B were used, the μ would be in the denominator.) If we do similar integrals to calculate the time-averaged energy per unit length in the waveguide, and we obtain

$$\boxed{\left\langle U_n^{\text{TM}} \right\rangle = \frac{1}{2} \frac{\omega^2}{\omega_{c,n}^2} \epsilon \int_{\mathcal{S}} da\, |E_{0,z,n}|^2 \qquad \left\langle U_n^{\text{TE}} \right\rangle = \frac{1}{2} \frac{\omega^2}{\omega_{c,n}^2} \mu \int_{\mathcal{S}} da\, |H_{0,z,n}|^2} \quad (9.318)$$

There are a few interesting comments to make:

▶ We do not have $\langle P_n \rangle = v_{\epsilon\mu} \langle U_n \rangle$ as we had in free space, or even $\langle P_n \rangle = v_n(\omega) \langle U_n \rangle$ as we might expect given the guided wave speed. This makes sense, as $v_n(\omega) > v_{\epsilon\mu}$, so $v_n(\omega)$ can exceed the speed of light. The quantity relating $\langle P_n \rangle$ and $\langle U_n \rangle$ is

$$v_{\epsilon\mu} \sqrt{1 - \frac{\omega_{c,n}^2}{\omega^2}} = \frac{d\omega}{dk_n(\omega)} \tag{9.319}$$

which one can see by differentiating the dispersion relation $\omega^2 = v_{\epsilon\mu}^2 \left([k_n(\omega)]^2 + \gamma_n^2\right)$:

$$2\omega \, d\omega = v_{\epsilon\mu}^2 \left(2 k_n \, dk_n + 0\right) \quad \Longrightarrow \quad \frac{d\omega}{dk_n} = v_{\epsilon\mu}^2 \frac{k_n(\omega)}{\omega} = v_{\epsilon\mu} \sqrt{1 - \frac{\omega_{c,n}^2}{\omega^2}} \tag{9.320}$$

This speed is the *group velocity*, $v_{g,n}(\omega)$. You are aware from Ph2a/12a how the group velocity is the speed at which a wave packet propagates. The group velocity here can never exceed $v_{\epsilon\mu}$, thus preventing the power flow from going faster than the speed of light. So, we have

$$\boxed{\langle P_n \rangle = v_{g,n}(\omega) \langle U_n \rangle \qquad v_{g,n}(\omega) = v_{\epsilon\mu} \sqrt{1 - \frac{\omega_{c,n}^2}{\omega^2}} \qquad v_{g,n}(\omega) \, v_n(\omega) = v_{\epsilon\mu}^2}$$

$$\tag{9.321}$$

▶ Let's consider the relative size of the longitudinal and transverse field energies per unit length, $\langle U_{z,n} \rangle$ and $\langle U_{\perp,n} \rangle$. It makes sense to talk separately about energies separately because of the dot products $\vec{E}^* \cdot \vec{E}$ and $\vec{H}^* \cdot \vec{H}$ in those expressions: there are no cross-terms between the longitudinal and transverse pieces. The energy expression implies that a fraction $\omega_{c,n}^2/\omega^2$ of the total energy is in $\langle U_{z,n} \rangle$ because we divide $\langle U_{z,n} \rangle$ by this quantity to get the total energy per unit length. That is already interesting.

Using this fraction, we see that, in the limit $\omega \to \omega_{c,n}$, the longitudinal field dominates the energy and the longitudinal field becomes large compared to the transverse fields.

Conversely, in the $\omega \to \infty$ limit, the longitudinal field energy becomes negligible compared to the transverse field energy and the longitudinal field becomes small compared to the transverse field. The latter behavior is consistent with the way the phase velocity, wavelength, and impedance approach the unguided values in the high-frequency limit. This fact was not quite evident in Equations 9.292 and 9.294 because of the $\vec{\nabla}_{\perp}$ in those equations.

▶ If we consider the energy in the transverse field, the amount in the electric and magnetic components is no longer equal as it is for unguided waves or TEM guided waves. For guided waves, we have

$$u_{m,\perp,n} = \frac{\mu}{2}|\vec{H}_{0,\perp}|^2 = \frac{\mu}{2}\frac{1}{Z_n^2}|\vec{E}_{0,\perp}|^2 = \frac{Z_{\epsilon\mu}^2}{Z_n^2}\frac{\epsilon}{2}|\vec{E}_{0,\perp}|^2 = \frac{Z_{\epsilon\mu}^2}{Z_n^2}u_{e,\perp,n} \quad (9.322)$$

$$u_{m,\perp,n}^{\mathsf{TM}} = \left(1 - \omega_{c,n}^2/\omega^2\right)^{-1}u_{e,\perp,n}^{\mathsf{TM}} \qquad u_{m,\perp,n}^{\mathsf{TE}} = \left(1 - \omega_{c,n}^2/\omega^2\right)u_{e,\perp,n}^{\mathsf{TE}} \quad (9.323)$$

(For unguided and TEM waves, $u_m = u_{m,\perp} = u_{e,\perp} = u_e$ at all frequencies.) This effect arises because the characteristic impedance is not $Z_{\epsilon\mu}$ and is frequency-dependent. We also see that one field dominates the transverse energy as $\omega \to \omega_{c,n}$ ($\vec{H}_{0,\perp}$ for TM, $\vec{E}_{0,\perp}$ for TE) while the energies in the two fields become equal in the $\omega \to \infty$, unguided limit. This behavior reflects that discussed in the context of the frequency-dependent impedance earlier: $\vec{E}_{0,\perp} \to 0$ for TM and $\vec{H}_{0,\perp} \to 0$ for TE as $\omega \to \omega_{c,n}$. Note also that it is the same field as the longitudinal field that becomes smaller in the low-frequency limit: $\vec{E}_{0,\perp}$ for TE ($E_z \neq 0$) and $\vec{H}_{0,\perp}$ for TM ($H_z \neq 0$).

▶ We thus see from all of the above that we can make general statements about relative energies:

$$\frac{\left\langle U_{e,\perp,n}^{\mathrm{TM}}\right\rangle}{\left\langle U_{m,\perp,n}^{\mathrm{TM}}\right\rangle} = 1 - \frac{\omega_{c,n}^2}{\omega^2} \qquad \frac{\left\langle U_{e,\perp,n}^{\mathrm{TM}}\right\rangle + \left\langle U_{m,\perp,n}^{\mathrm{TM}}\right\rangle}{\left\langle U_n^{\mathrm{TM}}\right\rangle} = 1 - \frac{\omega_{c,n}^2}{\omega^2} \qquad \frac{\left\langle U_{z,n}^{\mathrm{TM}}\right\rangle}{\left\langle U_n^{\mathrm{TM}}\right\rangle} = \frac{\omega_{c,n}^2}{\omega^2}$$

(9.324)

$$\frac{\left\langle U_{m,\perp,n}^{\mathrm{TE}}\right\rangle}{\left\langle U_{e,\perp,n}^{\mathrm{TE}}\right\rangle} = 1 - \frac{\omega_{c,n}^2}{\omega^2} \qquad \frac{\left\langle U_{m,\perp,n}^{\mathrm{TE}}\right\rangle + \left\langle U_{e,\perp,n}^{\mathrm{TE}}\right\rangle}{\left\langle U_n^{\mathrm{TE}}\right\rangle} = 1 - \frac{\omega_{c,n}^2}{\omega^2} \qquad \frac{\left\langle U_{z,n}^{\mathrm{TE}}\right\rangle}{\left\langle U_n^{\mathrm{TE}}\right\rangle} = \frac{\omega_{c,n}^2}{\omega^2}$$

(9.325)

We have to be careful about inferring from these relations statements about field magnitudes because the above relations rely on integration over the cross-sectional area.

What about when one has a wave with power in multiple modes? Do the above formulae hold? Yes. Suppose one has power in modes a and b. Then the time-averaged Poynting vector will be

$$\left\langle \vec{S} \right\rangle = \frac{1}{2} \mathcal{R} \left(\left\langle \left(c_a \vec{E}_a + c_b \vec{E}_b \right)^* \times \left(c_a \vec{H}_a + c_b \vec{H}_b \right) \right\rangle \right) \tag{9.326}$$

where c_a and c_b are the coefficients of the two modes. Consider one of the cross terms, under the assumption that both a and b are TM modes, and apply the same kind of algebra we used before to obtain:

$$\vec{E}_a^* \times \vec{H}_b \overset{\text{TM}}{=} -\frac{1}{Z_{\epsilon\mu}} \frac{i\,k_{\epsilon\mu}}{\gamma_b^2} E_{0,z,a}^* \vec{\nabla}_\perp E_{0,z,b} + \hat{z}\, \frac{k_a^{\text{TM}}(\omega)\, k_{\epsilon\mu}}{Z_{\epsilon\mu}\, \gamma_a^2\, \gamma_b^2} \vec{\nabla}_\perp E_{0,z,a}^* \cdot \vec{\nabla}_\perp E_{0,z,b} \tag{9.327}$$

Again, the first term does not yield propagating power, so we neglect it. When we integrate the second term over the waveguide cross section, we can see it vanishes via the same kind of application of the eigenvector-eigenvalue equation:

$$\int_{\mathcal{S}} \vec{\nabla}_\perp \psi_a^* \cdot \vec{\nabla}_\perp \psi_b = \gamma_b^2 \int_{\mathcal{S}} da\, \psi_a^* \psi_b = 0 \tag{9.328}$$

where the final vanishing occurs because the solutions of any eigenvector-eigenvalue equation form an orthonormal basis and the integration over the cross section is the appropriate inner product (like $\langle \psi_a | \psi_b \rangle$ in quantum mechanics).

Via a similar calculation, one can show the other cross term vanishes when integrated over the cross section $\mathcal{S}$. This leaves only the non-cross-terms, which yield the power in the individual modes. That is:

$$\left\langle P^{\mathrm{TM}} \right\rangle = |c_a|^2 \left\langle P_a^{\mathrm{TM}} \right\rangle + |c_b|^2 \left\langle P_b^{\mathrm{TM}} \right\rangle \tag{9.329}$$

One can show the same thing via a similar technique when a and b are both TE modes. When one is TE and one is TM, we apply the same technique yet again. Orthogonality still applies: the TE and TM modes solve the same eigenvector-eigenvalue equation but with different boundary conditions, so any pair of TE and TM modes must be orthogonal. Thus, we have the general result

$$\boxed{\langle P \rangle = \sum_m |c_m^{\mathrm{TM}}|^2 \left\langle P_m^{\mathrm{TE}} \right\rangle + \sum_n |c_n^{\mathrm{TE}}|^2 \left\langle P_n^{\mathrm{TE}} \right\rangle} \tag{9.330}$$

where c_m^{TM} and c_n^{TE} are the coefficients of the TM,m and TE,n modes. The same of course holds for the energy density:

$$\boxed{\langle U \rangle = \sum_m |c_m^{\mathrm{TM}}|^2 \left\langle U_m^{\mathrm{TE}} \right\rangle + \sum_n |c_n^{\mathrm{TE}}|^2 \left\langle U_n^{\mathrm{TE}} \right\rangle} \tag{9.331}$$

Waveguides with Finite Conductivity: Overview

Before considering imperfect conductors, let's summarize perfect conductors:

▶ The conductivity is infinite, so, even when $\vec{E}$ is varying, the surface charge redistributes itself so that $\vec{E}_c$, the field inside the conductor, vanishes. Our boundary conditions from Gauss's Law and Faraday's Law imply that there is a discontinuity in $\hat{n} \cdot \vec{E}$ due to this surface charge density while $\hat{t} \cdot \vec{E}$ is continuous, so the electric field outside the conductor is normal to its surface. Our TE and TM mode boundary conditions are consistent with these conditions.

▶ What about the magnetic field inside the conductor?

Since $\vec{E}_c$ vanishes, we know there can be no displacement current sourcing a magnetic field. However, trying to work out from Ohm's Law whether there is or is not a true current is difficult: $E_c \to 0$ but $\sigma \to \infty$, so $J = \sigma E_c$ is indeterminate.

Instead, let's take the $\sigma \to \infty$ limit of the behavior in good conductors as worked out in Section 4. There, we learned that, in a good conductor, both E and B fall off exponentially with decay length given by the skin depth $\delta = \sqrt{2/\mu\,\sigma\,\omega}$. The skin depth thus vanishes in a perfect conductor, implying that H (and thus B) also does not penetrate into the perfect conductor.

Coming back to the current, it is reasonable to conclude that, if both fields vanish in the conductor, there can be no J: if there were, it would generate B through Ampere's Law. (We are assured above that there is no cancelling displacement current term.)

Continuity of $\hat{n} \cdot \vec{B}$ implies that $\hat{n} \cdot \vec{B} = 0$ outside the conductor. On the other hand, a discontinuity in $\hat{t} \cdot \vec{H}$ is allowed if there is a surface current density $\vec{K}$. That current density, along with the vanishing of $\vec{B}$ in the conductor, implies that, just outside the conductor, $\hat{n} \times \vec{H} = \vec{K}$: the magnetic field outside the conductor is perfectly tangential.

Note that the conclusion that the magnetic field is tangential just outside the conductor — i.e., that $\hat{n} \cdot \vec{B} = 0$ at the boundary — is consistent with the boundary conditions we applied for perfect waveguides earlier.

▶ Let's revisit the continuity of, and thus vanishing of, the tangential component of $\vec{E}$ in light of the surface current density. This continuity continues to hold: Faraday's Law tells us that there will only be a discontinuity in $\widehat{t} \cdot \vec{E}$ if the magnetic field in the $\widehat{s}$ direction ($\widehat{s} = \widehat{t} \times \widehat{n}$) has δ-function behavior at the surface, and we have argued many times before that fields cannot have such singularities (though charge and current densities can).

The lack of a tangential $\vec{E}$ is consistent with the existence of the δ-function-like surface current density $\vec{K}$ because $\sigma \to \infty$. (Note the different result from the J case in spite of the same indeterminacy being present.)

▶ This surface current density's magnetic field cancels that of the external field to ensure no magnetic field propagates into the interior of the conductor (much the same way as the surface charge density does this for the electric field).

The above fact that magnetic fields vanish in perfect conductors when there is no free current source in the conductor has been hinted at before in homework and in our discussion of TEM modes for transmission lines, but we have never discussed it in as general a setting as this before. For our waveguide mode derivation, we only required $\widehat{n} \cdot \vec{\nabla}_{\perp} H_{0,z}$ vanish at the boundary.

Because of the importance of the surface current density, we will continue to work with $\vec{H}$ rather than $\vec{B}$, as we have been doing for some time.

Now, let's consider a waveguide with walls having *good* conductivity. There are two adjustments to be made:

▶ Our boundary conditions on the waveguide solutions no longer hold exactly because we may not assume that the electric and magnetic fields vanish perfectly inside the conductor: as we know, they decay over a skin depth. The condition $E_{0,z}|_{\mathcal{C}} = 0$ thus obviously does not hold. If we recall the derivation of $\widehat{n} \cdot \vec{\nabla}_{\perp} H_{0,z}\big|_{\mathcal{C}} = 0$, it started from the fact that $\widehat{t} \cdot \vec{E} = 0$ because $E_c = 0$. This condition is no longer true, so the conclusion $\widehat{n} \cdot \vec{\nabla}_{\perp} H_{0,z}\big|_{\mathcal{C}} = 0$ is no longer valid either.

▶ Moreover, because σ is finite, Ohm's Law is applicable and thus we also know that a δ-function-like surface current density is not physically allowed because it would require a δ-function-like electric field for finite σ. Therefore, there is no surface current to shield the conductor from the magnetic field in the waveguide medium, while there is nonzero J driven by E_c. Nonzero J and finite σ therefore imply there is power dissipation in the walls due to the currents flowing in a resistive material, and thus there must be attenuation of the wave.

It is quite difficult to include these effects exactly, but we can derive some approximate results in the case of a good conductor.

Specifically, we assume a successive approximation scheme where the fields outside the conductor may now have $\hat{t} \cdot \vec{E}$ and $\hat{n} \cdot \vec{B}$ components, but they are small compared to $\hat{n} \cdot \vec{E}$ and $\hat{t} \cdot \vec{B}$. We obtain the dominant components of the field outside the walls from the boundary value problem solutions for perfectly conductive walls (our waveguide solutions), we use the good-conductor boundary conditions to obtain from them the dominant components of the fields in the walls, and then we again use the good-conductor boundary conditions to obtain from these the corrections to the fields outside the walls, and finally we use the boundary conditions again to find the correction fields in the walls. We find these correction fields in the waveguide are indeed small compared to the perfect-conductor fields in the waveguide, justifying the approximation scheme. More importantly, finding the fields in the walls allows us to calculate the Joule dissipation and thus the attenuation of the wave in the waveguide. For the latter, after zooming in to find the fields in the conducting walls and the correction fields, we will zoom back out, treating the quickly decaying current density as a perfect sheet current and the fields as going to zero perfectly.

Fields Near the Walls for Waveguides with Finite Conductivity

The fields inside the conductor must be those we obtained for EM waves in good conductors (Section 4). We just need to know how to normalize them. The normalization is found by using the fact that $\hat{t} \cdot \vec{H}$ is continuous and $\hat{n} \cdot \vec{H} \approx 0$, so the EM wave in the conductor looks like a wave whose $\vec{H}$ field magnitude is given by $\hat{t} \cdot \vec{H}$ outside the conductor.

Let's also think about the propagation directions inside and outside the conductor. Outside the conductor, the $\vec{E}$ field is close to normal to the surface because $\hat{t} \cdot \vec{E} \approx 0$ while $\hat{n} \cdot \vec{E} \neq 0$. That is, $\theta_i \approx \pi/2$: we have glancing incidence! You showed in Ph106c Problem Set 4 #1 that the general expression for the propagation vector inside a good conductor is

$$\vec{k}_t = -\hat{n} \frac{1+i}{\delta} + \hat{w} \frac{\omega}{v_{\epsilon\mu}} \sin\theta_i \tag{9.332}$$

where the first term has $-\hat{n}$ rather than $\hat{n}$ because we use here a direction convention for $\hat{n}$ opposite to that in Problem Set 4. Since $\omega/v_{\epsilon\mu} = k_{\epsilon\mu}$ and $1/\delta \to k_{\epsilon\mu}\sqrt{\sigma/(2\,\epsilon\,\omega)} \gg k_{\epsilon\mu}$ in the good conductor limit (Equation 9.140), we see that $\hat{k}_t \approx -\hat{n}$ even if $\theta_i = \pi/2$! Thus, we are assured that the propagation direction in the conductor is normal to the interface. (Note, for our waveguide solutions, $k < k_{\epsilon\mu}$, so the above dominance of the normal component holds even accounting for the full frequency dependence of k from our discussion of dispersion relations.)

With this information about the propagation direction in the conductor, we define ξ to be the coordinate along $\widehat{k}_t = -\widehat{n}$ (with the minus sign because we have defined $\widehat{n}$ to point *out of* the conductor for waveguides) and η to be the coordinate perpendicular to $\widehat{k}_t$ (along the waveguide boundary). Then, we have (with $_0$ subscripts used so we can avoid writing everywhere the $e^{i(kz-\omega t)}$ factor)

$$\vec{H}_{0,c}(\xi, \eta) = \vec{H}_0(\xi = 0, \eta)\, e^{-\xi/\delta}\, e^{i\,\xi/\delta} \tag{9.333}$$

where $\vec{H}_0(\xi = 0, \eta)$ is the perfect waveguide field at the waveguide wall and $\delta = \sqrt{2/\mu_c\,\sigma\,\omega}$ is the skin depth in the conductor. We know we can neglect $\widehat{n} \cdot \vec{H}_c$ for two reasons: $\widehat{k}_t$ is approximately normal to the walls and even EM waves in conductors are transverse; and $\widehat{n} \cdot \vec{H}_0(\xi = 0) \approx 0$ due to the perfect conductor solution having $\widehat{n} \cdot \vec{B} = 0$ outside the conductor because $\vec{B}_c = 0$.

We can obtain the component of $\vec{E}_c$ parallel to the interface from the relation between $\vec{H}$ and $\vec{E}$ for good conductors, Equation 9.153, yielding (recall, $Z_\sigma = 1/(\sigma\,\delta)$)

$$\vec{E}_{0,c}(\xi, \eta) = \frac{2\,Z_\sigma}{1+i}\,\widehat{n} \times \vec{H}_{0,c}(\xi, \eta) = \frac{2\,Z_\sigma}{1+i}\,\widehat{n} \times \vec{H}_0(\xi = 0, \eta)\, e^{-\xi/\delta}\, e^{i\,\xi/\delta} \tag{9.334}$$

where again $\widehat{n}$ points *outward* from the conductor into the waveguide medium. We thus have the dominant components of the fields inside the conductor.

Rewriting these fields with the full spatial and time dependence:

$$\vec{H}_c(\xi, \eta, z, t) = \vec{H}_0(\xi = 0, \eta) \, e^{-\xi/\delta} \, e^{i\,\xi/\delta} \, e^{i(k\,z - \omega\,t)} \tag{9.335}$$

$$\vec{E}_c(\xi, \eta, z, t) = \frac{2\,Z_\sigma}{1 + i} \, \hat{n} \times \vec{H}_0(\xi = 0, \eta) \, e^{-\xi/\delta} \, e^{i\,\xi/\delta} \, e^{i(k\,z - \omega\,t)} \tag{9.336}$$

Using boundary conditions to connect the fields inside the conductor to those outside, we may now calculate the correction terms to the fields outside the conductor. The tangential electric field outside the conductor is given by $\vec{E}_c(\xi = 0, \eta, z, t)$ by continuity of this component. The normal component of $\vec{H}$ outside the conductor is found using Faraday's Law (now referring to the full fields, without $_0$ subscripts, because we need the full spatial and time dependence for Faraday's Law):

$$-\mu \, \hat{n} \cdot \frac{\partial \vec{H}}{\partial t}\bigg|_{\xi=0} = \hat{n} \cdot \left(\vec{\nabla} \times \vec{E}\big|_{\xi=0} \right) \tag{9.337}$$

For evaluating the curl, consider a temporary primed coordinate system, distinct from the one we have been using from the waveguide, with $\hat{z}' = \hat{n}$, $\hat{x}' \propto \vec{H}_{0,c}(\xi = 0, \eta)$, and $\hat{y}' \propto \vec{E}_{0,c}(\xi = 0), \eta) \propto \hat{n} \times \vec{H}_{0,c}(\xi = 0, \eta)$. In this coordinate system,

$$\hat{n} \cdot \left(\vec{\nabla} \times \vec{E} \right) = \hat{z}' \cdot \left(\vec{\nabla} \times \vec{E} \right) = \frac{\partial E_{y'}}{\partial x'} - \frac{\partial E_{x'}}{\partial y'} = \left(\hat{x}' \cdot \vec{\nabla} \right) \left(\hat{y}' \cdot \vec{E} \right) - \left(\hat{y}' \cdot \vec{\nabla} \right) \left(\hat{x}' \cdot \vec{E} \right) \tag{9.338}$$

Given how we have defined the primed unit vectors, $\widehat{x}' \cdot \vec{E} = 0$, so we only have the first term (now evaluating the time derivative $\partial/\partial t$ as $-i\,\omega$):

$$i\,\mu\,\omega\,\widehat{n} \cdot \vec{H}(\xi = 0, t) = \left(\widehat{x}' \cdot \vec{\nabla}\right) E(\xi = 0, t) \tag{9.339}$$

$$\widehat{n} \cdot \vec{H}(\xi = 0, t) = -\frac{\mu_c}{\mu}\, \frac{\delta}{1 - i}\, \left(\widehat{h} \cdot \vec{\nabla}\right) \left[\widehat{h} \cdot \vec{H}(\xi = 0)\right] \tag{9.340}$$

where we have defined $\widehat{h} = \widehat{x}'$ to be the tangential unit vector at the boundary that is parallel to $\vec{H}$ at the boundary (which will be along $\widehat{t}_\perp = \widehat{z} \times \widehat{n}$ for a TM mode but will be some linear combination of $\widehat{t}_\perp$ and $\widehat{z}$ for a TE mode).

Since $\widehat{n} \cdot \vec{B}$ is continuous, this equation also now gives the normal component of $\vec{B}_c$, the magnetic field inside the conductor, at the interface, which is a small correction to $\vec{B}_c$ (and thus $\vec{H}_c$).

What is the typical magnitude of the normal magnetic field? $\widehat{h} \cdot \vec{\nabla}$ has two components: one along $\widehat{t}_\perp$ and one along $\widehat{z}$. The $\widehat{z}$ piece will yield something proportional to k, while the $\widehat{t}_\perp$ piece will yield something proportional to γ_n, where $\gamma_n^2 = k_x^2 + k_y^2$. Since $k^2 + \gamma_n^2 = k_{\epsilon\mu}^2$, the $\widehat{h} \cdot \vec{\nabla}$ derivative will always bring in factor scaling like $k_{\epsilon\mu}$. (Note that it is $k_{\epsilon\mu}$ that appears, not k, so we need not worry about $k \to 0$ as $\omega \to \omega_{c,n}$.) Thus, the magnitude of the above gradient is given by $k_{\epsilon\mu} H_{0,z}$, so the ratio of the normal component of the magnetic field to the tangential component is $\delta/\lambda_{\epsilon\mu}$, which is small, as we expected.

Aside from seeing they are small, it is interesting to look at the magnitudes of the correction fields relative to the original waveguide fields more carefully. We have, with all fields evaluated right at the boundary ($\xi = 0$):

$$H_c = H_0 \tag{9.341}$$

$$E_c = Z_\sigma H_c \approx Z_\sigma H_0 = \frac{Z_\sigma}{Z_{\epsilon\mu}} Z_{\epsilon\mu} H_0 \approx \frac{Z_\sigma}{Z_{\epsilon\mu}} E_0 \ll E_0 \tag{9.342}$$

$$E_t \equiv \hat{t} \cdot \vec{E}\Big|_{\mathcal{C}} = E_c \ll E_0 \tag{9.343}$$

$$H_n \equiv \hat{n} \cdot \vec{H}\Big|_{\mathcal{C}} \approx \frac{\mu_c}{\mu} \delta\, k_{\epsilon\mu} H_0 \sim \delta\, k_{\epsilon\mu} H_0 \ll H_0 \tag{9.344}$$

Next, using
$$\frac{Z_\sigma}{Z_{\epsilon\mu,c}} = \frac{1}{\sigma\,\delta}\sqrt{\frac{\epsilon_c}{\mu_c}} = \frac{1}{\sigma}\sqrt{\frac{\mu_c\,\sigma\,\omega}{2}}\sqrt{\frac{\epsilon_c}{\mu_c}} = \sqrt{\frac{\epsilon_c\,\omega}{2\,\sigma}} = \sqrt{\frac{\omega\,\tau}{2}} = \frac{k_{\epsilon\mu,c}\,\delta}{2} \tag{9.345}$$

we see
$$\frac{E_t}{E_0} = \frac{Z_\sigma}{Z_{\epsilon\mu}} = \frac{Z_\sigma}{Z_{\epsilon\mu,c}}\frac{Z_{\epsilon\mu,c}}{Z_{\epsilon\mu}} = \frac{k_{\epsilon\mu,c}\,\delta}{2}\frac{Z_{\epsilon\mu,c}}{Z_{\epsilon\mu}} = \frac{k_{\epsilon\mu}\,\delta}{2}\frac{k_{\epsilon\mu,c}}{k_{\epsilon\mu}}\frac{Z_{\epsilon\mu,c}}{Z_{\epsilon\mu}} \tag{9.346}$$

$$= \frac{k_{\epsilon\mu}\,\delta}{2}\sqrt{\frac{\epsilon_c\,\mu_c}{\epsilon\,\mu}\frac{\mu_c}{\epsilon_c}\frac{\epsilon}{\mu}} = \frac{\mu_c}{\mu}\frac{k_{\epsilon\mu}\,\delta}{2} \approx \frac{H_n}{H_0} \qquad \text{as we might have expected!}$$

Power Flow for Waveguides with Finite Conductivity

In the good conductor approximation, we expect the wave in the waveguide to maintain its functional form but to decay with some decay length that is large compared to the wavelength (in the waveguide, not the conductor!). We can determine this decay constant by considering the (mean) Joule dissipation in the conductor obtained from knowing the (approximate) fields in the conductor. This power dissipated per unit volume in the conductor is

$$\frac{d\langle P_J \rangle}{d\tau} = \left\langle \vec{J}^* \cdot \vec{E} \right\rangle = \frac{\sigma}{2} \left| \vec{E}_{0,c}(\xi, \eta) \right|^2 = \frac{\sigma}{2} \frac{\omega^2 \delta^2}{2} \mu_c^2 \left| \vec{H}_0(\xi = 0, \eta) \right|^2 e^{-2\xi/\delta} \quad (9.347)$$

where the z and t dependence have been eliminated when we take the magnitude, leaving only the normalizing $_0$ fields that depend on the transverse coordinates (ξ, η). (Note that we have used the definitions of Z_σ and δ to manipulate the prefactor.)

When we integrate this over depth to get the Joule dissipation per unit area, $\int d\xi \, e^{-2\xi/\delta}$ yields $\delta/2$. After some algebra (recall, $\delta = \sqrt{2/\mu_c \, \sigma \, \omega}$ and $Z_\sigma = 1/(\sigma \, \delta)$), we find

$$\frac{d\langle P_J \rangle}{da} = \frac{1}{2 \, \sigma \, \delta} \left| \vec{H}_0(\xi = 0, \eta) \right|^2 = \frac{Z_\sigma}{2} \left| \vec{K}(\eta) \right|^2 \qquad (9.348)$$

where we have used the fact that, for the perfect conductor waveguide solution, the tangential component of the auxiliary field gives the surface current density. (There is no inconsistency in reintroducing $\vec{K}$ at this point: recall that $\vec{K}$ is the $\sigma \to \infty$, $\delta \to 0$ limit of $\delta \, \vec{J}$. Or, looked at from a length scale large compared to δ, the quickly decaying $\vec{J}$ looks like an infinitesimally thin sheet of current. From that same perspective, we recover the perfect conductor limit because the exponential decay of the fields is compressed to zero thickness, yielding an apparent discontinuity in the fields at the conductor boundary.) Looks like $I^2 R/2$!

Integrating over the boundary yields the rate of power loss per unit length along the waveguide

$$-\frac{d\langle P \rangle}{dz} = \oint_{\mathcal{C}(\mathcal{S})} d\eta \, \frac{d\langle P_J(\eta) \rangle}{da} = \oint_{\mathcal{C}(\mathcal{S})} d\eta \, \frac{Z_\sigma}{2} \left| \vec{K}(\eta) \right|^2 \qquad (9.349)$$

Now, assuming a TM mode (so $H_{0,z} = 0$, $\widehat{h} = \widehat{t}_\perp$, and $\vec{K} = K_z\,\widehat{z}$) and using the tangential boundary condition on the magnetic field:

$$K_{0,z}(\eta) = H(\xi = 0, \eta) = \frac{i\,k_{\epsilon\mu}}{Z_{\epsilon\mu}\,\gamma_n^2}\,\widehat{t}_\perp \cdot \left(\widehat{z} \times \vec{\nabla}_\perp E_{0,z}\right)\Bigg|_{\xi=0,\eta} \tag{9.350}$$

$$= \frac{i\,k_{\epsilon\mu}}{Z_{\epsilon\mu}\,\gamma_n^2}\,\left(\widehat{t}_\perp \times \widehat{z}\right) \cdot \vec{\nabla}_\perp E_{0,z}\Bigg|_{\xi=0,\eta} = \frac{i\,k_{\epsilon\mu}}{Z_{\epsilon\mu}\,\gamma_n^2}\,\widehat{n} \cdot \vec{\nabla}_\perp E_{0,z}\Bigg|_{\xi=0,\eta} \tag{9.351}$$

where we used the cyclicity of the triple scalar product and $\widehat{t}_\perp = \widehat{z} \times \widehat{n}$. Plugging into Equation 9.349 and using $Z_{\epsilon\mu} = \sqrt{\mu/\epsilon} = \mu\,v_{\epsilon\mu}$, $k_{\epsilon\mu} = \omega/v_{\epsilon\mu}$, and $\omega_{c,n} = v_{\epsilon\mu}\,\gamma_n$ (putting the n subscripts in now and not showing the (ξ, η) dependence for brevity):

$$-\frac{d\langle P_n^{\mathrm{TM}}\rangle}{dz} = \frac{Z_\sigma}{2}\,\frac{\omega^2}{\omega_{c,n}^2}\,\frac{1}{\mu^2\,\omega_{c,n}^2}\oint_{\mathcal{C}(\mathcal{S})} d\eta\,\left|\widehat{n}\cdot\vec{\nabla}_\perp E_{0,z,n}\right|^2 \tag{9.352}$$

One can obtain the corresponding expression for TE modes (which has two terms, one each for the currents associated with the transverse and longitudinal $\vec{H}$ components):

$$-\frac{d\langle P_n^{\mathrm{TE}}\rangle}{dz} = \frac{Z_\sigma}{2}\,\frac{\omega^2}{\omega_{c,n}^2}\oint_{\mathcal{C}(\mathcal{S})} d\eta\,\left[\frac{v_{\epsilon\mu}^2}{\omega_{c,n}^2}\left(1 - \frac{\omega_{c,n}^2}{\omega^2}\right)\left|\widehat{n}\times\vec{\nabla}_\perp H_{0,z,n}\right|^2 + \frac{\omega_{c,n}^2}{\omega^2}\,\left|H_{0,z,n}\right|^2\right] \tag{9.353}$$

To calculate the power loss per unit length explicitly, we must evaluate the above integrals using the known waveguide solutions, so the exact result will depend on the waveguide geometry. However, we can determine the approximate size of the above expressions generically. Recall that, for perfectly conducting walls, the eigenvalue-eigenvector equation implies (see Equations 9.314 and 9.315)

$$\int da \left|\vec{\nabla}_\perp \psi_n\right|^2 = \gamma_n^2 \int_{\mathcal{S}} da \left|\psi_n\right|^2 \qquad (9.354)$$

The above exact equality implies the approximate equality

$$\left\langle \left|\hat{n} \cdot \vec{\nabla}_\perp E_{0,z,n}\right|^2 \right\rangle_{\mathcal{C}(\mathcal{S})} \approx \gamma_n^2 \left\langle \left|E_{0,z,n}\right|^2 \right\rangle_{\mathcal{S}} \qquad \left\langle \left|\hat{n} \times \vec{\nabla}_\perp H_{0,z,n}\right|^2 \right\rangle_{\mathcal{C}(\mathcal{S})} \approx \gamma_n^2 \left\langle \left|H_{0,z,n}\right|^2 \right\rangle_{\mathcal{S}}$$

where the averages are over the boundary or the surface, not time. Thus, we can relate the line integral over the boundary to an area integral over the cross section:

$$\oint_{\mathcal{C}(\mathcal{S})} \frac{d\eta}{\gamma_n^2} \left|\hat{n} \cdot \vec{\nabla}_\perp E_{0,z,n}\right|^2 = \xi_n \frac{C}{A} \int_{\mathcal{S}} da \left|E_{0,z,n}\right|^2 \quad \text{with} \quad C = \oint_{\mathcal{C}(\mathcal{S})} d\eta \quad A = \int_{\mathcal{S}} da \quad (9.355)$$

where ξ_n is a fudge factor that has to be calculated for the specific geometry and is defined by the above equations. There is a similar expression for the TE modes. The area integral on the right is related to the power flow down the waveguide, Equation 9.316.

In fact, by combining Equations 9.316 and Equation 9.317 and the analogue of the above equation for TE modes, one can show that the *field* decay constant is

$$\kappa_n = -\frac{1}{2\langle P_n \rangle} \frac{d\langle P_n \rangle}{dz} = \frac{Z_\sigma(\omega_{c,n})}{Z_{\epsilon\mu}} \frac{C}{2A} \frac{\sqrt{\frac{\omega}{\omega_{c,n}}}}{\sqrt{1 - \frac{\omega_{c,n}^2}{\omega^2}}} \left[\xi_n + \eta_n \frac{\omega_{c,n}^2}{\omega^2} \right] \qquad (9.356)$$

where Z_σ is evaluated at $\omega_{c,n}$, ξ_n is the order unity quantity defined above for TM modes (and can be defined analogously for TE modes), and η_n is an analogous quantity that is nonzero only for TE modes (corresponding to the second term in the TE expression, proportional to $\omega_{c,n}^2/\omega^2$, which comes from the current associated with the longitudinal field). For any given waveguide geometry, one can find ξ_n, η_n, and κ_n by evaluating the explicit $d\langle P_n \rangle/dz$ and $\langle P_n \rangle$ formulae. Note that ξ_n and η_n are unrelated to the earlier (ξ, η) coordinates.

The useful and interesting thing about the above expression is that it exhibits all the frequency dependence. ξ_n and η_n are unitless, $\mathcal{O}(1)$ factors that depend on the solution to the eigenvector-eigenvalue equation, which is geometry- but not frequency-dependent.

Summing everything up, the fields in the waveguide medium including the decay along the waveguide are (k_n and κ_n functions of ω):

$$\vec{E}_n(\vec{r}, t) = \vec{E}_{0,n}(\vec{r}_\perp) \, e^{i(k_n(\omega)\, z - \omega\, t)} \, e^{-\kappa_n(\omega)\, z} \tag{9.357}$$

$$\vec{H}_n(\vec{r}, t) = \vec{H}_{0,n}(\vec{r}_\perp) \, e^{i(k_n(\omega)\, z - \omega\, t)} \, e^{-\kappa_n(\omega)\, z} \tag{9.358}$$

The fields in the walls are ($\vec{r}_\mathcal{C}$ is a vector $\vec{r} \in \mathcal{C}$)

$$\vec{E}_{c,n}(\vec{r}_\mathcal{C} - \hat{n}\xi, z, t) = \frac{2\, Z_\sigma(\omega)}{1 + i} \, \hat{n} \times \vec{H}_{0,n}(\vec{r}_\mathcal{C}, z = 0) \, e^{i(\xi/\delta(\omega))} \, e^{-\xi/\delta(\omega)} \, e^{i(k_n(\omega)\, z - \omega\, t)} \, e^{-\kappa_n(\omega)\, z} \tag{9.359}$$

$$\vec{H}_{c,n}(\vec{r}_\mathcal{C} - \hat{n}\xi, z, t) = \vec{H}_{0,n}(\vec{r}_\mathcal{C}, z = 0) \, e^{i(\xi/\delta(\omega))} \, e^{-\xi/\delta(\omega)} \, e^{i(k_n(\omega)\, z - \omega\, t)} \, e^{-\kappa_n(\omega)\, z} \tag{9.360}$$

We do not work out the full spatial dependences of the correction fields in the waveguide medium, *i.e.*, the correction fields that yield nonzero $(\hat{n} \times \hat{h}) \cdot \vec{E}$ and $\hat{n} \cdot \vec{H}$ at the waveguide walls, because that would require an additional step of re-solving the eigenvalue-eigenvector equations with a correction term to the boundary conditions, so the transverse spatial dependences $\vec{E}_{0,n}$ and $\vec{H}_{0,n}$ in the waveguide medium remain the perfect waveguide solutions.

Section 10
Potentials Revisited

Lecture 31:

Potentials Revisited:
Potential Formulation
Retarded Potentials
Date Revised: 2024/05/09 05:00
Date Given: 2024/05/09

Potential Formulation

Revisiting Scalar and Vector Potentials

Recall how we arrived at Maxwell's Equations. We first developed Faraday's Law by incorporating both empirical information (Faraday's observations) and the requirement of the Lorentz Force being consistent with Galilean relativity. We then found an inconsistency that required the introduction of the displacement current, yielding the full set of Maxwell's Equations.

However, we have not revisited the static potentials that we developed in the static field cases. Let's do that because they will provide a much faster way to calculate the fields radiated by accelerating charges.

We still have that $\vec{\nabla} \cdot \vec{B} = 0$, so the Helmholtz Theorem guarantees that we may continue to write

$$\boxed{\vec{B} = \vec{\nabla} \times \vec{A}} \qquad (10.1)$$

However, Faraday's Law implies that $\vec{\nabla} \times \vec{E} \neq 0$ when $\vec{B}$ has time dependence. Therefore, we cannot assume $\vec{E} = -\vec{\nabla}V$. However, using $\vec{B} = \vec{\nabla} \times \vec{A}$, we see that

$$\vec{\nabla} \times \vec{E} = -\frac{\partial \vec{B}}{\partial t} = -\frac{\partial}{\partial t}\left(\vec{\nabla} \times \vec{A}\right) \qquad \Longrightarrow \qquad \vec{\nabla} \times \left(\vec{E} + \frac{\partial \vec{A}}{\partial t}\right) = 0 \qquad (10.2)$$

Thus, the Helmholtz Theorem implies

$$\vec{E} + \frac{\partial \vec{A}}{\partial t} = -\vec{\nabla}V \qquad \Longrightarrow \qquad \boxed{\vec{E} = -\vec{\nabla}V - \frac{\partial \vec{A}}{\partial t}} \qquad (10.3)$$

We had in fact already seen the need for this correction when we calculated in Equation 7.58 that the Faraday's Law induced electric field is given by the second term, though we had not considered at the time its implications for the usual $\vec{E}_{coul} = -\vec{\nabla}V$ relation. The above is therefore just the statement that the total electric field is the sum of the Coulomb's Law conservative piece, which can be derived from a scalar potential, and the Faraday's Law non-conservative piece, which is directly given by $-\partial \vec{A}/\partial t$. The former part is sourced by charges, whether they are moving or not, and the latter part is sourced by anything time variable — currents and their time variation, which includes any movement of charges that involves acceleration. This statement also proves more directly Equation 7.44, which we had previously proven by noting that Coulomb's Law is independent of inertial reference frame (i.e., does not care whether the charges are moving or not).

By design, the above relations are consistent with the *homogeneous* Maxwell Equations (the ones that have no sources):

$$\vec{\nabla} \cdot \vec{B} = 0 \qquad \vec{\nabla} \times \vec{E} + \frac{\partial \vec{B}}{\partial t} = 0 \qquad (10.4)$$

Let's check whether Gauss's Law and Ampere's Law can be satisfied. In terms of the potentials, Gauss's Law becomes

$$\frac{\rho}{\epsilon_o} = \vec{\nabla} \cdot \vec{E} = -\nabla^2 V - \frac{\partial}{\partial t}\left(\vec{\nabla} \cdot \vec{A}\right) \qquad (10.5)$$

and Ampere's Law becomes

$$\vec{\nabla} \times \left(\vec{\nabla} \times \vec{A}\right) = \mu_o \vec{J} + \epsilon_o \mu_o \frac{\partial \vec{E}}{\partial t} = \mu_o \vec{J} - \epsilon_o \mu_o \vec{\nabla}\frac{\partial V}{\partial t} - \epsilon_o \mu_o \frac{\partial^2 \vec{A}}{\partial t^2} \qquad (10.6)$$

which we may rewrite using the usual $\vec{\nabla} \times (\vec{\nabla} \times \vec{a}) = \vec{\nabla}(\vec{\nabla} \cdot \vec{a}) - \nabla^2 \vec{a}$ identity:

$$\left(\nabla^2 \vec{A} - \epsilon_o \mu_o \frac{\partial^2 \vec{A}}{\partial t^2}\right) - \vec{\nabla}\left(\vec{\nabla} \cdot \vec{A} + \epsilon_o \mu_o \frac{\partial V}{\partial t}\right) = -\mu_o \vec{J} \qquad (10.7)$$

Equations 10.5 and 10.7 are second-order coupled partial differential equations for V and $\vec{A}$ and, so far, we see no reason they cannot be solved.

We can write Gauss' Law (Equation 10.5) and Ampere's Law (Equation 10.7) in terms of the potentials in the symmetrical form (defining the d'Alembertian $\Box^2$ with the opposite sign convention as Griffiths to be consistent with our later discussion of special relativity):

$$\Box^2 V - \frac{\partial L}{\partial t} = \frac{\rho}{\epsilon_o} \qquad\qquad \Box^2 \vec{A} + \vec{\nabla} L = \mu_o \, \vec{J} \qquad\qquad (10.8)$$

$$\text{with} \qquad \Box^2 \equiv \epsilon_o \, \mu_o \, \frac{\partial^2}{\partial t^2} - \nabla^2 \qquad L \equiv \vec{\nabla} \cdot \vec{A} + \epsilon_o \, \mu_o \, \frac{\partial V}{\partial t} \qquad\qquad (10.9)$$

Gauge Freedom and Transformations

The new equations for the potentials reduce the number of degrees of freedom from six to four. There remains what is called a *gauge freedom*, which we will first describe and then make choices for. This is the generalization of the freedom we had to set the constant offset for the potential in electrostatics and to choose the value of $\vec{\nabla} \cdot \vec{A}$ in magnetostatics.

Gauge freedom answers the question: how much can we change $\vec{A}$ and V while still obtaining the same fields? That is, let us assume two sets of potentials $(V, \vec{A})$ and $(V', \vec{A}')$ that differ by functions $\vec{\alpha}$ and β:

$$\vec{A}'(\vec{r}, t) = \vec{A}(\vec{r}, t) + \vec{\alpha}(\vec{r}, t) \qquad\qquad V'(\vec{r}, t) = V(\vec{r}, t) + \beta(\vec{r}, t) \qquad (10.10)$$

The requirement that these two sets of potentials yield the same fields gives the equations:

$$0 = \vec{\nabla} \times \left[\vec{A}'(\vec{r}, t) - \vec{A}(\vec{r}, t) \right] = \vec{\nabla} \times \vec{\alpha}(\vec{r}, t) \qquad (10.11)$$

$$0 = -\vec{\nabla} \left[V'(\vec{r}, t) - V(\vec{r}, t) \right] - \frac{\partial}{\partial t} \left[\vec{A}'(\vec{r}, t) - \vec{A}(\vec{r}, t) \right]$$

$$= -\vec{\nabla}\beta(\vec{r}, t) - \frac{\partial \vec{\alpha}(\vec{r}, t)}{\partial t} \qquad (10.12)$$

The first equation implies that $\vec{\alpha}$ can be written as the gradient of a scalar

$$\vec{\alpha}(\vec{r}, t) = \vec{\nabla}\lambda(\vec{r}, t) \tag{10.13}$$

which we then plug into the other equation

$$\vec{\nabla}\left[\beta(\vec{r}, t) + \frac{\partial\lambda(\vec{r}, t)}{\partial t}\right] = 0 \tag{10.14}$$

The above equation implies the quantity in parentheses must be position-independent, which yields

$$\beta(\vec{r}, t) + \frac{\partial\lambda(\vec{r}, t)}{\partial t} = k(t) \tag{10.15}$$

We can absorb $k(t)$ into λ: add to λ the position-independent term $\int_0^t dt'\, k(t')$, whose gradient vanishes and thus does not affect $\alpha(\vec{r}, t)$. Thus $\beta(\vec{r}, t) = -\partial\lambda(\vec{r}, t)/\partial t$ and we have the relations

$$\vec{A}'(\vec{r}, t) = \vec{A}(\vec{r}, t) + \vec{\nabla}\lambda(\vec{r}, t) \qquad\qquad V'(\vec{r}, t) = V(\vec{r}, t) - \frac{\partial\lambda(\vec{r}, t)}{\partial t} \tag{10.16}$$

This kind of change in $(V, \vec{A})$ that has no effect on the fields is called a *gauge transformation*. The choice of λ (perhaps implicit through specifying properties of $\vec{A}$ and V) is called the choice of *gauge*. With the fourth degree of freedom in the potentials identified with this gauge freedom, we thus see that Maxwell's Equations imply there are only three *physical* degrees of freedom in the electromagnetic field.

Coulomb Gauge

This is the gauge we chose earlier for magnetostatics,

$$\boxed{\text{Coulomb Gauge} \qquad \vec{\nabla} \cdot \vec{A} = 0} \tag{10.17}$$

In this gauge, Poisson's Equation simplifies to the electrostatic form

$$\boxed{\text{Coulomb Gauge} \qquad \nabla^2 V = -\frac{\rho}{\epsilon_o}} \tag{10.18}$$

That is, the charge density sets the potential in the same way as in electrostatics, so changes in charge density propagate into the potential *instantaneously*. Of course, you know from special relativity that this is not possible. We will see that there are corrections from $\partial \vec{A}/\partial t$ that prevent $\vec{E}$ from responding instantaneously to such changes. The differential equation for $\vec{A}$ becomes

$$\boxed{\text{Coulomb Gauge} \qquad \nabla^2 \vec{A} - \epsilon_o\,\mu_o\,\frac{\partial^2 \vec{A}}{\partial t^2} = -\mu_o\,\vec{J} + \epsilon_o\,\mu_o\,\vec{\nabla}\left(\frac{\partial V}{\partial t}\right)} \tag{10.19}$$

which is a wave equation for $\vec{A}$ with source terms from $\vec{J}$ and $\vec{\nabla}\left(\partial V/\partial t\right)$. The latter is determined by $\partial \rho/\partial t$.

Lorenz Gauge

The missing "t" is not a typo: different person!

Here the choice is

$$\boxed{\text{Lorenz Gauge} \qquad \vec{\nabla} \cdot \vec{A} = -\epsilon_o\,\mu_o\,\frac{\partial V}{\partial t}} \tag{10.20}$$

The reason this choice is made is that it sets $L = 0$ in Equation 10.9 so that we obtain symmetric, *inhomogeneous* (have source terms) wave equations for V and $\vec{A}$:

$$\boxed{\text{Lorenz Gauge} \qquad \Box^2 V = \frac{\rho}{\epsilon_o} \qquad \Box^2 \vec{A} = \mu_o\,\vec{J} \qquad \Box^2 \equiv \epsilon_o\,\mu_o\,\frac{\partial^2}{\partial t^2} - \nabla^2} \tag{10.21}$$

with, again, the opposite sign convention as Griffiths for $\Box^2$. This will be the natural gauge for calculating radiation of EM waves from moving charges and currents, especially in a manner that is manifestly consistent with relativity.

Lecture 38:

Advanced Potential Formulation (Optional Material):
Lorentz Force Law in Potential Form and
Implications for Classical and Quantum Mechanics
Gauge Transformations and Coupling of Matter to EM Fields

Date Revised: 2023/06/06 06:45
Date Given: Not given in 2024

These two sections,
consisting of material colored green and orange,
is optional!

The Lorentz Force Law in Potential Form and Implications for Classical and Quantum Mechanics (Optional)

Let's write the Lorentz Force Law in terms of the potentials:

$$\vec{F} = \frac{d\vec{p}}{dt} = q \left[-\vec{\nabla} V - \frac{\partial \vec{A}}{\partial t} + \vec{v} \times \left(\vec{\nabla} \times \vec{A} \right) \right] \tag{10.22}$$

Using the $BAC - CAB$ rule for the triple vector product and recognizing $\vec{v}$ is not a function of position, we may rewrite as

$$m \frac{d\vec{v}}{dt} = -q \left[\frac{\partial \vec{A}}{\partial t} + \left(\vec{v} \cdot \vec{\nabla} \right) \vec{A} + \vec{\nabla} \left(V - \vec{v} \cdot \vec{A} \right) \right] \tag{10.23}$$

We may write

$$\frac{\partial \vec{A}}{\partial t} + \left(\vec{v} \cdot \vec{\nabla} \right) \vec{A} = \frac{d\vec{A}}{dt} \tag{10.24}$$

where d/dt is the *total derivative* of $\vec{A}$, taking into account both the explicit time dependence of $\vec{A}$ (the $\partial \vec{A}/\partial t$ term) and the time dependence of $\vec{A}$ to which the particle is subject due to its motion (the $(\vec{v} \cdot \vec{\nabla})\vec{A}$ term). This kind of derivative should be familiar to you from classical mechanics when considering the total time derivative of the Lagrangian. The quantity on the left side of the above equation is called the *convective derivative* of $\vec{A}$ (*convective* because it is particularly important in fluid mechanics, where such a derivative accounts for the motion of a fluid element.)

Let's be explicit about why the convective derivative correctly gives the total derivative of $\vec{A}$. Consider a time interval dt in which the particle moves $d\vec{r} = \vec{v}\,dt$:

$$d\vec{A} = \vec{A}(\vec{r} + \vec{v}\,dt, t + dt) - \vec{A}(\vec{r}, t) \tag{10.25}$$

$$= v_x\,dt\,\frac{\partial \vec{A}}{\partial x} + v_y\,dt\,\frac{\partial \vec{A}}{\partial y} + v_z\,dt\,\frac{\partial \vec{A}}{\partial z} + dt\,\frac{\partial \vec{A}}{\partial t} = dt\left[\left(\vec{v} \cdot \vec{\nabla}\right)\vec{A} + \frac{\partial \vec{A}}{\partial t}\right] \tag{10.26}$$

With the appropriate definition of the total derivative of $\vec{A}$, we may rewrite the Lorentz Force Law as

$$\frac{d}{dt}\left(m\vec{v} + q\vec{A}\right) = -\vec{\nabla}\left[q\left(V - \vec{v} \cdot \vec{A}\right)\right] \tag{10.27}$$

which may be rewritten using the *canonical momentum*, $\vec{\pi}$, and a *velocity-dependent potential energy*, $U_v(\vec{r}, \vec{v}, t)$:

$$\boxed{\frac{d\vec{\pi}}{dt} = -\vec{\nabla}U_v \qquad \vec{\pi} = m\vec{v} + q\vec{A} \qquad U_v(\vec{r}, \vec{v}, t) = q\left[V(\vec{r}, t) - \vec{v} \cdot \vec{A}(\vec{r}, t)\right]} \tag{10.28}$$

You may have seen these ideas of a velocity-dependent potential and canonical momentum distinct from mechanical momentum in Ph106a. Recall that the Lagrangian formulation of mechanics is derivable as a special case, for conservative forces, of the *generalized equation of motion*:

$$\sum_i \vec{F}_i^{(nc)} \cdot \frac{\partial \vec{r}_i}{\partial q_k} = \mathcal{F}_k = \frac{d}{dt}\left(\frac{\partial T}{\partial \dot{q}_k}\right) - \frac{\partial T}{\partial q_k} \tag{10.29}$$

where the $\vec{F}_i^{(nc)}$ are the set of *non-constraint forces*, the q_k are generalized coordinates, $\mathcal{F}_k$ is the generalized force in the kth generalized coordinate, and T is the kinetic energy in terms of the generalized coordinates and velocities. For a conservative force, the force term on the left side can be written as a gradient of a potential energy U that is a function only of the coordinates. This form permits one to define the Lagrangian function $L = T - U$ and incorporate the gradient of U into the second term on the right side. (Such a U does not contribute to the first term because it does not depend on the generalized velocities.) The result is the Euler-Lagrange Equation

$$\frac{d}{dt}\left(\frac{\partial L}{\partial \dot{q}_k}\right) - \frac{\partial L}{\partial q_k} = 0 \tag{10.30}$$

For nonconservative, velocity-dependent forces like the Lorentz Force, it is clear that, if one can construct the appropriate function U_v such that its partial derivatives with respect to q_k and $\dot{q}_k$ generate the appropriate force terms on the left side, then the same construction applies, now with $L = T - U_v$ and U_v being a function generalized coordinates and velocities, a velocity-dependent potential (energy).

Let's see how it works out in detail in this case, which will also lead us to the appropriate Hamiltonian operator for quantum mechanics in the presence of electromagnetic fields. Using the velocity-dependent potential energy we defined in Equations 10.28, we write down the corresponding Lagrangian:

$$L = T - U_v = \frac{1}{2} m v^2 - q \left[V(\vec{r}, t) - \vec{v} \cdot \vec{A}(\vec{r}, t) \right] \tag{10.31}$$

The generalized coordinates are $q_k = r_k$ and the generalized velocities are $\dot{q}_k = \dot{r}_k = v_k$, so the Euler-Lagrange Equations become (recall, $\partial \dot{r}_k / \partial r_j = 0$ as we take the derivatives to obtain the Euler-Lagrange Equations; they do not become tied to each other until we have those equations):

$$\frac{d}{dt} \left[m \dot{r}_k + q A_k \right] + q \left[\frac{\partial V}{\partial r_k} - \vec{v} \cdot \frac{\partial \vec{A}}{\partial r_k} \right] = 0 \tag{10.32}$$

$$m \frac{d\dot{r}_k}{dt} + q \frac{\partial A_k}{\partial t} + q \left(\vec{v} \cdot \vec{\nabla} \right) A_k + q \frac{\partial V}{\partial r_k} - q \vec{v} \cdot \frac{\partial \vec{A}}{\partial r_k} = 0 \tag{10.33}$$

where we used the expression we derived earlier for the total derivative of $\vec{A}$.

Moving terms around and being careful not to reorder derivatives now that $\vec{r}$ and $\vec{v} = \dot{\vec{r}}$ are no longer independent, we have

$$m\ddot{r}_k = q\left[-\frac{\partial V}{\partial r_k} - \frac{\partial A_k}{\partial t}\right] + q\left[\sum_{n=1}^{3}\dot{r}_n\frac{\partial A_n}{\partial r_k} - \sum_{n=1}^{3}\dot{r}_n\frac{\partial A_k}{\partial r_n}\right] \tag{10.34}$$

If we put the above in vector notation, we obtain:

$$m\ddot{\vec{r}} = q\left[-\vec{\nabla}V - \frac{\partial\vec{A}}{\partial t} + \vec{\nabla}\left(\dot{\vec{r}}\cdot\vec{A}\right) - \left(\dot{\vec{r}}\cdot\vec{\nabla}\right)\vec{A}\right] \tag{10.35}$$

which recovers Equation 10.23 with $\dot{\vec{r}} = \vec{v}$ and $\ddot{\vec{r}} = d\vec{v}/dt$. We may thus invert the use of the $BAC - CAB$ rule that took us from Equation 10.22 to Equation 10.23, yielding

$$m\ddot{\vec{r}} = q\left[-\vec{\nabla}V - \frac{\partial\vec{A}}{\partial t} + \dot{\vec{r}}\times\left(\vec{\nabla}\times\vec{A}\right)\right] \tag{10.36}$$

as we expect: both the Lorentz Force Law and a second-order ODE for the particle trajectory $\vec{r}(t)$.

Let's also calculate the canonical momentum and the Hamiltonian. Normally, one writes the Hamiltonian in terms of the canonical momentum and the generalized coordinates, not the generalized coordinates and velocities, but we will do the latter first so we can see that it yields the correct total energy:

$$\pi_k = \frac{\partial L}{\partial \dot{q}_k} = m \, \dot{r}_k + q \, A_k \tag{10.37}$$

$$H = \sum_k \pi_k \, \dot{q}_k - L = \left(m \, \dot{\vec{r}} + q \, \vec{A} \right) \cdot \dot{\vec{r}} - L \tag{10.38}$$

$$= \left(m \, \dot{\vec{r}} + q \, \vec{A} \right) \cdot \dot{\vec{r}} - \left\{ \frac{1}{2} \, m \, \dot{\vec{r}} \cdot \dot{\vec{r}} - q \left[V - \dot{\vec{r}} \cdot \vec{A} \right] \right\}$$

$$= \frac{1}{2} \, m \, \dot{\vec{r}} \cdot \dot{\vec{r}} + q \, V = T + q \, V \tag{10.39}$$

which is the energy, as we expect.

Let's rewrite the Hamiltonian as it should be done, in terms of the canonical momentum, so that we can see how one again recovers the Lorentz Force Law as well as obtains the quantum mechanical Hamiltonian operator. Start from the second line above, using $\vec{v} = \left(\vec{\pi} - q\,\vec{A} \right) / m$:

$$H = \vec{\pi} \cdot \frac{1}{m} \left(\vec{\pi} - q\,\vec{A} \right) - \left\{ \frac{1}{2\,m} \left| \vec{\pi} - q\,\vec{A} \right|^2 - q \left[V - \frac{1}{m} \left(\vec{\pi} - q\,\vec{A} \right) \cdot \vec{A} \right] \right\} \quad (10.40)$$

$$= \frac{1}{2\,m} \left| \vec{\pi} - q\,\vec{A} \right|^2 + q\,V \quad (10.41)$$

Hamilton's Equations of Motion relate the rate of change of the canonical momentum to the gradient of the Hamiltonian and the rate of change of the generalized coordinates to the canonical momenta:

$$\dot{\pi}_k = \frac{d\pi_k}{dt} = -\frac{\partial H}{\partial r_k} = -\frac{1}{m} \sum_{n=1}^{3} (\pi_n - q\,A_n) \frac{\partial}{\partial r_k} (\pi_n - q\,A_n) - q \frac{\partial V}{\partial r_k}$$

$$= \frac{q}{m} \sum_{n=1}^{3} (\pi_n - q\,A_n) \frac{\partial A_n}{\partial r_k} - q \frac{\partial V}{\partial r_k}$$

$$\dot{r}_k = \frac{dr_k}{dt} = \frac{\partial H}{\partial \pi_k} = \frac{1}{m} (\pi_k - q\,A_k)$$

In calculating the above, one must remember that π_k and r_j are independent variables, $\partial \pi_k / \partial r_j = 0$, during the process of taking the derivatives in Hamilton's Equations (similar to the way r_k and $\dot{r}_k$ are independent in the process of taking the derivatives to obtain the Euler-Lagrange Equation).

After we take the derivatives, though, we drop this independence and substitute the second equation in the first one to eliminate π_k and obtain a second-order ODE in r_k:

$$\frac{d}{dt}\left(m\dot{r}_k + q A_k\right) = q\sum_{n=1}^{3}\dot{r}_n\frac{\partial A_n}{\partial r_k} - q\frac{\partial V}{\partial r_k}$$

$$\frac{d}{dt}\left(m\dot{r}_k + q A_k\right) = -\frac{\partial}{\partial r_k}\left[q\left(V - \dot{\vec{r}}\cdot\vec{A}\right)\right] = -\frac{\partial U_v}{\partial r_k}$$

which recovers Equation 10.28 with the understanding that derivatives with respect to r_k should not be interpreted as acting on $\dot{r}_k$ in order to get the last equation.

The greater point of interest about writing out the Hamiltonian is to see the motivation for the quantum mechanical Hamiltonian. To go from classical mechanics to quantum mechanics, we promote the *canonical momentum* to the momentum operator $-i\hbar\vec{\nabla}$ (or, more correctly, we promote $\vec{\pi}$ to the quantum mechanical momentum operator $\vec{\Pi}$, whose representation in the position basis is $-i\hbar\vec{\nabla}$). The above derivation shows why the Hamiltonian in the presence of an arbitrary EM field has $\vec{A}$ in it, which otherwise goes unexplained in quantum mechanics.

The presence of $\vec{A}$ in the Hamiltonian is also why the vector potential plays a more important role in quantum mechanics than it does in classical mechanics. Its presence in H gives it the opportunity to affect the phase of the quantum mechanical state.

We can also find an equation relating the time derivative of the particle energy to the potentials. This will become useful when we consider the relativistic generalization of the Lorentz Force Law, which will include an energy component. We obtain this by calculating the work done per unit time by the Lorentz Force:

$$\frac{dT}{dt} = \vec{v} \cdot \vec{F} = q \left[\vec{v} \cdot \left(-\vec{\nabla} V \right) - \vec{v} \cdot \frac{\partial \vec{A}}{\partial t} + \vec{v} \cdot \left\{ \vec{v} \times \left(\vec{\nabla} \times \vec{A} \right) \right\} \right] \qquad (10.42)$$

The last term vanishes because the triple cross product is perpendicular to $\vec{v}$, reflecting the fact that magnetic fields can do no work (except on fundamental magnetic dipoles). Now, let's add the total derivative of $q\,V$,

$$\frac{d}{dt} \left(q\,V \right) = \left(\vec{v} \cdot \vec{\nabla} \right) \left(q\,V \right) + \frac{\partial}{\partial t} \left(q\,V \right) \qquad (10.43)$$

to both sides, yielding

$$\boxed{\frac{d}{dt} \left(T + q\,V \right) = \frac{\partial}{\partial t} \left[q \left(V - \vec{v} \cdot \vec{A} \right) \right] = \frac{\partial}{\partial t} U_v} \qquad (10.44)$$

This gives us an expression for conservation of energy in terms of the potentials (rather than the fields) for fully time-dependent (but non-relativistic) situations. The latter term on the RHS does not imply magnetic fields do work, as $\vec{B}$ is always perpendicular to $\vec{A}$, so $\vec{v} \cdot \vec{A}$ has no contribution from $\vec{B}$. That term is present because of the portion of $\vec{A}$ that produces an electric field.

Gauge Transformations and Coupling of Matter to Electromagnetic Fields (Optional)

Requiring invariance under gauge transformations provides another interesting way to introduce the coupling of matter to electromagnetic fields in quantum mechanics, differing from the approach we just provided.

We begin by showing that one can derive the Schrödinger Equation by considering the wavefunction $\psi(\vec{r}, t)$ to be a classical field with Lagrangian density (Lagrangian per unit volume):

$$\mathcal{L} = \frac{1}{2} \left[\psi^* \left(i\hbar \frac{\partial}{\partial t} \right) \psi + \psi \left(i\hbar \frac{\partial}{\partial t} \right)^* \psi^* \right] - \frac{1}{2m} \psi^* \left(-i\hbar \vec{\nabla} \right)^2 \psi \qquad (10.45)$$

With a Lagrangian density, we consider the field at every point in space $\psi(\vec{r}, t)$ to be an independent trajectory in time, analogous to $\vec{x}(t)$ for a single particle. Therefore, our generalized coordinate is $\psi(\vec{r}, t)$ with $\vec{r}$ acting as a label or index, with each point in space corresponding to an independent variable. When one has a complex field, one can show that one should actually use ψ^* as the independent variable for the Euler-Lagrange Equation; that is, the Euler-Lagrange Equation is

$$\frac{d}{dt} \left(\frac{\partial \mathcal{L}}{\partial \dot{\psi}^*} \right) - \frac{\partial \mathcal{L}}{\partial \psi^*} = 0 \qquad (10.46)$$

where $\dot{\psi}^* = \partial \psi^* / \partial t$ and the equation should be evaluated independently at every $\vec{r}$.

Applying the above, we obtain

$$\frac{d}{dt}\left(-\frac{i\,\hbar}{2}\,\psi\right) - \left[\frac{i\,\hbar}{2}\,\frac{\partial\psi}{\partial t} + \frac{\hbar^2}{2\,m}\,\nabla^2\psi\right] = 0 \qquad (10.47)$$

$$\frac{1}{2\,m}\left(-i\,\hbar\,\vec{\nabla}\right)^2\psi = i\,\hbar\,\frac{\partial\psi}{\partial t} \qquad (10.48)$$

which is Schrödinger's Equation for a free particle.

The above Lagrangian density is invariant under the phase transformation

$$\psi(\vec{r},t) \to \psi'(\vec{r},t) = \psi(\vec{r},t)\,e^{i\,\chi} \qquad \text{with } \chi \text{ a constant} \qquad (10.49)$$

simply because every time ψ appears, so does ψ^*: this phase transformation is a symmetry transformation of the Lagrangian. Therefore, by Noether's Theorem of classical mechanics, there is a conserved quantity found by summing the product of the canonical momentum and the derivative of the transformation with respect to its parameter (χ) over all degrees of freedom (all values of $\vec{r}$; *i.e.*, integrating over space):

$$\mathrm{I} = \int d\tau\, p_\psi\,\frac{\partial\psi'}{\partial\chi} = \int d\tau\,\frac{\partial\mathcal{L}}{\partial\dot{\psi}}\,\frac{\partial\psi'}{\partial\chi} = \int d\tau\left(\frac{i\,\hbar}{2}\,\psi^*\right)(i\,\psi) = -\frac{\hbar^2}{2}\int d\tau\,|\psi(\vec{r},t)|^2 \qquad (10.50)$$

That is, the fact that the phase transformation is a symmetry transformation of the Lagrangian density implies that the total probability of the wavefunction is conserved (the negative sign is unimportant, as are the other prefactors).

It is a general principle of modern field theory that we should expect global symmetry transformations like the one above to also be symmetry transformations of the Lagrangian when the transformation is made local. One reason we make this requirement is that a symmetry transformation that is the same everywhere in space is inconsistent with special relativity: how does the information propagate instantaneously to all points in space? We also know simultaneity is frame-dependent in special relativity, which also calls into question the idea of applying a transformation that is the same at all points in space. In this case, that would imply the position-dependent phase transformation

$$\psi(\vec{r}, t) \rightarrow \psi'(\vec{r}, t) = \psi(\vec{r}, t)\, e^{i\,\chi(\vec{r}, t)} \tag{10.51}$$

is a symmetry transformation of the Lagrangian density.

However, we quickly see that, when we try to apply such a local transformation to the Lagrangian density, the spatial and time derivatives act on $\chi(\vec{r}, t)$ and create all kinds of additional terms such that $\mathcal{L}$ is *not* invariant under the local version of the transformation. Specifically:

$$\frac{\partial \psi'}{\partial t} = \frac{\partial \psi}{\partial t}\, e^{i\chi} + i\,\psi\, e^{i\chi}\, \frac{\partial \chi}{\partial t} \tag{10.52}$$

$$\nabla^2 \psi' = \vec{\nabla} \cdot \vec{\nabla} \left[\psi\, e^{i\chi} \right] = \vec{\nabla} \cdot \left[e^{i\chi} \vec{\nabla}\psi + i\,\psi\, e^{i\chi}\, \vec{\nabla}\chi \right] \tag{10.53}$$

$$= e^{i\chi} \nabla^2 \psi + 2\, i\, e^{i\chi}\, \vec{\nabla}\psi \cdot \vec{\nabla}\chi - e^{i\,\chi}\, \psi\, \vec{\nabla}\chi \cdot \vec{\nabla}\chi + i\,\psi\, e^{i\chi} \nabla^2 \chi \tag{10.54}$$

In each expression, the first term is what we had for the global version of the transformation, but the additional terms break the invariance.

There is a standard prescription for fixing this kind of problem. We need counterterms to cancel the additional terms that appeared above. We make the replacement

$$\vec{p}_{\vec{r}} = \vec{\pi}_{\vec{r}} = -i\,\hbar\,\vec{\nabla} \qquad \longrightarrow \qquad \vec{p}_{\vec{r}} = \vec{\pi}_{\vec{r}} - q\,\vec{A}(\vec{r}, t) = -i\,\hbar\,\vec{\nabla} - q\,\vec{A}(\vec{r}, t) \qquad (10.55)$$

$$p_t = \pi_t = i\,\hbar\,\frac{\partial}{\partial t} \qquad \longrightarrow \qquad p_t = \pi_t - q\,V(\vec{r}, t) = i\,\hbar\,\frac{\partial}{\partial t} - q\,V(\vec{r}, t) \qquad (10.56)$$

where the new functions $\vec{A}(\vec{r}, t)$ and $V(\vec{r}, t)$ also participate in the local phase transformation, now called, for obvious reasons, a *gauge transformation*:

$$V'(\vec{r}, t) = V(\vec{r}, t) - \frac{\partial \lambda(\vec{r}, t)}{\partial t} \qquad (10.57)$$

$$\vec{A}'(\vec{r}, t) = \vec{A}(\vec{r}, t) + \vec{\nabla}\lambda(\vec{r}, t) \qquad (10.58)$$

with $\chi(\vec{r}, t) = q\,\lambda(\vec{r}, t)/\hbar$ the same function used in the transformation of the wavefunction up to a constant factor, so the coupled gauge transformation of the wavefunction is

$$\psi(\vec{r}, t) \rightarrow \psi'(\vec{r}, t) = \psi(\vec{r}, t)\,e^{i\,q\,\lambda(\vec{r}, t)/\hbar} \qquad (10.59)$$

With this replacement, the Lagrangian is

$$\mathcal{L} = \frac{1}{2} \left[\psi^* \left(i\,\hbar\,\frac{\partial}{\partial t} - q\,V \right) \psi + \psi \left(i\,\hbar\,\frac{\partial}{\partial t} - q\,V \right)^* \psi^* \right] - \frac{1}{2\,m}\,\psi^* \left(-i\,\hbar\,\vec{\nabla} - q\,\vec{A} \right)^2 \psi \tag{10.60}$$

We can check this Lagrangian density is invariant under the gauge transformation. The space piece is:

$$\left(-i\,\hbar\,\vec{\nabla} - q\,\vec{A} \right) \psi \quad \longrightarrow$$

$$\left(-i\,\hbar\,\vec{\nabla} - q\,\vec{A}' \right) \psi' \tag{10.61}$$

$$= -i\,\hbar \left[e^{i\,q\,\lambda/\hbar}\vec{\nabla}\psi + e^{i\,q\,\lambda/\hbar}\,\psi \left(i\,\frac{q}{\hbar} \right) \vec{\nabla}\lambda \right] - q \left[\vec{A} + \vec{\nabla}\lambda \right] e^{i\,q\,\lambda/\hbar}\,\psi$$

$$= e^{i\,q\,\lambda/\hbar} \left(-i\,\hbar\,\vec{\nabla} - q\,\vec{A} \right) \psi \tag{10.62}$$

$$\implies \quad \left(-i\,\hbar\,\vec{\nabla} - q\,\vec{A} \right)^2 \psi \quad \longrightarrow$$

$$\left(-i\,\hbar\,\vec{\nabla} - q\,\vec{A}' \right)^2 \psi' = \left(-i\,\hbar\,\vec{\nabla} - q\,\vec{A}' \right) \cdot \left[e^{i\,q\,\lambda/\hbar} \left(-i\,\hbar\,\vec{\nabla} - q\,\vec{A} \right) \psi \right]$$

$$= e^{i\,q\,\lambda/\hbar} \left[q\,\vec{\nabla}\lambda - i\,\hbar\,\vec{\nabla} - q\,\vec{A} - q\,\vec{\nabla}\lambda \right] \cdot \left(-i\,\hbar\,\vec{\nabla} - q\,\vec{A} \right) \psi$$

$$= e^{i\,q\,\lambda/\hbar} \left(-i\,\hbar\,\vec{\nabla} - q\,\vec{A} \right)^2 \psi \tag{10.63}$$

The transformation yields just the simple phase factor in front that we need for the cancellation to work.

The time pieces are

$$\left(i\,\hbar\,\frac{\partial}{\partial t} - q\,V \right)\psi \quad \longrightarrow$$

$$\left(i\,\hbar\,\frac{\partial}{\partial t} - q\,V' \right)\psi' \tag{10.64}$$

$$= i\,\hbar\left[e^{i\,q\,\lambda/\hbar}\,\frac{\partial\psi}{\partial t} + e^{i\,q\,\lambda/\hbar}\,\psi\left(i\,\frac{q}{\hbar} \right)\frac{\partial\lambda}{\partial t} \right] - q\left[V - \frac{\partial\lambda}{\partial t} \right]e^{i\,q\,\lambda/\hbar}\,\psi$$

$$= e^{i\,q\,\lambda/\hbar}\left(i\,\hbar\,\frac{\partial}{\partial t} - q\,V \right)\psi \tag{10.65}$$

$$\left(i\,\hbar\,\frac{\partial}{\partial t} - q\,V \right)^{*}\psi^{*} \quad \longrightarrow$$

$$\left(i\,\hbar\,\frac{\partial}{\partial t} - q\,V' \right)^{*}(\psi^{*})' \tag{10.66}$$

$$= -i\,\hbar\left[e^{-i\,q\,\lambda/\hbar}\,\frac{\partial\psi^{*}}{\partial t} + e^{-i\,q\,\lambda/\hbar}\,\psi^{*}\left(-i\,\frac{q}{\hbar} \right)\frac{\partial\lambda}{\partial t} \right] - q\left[V - \frac{\partial\lambda}{\partial t} \right]e^{-i\,q\,\lambda/\hbar}\,\psi^{*}$$

$$= e^{-i\,q\,\lambda/\hbar}\left(i\,\hbar\,\frac{\partial}{\partial t} - q\,V \right)^{*}\psi^{*} \tag{10.67}$$

Again, the transformation yields just the simple phase factor in front that we need for the cancellation to work.

If we apply the Euler-Lagrange equation to the new Lagrangian, we get

$$\frac{d}{dt}\left(-\frac{i\,\hbar}{2}\,\psi\right) - \left[\frac{i\,\hbar}{2}\,\frac{\partial\psi}{\partial t} - \frac{1}{2}\,q\,V\,\psi - \frac{1}{2}\,q\,V\,\psi - \frac{1}{2\,m}\left(-i\,\hbar\,\vec{\nabla} - q\,\vec{A}\right)^2\psi\right] = 0$$

(10.68)

$$\frac{1}{2\,m}\left(-i\,\hbar\,\vec{\nabla} - q\,\vec{A}\right)^2\psi + q\,V\,\psi = i\,\hbar\,\frac{\partial\psi}{\partial t}$$

(10.69)

which is exactly the Schrödinger Equation with coupling of the particle of charge q to an electrostatic potential $V(\vec{r}, t)$ and a vector potential $\vec{A}(\vec{r}, t)$!

This is the same Schrödinger Equation that get if we promote the classical Hamiltonian from Equation 10.41 to a quantum mechanical Hamiltonian operator by replacing $\vec{\pi}$ with $-i\,\hbar\,\vec{\nabla}$ (in the position basis), as indicated earlier .

Lastly, with the coupling between charge and the potentials now present, we can reinterpret our Noether's Theorem result: not only does it imply conservation of probability but, because the particle now has a charge associated with it, it implies conservation of charge, too! This is conservation of integrated charge over all of space for each particle, not local conservation of charge.

To make the next step to local conservation of charge — *i.e.*, the continuity equation — one must demonstrate local conservation of probability. One does this by defining a probability current

$$\vec{j}(\vec{r}, t) = -\frac{i}{2} \frac{\hbar}{m} \left[\psi^* \vec{\nabla} \psi - \psi \vec{\nabla} \psi^* \right]$$

and showing, using the Schrödinger Equation, that

$$\vec{\nabla} \cdot \vec{j} + \frac{\partial}{\partial t} |\psi|^2 = 0$$

which is the continuity equation for probability. If one multiplies all of the above by q, one obtains the continuity equation for charge density too!

Retarded Potentials and Fields

Retarded Potentials

We would like to obtain solutions to the inhomogeneous wave equations, Equation 10.21. Those equations have the structure of Poisson's Equation aside from the addition of the time derivative term. In the case of static fields, they in fact reduce to Poisson's Equation exactly. We know from the homogeneous wave equation that this time derivative implies that the solution propagates at speed $c = 1/\sqrt{\epsilon_o \mu_o}$. It is thus plausible that the solutions are given by generalizing the static equations to account for the propagation time:

$$\boxed{\text{retarded time} \qquad t_r = t - \frac{|\vec{r} - \vec{r}\,'|}{c}} \qquad (10.70)$$

$$\boxed{\begin{array}{l} \text{retarded} \\ \text{potentials} \end{array} \quad V(\vec{r}, t) = \frac{1}{4\pi\epsilon_o} \int_{\mathcal{V}} d\tau' \frac{\rho(\vec{r}\,', t_r)}{|\vec{r} - \vec{r}\,'|} \qquad \vec{A}(\vec{r}, t) = \frac{\mu_o}{4\pi} \int_{\mathcal{V}} d\tau' \frac{\vec{J}(\vec{r}\,', t_r)}{|\vec{r} - \vec{r}\,'|}}$$
$$(10.71)$$

We check below that these forms in fact solve the inhomogeneous wave equations.

We would like to act with $\Box^2$ on integrals of the following type:

$$\phi(\vec{r}, t) = \int_{\mathcal{V}} d\tau' \frac{f(\vec{r}\,', t_r)}{|\vec{r} - \vec{r}\,'|} \tag{10.72}$$

Since $\Box^2$ acts on the $\vec{r}$ (rather than the $\vec{r}\,'$) coordinate, we can pass it through the integral to act on the integrand. The time term of $\Box^2$ is easy:

$$\frac{\partial^2}{\partial t^2} \left[\frac{f(\vec{r}\,', t_r)}{|\vec{r} - \vec{r}\,'|} \right] = \frac{1}{|\vec{r} - \vec{r}\,'|} \frac{\partial^2}{\partial t^2} f(\vec{r}\,', t_r) = \frac{1}{|\vec{r} - \vec{r}\,'|} \frac{\partial^2}{\partial t_r^2} f(\vec{r}\,', t_r) \tag{10.73}$$

because $\frac{\partial t_r}{\partial t} = 1$. For the space term in $\Box^2$, we will need $\nabla^2 f(\vec{r}\,', t_r)$. Calculating this by brute force is difficult because $\vec{\nabla}$ is with respect to $\vec{r}$ but t_r depends on $|\vec{r} - \vec{r}\,'|$. We'll employ a similar trick as we did when doing such manipulations in electrostatics, which is to define $\vec{s} = \vec{r} - \vec{r}\,'$ (so that $t_r = t - s/c$) and recognize that

$$\nabla_{\vec{r}}^2 f(\vec{r}\,', t_r)\big|_{\vec{r}} = \nabla_{\vec{s}}^2 f(\vec{r}\,', t_r)\big|_{\vec{s} = \vec{r} - \vec{r}\,'} \tag{10.74}$$

because the expression is a scalar and thus cannot depend on the coordinate system origin aside from evaluating it at the correct physical location (the business with the arguments in the subscripts). Note that $\vec{s}$ and $\vec{r}\,'$ are independent variables, just as $\vec{r}$ and $\vec{r}\,'$ were independent, so $\nabla_{\vec{s}}^2$ does not act on the $\vec{r}\,'$ dependence of $f(\vec{r}\,', t_r)$!

Let's evaluate the easier expression (recognizing t_r only depends on $s = |\vec{s}|$):

$$\nabla_{\vec{s}}^2 f(\vec{r}', t_r) = \frac{1}{s^2} \frac{\partial}{\partial s} \left(s^2 \frac{\partial}{\partial s} f(\vec{r}', t_r) \right) = \frac{1}{s^2} \frac{\partial}{\partial s} \left(s^2 \frac{\partial f}{\partial t_r} \frac{\partial t_r}{\partial s} \right) \tag{10.75}$$

$$= \frac{1}{s^2} \frac{\partial}{\partial s} \left(s^2 \frac{\partial f}{\partial t_r} \left(-\frac{1}{c} \right) \right) = \frac{1}{c^2} \frac{\partial^2 f}{\partial t_r^2} - \frac{2}{s\,c} \frac{\partial f}{\partial t_r} \tag{10.76}$$

Again, $\partial/\partial s$ did not act on the $\vec{r}'$ dependence of $f(\vec{r}', t_r)$ because the transformation of variables was from $(\vec{r}, \vec{r}', t)$ to $(\vec{s}, \vec{r}', t)$ and so the independence of $\vec{r}$ and $\vec{r}'$ translates to independence of $\vec{s}$ and $\vec{r}'$. From the above, we obtain the expression we will need below:

$$\frac{\nabla_{\vec{r}}^2 f(\vec{r}', t_r)}{|\vec{r} - \vec{r}'|} = \frac{1}{c^2} \frac{\partial^2 f}{\partial t_r^2} \frac{1}{|\vec{r} - \vec{r}'|} - \frac{2}{c} \frac{\partial f}{\partial t_r} \frac{1}{|\vec{r} - \vec{r}'|^2} \tag{10.77}$$

In a similar fashion, again because it is a scalar, we may evaluate another expression we will need below:

$$2 \left[\vec{\nabla}_{\vec{r}} f(\vec{r}', t_r) \right] \cdot \vec{\nabla}_{\vec{r}} \left(\frac{1}{|\vec{r} - \vec{r}'|} \right) = 2 \left[\vec{\nabla}_{\vec{s}} f(\vec{r}', t_r) \right] \cdot \vec{\nabla}_{\vec{s}} \left(\frac{1}{s} \right) \tag{10.78}$$

$$= \frac{2}{c\,s^2} \frac{\partial f}{\partial t_r} = \frac{2}{c} \frac{\partial f}{\partial t_r} \frac{1}{|\vec{r} - \vec{r}'|^2} \tag{10.79}$$

Now, to proceed with the second term in $\square^2$ involving ∇^2. Let's work on it using the product rule (this is a bit cleaner than the way Griffiths does it):

$$\nabla_{\vec{r}}^2 \left[\frac{f(\vec{r}\,', t_r)}{|\vec{r} - \vec{r}\,'|} \right] = \left[\vec{\nabla}_{\vec{r}} \cdot \frac{\vec{\nabla}_{\vec{r}} f(\vec{r}\,', t_r)}{|\vec{r} - \vec{r}\,'|} + \vec{\nabla}_{\vec{r}} \cdot \left(f(\vec{r}\,', t_r) \vec{\nabla}_{\vec{r}} \left(\frac{1}{|\vec{r} - \vec{r}\,'|} \right) \right) \right] \quad (10.80)$$

$$= \left[\frac{\nabla_{\vec{r}}^2 f(\vec{r}\,', t_r)}{|\vec{r} - \vec{r}\,'|} + f(\vec{r}\,', t_r) \nabla_{\vec{r}}^2 \left(\frac{1}{|\vec{r} - \vec{r}\,'|} \right) \right.$$

$$\left. + 2 \left[\vec{\nabla}_{\vec{r}} f(\vec{r}\,', t_r) \right] \cdot \vec{\nabla}_{\vec{r}} \left(\frac{1}{|\vec{r} - \vec{r}\,'|} \right) \right] \quad (10.81)$$

$$= \left[\frac{1}{|\vec{r} - \vec{r}\,'|} \frac{1}{c^2} \frac{\partial^2 f}{\partial t_r^2} + f(\vec{r}\,', t_r) \left[4\,\pi \delta(\vec{r} - \vec{r}\,') \right] \right] \quad (10.82)$$

where we plugged in the expressions we derived above for the first and third terms on the second line and discarded the cancelling terms, and we used the usual relation between the delta function and the Laplacian of the inverse distance. In the end result we have, we see that the first term will cancel the $(1/c^2)\,\partial^2/\partial t^2$ term from $\square^2$. The second term with the delta function can be easily evaluated (note that doing so gives $t_r = t$, also).

Combining the two terms and applying the integral, we have

$$\Box^2 \phi(\vec{r}, t) = \int_{\mathcal{V}} d\tau' \left[\frac{1}{c^2} \frac{\partial^2}{\partial t^2} - \nabla_{\vec{r}}^2 \right] \left(\frac{f(\vec{r}', t_r)}{|\vec{r} - \vec{r}'|} \right) \tag{10.83}$$

$$= 4\pi \int_{\mathcal{V}} d\tau' \, f(\vec{r}', t_r) \, \delta(\vec{r} - \vec{r}') = 4\pi \, f(\vec{r}, t) \tag{10.84}$$

Applying this to our proposed expressions for $V(\vec{r}, t)$ and $\vec{A}(\vec{r}, t)$, we obtain

$$\Box^2 V(\vec{r}, t) = \frac{\rho(\vec{r}, t)}{\epsilon_o} \qquad\qquad \Box^2 \vec{A}(\vec{r}, t) = \mu_o \, \vec{J}(\vec{r}, t) \tag{10.85}$$

as desired.

We need to confirm the Lorenz gauge condition is satisfied since we assumed it. We will need to calculate the divergence of the integrand in the expression for $\vec{A}$. Recall:

$$\vec{\nabla}_{\vec{r}} \frac{1}{|\vec{r} - \vec{r}'|} = -\frac{\vec{r} - \vec{r}'}{|\vec{r} - \vec{r}'|^3} = \frac{\vec{r}' - \vec{r}}{|\vec{r} - \vec{r}'|^3} = -\vec{\nabla}_{\vec{r}'} \frac{1}{|\vec{r} - \vec{r}'|} \qquad (10.86)$$

Therefore,

$$\vec{\nabla}_{\vec{r}} \cdot \left(\frac{\vec{J}(\vec{r}', t_r)}{|\vec{r} - \vec{r}'|} \right) = \frac{\vec{\nabla}_{\vec{r}} \cdot \vec{J}(\vec{r}', t_r)}{|\vec{r} - \vec{r}'|} + \vec{J}(\vec{r}', t_r) \cdot \vec{\nabla}_{\vec{r}} \left(\frac{1}{|\vec{r} - \vec{r}'|} \right) \qquad (10.87)$$

$$= \frac{\vec{\nabla}_{\vec{r}} \cdot \vec{J}(\vec{r}', t_r)}{|\vec{r} - \vec{r}'|} - \vec{J}(\vec{r}', t_r) \cdot \vec{\nabla}_{\vec{r}'} \left(\frac{1}{|\vec{r} - \vec{r}'|} \right) \qquad (10.88)$$

$$= \frac{\vec{\nabla}_{\vec{r}} \cdot \vec{J}(\vec{r}', t_r)}{|\vec{r} - \vec{r}'|} - \vec{\nabla}_{\vec{r}'} \cdot \left(\frac{\vec{J}(\vec{r}', t_r)}{|\vec{r} - \vec{r}'|} \right) + \frac{\vec{\nabla}_{\vec{r}'} \cdot \vec{J}(\vec{r}', t_r)}{|\vec{r} - \vec{r}'|} \qquad (10.89)$$

$$= -\vec{\nabla}_{\vec{r}'} \cdot \left(\frac{\vec{J}(\vec{r}', t_r)}{|\vec{r} - \vec{r}'|} \right) \qquad (10.90)$$

$$+ \frac{1}{|\vec{r} - \vec{r}'|} \left[\frac{\partial \vec{J}}{\partial t_r} \cdot \vec{\nabla}_{\vec{r}} t_r - \frac{\partial \rho}{\partial t_r} + \frac{\partial \vec{J}}{\partial t_r} \cdot \vec{\nabla}_{\vec{r}'} t_r \right]$$

where the continuity equation was used when evaluating $\vec{\nabla}_{\vec{r}'} \cdot \vec{J}(\vec{r}', t_r)$ but not for $\vec{\nabla}_{\vec{r}} \cdot \vec{J}(\vec{r}', t_r)$ because the argument of $\vec{J}$ must match the divergence's variable for continuity to apply. Note that the $\frac{\partial}{\partial t_r}$ are evaluated at $(\vec{r}', t_r)$.

Using the same kind of argument as we used on the previous page for showing $\vec{\nabla}_{\vec{r}}(1/|\vec{r}-\vec{r}'|) = -\vec{\nabla}_{\vec{r}'}(1/|\vec{r}-\vec{r}'|)$, we have

$$\vec{\nabla}_{\vec{r}}\, t_r = -\frac{1}{c}\,\vec{\nabla}_{\vec{r}}|\vec{r}-\vec{r}'| = \frac{1}{c}\,\vec{\nabla}_{\vec{r}'}|\vec{r}-\vec{r}'| = -\vec{\nabla}_{\vec{r}'}\, t_r \qquad (10.91)$$

Therefore, those terms cancel each other and we have

$$\vec{\nabla}_{\vec{r}}\cdot\left(\frac{\vec{J}(\vec{r}',t_r)}{|\vec{r}-\vec{r}'|}\right) = -\vec{\nabla}_{\vec{r}'}\cdot\left(\frac{\vec{J}(\vec{r}',t_r)}{|\vec{r}-\vec{r}'|}\right) - \frac{1}{|\vec{r}-\vec{r}'|}\frac{\partial\rho}{\partial t_r} \qquad (10.92)$$

We may now calculate the divergence of $\vec{A}$:

$$\vec{\nabla}_{\vec{r}}\cdot\vec{A}(\vec{r},t) = \vec{\nabla}_{\vec{r}}\cdot\frac{\mu_o}{4\,\pi}\int_{\mathcal{V}} d\tau'\,\frac{\vec{J}(\vec{r}',t_r)}{|\vec{r}-\vec{r}'|} = \frac{\mu_o}{4\,\pi}\int_{\mathcal{V}} d\tau'\,\vec{\nabla}_{\vec{r}}\cdot\left(\frac{\vec{J}(\vec{r}',t_r)}{|\vec{r}-\vec{r}'|}\right) \qquad (10.93)$$

$$= \frac{\mu_o}{4\,\pi}\int_{\mathcal{V}} d\tau'\left[-\vec{\nabla}_{\vec{r}'}\cdot\left(\frac{\vec{J}(\vec{r}',t_r)}{|\vec{r}-\vec{r}'|}\right) - \frac{1}{|\vec{r}-\vec{r}'|}\frac{\partial\rho}{\partial t_r}\right] \qquad (10.94)$$

$$= -\epsilon_o\,\mu_o\,\frac{\partial}{\partial t}\frac{1}{4\,\pi\,\epsilon_o}\int_{\mathcal{V}} d\tau'\,\frac{\rho(\vec{r}',t_r)}{|\vec{r}-\vec{r}'|} = -\epsilon_o\,\mu_o\,\frac{\partial}{\partial t}\,V(\vec{r},t) \qquad (10.95)$$

where the first term in the second line has been transformed into a surface integral at infinity that vanishes because the current distribution is assumed to be finite in extent. We convert $\partial/\partial t_r$ to $\partial/\partial t$ by using $\partial t/\partial t_r = 1$. The Lorenz condition is thus satisfied.

Fields from Retarded Potentials: Jefimemko's Equations

The natural next step is to calculate the fields from the above retarded potentials. This is straightforward using calculations similar to what we did above (easier, in fact, because we are only evaluating first derivatives rather than $\square^2$). First, let's write down a generic expression for a first derivative:

$$\frac{\partial}{\partial r_i}\left(\frac{f(\vec{r}\,', t_r)}{|\vec{r} - \vec{r}\,'|}\right) = \frac{1}{|\vec{r} - \vec{r}\,'|}\frac{\partial f}{\partial t_r}\frac{\partial t_r}{\partial r_i} + f(\vec{r}\,', t_r)\frac{\partial}{\partial r_i}\left(\frac{1}{|\vec{r} - \vec{r}\,'|}\right) \tag{10.96}$$

$$= \left[-\frac{1}{c}\frac{1}{|\vec{r} - \vec{r}\,'|}\frac{\partial f}{\partial t_r} - \frac{f(\vec{r}\,', t_r)}{|\vec{r} - \vec{r}\,'|^2}\right]\frac{(\vec{r} - \vec{r}\,')\cdot\hat{r}_i}{|\vec{r} - \vec{r}\,'|} \tag{10.97}$$

$$\text{where}\quad \frac{\partial t_r}{\partial r_i} = -\frac{1}{c}\frac{\partial}{\partial r_i}|\vec{r} - \vec{r}\,'| = -\frac{1}{c}\frac{(\vec{r} - \vec{r}\,')\cdot\hat{r}_i}{|\vec{r} - \vec{r}\,'|} \tag{10.98}$$

$$\text{Then,}\quad \vec{E}(\vec{r}, t) = -\vec{\nabla}_{\vec{r}}\,V(\vec{r}, t) - \frac{\partial}{\partial t}\vec{A}(\vec{r}, t) \tag{10.99}$$

$$= -\frac{1}{4\pi}\int_{\mathcal{V}}d\tau'\left[\frac{1}{\epsilon_o}\vec{\nabla}_{\vec{r}}\left(\frac{\rho(\vec{r}\,', t_r)}{|\vec{r} - \vec{r}\,'|}\right) + \mu_o\frac{\partial}{\partial t}\left(\frac{\vec{J}(\vec{r}\,', t_r)}{|\vec{r} - \vec{r}\,'|}\right)\right] \tag{10.100}$$

$$= \frac{1}{4\pi\epsilon_o}\int_{\mathcal{V}}d\tau'\left[\rho(\vec{r}\,', t_r)\frac{\vec{r} - \vec{r}\,'}{|\vec{r} - \vec{r}\,'|^3} + \frac{1}{c}\frac{\partial\rho}{\partial t_r}\frac{\vec{r} - \vec{r}\,'}{|\vec{r} - \vec{r}\,'|^2} - \frac{1}{c^2}\frac{\partial\vec{J}}{\partial t_r}\frac{1}{|\vec{r} - \vec{r}\,'|}\right]$$
$$\tag{10.101}$$

where, as usual, ρ and $\vec{J}$ are evaluated at $(\vec{r}\,', t_r)$ and we used $\epsilon_o\,\mu_o = 1/c^2$.

Calculating the magnetic field is straightforward:

$$\vec{B}(\vec{r}, t) = \vec{\nabla}_{\vec{r}} \times \vec{A}(\vec{r}, t) = \frac{\mu_o}{4\pi} \int_{\mathcal{V}} d\tau' \, \vec{\nabla}_{\vec{r}} \times \left(\frac{\vec{J}(\vec{r}', t_r)}{|\vec{r} - \vec{r}'|} \right) \qquad (10.102)$$

$$= \frac{\mu_o}{4\pi} \int_{\mathcal{V}} d\tau' \sum_{i,j,k=1}^{3} \epsilon_{ijk} \, \widehat{r}_i \, \frac{\partial}{\partial r_j} \left(\frac{J_k(\vec{r}', t_r)}{|\vec{r} - \vec{r}'|} \right) \qquad (10.103)$$

$$= \frac{\mu_o}{4\pi} \int_{\mathcal{V}} d\tau' \sum_{i,j,k=1}^{3} \epsilon_{ijk} \, \widehat{r}_i \left[-\frac{1}{c} \frac{1}{|\vec{r} - \vec{r}'|} \frac{\partial J_k}{\partial t_r} - \frac{J_k(\vec{r}', t_r)}{|\vec{r} - \vec{r}'|^2} \right] \frac{(\vec{r} - \vec{r}') \cdot \widehat{r}_j}{|\vec{r} - \vec{r}'|} \qquad (10.104)$$

$$= \frac{\mu_o}{4\pi} \int_{\mathcal{V}} d\tau' \left[\frac{\vec{J}(\vec{r}', t_r)}{|\vec{r} - \vec{r}'|^2} + \frac{1}{c} \frac{1}{|\vec{r} - \vec{r}'|} \frac{\partial \vec{J}}{\partial t_r} \right] \times \frac{\vec{r} - \vec{r}'}{|\vec{r} - \vec{r}'|} \qquad (10.105)$$

where we obtained a sign flip in the last step by exchanging the j and k indices.

We thus have *Jefimenko's Equations*:

$$\vec{E}(\vec{r}, t) = \frac{1}{4\,\pi\,\epsilon_o} \int_{\mathcal{V}} d\tau' \left[\rho(\vec{r}', t_r) \frac{\vec{r} - \vec{r}'}{|\vec{r} - \vec{r}'|^3} + \frac{1}{c} \frac{\partial \rho}{\partial t_r} \frac{\vec{r} - \vec{r}'}{|\vec{r} - \vec{r}'|^2} - \frac{1}{c^2} \frac{\partial \vec{J}}{\partial t_r} \frac{1}{|\vec{r} - \vec{r}'|} \right]$$

(10.106)

$$\vec{B}(\vec{r}, t) = \frac{\mu_o}{4\,\pi} \int_{\mathcal{V}} d\tau' \left[\frac{\vec{J}(\vec{r}', t_r)}{|\vec{r} - \vec{r}'|^2} + \frac{1}{c} \frac{1}{|\vec{r} - \vec{r}'|} \frac{\partial \vec{J}}{\partial t_r} \right] \times \frac{\vec{r} - \vec{r}'}{|\vec{r} - \vec{r}'|}$$

(10.107)

These equations are not incredibly useful in practice because it is usually easier to calculate the retarded potentials and differentiate to find the fields, but they give a clear physical understanding of how disturbances in the source distributions determine the fields at a distance.

The terms that fall off like $1/|\vec{r} - \vec{r}'|^2$ are just Coulomb's Law and the Biot-Savart Law, evaluated at t_r. They simply account for the time delay in propagation of the information about the source position.

We note that the $\partial/\partial t_r$ terms in each equation fall off like $1/|\vec{r} - \vec{r}'|$ while the other terms fall off like $1/|\vec{r} - \vec{r}'|^2$, so it is these $\partial/\partial t_r$ terms that dominate at large distances. We won't do the calculations using the above expressions (we will use the retarded potentials), but this is an interesting fact because it tells us that it is the *time variation* of the sources that generates the strongest fields at large distances (the *radiated* fields). A heuristic (*i.e.*, handwaving) way of understanding this is to attribute it to the fact that time variation generates EM waves, which are self-propagating, while static sources and charges must generate the distant field directly.

Heald and Marion §8.2 has some nice discussion of the interpretation of these equations as well as references to some papers that discuss the topic further. We note two of their points here:

▶ It is frequently said (as we have said) that changing electric and magnetic fields induce magnetic and electric fields, respectively, and this is how EM waves can propagate. While this is true mathematically in Maxwell's Equations, this is something of an artifact. Maxwell's Equations are *local* equations, meaning they relate the fields, their derivatives, and the sources at a particular point in space and time. But if we look at Jefimemko's Equations, *which are exactly true*, we see that fields at a distance are sourced by charges and currents. In effect, Faraday's Law and the displacement current term in Ampere's Law are intermediaries necessary to make Maxwell's Equations local.

▶ There is also an interesting asymmetry in the above equations. The $\partial \vec{J}/\partial t_r$ term in the expression for the electric field is the term for which Faraday's Law is the intermediary: calculate a contribution to the electric field at a given space-time point $(\vec{r}, t)$ from the rate of change of the magnetic field at that point, which depends on the rate of change of the sourcing current at $(\vec{r}\,', t_r)$. (It is the retarded version of Equation 7.57.) Yet there is no corresponding $\partial \rho/\partial t_r$ term in the magnetic field expression to handle the displacement current term! The retardation alone calculates the displacement current term in Ampere's Law. This is not explicitly obvious, but can be seen by inference by

the following argument. Let's Taylor expand the source charge and current distributions around $t = t_r$:

$$\rho(\vec{r}', t) = \rho(\vec{r}', t_r) + \left.\frac{\partial \rho}{\partial t}\right|_{\vec{r}', t_r} \left(\frac{|\vec{r} - \vec{r}'|}{c}\right) + \frac{1}{2}\left.\frac{\partial^2 \rho}{\partial t^2}\right|_{\vec{r}', t_r} \left(\frac{|\vec{r} - \vec{r}'|}{c}\right)^2 + \cdots$$

$$\vec{J}(\vec{r}', t) = \vec{J}(\vec{r}', t_r) + \left.\frac{\partial \vec{J}}{\partial t}\right|_{\vec{r}', t_r} \left(\frac{|\vec{r} - \vec{r}'|}{c}\right) + \frac{1}{2}\left.\frac{\partial^2 \vec{J}}{\partial t^2}\right|_{\vec{r}', t_r} \left(\frac{|\vec{r} - \vec{r}'|}{c}\right)^2 + \cdots$$

That is, we can obtain the first two terms in both of Jefimenko's Equations simply by accounting for retardation. The third term in the electric field equation is the manifestation of Faraday's Law, as we said above. But there is no corresponding term in the magnetic field expression for the displacement current, so we are forced to conclude that the retardation calculation accounts for it. Amazing!

(We also see from the above Taylor expansion how, again, the first terms in Jefimenko's Equations are Coulomb's Law and the Biot-Savart Law evaluated at t_r.)

Another nice feature of Jefimemko's Equations is that they make it possible to formally define the quasistatic approximation. Let τ_s be the characteristic timescale on which a source term varies, as defined by

$$\frac{1}{\tau_s} = \frac{1}{f}\frac{\partial f}{\partial t_r} \tag{10.108}$$

where f is any source term (ρ or a component of $\vec{J}$). Then the second terms are smaller than the first terms by the factor

$$\alpha = \frac{|\vec{r} - \vec{r}'|}{c}\frac{1}{f}\frac{\partial f}{\partial t_r} = \frac{|\vec{r} - \vec{r}'|/c}{\tau_s} \tag{10.109}$$

which is the ratio of the light travel time to the source variation timescale. The quasistatic approximation consists of assuming that $\alpha \ll 1$, which justifies neglecting these terms and setting $t_r = t$.

The third term in the $\vec{E}$ equation requires a bit more analysis. It involves the current, but we cannot compare it to the current terms in the $\vec{B}$ equation because $\vec{E}$ and $\vec{B}$ have different units. So we must compare it to the other terms in the $\vec{E}$ equation, which involve ρ. Let's write $J = \rho\, v$ (drop the vector nature, it's not important for this analysis), so then the ratio of the third to the first term is

$$\frac{|\vec{r} - \vec{r}\,'|}{c} \frac{1}{\rho\, c} \frac{\partial J}{\partial t} = \frac{|\vec{r} - \vec{r}\,'|}{c} \frac{1}{\rho\, c} \frac{\partial(\rho v)}{\partial t} = \frac{|\vec{r} - \vec{r}\,'|}{c} \left[\beta \frac{1}{\rho} \frac{\partial \rho}{\partial t} + \frac{1}{c} \frac{\partial v}{\partial t} \right] \qquad (10.110)$$

$$= \alpha\, \beta + \frac{|\vec{r} - \vec{r}\,'|/c}{\tau_a}$$

where $\beta = v/c < 1$ and $\tau_a^{-1} = (1/c)\, \partial v/\partial t$ is a *relativistic* (because it is c, not v in the denominator) acceleration timescale. The first of the two above terms thus only becomes important when both α and β are near unity; *i.e.*, when the source variation timescale approaches the light travel time *and* when the particles in the current are moving relativistically. The second of the two above terms only becomes important when the particles' *relativistic* acceleration timescale τ_a approaches the light travel time.

Section 11
Relativity and Electrodynamics

Lecture 32:

Review of Special Relativity I:
Review of Postulates and Lorentz Transformation
Implications of Lorentz Transformation
Four-Vectors and Tensors
Covariant and Contravariant Indices
Covariant Gradient
Proper Time, Four-Velocity, and Velocity Addition

Date Revised: 2024/05/16 07:00
Revised lecture break
(added Proper Time, Four-Velocity, and Velocity Addition)
Date Given: 2024/05/14

Study Guide

Since you already studied special relativity in Ph1b, we will not repeat the discussion of all the thought experiments that lead one to it (Griffiths §12.1). The notes include a more rigorous derivation of the Lorentz Transformation than in Ph1b, but the result is the same, so we will leave that for you to read outside of class. We will quickly review the basic results of special relativity you learned in Ph1b (Griffiths §12.1.3). We will then introduce four-vectors, the metric, and tensors (Griffith §12.1.4, with added rigor), and proceed to discuss relativistic velocity (Griffiths §12.2.1). We skip relativistic collisions (Griffiths §12.2.3) because you have seen it before in Ph1a and we do not need that material for what follows. We will then switch to a discussion of how to rewrite electrodynamics in relativistic notation, including writing the total electromagnetic force (electric + magnetic) in a form that respects special relativity and uses the relativistic notation (Griffiths §12.3.3–12.3.5), including derivation of the separate transformation properties of the electric and magnetic fields (yielding the results in Griffiths §12.3.2 via a different approach). In the latter, you will see that we derive no new physics but rather that electrodynamics is written far more naturally and concisely in this notation. We end by discussing relativistic energy-momentum and dynamics (Griffiths §12.2.4) in the context of electrodynamics. There is a lot of optional material on the relativistic stress-energy tensor, angular momentum tensor, and conservation laws at the end of the section.

Griffiths' discussion in §12.3.1-12.3.2 is largely a duplication of what you saw in Ph1bc, so we will not reproduce that discussion here. §12.3.1 demonstrates, using the movement of a charged particle near a current-carrying wire, that a force that we consider magnetic in one frame of reference is purely electric in the rest frame of the charged particle. §12.3.2 derives the transformation of the fields by applying appropriate Lorentz contractions and time dilations to specific cases like capacitor plates and solenoids. Of course, you should review these discussions, which show that the electric and magnetic fields are tied to each other not just through Maxwell's Equations but also through relativistic transformations.

Fundamentals of Special Relativity

Reference Frames, Coordinate Representations, Transformation Properties, Invariance, and Covariance

So far in this course, we have not concerned ourselves much with how the physical objects we have defined — *e.g.*, charge, currents, fields, forces, etc. — depend on the frame in which they are written. That is generally a topic more relevant for mechanics. But special relativity is mechanics and, more importantly, transformation properties of these various quantities will play a central role in determining how we write down electrodynamics in a way that is manifestly consistent with special relativity. So let's define a few terms that we will use in that discussion.

▶ **Reference Frame**
  As you know, special relativity is concerned with how physical quantities behave when we look at them in different *inertial reference frames* — frames that are moving at fixed velocity. A frame that moves at fixed velocity is simply one in which Newton's Laws are obeyed. We will denote such frames by F, $\tilde{F}$, etc.

▶ **Coordinate Representation**
Any physical quantity has an existence independent of any specific coordinate
system or reference frame. (We use both terms because one can imagine
different coordinate systems that describe the same reference frame, and one
might not consider the coordinate systems of two reference frames to be
different if they are aligned and they coincide at one instant in time.) When we
write down physical laws, such as Newton's Law or Coulomb's Law or the
Biot-Savart Law or Maxwell's Equations, those are statements about
relationships between physical quantities that are independent of coordinate
system or reference frame. Frequently, though, to actually execute those
relationships, we do need to pick a coordinate system or reference frame. The
coordinate representation of a physical quantity, such as an electric current, is
the set of numbers that define it in a given coordinate system or reference
frame. This representation can vary between coordinate systems or reference
frames. But that variation does not change the physical meaning of the quantity
to which one is referring. For the example of a current, which is a vector, its
coordinate representation will depend on coordinate system orientation. But it is
still the same current independent of that choice.

▶ **Transformation Properties; Invariance and Covariance**
The term we use for the way a quantity behaves under such a change of
coordinate system or reference frame is **transformation properties**. (Note: when
we say "coordinate system or reference frame" here, we are not implying both
together, but rather one or the other. A quantity's transformation properties are
generally different under the two when special relativity is considered!) A
physical quantity that does not change under a coordinate system or reference
frame change is called **invariant**. An electrostatic potential difference between
two points is, for example, invariant under translational or rotational
transformations of the coordinate system because its coordinate representation is
the same in all such coordinate systems. (We shall see that, in special relativity,
electrostatic potential difference is not independent of inertial reference frame
velocity!) On the other hand, an electric field's coordinate representation — the
three numbers representing it — does depend on coordinate system rotation.
When there are well-defined rules for how the representations in different
coordinate systems or reference frames are related, the quantity is called
covariant to indicate the existence of such rules. To provide a counterexample,
the z-component of an electric current is neither invariant nor covariant, but the
three components of the electric field together are covariant.

There is clearly great value in writing down the laws of physics in terms of invariant and covariant quantities: this ensures our laws can be written down in a manner that is manifestly independent of coordinate system or reference frame. After reviewing special relativity, out goal is to formulate electrodynamics in such a way that the transformation properties of the various quantities we deal with — source distributions, potentials, fields, etc. — are made clear. We will see that electrodynamics is deeply related to relativity, a realization that in fact initiated the development of special relativity by Poincaré, Lorentz, Einstein, and others. In fact, Einstein's first paper on special relativity is titled "On the Electrodynamics of Moving Bodies" (Annalen der Physik, **17**: 891 (1905)).

Sign Convention

As with our definition of $\Box^2$, we deviate from Griffiths and follow the more standard sign convention for relativity where the inner product of a four-vector is of the form $(\text{time component})^2 - (\text{space components})^2$.

Postulates of Special Relativity

The two basic postulates of special relativity are

1. Physics is the same in all inertial frames.
2. The speed of light is the same in all inertial frames.

The first postulate is just the principle of Galilean relativity that we discussed in the context of Faraday's Law. At that point, we defined an inertial frame to be a frame in which Newton's Laws hold. The implication of Galilean relativity is that there is no absolute frame of reference; every inertial reference frame is as good as every other one.

It is the second postulate that was Einstein's brilliant leap and that leads to all the nonintuitive implications of special relativity. To some extent, the second postulate is a corollary of the first once one realizes that electromagnetic waves do not travel in a medium. If the laws of electromagnetism, which give rise to the speed of light, are to be the same in all frames, then the speed of light must of necessity be the same in all frames. But it took the Michelson-Morley experiment to kill the concept of the ether, a medium in which light propagates.

Derivation of the Lorentz Transformation

Consider two inertial reference frames F and $\widetilde{F}$. Let them have coordinate axes (x, y, z, t) and $(\widetilde{x}, \widetilde{y}, \widetilde{z}, \widetilde{t})$. We include t as a coordinate and allow it to be different in the $\widetilde{F}$ frame because it will be necessary to avoid a contradiction with the second postulate. Let the two systems' axes and origins coincide at $t = \widetilde{t} = 0$. Let the $\widetilde{F}$ frame be moving at speed $v = c\,\beta$ along the $+x$ axis with respect to F, which means that the position of the $\widetilde{F}$ origin obeys $x = \beta\,c\,t$ in the F frame.

path of origin
of $\tilde{F}$ system,
slope = β^{-1}

$c\,t$

path of light
(slope = 1)

x
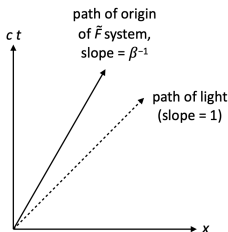
This information is summarized in our **space-time diagram** in the F system, a plot of t vs. x with the origin of the $\widetilde{F}$ system represented by the solid line of slope β^{-1} and the path of a light ray emitted from the origin shown by the dashed line of slope 1. Points in the space-time diagram are referred to as **events** because they are not just points in space, but in time also. Obviously, one can generalize space-time to more than one spatial dimension, it just becomes hard to visualize.

Coordinate Axis Parallel to Motion

Let us consider the path of a light ray in the two frames, requiring that the second postulate hold. Suppose the light ray is emitted from the $\widetilde{F}$ origin at time $c\,\widetilde{t}_e = -\widetilde{L}$ in the $+x$ direction, it hits a mirror at $\widetilde{x}_r = \widetilde{L}$ at $c\,\widetilde{t}_r = 0$ and is reflected, returning to the $\widetilde{F}$ origin at $c\,\widetilde{t}_a = \widetilde{L}$. (We use $\widetilde{L}$ instead of L to avoid confusion when we later discuss length contraction.) The light ray has $\widetilde{y} = \widetilde{z} = 0$ for all time. In a space-time diagram of the $\widetilde{F}$ frame, the reflection event $(\widetilde{x}_r, c\,\widetilde{t}_r)$ is obtained by the intersection of light rays propagating forward in time from the emission event $(\widetilde{x}_e, c\,\widetilde{t}_e) = (0, -\widetilde{L})$ and backward in time from the absorption event $(\widetilde{x}_a, c\,\widetilde{t}_a) = (0, \widetilde{L})$. The intersection is at $(\widetilde{x}_r, c\,\widetilde{t}_r) = (\widetilde{L}, 0)$.
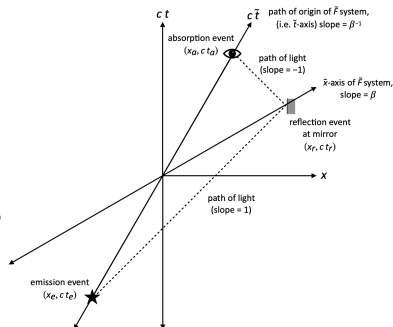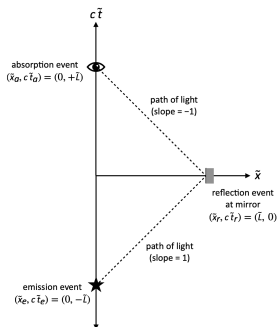
Left: $\widetilde{F}$ system. Right: F system.

Let's examine the path in the F system, assuming we do not yet know the transformation law for coordinates between reference frames. The light ray travels only along the x direction, so it satisfies $y = z = 0$ for all time also. Let $(x_e, c\, t_e)$, $(x_r, c\, t_r)$ and $(x_a, c\, t_a)$ indicate the coordinates of the three events in the F frame. The emission and absorption events must occur on the solid line representing the position of the $\widetilde{F}$ origin, which means their space and time coordinates are related by the slope β^{-1}. The line of this slope through the origin thus gives the space-time location of the $c\,\widetilde{t}$ axis in F. Time reversal invariance says that these events' symmetric occurence in $\widetilde{F}$ implies they occur at symmetric times in F. So the two events must be at $(-c\, t_a\, \beta, -c\, t_a)$ and $(c\, t_a\, \beta, c\, t_a)$. Let's calculate where light rays from these two events would intersect, which will give us the position of the reflection event in F, which yields the space-time location of the $\widetilde{x}$ axis in F. In the following, r and s begin as undetermined parameters indicating how much time passes in the F frame between the emission or absorption event and the reflection event, respectively:

$$(-c\, t_a\, \beta, -c\, t_a) + r\, c\, (1, 1) = (c\, t_a\, \beta, c\, t_a) + s\, c\, (1, -1) \tag{11.1}$$

$$r - s = 2\beta\, t_a \qquad r + s = 2\, t_a \tag{11.2}$$

$$r = (1 + \beta)\, t_a \qquad s = (1 - \beta)\, t_a \tag{11.3}$$

$$(x_r, c\, t_r) = (-c\, t_a\, \beta, -c\, t_a) + r\, c\, (1, 1) = (c\, t_a, \beta\, c\, t_a) \tag{11.4}$$

The last line implies that the reflection event $(x_r, c\, t_r)$ sits on the line through the origin with slope β. Its position on that line depends on the value of t_a, the time of the absorption event — that is, on $\widetilde{L}$ and β. Since the reflection event $(\widetilde{x}_r, c\,\widetilde{t}_r)$ is always at $c\,\widetilde{t}_r = 0$, the line thus tells us where the $\widetilde{x}$ axis sits in the F space-time diagram. So, in sum, we have that the $\widetilde{t}$ axis is a line of slope β^{-1} and the $\widetilde{x}$ axis is a line of slope β, both going through the F origin.

We may construct the general form for the transformation of coordinates from $\widetilde{F}$ to F by using the above information along with expected properties of the transformation:

1. **Linearity**: We have seen that events on the $\widetilde{x}$ axis lie on a line of slope β and events on the $c\,\widetilde{t}$ axis lie on a line of slope β^{-1} in F. If we assume the transformation is linear in the space-time coordinates $(\widetilde{x}, c\,\widetilde{t})$ then the transformation for an arbitrary space time event with $\widetilde{F}$ frame coordinates $(\widetilde{x}, c\,\widetilde{t})$ to the F frame can be written

$$(x, c\,t) = \gamma(\beta)\,(1, \beta)\,\widetilde{x} + \tilde{\gamma}(\beta)\,(\beta, 1)\,c\,\widetilde{t} \qquad (11.5)$$

   Note that we only require linearity in $(\widetilde{x}, c\,\widetilde{t})$, not in β.

2. **Equivalence of the two frames:** Since one of our postulates is that physics is the same in any reference frame, the transformation for going from F to $\widetilde{F}$ must have the same form as the transformation from $\widetilde{F}$ to F, modulo the change in sign of β, so we also have

$$(\widetilde{x}, c\,\widetilde{t}) = \gamma(-\beta)\,(1, -\beta)\,x + \tilde{\gamma}(-\beta)\,(-\beta, 1)\,c\,t \qquad (11.6)$$

3. **Symmetry of the two frames:** By symmetry, the slope of the x and t axes in the $\widetilde{F}$ space-time diagram must be the same as the slope of the $\widetilde{x}$ and $\widetilde{t}$ axes in the F space-time diagram, so $\gamma(\beta) = \gamma(-\beta)$ and $\tilde{\gamma}(\beta) = \tilde{\gamma}(-\beta)$ is required (the sign flip has already been accounted for). So we have

$$\widetilde{x} = \gamma(|\beta|)\, x - \tilde{\gamma}(|\beta|)\, \beta\, c\, t \qquad\qquad c\, \widetilde{t} = -\gamma(|\beta|)\, \beta\, x + \tilde{\gamma}(|\beta|)\, c\, t \qquad (11.7)$$

$$x = \gamma(|\beta|)\, \widetilde{x} + \tilde{\gamma}(|\beta|)\, \beta\, c\, \widetilde{t} \qquad\qquad c\, t = \gamma(|\beta|)\, \beta\, \widetilde{x} + \tilde{\gamma}(|\beta|)\, c\, \widetilde{t} \qquad (11.8)$$

4. **The transformation be invertible by the reverse transformation:** We require that if we transform from $\widetilde{F}$ to F and then from F to $\widetilde{F}$, the overall transformation should return the original $(\widetilde{x}, c\, \widetilde{t})$. We apply this by using the first pair of formulae for x and $c\, t$ in the second pair of formulae and requiring that we recover $\widetilde{x}$ and $c\, \widetilde{t}$:

$$[\gamma(|\beta|)]^2 - \tilde{\gamma}(|\beta|)\, \gamma(|\beta|) = 0 \qquad [\gamma(|\beta|)]^2 - \tilde{\gamma}(|\beta|)\, \gamma(|\beta|)\, \beta^2 = 1 \qquad (11.9)$$

which is solved by $\qquad \tilde{\gamma}(|\beta|) = \gamma(|\beta|) = \dfrac{1}{\sqrt{1 - \beta^2}} \qquad\qquad\qquad (11.10)$

With these conditions, the transformation law is

$$\boxed{\begin{array}{llc} \widetilde{x} = \gamma\,(x - \beta\, c\, t) & c\, \widetilde{t} = -\gamma\,(\beta\, x - c\, t) & \text{Lorentz} \\[4pt] x = \gamma\,(\widetilde{x} + \beta\, c\, \widetilde{t}) & c\, t = \gamma\,(\beta\, \widetilde{x} + c\, \widetilde{t}) & \text{Transformation} \end{array}} \qquad (11.11)$$

This nomenclature "Lorentz Transformation" is historical: Lorentz derived these before Einstein proposed special relativity.

Coordinates Perpendicular to the Direction of Motion

What happens to the coordinates perpendicular to the boost direction, y and z or $\widetilde{y}$ and $\widetilde{z}$? It turns out they are unaffected. We use an argument similar to the one that we used to deduce the Lorentz transformation for x and t.

Consider again two frames F and $\widetilde{F}$, with $\widetilde{F}$ moving at speed β in the $+x$ direction relative to F, and assume the origins of the two frames coincide at $t = \widetilde{t} = 0$. In this case, we will emit a light ray from the origin in the $+y$ direction and reflect it back to the origin. The emission, reflection, and absorption events are given by (neglecting the z coordinate, which is not involved at all):

$$(\widetilde{x}_e, \widetilde{y}_e, c\,\widetilde{t}_e) = (0, 0, -\widetilde{L}) \tag{11.12}$$

$$(\widetilde{x}_r, \widetilde{y}_r, c\,\widetilde{t}_r) = (0, \widetilde{L}, 0) \tag{11.13}$$

$$(\widetilde{x}_a, \widetilde{y}_a, c\,\widetilde{t}_a) = (0, 0, \widetilde{L}) \tag{11.14}$$

Let's now repeat our argument regarding the coordinates of the three events in F, with the complication that our space-time now has three dimensions, though the motion is only in two of them.

Walking through the argument:

1. The position of the origin of $\widetilde{F}$ in F is described by the same line as in our original argument, except it is now a line in the x-ct plane in three dimensions. Its coordinates in F are $(x_o, y_o, ct) = (\beta\, ct, 0, ct)$ as a function of ct. We know the origin's y coordinate does not change because there is no motion in that direction; whatever nonintuitive there may be about relativity, relative motion of two points in a given frame is always well-defined.

2. As before, the emission and absorption events both occur at the position of the origin of the $\widetilde{F}$ system in F, and they must occur symmetrically about the time origin. So we have

$$(x_e, y_e, c\, t_e) = -c\, t_a\, (\beta, 0, 1) \qquad (11.15)$$

$$(x_a, y_a, c\, t_a) = c\, t_a\, (\beta, 0, 1) \qquad (11.16)$$

Symmetry about the origin relates the time coordinates and the known velocity vector of $\widetilde{F}$ in F relates the space coordinates of the two events.

3. Though the reflection event occurs at $\widetilde{y} \neq 0$ and $y \neq 0$, the emission and absorption events occur at the origin and thus are unaffected by the existence of the y dimension. They must thus obey the Lorentz transformation rule, which implies

$$(x_e, y_e, c\, t_e) = -\gamma\, \widetilde{L}\, (\beta, 0, 1) \qquad (11.17)$$

$$(x_a, y_a, c\, t_a) = \gamma\, \widetilde{L}\, (\beta, 0, 1) \qquad (11.18)$$

4. Let us now calculate the position of the intersection of light rays sent out from the emission and absorption events. We use the postulate that the speed of light is c in F. The space-time displacement of a light ray in one unit of $c\,t$ is $(\sqrt{1-\delta^2}, \delta, 1)$, where δ allows for freedom in the direction (in the xy plane). δ carries the same sign as y_r. We have:

$$(x_e, y_e, c\,t_e) + r\left(\sqrt{1-\delta^2}, \delta, 1\right) = (x_a, y_a, c\,t_a) + s\left(-\sqrt{1-\delta^2}, \delta, -1\right) \quad (11.19)$$

$$-\gamma\,\widetilde{L}\left(\beta, 0, 1\right) + r\left(\sqrt{1-\delta^2}, \delta, 1\right) = \gamma\,\widetilde{L}\left(\beta, 0, 1\right) + s\left(-\sqrt{1-\delta^2}, \delta, -1\right) \quad (11.20)$$

Let's explain this a bit more. We may assume the velocity vector has the same components on the two sides, with just a sign flip in the y component, because the light path must be symmetric about $t = \tilde{t} = 0$ because the motion is along x, not y. The signs are obtained as follows:

▶ The signs on the left-side velocity term are obtained by simple arguments:

  ▶ If $y_e = 0$ and $y_r > 0$, then the y light velocity must be positive between the two events.
  ▶ In the limit $\beta \ll 1$, we must recover the nonrelativistic limit, and in that limit we know that if the light ray always has $\tilde{x} = 0$ and the $\tilde{x}$ origin is moving in the $+x$ direction, then $x \geq 0$ is required for the position of the light ray.

- ▶ The signs on the right-side are obtained by similar arguments:
  - ▶ If $y_r > 0$ and $y_a = 0$, then the y light velocity must be negative between the two events.
  - ▶ Again, the x light velocity must be nonnegative to obtain the nonrelativistic limit.

After a bit of manipulation, the equations for the three coordinates are:

$$r + s = 2\,\gamma\,\widetilde{L} \qquad (r + s)\,\sqrt{1 - \delta^2} = 2\,\beta\,\gamma\,\widetilde{L} \qquad (r - s)\,\delta = 0 \qquad (11.21)$$

Simplifying, we obtain

$$r = s = \gamma\,\widetilde{L} \qquad \sqrt{1 - \delta^2} = \beta \qquad \delta = \sqrt{1 - \beta^2} = \gamma^{-1} \qquad (11.22)$$

The reflection event thus satisfies

$$(x_r, y_r, t_r) = (x_e, y_e, c\, t_e) + r\,(\sqrt{1 - \delta^2}, \delta, 1) \qquad (11.23)$$

$$= -\gamma\,\widetilde{L}\,(\beta, 0, 1) + \gamma\,\widetilde{L}\,(\beta, \gamma^{-1}, 1) = (0, \widetilde{L}, 0) \qquad (11.24)$$

We thus see that the transverse coordinate is unchanged by the Lorentz transformation. This would hold for z also.

Implications of the Lorentz Transformation

We can derive a number of the most shocking implications of relativity from the simple Lorentz transformation laws:

▶ **Time dilation**
Consider two events occuring at a fixed point in space in the frame $\widetilde{F}$; for example, two ticks of a clock. They are separated by the vector $(\widetilde{x}, c\,\widetilde{t}) = (0, c\,\tau)$. What is the separation of the two events in the frame F, relative to which $\widetilde{F}$ is moving at speed β? The Lorentz transformation tells us

$$x = \gamma\,\beta\,c\,\tau \qquad c\,t = \gamma\,c\,\tau \qquad (11.25)$$

The time between the events in the F frame is larger than in $\widetilde{F}$. Hence, the term "time dilation": time "slows down" in the moving frame, the two events have a smaller time separation in their rest frame than in any other frame. One is not obtaining something for nothing, though, because the spatial separation of the two events has become nonzero. That is, we are no longer measuring just the time separation of two events that occur at the same point in space; we are measuring a separation with both time and space components.

▶ **Length contraction**
The length $\widetilde{L}$ of an object at rest in $\widetilde{F}$ can be viewed as two events with separation $(\Delta \widetilde{x}, c \, \Delta \widetilde{t}) = (\widetilde{L}, 0)$, representing the left and right ends of the object at some common time $c \, \widetilde{t}$. The time separation between these two events will become nonzero in the F frame because of their nonzero spatial separation, so these two events are not valid as a length measurement in F. Explicitly, and without loss of generality, we let the two events corresponding to the $\widetilde{F}$ length measurement be

$$(\widetilde{x}_1, c \, \widetilde{t}_1) = (0, 0) \qquad (\widetilde{x}_2, c \, \widetilde{t}_2) = (\widetilde{L}, 0) \qquad (11.26)$$

The Lorentz transformation of the trajectories is

$$(x_1, c \, t_1) = (0, 0) \qquad (x_2, c \, t_2) = (\gamma \, \widetilde{L}, \gamma \, \beta \, \widetilde{L}) \qquad (11.27)$$

$$\implies \qquad (\Delta x, \Delta t) = (\gamma \, \widetilde{L}, \gamma \, \beta \, \widetilde{L}) \qquad (11.28)$$

confirming the expectation that the length measurement events in $\widetilde{F}$ do not give a length measurement in F. To make a length measurement in F, we must pick points on the two trajectories that are separated by $\Delta t = 0$, which the above expression implies is not possible if $\widetilde{t}_1 = \widetilde{t}_2$. Let us allow $\widetilde{t}_1 \neq \widetilde{t}_2$ and apply the requirement $\Delta t = 0$, using the Lorentz transformation to write this condition in terms of $\widetilde{F}$ coordinates:

$$0 = c \, \Delta t = c \, t_2 - c \, t_1 = \gamma \left[\beta \left(\widetilde{x}_2 - \widetilde{x}_1 \right) + \left(c \, \widetilde{t}_2 - c \, \widetilde{t}_1 \right) \right] \qquad (11.29)$$

Without loss of generality, we again assume event 1 occurs when the two origins intersect, so $\widetilde{x}_1 = x_1 = c\,\widetilde{t}_1 = c\,t_1 = 0$. We choose $c\,t_2 = 0$ also in order to obtain a length measurement in F. The assumption about the origins implies that $\widetilde{x}_2 = \widetilde{L}$ for all time because the object is at rest in $\widetilde{F}$. With these assumptions, the $c\,\Delta t = 0$ requirement reduces to

$$c\,\widetilde{t}_2 = -\beta\,\widetilde{x}_2 = -\beta\,\widetilde{L} \tag{11.30}$$

With the same assumptions, the length as measured in F corresponds to the Lorentz transformation of event 2:

$$x_2 = \gamma\left(\widetilde{x}_2 + \beta\,c\,\widetilde{t}_2\right) = \gamma\left(\widetilde{x}_2 - \beta^2\widetilde{x}_2\right) = \gamma^{-1}\widetilde{x}_2 = \gamma^{-1}\widetilde{L} \tag{11.31}$$

We see a decrease in the apparent length. The F-frame length measurement has space-time coordinates

$$(x_1, c\,t_1) = (0, 0) \qquad\qquad (\widetilde{x}_1, c\,\widetilde{t}_1) = (0, 0) \tag{11.32}$$

$$(x_2, c\,t_2) = (\gamma^{-1}\widetilde{L}, 0) \qquad\qquad (\widetilde{x}_2, c\,\widetilde{t}_2) = (\widetilde{L}, -\beta\,\widetilde{L}) \tag{11.33}$$

We see that in order to be simultaneous in the F frame, the two events must occur with negative time separation, i.e., with $c\,\Delta\widetilde{t} = c\,\widetilde{t}_2 - c\,\widetilde{t}_1 = -\beta\,\widetilde{L}$, in the rest frame of the object. The object is moving to the right with the $\widetilde{F}$ frame, so the right end of the object is not at $x = \widetilde{L}$ yet in the F frame when the measurement event occurs in F, hence the apparent length contraction.

▶ **Relativity of Simultaneity**
  The length contraction example shows that two events that are temporally
  simultaneous but spatially separated in one frame may not be temporally
  simultaneous in another; *i.e.*, $c\,\Delta\widetilde{t} = 0$ does not imply $c\,\Delta t = 0$ unless $\Delta\widetilde{x} = 0$
  also. Generically, then, simultaneity of physically separated events is no longer
  well defined. Discussion below of light cones extends this point to show that
  time-ordering can be frame dependent, but only for events that cannot be in
  causal contact with each other, eliminating any issue of non-physicality.

▶ **Transformations of Areas**
  The matrix of partial derivatives of our transformation is

  $$\mathbf{J}(\beta) = \begin{bmatrix} \gamma & -\beta\,\gamma \\ -\beta\,\gamma & \gamma \end{bmatrix} \tag{11.34}$$

  So, space-time areas are preserved:

  $$d\widetilde{x}\,c\,d\widetilde{t} = |\mathbf{J}|\,dx\,c\,dt = dx\,c\,dt \tag{11.35}$$

  because $\gamma^2 - \beta^2\gamma^2 = 1$. This is a necessity, as there must be symmetry between
  the two directions for the Lorentz transformation, which would not hold if the
  Jacobian determinant were not unity.

▶ **Invariant Interval**
  From the Lorentz transformation and again using $\gamma^2 - \beta^2\gamma^2 = 1$, one can show

  $$(c\,t)^2 - x^2 = (c\,\widetilde{t})^2 - (\widetilde{x})^2 \equiv s^2 \tag{11.36}$$

  The quantity s^2 is the **invariant interval** associated with the space-time vector
  $(x, c\,t)$ and $(\widetilde{x}, c\,\widetilde{t})$. It can be thought of like the magnitude of a vector in
  space, which is invariant under spatial rotations. The invariant interval is
  invariant under Lorentz transformations, also known as **boosts**.

Geometric Interpretation of Lorentz Transformation

We may develop a geometrical interpretation of how the Lorentz transformation gives the shape of one set of space-time axes in another frame. Define the **rapidity** or **boost angle** or **boost parameter** η by $\tanh \eta = \beta$ (any η is possible because $\beta = \tanh \eta$ goes to ± 1 as $\eta \to \pm\infty$). Then we have

$$\beta = \tanh \eta \qquad \gamma = \cosh \eta \qquad \beta \gamma = \sinh \eta \qquad (11.37)$$

$$x = \widetilde{x} \cosh \eta + c\,\widetilde{t} \sinh \eta \qquad c\,t = \widetilde{x} \sinh \eta + c\,\widetilde{t} \cosh \eta \qquad (11.38)$$

So, the Lorentz transformation looks something like a coordinate rotation, except by an imaginary angle $i\,\eta$. More importantly, though, we see the contours of constant $\widetilde{x}$ or $c\,\widetilde{t}$ as β (*i.e.*, η) varies form hyperbolic curves in the $x\,t$ plane. That is, the event $(\widetilde{x}, 0)$ in the $\widetilde{F}$ frame appears on the hyperbola $(x, c\,t) = \widetilde{x}\,(\cosh \eta, \sinh \eta)$ in F, with η increasing as β increases. Similarly, the event $(0, c\,\widetilde{t})$ appears on the hyperbola $(x, c\,t) = c\,\widetilde{t}\,(\sinh \eta, \cosh \eta)$ in F. This is illustrated in the figure on the following slide.

NOTE: Even though we have made the above geometrical intepretation, one has to be a bit careful about overinterpreting it. The difficulty is that the quantity left constant by a Lorentz transformation, the invariant interval, does not correspond to a curve of fixed distance from the origin on the above plots. Rather, it corresponds to the hyperbolic curves that the fiducial points follow as β is changed. Another way of seeing this is that the entire first and third quadrants of the $\widetilde{F}$ frame occupy the area between the two corresponding slanted space-time axes displayed in the right plot. Therefore, spatial distances and areas are not preserved by the mapping. Invariant interval is preserved but does not correspond to a fixed distance on the plot!
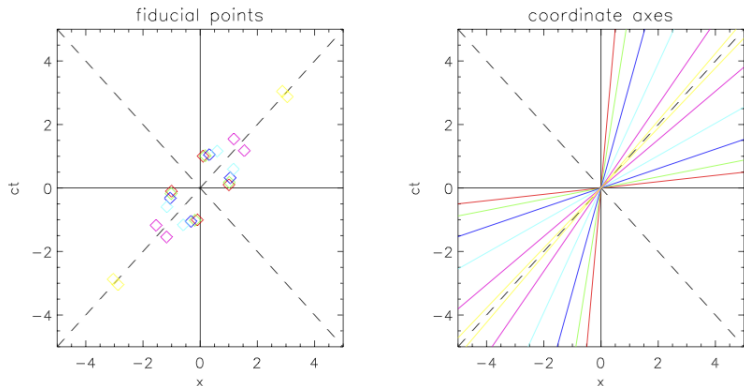
Left: position of $(\widetilde{x}, c\,\widetilde{t}) = (1,0), (0,1), (-1,0), (0,-1)$ in F for $\eta = 10^{-1}$, $10^{-3/4}$, $10^{-1/2}$, $10^{-1/4}$, 1, $10^{1/4}$ (moving sequentially outward). This shows how a particular event is seen in another frame as the relative speed β is increased. Right: $\widetilde{F}$ space-time axes in F for same values of η (larger $\eta \Longrightarrow$ increasingly oblique). This plot shows how the $\widetilde{F}$ space-time axes appear "squeezed together" when seen in the F frame. Both plots are for positive β and η. Negative β and η would occupy the other two quadrants.

Light Cones, Causality, and Simultaneity

The geometrical interpretation and invariant interval will make it possible to prove that the relativity of simultaneity can never produce causality problems.

First, let us define the **light cone** as the region in space-time that can be reached from a given event; refer to the figure on the slide at the end of this section. The light cone of the event at the origin of a space-time diagram consists of all events with $|c\,t| > |x|$. The region $c\,t < 0$, the **past light cone**, consists of all events that could have causal influence on the event $(x, c\,t) = (0,0)$. The region $c\,t > 0$, the **future light cone**, consists of all events that the event $(x, c\,t) = (0,0)$ can have causal influence on. Since we know that, under Lorentz transformations, an event slides along a hyperbolic curve in space-time as the speed β is varied, we are assured that events that are in the future light cone of the event at the origin in one reference frame are also in that event's future light cone in any other reference frame. Similarly for past light cones.

What about simultaneous events? Two events are simultaneous in a particular frame $\widetilde{F}$ if they occur at the same value of $\widetilde{t}$, the time in the frame $\widetilde{F}$. Let one event be at the origin $(0,0)$ and the other event at $(\widetilde{x}, 0)$, so the space-time vector separating them is $(\widetilde{x}, 0)$. This event is outside the light cone of the origin, so, in this frame, the two events are out of causal contact. When we transform to a different frame F, but one whose origin coincides with that of $\widetilde{F}$ at $t = \widetilde{t} = 0$, the space-time vector between the two events will slide on the hyperbolic curve $(\cosh \eta, \sinh \eta)$. The event at $(\widetilde{x}, 0)$ in $\widetilde{F}$ may move to $t < 0$ or $t > 0$ in other frames, raising causality questions. The causality worries are put to rest by the fact that the hyperbola is entirely outside the light cone of the first event, so the second event is always outside of causal contact with the first event.

Whether two events are causally connected is determined by the sign of the invariant interval of the space-time vector separating them. If $s^2 > 0$, then $|c\,t| > |x|$ and the two events connected by the vector are in causal contact. The vector is called **time-like**. Our argument about hyperbolic curves ensures that the sign of the time component of the vector does not change, preserving causal relationships. If $s^2 < 0$, then $|c\,t| < |x|$ and the two events are never in causal contact, regardless of frame. The vector is called **space-like**. The sign of the time component of the vector may depend on the reference frame. If $s^2 = 0$, the vector is called **null** or **light-like** since only light (or, as we shall see, any other massless particle) can travel on such a path in space-time. The concepts are illustrated on the following slide.
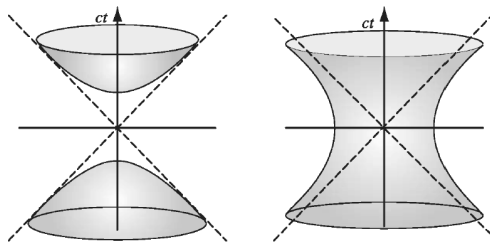
© 2013 Griffiths, *Introduction to Electrodynamics*

Space-time diagrams illustrating light cones, invariant intervals, and the preservation of causality. The dashed lines indicate the light cone of the event at $(x, c\,t) = (0,0)$. (Left) Surface of constant $s^2 > 0$, corresponding to a time-like interval. Lorentz transformations move space-time events along each half of this surface but cannot cause events to move from one half to the other half and thus cannot change the sign of t or affect causality. (Right) Surface of constant $s^2 < 0$, corresponding to a space-like interval. Lorentz transformations also move space-time events along this surface, and the sign of t may change. However, none of these events are in the light cone of $(0,0)$ and thus sign changes in t cannot affect causality.

Four-Vectors in Special Relativity

A *four-vector* $\overset{\Rightarrow}{r}$ is an object whose *coordinate representation* in a given reference frame F consists of four numbers r^μ and for which the transformation of this coordinate representation between inertial references frames F and $\widetilde{F}$, with F having relative velocity $c\,\vec{\beta} = c\,\beta\,\widehat{x}$ with respect to $\widetilde{F}$ (F and $\widetilde{F}$ are exchanged relative to our previous derivations for reasons of convention), is given by the Lorentz transformation law

$$\widetilde{r}^\mu = \sum_{\mu,\nu=0}^{3} \Lambda^\mu_{\ \nu}\, r^\nu \equiv \Lambda^\mu_{\ \nu}\, r^\nu \qquad \text{with} \quad \Lambda^\mu_{\ \nu} = \begin{bmatrix} \gamma & \gamma\beta & 0 & 0 \\ \gamma\beta & \gamma & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \qquad (11.39)$$

We introduce Greek indices to indicate 0, 1, 2, 3 (as opposed to just 1, 2, 3) and the *Einstein summation convention* wherein a repeated index implies a sum over that index. The term *contraction* is used to refer to matching up an index and summing over it. We will use Latin indices to indicate the space components 1, 2, 3 only.

The prototypical four-vector is the space-time position vector

$$r^\mu = (r^0, r^1, r^2, r^3) = (ct, x, y, z) \tag{11.40}$$

Such four-vectors are termed "Lorentz covariant" to indicate that, while they do vary between reference frames, there are rules for how they vary and those rules are given by the Lorentz transformation law.

A quick way to check the sign of the matrix is to check the transformation properties of the origin of the F frame, with $r^\mu = (ct, 0, 0, 0)$. Using the above matrix, we get $\widetilde{r}^\mu = (\gamma ct, \gamma \beta ct, 0, 0)$: the $\widetilde{x} = \widetilde{r}^1$ coordinate is positive, as expected if F moves at $c\,\beta\,\widehat{x}$ with respect to $\widetilde{F}$. Note how time dilation manifests itself in the transformation.

By application of three-dimensional rotation matrices, one can show that, for an arbitrary direction of motion $\vec{\beta}$, the Lorentz transformation matrix has the form

$$\Lambda^0{}_0 = \gamma \qquad \Lambda^0{}_i = \Lambda^i{}_0 = \gamma\,\beta_i \qquad \Lambda^i{}_j = \delta_{ij} + (\gamma - 1)\,\frac{\beta_i\,\beta_j}{\beta^2} \qquad i, j = 1, 2, 3 \tag{11.41}$$

where δ_{ij} is the usual Kronecker delta.

The Invariant Norm of a Four-Vector and the Metric

We saw above in Equation 11.36 that we have a quantity connected to $\overrightarrow{r}$ that is invariant under Lorentz transformations (change of reference frame), the **invariant interval**, more generally called the **invariant norm** in the context of a general four-vector (as opposed to a four-vector corresponding to an interval between space-time events). We may write it using our four-vector notation as follows, also defining the **Lorentz scalar product**:

$$\overrightarrow{r} \cdot \overrightarrow{r} \equiv |\overrightarrow{r}|^2 = |r^\mu|^2 = g_{\mu\nu}\, r^\mu\, r^\nu \quad \text{with} \quad g_{\mu\nu} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{bmatrix} \tag{11.42}$$

g is the *metric*. One can check that this particular definition of g and the resulting definition of $|\overrightarrow{r}|^2$ is indeed independent of reference frame by applying the Lorentz transformation to show explicitly

$$\widetilde{g}_{\mu\nu}\, \widetilde{r}^\mu\, \widetilde{r}^\nu = g_{\mu\nu}\, \widetilde{r}^\mu\, \widetilde{r}^\nu = g_{\mu\nu}\, r^\mu\, r^\nu \tag{11.43}$$

The chain of logic is important: g cannot be *derived* from some more fundamental principle; it has this form because this form provides the invariant quantity we wanted.

Note that the $\cdot$ notation for the Lorentz scalar product will quickly become useless for tensors, which may have more than two indices. We introduce it for analogy to the three-dimensional dot product. The contraction notation is Lorentz-covariant and is thus entirely sufficient.

Tensors

More generally, a *n-th rank tensor* $\mathcal{T}$ is an object that has 4^n components organized with n indices, $T^{\mu_1 \cdots \mu_n}$, and that transforms under a change of reference frame via the rule

$$\boxed{\widetilde{T}^{\mu_1 \cdots \mu_n} = \Lambda^{\mu_1}{}_{\nu_1} \cdots \Lambda^{\mu_n}{}_{\nu_n} \, T^{\nu_1 \cdots \nu_n}} \tag{11.44}$$

i.e., contraction with Lorentz transformation matrices in all the indices. A four-vector is a first-rank tensor. We will see second-rank tensors soon enough. A *four-scalar* (or just *scalar*) is a zeroth-rank tensor and thus is invariant under Lorentz transformations; conversely, any quantity that is invariant under Lorentz transformations is a scalar. From now on, when we talk about tensors, we also include four-vectors in the category. Given its transformation properties, the metric is a second-rank tensor.

Covariant and Contravariant Indices

With $g_{\mu\nu}$ defined as it is, we introduce the concept of "lowering" and "raising" indices. For tensors, we have only defined "raised" indices. We *define* "lowered indices" via contraction with the metric:

$$r_\mu = g_{\mu\nu} \, r^\nu \qquad T^{\mu_1 \cdots \mu_{j-1}}{}_\nu{}^{\mu_{j+1} \cdots \mu_n} = g_{\nu\mu_j} \, T^{\mu_1 \cdots \mu_{j-1} \mu_j \mu_{j+1} \cdots \mu_n} \tag{11.45}$$

It is straightforward to see that $r_0 = r^0$ and $r_i = -r^i$ for $i = 1, 2, 3$.

If we are going to lower indices, we need a way to raise them, so we *define* in any frame

$$g^{\mu\nu} \, g_{\nu\sigma} = g_{\mu\nu} \, g^{\nu\sigma} = \delta^\mu_\sigma = \delta_\mu{}^\sigma \qquad \Longleftrightarrow \qquad g^{\mu\nu} \equiv (g^{-1})_{\mu\nu} \tag{11.46}$$

where the $^{-1}$ implies the matrix inversion operation and where both versions of δ are the identity matrix with ones along the diagonal in any reference frame. We can check that this definition of $g^{\mu\nu}$ raises indices in a manner consistent with our definition of how indices are lowered:

$$g^{\mu\nu} r_\nu = g^{\mu\nu} \, g_{\nu\sigma} \, r^\sigma = \delta^\mu_\sigma \, r^\sigma = r^\mu \tag{11.47}$$

That is, we recover the raised-index four-vector we started with if we use the above definitions.

Now, because g is so simple, it turns out that $g^{\mu\nu} = g_{\mu\nu}$ (just try it out in any given reference frame). Note also that

$$g^{\mu}_{\ \nu} = g^{\mu\sigma}\, g_{\sigma\nu} = \delta^{\mu}_{\ \nu} \tag{11.48}$$

Moreover, it doesn't make much sense to write $\delta^{\mu\nu}$ or $\delta_{\mu\nu}$ because they are the metric,

$$\delta_{\mu\sigma} = g_{\mu\nu}\delta^{\nu}_{\ \sigma} = g_{\mu\sigma} \qquad \delta^{\mu\sigma} = g^{\mu\nu}\delta_{\nu}^{\ \sigma} = g^{\mu\sigma} \tag{11.49}$$

and thus writing $\delta_{\mu\sigma}$ or $\delta^{\mu\sigma}$ would be confusing: some of their elements would be -1! It is ok to write $g^{\mu}_{\ \nu}$ or $g_{\mu}^{\ \nu}$, but it is preferred to use $\delta^{\mu}_{\ \nu}$ or $\delta_{\mu}^{\ \nu}$ to avoid obfuscation.

We can derive Lorentz transformation properties for "lowered" indices by using the known properties of the raised indices and the metric:

$$\widetilde{r}_{\mu} = \widetilde{g}_{\mu\nu}\, \widetilde{r}^{\nu} = g_{\mu\nu}\, \Lambda^{\nu}_{\ \sigma}\, r^{\sigma} = g_{\mu\nu}\, \Lambda^{\nu}_{\ \sigma}\, g^{\sigma\lambda}\, r_{\lambda} \tag{11.50}$$

$$\implies \boxed{\widetilde{r}_{\mu} = \Lambda_{\mu}^{\ \lambda}\, r_{\lambda} \quad \text{with} \quad \Lambda_{\mu}^{\ \lambda} = g_{\mu\nu}\, g^{\sigma\lambda}\, \Lambda^{\nu}_{\ \sigma}} \tag{11.51}$$

where we used $\widetilde{g}_{\mu\nu} = g_{\mu\nu}$ (definition of metric) in the first step, the definition of the raising operation in the second step, and then we applied a self-consistent use of raising and lower operations on $\Lambda^{\nu}_{\ \sigma}$ in the third step. If one works it out in detail, one can see that the components $\Lambda_{\mu}^{\ \nu}$ are given by $\Lambda^{\mu}_{\ \nu}$ with a sign flip on the velocity.

Raised indices are called *contravariant* indices and lowered indices are called *covariant*. The rationale is as follows.

For the sake of developing intuition, consider a coordinate system rotation instead of Lorentz transformation. The rotation changes the coordinate axes. But a point in space remains fixed — the rotation is just a relabeling of points in space. So its coordinates in the new coordinate system are different than those in the old coordinate system. The amount by which the coordinates must change is exactly the opposite of the amount that the axes changed. For example, suppose one rotates about the *z*-axis by $+30°$ so that the new *x*-axis is in the first quadrant of the old coordinate system. Then a point that was on the *x*-axis of the old coordinate system is now in the fourth quadrant of the new coordinate system. When the unit vectors are rotated in the positive direction, the coordinates of points *appear* to rotate in the negative direction.

There is an analogous discussion for Lorentz transformations, leading us to conclude that the coordinates r^μ of a space-time point $\overset{\Rightarrow}{r}$ transform in a manner opposite to that by which the coordinate axes transform. Hence the term *contravariant* — "contra" meaning "opposite to". Lower indices transform like the coordinate axes, so they are called *covariant*.

A four-vector can be covariant or contravariant because it has only one index. It is more complicated for rank-*n* tensors: each index can be covariant or contravariant.

There is no deep physical meeting to covariant/contravariant indices *in special relativity* except for the "sign-flip" issue noted above. Respecting the raising/lowering conventions is necessary for a self-consistent mathematical scheme. Physically, the only implication is consistency about signs of velocities in Lorentz transformations.

Covariant Gradient

The next obvious tool we need is a four-vector version of the gradient operator. The natural thing to want is for this gradient operator to do the right thing for Taylor expansions. Given a scalar function $S\left(\vec{\vec{r}}\right)$, we would like

$$S\left(\vec{\vec{r}} + d\vec{\vec{r}}\right) - S\left(\vec{\vec{r}}\right) = \left[\vec{\vec{\nabla}}S\right] \cdot d\vec{\vec{r}} = \left[\vec{\vec{\nabla}}S\right]_\mu dr^\mu \qquad (11.52)$$

where we have used the Lorentz scalar product to ensure both sides are Lorentz scalars even though the right side is made of coordinate-system-dependent — contravariant and covariant under Lorentz transformation — quantities. We know that, since r^μ is a contravariant four-vector, its differential version dr^μ is also a contravariant four-vector. This, along with the known form of Taylor expansions for the coordinate representation in a particular reference frame, implies that

$$\left[\vec{\vec{\nabla}}S\right]_\mu = \frac{\partial S}{\partial r^\mu} \equiv \partial_\mu S \qquad \Longrightarrow \qquad \boxed{\partial_\mu \equiv \left(\frac{\partial}{\partial r^0}, \frac{\partial}{\partial r^1}, \frac{\partial}{\partial r^2}, \frac{\partial}{\partial r^3}\right)} \qquad (11.53)$$

would yield the desired result. Thus, we use the above as the definition of the *four-gradient* or *covariant gradient*.

Note that the contravariant version is

$$\partial^\mu \equiv \left(\frac{\partial}{\partial r^0}, -\frac{\partial}{\partial r^1}, -\frac{\partial}{\partial r^2}, -\frac{\partial}{\partial r^3} \right) \tag{11.54}$$

The Lorentz-invariant second derivative is called the *d'Alembertian*, which we introduced earlier:

$$\Box^2 = \vec{\overrightarrow{\nabla}} \cdot \vec{\overrightarrow{\nabla}} = \partial_\mu \, \partial^\mu = \frac{\partial}{\partial (r^0)^2} - \frac{\partial}{\partial (r^1)^2} - \frac{\partial}{\partial (r^2)^2} - \frac{\partial}{\partial (r^3)^2} \tag{11.55}$$

$$\implies \boxed{\Box^2 = \vec{\overrightarrow{\nabla}} \cdot \vec{\overrightarrow{\nabla}} = \partial_\mu \, \partial^\mu = \frac{1}{c^2} \frac{\partial}{\partial t^2} - \vec{\nabla}^2} \tag{11.56}$$

Proper Time, the Covariant Four-Velocity, and Velocity Addition

It is natural to ask whether it is possible to create a space-time vector for the velocity, an entity that transforms like the space-time position four-vector. This would provide a "unified" treatment of velocity and position. The trick, of course, is to find a set of four numbers that are covariant, that transform in the appropriate way under Lorentz transformations.

First, we must define the **proper time**. Consider a frame that is moving with the particle whose velocity we want to specify. In that frame, the particle's space-time position is always $r^\mu = (c\,\tau, 0, 0, 0)$ because the particle is at the origin. $c\,\tau$ specifies the time in the frame moving with the particle. That may seem frame-specific; but remember that $c^2\tau^2$ is the invariant interval of the particle's position. This will be the same in all reference frames that share the same spatial origin at their time origin, so it is reasonable to think of it as a quantity available in all frames.

Given the proper time, a reasonable coordinate-free definition of a **four-velocity** is

$$\overset{\Rightarrow}{v} = \frac{d}{d\tau}\,\overset{\Rightarrow}{r} \tag{11.57}$$

or, if we want the representation in a given frame

$$v^\mu = \frac{d}{d\tau}\,r^\mu \tag{11.58}$$

where r^μ is the position of the particle in the frame in which one wants to know the four-velocity (at some time t in that frame) and τ is the invariant interval associated with the particle's position at that same time t. The notation is somewhat confusing because τ seems like a rest-frame quantity, but it is also an invariant quantity. Perhaps a more obvious, though more cumbersome, definition would be

$$v^\mu = c\,\frac{d}{d\sqrt{r_\mu r^\mu}}\,r^\mu \tag{11.59}$$

where $r_\mu r^\mu = (r^0)^2 - (r^1)^2 - (r^2)^2 - (r^3)^2$ is the invariant interval associated with the position of the particle. The definition of four-velocity clearly transforms like the space-time position Lorentz vector because it is, essentially, the ratio of the space-time position Lorentz vector to a Lorentz scalar.

So, we have a good formal definition. What does it look like in terms of quantities we have easy access to, three-velocities? Consider a particle moving at velocity $c\,\vec{\beta}_p$ (possibly a function of time) in some frame F. (We use the subscript $_p$ to distinguish the particle velocity from that of the frame F relative to another frame $\widetilde{F}$.) The particle's trajectory is $r^\mu(t)$ in this frame. We have the obvious differential relations

$$dr^\mu = c\,(1, \beta_{px}, \beta_{py}, \beta_{pz})\,dt \tag{11.60}$$

$$d\tau = \gamma_p^{-1}\,dt \tag{11.61}$$

where the second relation is simply time dilation. These relations are instantaneously true even if the particle is accelerating. Thus, the four-velocity in F is

$$v^\mu = \frac{dr^\mu}{d\tau} = \frac{c\,(1, \beta_{px}, \beta_{py}, \beta_{pz})\,dt}{\gamma_p^{-1}dt} = \gamma_p c\,(1, \beta_{px}, \beta_{py}, \beta_{pz}) = \gamma_p c\left(1, \vec{\beta}_p\right) \tag{11.62}$$

where $\gamma_p = \left(1 - \left|\vec{\beta}_p\right|^2\right)^{-1/2}$ is the γ factor associated with the particle velocity $\vec{\beta}_p$. Thus, given a particle trajectory, we can easily calculate the four-velocity in any frame.

We note that we could have also obtained the four-velocity by Lorentz transformation of the four-vector $(c, \vec{0})$ from one frame F to a frame $\widetilde{F}$ in which F is moving a velocity $c\,\vec{\beta}_p$, making the argument that $(c, \vec{0})$ is the four-velocity of the F frame origin in F. It is hardly obvious that that would have been the right thing to do, though!

We may now use the Lorentz transformation of the four-velocity to derive the rule for **velocity addition**. Start with a particle with four-velocity $\vec{v}$ with coordinate representation v_1^μ in F and suppose F is moving at velocity $c\,\beta_2\hat{x}$ with respect to the frame $\widetilde{F}$. This will give us the addition of $\vec{v}$ and the four-velocity $\vec{u} = \gamma_2 c\,(1, \beta_2, 0, 0)$. It is:

$$\widetilde{v}^\mu = \left[\begin{array}{cccc} \gamma_2 & \gamma_2\,\beta_2 & 0 & 0 \\ \gamma_2\,\beta_2 & \gamma_2 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{array}\right] \gamma_1\,c \left[\begin{array}{c} 1 \\ \beta_{1,x} \\ \beta_{1,y} \\ \beta_{1,z} \end{array}\right] = \gamma_1\,c \left[\begin{array}{c} \gamma_2\,(1 + \beta_2\,\beta_{1,x}) \\ \gamma_2\,(\beta_2 + \beta_{1,x}) \\ \beta_{1,y} \\ \beta_{1,z} \end{array}\right] \quad (11.63)$$

Because $\vec{v}$ is a four-vector, we expect $\widetilde{v}^\mu = \gamma_{12}\,c\left(1, \vec{\beta}_{12}\right)$ where $\vec{\beta}_{12}$ is the velocity relative to the $\widetilde{F}$ frame. Therefore we may conclude

$$\gamma_{12} = \frac{\widetilde{v}^0}{c} = \gamma_1\gamma_2\,(1 + \beta_2\,\beta_{1,x}) \qquad \beta_{12,y} = \frac{\widetilde{v}^2}{\widetilde{v}^0} = \frac{1}{\gamma_2}\frac{\beta_{1,y}}{1 + \beta_2\,\beta_{1,x}} \qquad (11.64)$$

$$\beta_{12,x} = \frac{\widetilde{v}^1}{\widetilde{v}^0} = \frac{\beta_2 + \beta_{1,x}}{1 + \beta_2\,\beta_{1,x}} \qquad \beta_{12,z} = \frac{\widetilde{v}^3}{\widetilde{v}^0} = \frac{1}{\gamma_2}\frac{\beta_{1,z}}{1 + \beta_2\,\beta_{1,x}}$$

The expression for γ_{12} could be derived more algebraically, but the above is clearly the simplest way to obtain it.

One could of course consider a more general velocity addition by allowing F to move in an arbitrary direction with respect to $\widetilde{F}$. The algebra is worse but there is no conceptual difference.

It is possible to derive the above, or understand it, by considering all the length contractions and time dilations involved. Clearly, using the Lorentz transformation is much faster.

Lecture 33:

Relativity and Electrodynamics:
Covariance Source Density and Potential
Electromagnetic Field Tensor
Transformation of Fields, Field of a Moving Point Charge
Maxwell's Equations
Relativistic Dynamics with Electromagnetic Fields

Date Revised: 2024/05/16 07:00
Date Given: 2024/05/16

Relativistically Covariant Formulation of Electrodynamics

The Covariant Source Density

The natural way to define the charge/current source four-vector is to consider a charge distribution in its rest frame F and Lorentz transform it. If ρ_o is the rest-frame charge density, and there is no current in this rest frame because no charges are moving, then, going to a new frame $\widetilde{F}$ relative to which F is moving at velocity $c\,\vec{\beta}$, we know the charge density increases due to length contraction in the direction of motion and a current appears, given by the charge density in $\widetilde{F}$ and the velocity $c\,\vec{\beta}$:

$$\rho = \gamma\,\rho_o \qquad \vec{J} = \rho\,\vec{v} = \rho_o\,\gamma\,\vec{v} \tag{11.65}$$

This is easily summarized using the four-velocity $\overset{\Rightarrow}{v}$ we just defined:

$$\boxed{\overset{\Rightarrow}{J} = \rho_o\,\overset{\Rightarrow}{v} \qquad J^\mu = \rho_o\,v^\mu = \gamma\,(\rho_o\,c,\,\rho_o\,\vec{v}) = (\rho\,c,\,\rho\,\vec{v})} \tag{11.66}$$

Since we are multiplying a four-vector by the invariant quantity ρ_o, we are assured $\overset{\Rightarrow}{J}$ is also a four-vector, which we call the *covariant source density*. Note: ρ itself is not an invariant quantity, but the rest-frame charge density ρ_o is invariant because the total amount of charge cannot depend on the reference frame (it just consists of counting up charges!). The continuity equation is then written in manifestly Lorentz-invariant form (using the Lorentz scalar product and covariant gradient):

$$\boxed{0 = \frac{\partial \rho}{\partial t} + \vec{\nabla} \cdot \vec{J} = \partial_\mu J^\mu = \overset{\Rightarrow}{\vec{\nabla}} \cdot \overset{\Rightarrow}{J}} \tag{11.67}$$

The Covariant Potential

Recall Equations 10.21, our inhomogeneous wave equations in the Lorenz gauge:

$$\Box^2 V = \frac{\rho}{\epsilon_o} \qquad\qquad \Box^2 \vec{A} = \mu_o \vec{J} \qquad\qquad \Box^2 \equiv \frac{1}{c^2}\frac{\partial^2}{\partial t^2} - \vec{\nabla}^2 \qquad (11.68)$$

If we define the *covariant potential* $\overset{\Rightarrow}{A}$ by

$$\boxed{A^\mu \equiv \left(\frac{V}{c}, \vec{A}\right)} \qquad (11.69)$$

then the above equations can be rewritten as the *covariant wave equation*

$$\boxed{\Box^2 \overset{\Rightarrow}{A} = \mu_o \overset{\Rightarrow}{J}} \qquad (11.70)$$

Since $\overset{\Rightarrow}{J}$ is a four-vector and $\Box^2$ and μ_o are scalars, we are assured that $\overset{\Rightarrow}{A}$ is a four-vector as desired. Of course, the above equations only apply in Lorenz gauge, the condition for which can be written in a manifestly Lorentz-invariant form:

$$\boxed{0 = \epsilon_o\,\mu_o\,\frac{\partial V}{\partial t} + \vec{\nabla}\cdot\vec{A} = \partial_\mu A^\mu = \overset{\Rightarrow}{\vec{\nabla}}\cdot\overset{\Rightarrow}{A}} \qquad (11.71)$$

The Lorenz gauge condition is invariant under change of reference frame because it is a four-scalar by definition. Note how it is the natural extension of $\vec{\nabla}\cdot\vec{A} = 0$ (which is a scalar under rotations and thus independent of coordinate axis orientation.)

Lienard-Wiechert Potential via Lorentz Transformation

The definition of $\overset{\Rightarrow}{A}$ makes derivation of the potentials of a moving point charge trivial. These are called the Lienard-Wiechert potentials and their derivation without special relativity is highly nontrivial, as seen in Griffiths §10.3.1.

In the rest frame F of a point charge q, the covariant potential has components

$$V(\vec{r}, t) = \frac{1}{4\,\pi\,\epsilon_o} \frac{q}{[x^2 + y^2 + z^2]^{1/2}} \qquad \vec{A}(\vec{r}, t) = 0 \qquad (11.72)$$

Now, to have the charge move with velocity $\vec{v} = c\,\vec{\beta} = c\,\beta\,\hat{x}$, we simply need to obtain $\overset{\Rightarrow}{A}$ in a lab frame $\widetilde{F}$ relative to which the charge's rest frame F moves at $\vec{v}$. That is, we use the Lorentz transformation, Equation 11.39, on $\overset{\Rightarrow}{A}$:

$$\frac{1}{c}\,\widetilde{V} = \gamma\left[\frac{1}{c}\,V + \beta\,A_x\right] = \frac{\gamma}{c}\,V \qquad \widetilde{A}_x = \gamma\left[A_x + \beta\,\frac{1}{c}\,V\right] = \frac{\gamma\,\beta}{c}\,V \quad (11.73)$$

and $\widetilde{A}_y = A_y = 0$, $\widetilde{A}_z = A_z = 0$.

This is half the work. V is of course written in terms of the rest frame coordinates $r^\mu = (c\,t, x, y, z)$, so we need to rewrite it in terms of the lab frame coordinates $\widetilde{r}^\mu = (c\,\widetilde{t}, \widetilde{x}, \widetilde{y}, \widetilde{z})$. These coordinates are also related by Lorentz transformation, but we need the one going in the opposite direction (because we are now writing F coordinates in terms of $\widetilde{F}$ coordinates)

$$c\,t = \gamma \left[c\,\widetilde{t} - \beta\,\widetilde{x}\right] \qquad\qquad x = \gamma\left[-\beta\,c\,\widetilde{t} + \widetilde{x}\right] \qquad (11.74)$$

and $y = \widetilde{y}$ and $z = \widetilde{z}$. (You can check that this is the correct direction for the transformation by considering the position of the F frame origin in $\widetilde{F}$ frame coordinates: $(ct, x = 0)$ should obey $\widetilde{x} = \beta\,c\,\widetilde{t}$.) Combining the Lorentz transformation of the potential with the above transformation of the coordinates, we obtain:

$$\widetilde{V}(\widetilde{r}^\mu) = \frac{1}{4\,\pi\,\epsilon_o}\,\frac{\gamma\,q}{\left[\gamma^2\left(\beta\,c\,\widetilde{t} - \widetilde{x}\right)^2 + \widetilde{y}^2 + \widetilde{z}^2\right]^{1/2}} \qquad (11.75)$$

$$= \frac{1}{4\,\pi\,\epsilon_o}\,\frac{q}{\left[\left\{\left(\widetilde{x} - v\,\widetilde{t}\right)^2 + \widetilde{y}^2 + \widetilde{z}^2\right\} - \beta^2\left(\widetilde{y}^2 + \widetilde{z}^2\right)\right]^{1/2}} \qquad (11.76)$$

The coordinates in the above are the current position of the charge in $\widetilde{F}$.

Let's generalize the above expression for an arbitrary direction of motion by defining

$$R(\widetilde{r}^{\mu}) = \left[\left(\widetilde{x} - v\,\widetilde{t} \right)^2 + \widetilde{y}^2 + \widetilde{z}^2 \right]^{1/2} \qquad \sin^2\theta = \frac{\widetilde{y}^2 + \widetilde{z}^2}{\left(\widetilde{x} - v\,\widetilde{t} \right)^2 + \widetilde{y}^2 + \widetilde{z}^2} \qquad (11.77)$$

Then we have
$$\widetilde{V}(\widetilde{r}^{\mu}) = \frac{1}{4\,\pi\,\epsilon_o} \frac{q}{R(\widetilde{r}^{\mu})} \frac{1}{\left[1 - \beta^2 \sin^2\theta \right]^{1/2}} \qquad (11.78)$$

$$\widetilde{A}_i(\widetilde{r}^{\mu}) = \frac{1}{c^2} \frac{1}{4\,\pi\,\epsilon_o} \frac{q\,v_i}{R(\widetilde{r}^{\mu})} \frac{1}{\left[1 - \beta^2 \sin^2\theta \right]^{1/2}} \qquad (11.79)$$

where θ is now the angle between the direction of motion $\vec{v}$ and the vector $\vec{R}$ from the particle's current position to the position at which we want to know the potentials.

We see that the potentials are unmodified along the particle's trajectory (ahead of or behind) but that the potential strengthens as one moves away from that trajectory, becoming strongest transverse to the direction of motion.

Notice how the above expression shows us that the vector potential of the moving particle is related to the scalar potential by a factor of β/c. When we calculate forces, there will be another factor of v that shows up for magnetic force, making it smaller than the electric force by a factor of $\beta_1\,\beta_2$ where β_1 and β_2 are the factors for the source charge and the charge feeling the force.

We will see later it will be useful to rewrite these expressions in terms of the retarded position of the particle — the position at the retarded time — but we will avoid that digression for now.

Electric and Magnetic Fields

We define the *electromagnetic field tensor* or the *Faraday tensor* $\mathcal{F}$ by

$$\boxed{\mathcal{F} = \overset{\Rightarrow}{\nabla} \wedge \overset{\Rightarrow}{A} \quad \Longleftrightarrow \quad F^{\mu\nu} = \partial^\mu A^\nu - \partial^\nu A^\mu \quad \Longleftrightarrow \quad -F^{0j} = F^{j0} = \frac{1}{c} E_j \quad F^{ij} = -\epsilon_{ijk} B_k}$$
(11.80)

where we have introduced the **wedge product**, $\wedge$, which is the four-dimensional generalization of the cross-product, and where i, j, k run over $1, 2, 3$ and ϵ_{ijk} is the completely antisymmetric Levi-Civita symbol. (We define $E_j \equiv E^j$ and $B_j \equiv B^j$; we are free to do this because they are not components of a four-vector, for which the relation between contravariant and covariant components is fixed). From the above,

$$\boxed{F^{\mu\nu} = \begin{bmatrix} 0 & -E_x/c & -E_y/c & -E_z/c \\ E_x/c & 0 & -B_z & B_y \\ E_y/c & B_z & 0 & -B_x \\ E_z/c & -B_y & B_x & 0 \end{bmatrix}}$$
(11.81)

Thanks to the definition of $\mathcal{F}$ in terms of the covariant gradient and the covariant potential, we are assured it is a second-rank tensor. $\mathcal{F}$ is clearly the Lorentz-covariant way to represent the electric and magnetic fields: its second-rank tensor transformation properties provide all the information we need:

$$\boxed{\widetilde{F}^{\mu\nu} = \Lambda^\mu{}_\lambda \Lambda^\nu{}_\sigma F^{\lambda\sigma}}$$
(11.82)

Let's write these transformation laws in terms of the fields explicitly so we can see the transformation rules more clearly. As noted above, we *define* lowered roman indices *for three-tensors* by $E_j \equiv E^j$, $B_j \equiv B^j$, $\beta_i \equiv \beta^i$, $v_i \equiv v^i = c\,\beta^i$, and $\delta_{jk} \equiv \delta^j{}_k$ (we will not use v^μ here so there is no confusion); for the space components of *four-tensors*, $r_i = -r^i$, $g^j{}_k = -g^{jk}$ continue to hold. We have:

$$\widetilde{E}_j = -c\,\widetilde{F}^{0j} = -c\Lambda^0{}_\lambda\,N^j{}_\sigma\,F^{\lambda\sigma} \tag{11.83}$$

$$= -c\left[\Lambda^0{}_0\,N^j{}_0\,F^{00} + \Lambda^0{}_0\,N^j{}_k\,F^{0k} + \Lambda^0{}_k\,N^j{}_0\,F^{k0} + \Lambda^0{}_k\,N^j{}_\ell\,F^{k\ell}\right] \tag{11.84}$$

where we have expanded out the sums into space and time components. We use $F^{00} = 0$ and the generic forms of the components of Λ, Equation 11.41,

$$\widetilde{E}_j = -c\left[\gamma\left(\delta_{jk} + (\gamma - 1)\frac{\beta_j\,\beta_k}{\beta^2}\right)F^{0k} + \gamma\,\beta_k\,\gamma\,\beta_j\,F^{k0} + \gamma\,\beta_k\left(\delta_{j\ell} + (\gamma - 1)\frac{\beta_j\,\beta_\ell}{\beta^2}\right)F^{k\ell}\right] \tag{11.85}$$

Next, we use the antisymmetry of $\mathcal{F}$: $F^{k0} = -F^{0k}$ and $\beta_k\,\beta_\ell\,F^{k\ell} = 0$ in the last term of the last term. This yields

$$\widetilde{E}_j = -c\left[\gamma\,F^{0j} + \frac{\beta_j\,\beta_k}{\beta^2}\left[\gamma^2 - \gamma - \beta^2\,\gamma^2\right]F^{0k} + \gamma\,\beta_k\,F^{kj}\right] \tag{11.86}$$

Next, we use $F^{0j} = -E_j/c$ and $F^{kj} = -F^{jk} = \epsilon_{jk\ell} B_\ell$:

$$\widetilde{E}_j = \gamma\, E_j + (1 - \gamma)\, \frac{\beta_j\, \beta_k}{\beta^2}\, E_k - \gamma\, v_k\, \epsilon_{jk\ell}\, B_\ell \tag{11.87}$$

We will reduce this a simpler form soon, but let's obtain the analogous result for the magnetic field first. To extract $\vec{B}$ from $\mathcal{F}$, we need the identity $\epsilon_{ijk}\, \epsilon_{ijm} = 2\, \delta_{km}$. With this,

$$-\frac{1}{2}\, \epsilon_{jk\ell}\, F^{k\ell} = \frac{1}{2}\, \epsilon_{jk\ell}\, \epsilon_{k\ell m}\, B_m = \frac{1}{2}\, \epsilon_{k\ell j}\, \epsilon_{k\ell m}\, B_m = \delta_{jm}\, B_m = B_j \tag{11.88}$$

Therefore,

$$\widetilde{B}_j = -\frac{1}{2}\, \epsilon_{jk\ell}\, \widetilde{F}^{k\ell} = -\frac{1}{2}\, \epsilon_{jk\ell} \Lambda^k{}_\lambda\, \Lambda^\ell{}_\sigma\, F^{\lambda\sigma} \tag{11.89}$$

$$= -\frac{1}{2}\, \epsilon_{jk\ell} \left[\gamma\, \beta_k \left(\delta_{\ell m} + (\gamma - 1) \frac{\beta_\ell\, \beta_m}{\beta^2} \right) F^{0m} + \left(\delta_{km} + (\gamma - 1) \frac{\beta_k\, \beta_m}{\beta^2} \right) \gamma\, \beta_\ell\, F^{m0} \right.$$

$$\left. + \left(\delta_{km} + (\gamma - 1) \frac{\beta_k\, \beta_m}{\beta^2} \right) \left(\delta_{\ell n} + (\gamma - 1) \frac{\beta_\ell\, \beta_n}{\beta^2} \right) F^{mn} \right] \tag{11.90}$$

where we have used $F^{00} = 0$ already.

Using $F^{m0} = -F^{0m}$, $F^{m\ell} = -F^{\ell m}$, $F^{kn} = -F^{nk}$, $\epsilon_{jk\ell}\,\beta_k\,\beta_\ell = 0$ and $\beta_m\,\beta_n\,F^{mn} = 0$:

$$\widetilde{B}_j = -\frac{1}{2}\,\epsilon_{jk\ell}\left[\gamma\,\beta_k\,F^{0\ell} - \gamma\,\beta_\ell\,F^{0k} + F^{k\ell} - \frac{\gamma-1}{\beta^2}\left(\beta_k\,\beta_m\,F^{\ell m} + \beta_\ell\,\beta_n\,F^{nk}\right)\right]$$

$$(11.91)$$

then, using $\epsilon_{jk\ell} = -\epsilon_{j\ell k}$, $F^{0a} = -E_a/c$, $-\epsilon_{jk\ell}\,F^{k\ell}/2 = B_j$ and $F^{ab} = -\epsilon_{abc}B_c$:

$$\widetilde{B}_j = \frac{1}{2c}\,\gamma\left[\epsilon_{jk\ell}\,\beta_k\,E^\ell + \epsilon_{j\ell k}\,\beta_\ell\,E^k\right] + B_j - \frac{1}{2}\,\epsilon_{jk\ell}\,\frac{\gamma-1}{\beta^2}\left[\beta_k\,\beta_m\,\epsilon_{\ell mn}\,B_n + \beta_\ell\,\beta_n\,\epsilon_{nkm}\,B_m\right]$$

$$(11.92)$$

and, finally, using the cyclicity and antisymmetry of ϵ_{abc} in its indices followed by the identity $\epsilon_{abc}\,\epsilon_{dec} = \delta_{ad}\,\delta_{be} - \delta_{ae}\,\delta_{bd}$ and then apply the δ's and rearranging:

$$\widetilde{B}_j = \frac{\gamma}{c^2}\,\epsilon_{jk\ell}\,v_k\,E_\ell + B_j - \frac{1}{2}\,\frac{\gamma-1}{\beta^2}\left[\beta_k\,\beta_m\,B_n\,\epsilon_{jk\ell}\,\epsilon_{mn\ell} + \beta_\ell\,\beta_n\,B_m\,\epsilon_{j\ell k}\,\epsilon_{nmk}\right] \quad (11.93)$$

$$= \frac{\gamma}{c^2}\,\epsilon_{jk\ell}\,v_k\,E_\ell + B_j - \frac{\gamma-1}{\beta^2}\,\beta_k\,\beta_m\,B_n\left[\delta_{jm}\,\delta_{kn} - \delta_{jn}\,\delta_{km}\right] \quad (11.94)$$

$$= \frac{\gamma}{c^2}\,\epsilon_{jk\ell}\,v_k\,E_\ell + B_j - \frac{\gamma-1}{\beta^2}\left[\beta_j\,\beta_k\,B_k - \beta^2\,B_j\right] \quad (11.95)$$

$$= \gamma\,B_j + (1-\gamma)\,\frac{\beta_j\,\beta_k}{\beta^2}\,B_k + \frac{\gamma}{c^2}\,\epsilon_{jk\ell}\,v_k\,E_\ell \quad (11.96)$$

which we see is a form very similar to the one we obtained for the transformation of the electric field, Equation 11.87.

We may write Equations 11.87 and 11.96 in three-vector form:

$$\tilde{\vec{E}} = \gamma \vec{E} + (1 - \gamma) \widehat{\beta} \left(\widehat{\beta} \cdot \vec{E} \right) - \gamma \vec{\beta} \times c \vec{B} \tag{11.97}$$

$$\tilde{\vec{B}} = \gamma \vec{B} + (1 - \gamma) \widehat{\beta} \left(\widehat{\beta} \cdot \vec{B} \right) + \gamma \vec{\beta} \times \frac{\vec{E}}{c} \tag{11.98}$$

Let us resolve the fields into pieces parallel and perpendicular to the velocity:

$$\tilde{\vec{E}}_{||} = \widehat{\beta}\widehat{\beta} \cdot \vec{E} \qquad \vec{E}_{\perp} = \vec{E} - \widehat{\beta}\widehat{\beta} \cdot \vec{E} \qquad \tilde{\vec{B}}_{||} = \widehat{\beta}\widehat{\beta} \cdot \vec{B} \qquad \vec{B}_{\perp} = \vec{B} - \widehat{\beta}\widehat{\beta} \cdot \vec{B} \tag{11.99}$$

Thus, we have

$$\widetilde{E}_{||} = E_{||} \qquad \tilde{\vec{E}}_{\perp} = \gamma \left[\vec{E}_{\perp} - \vec{\beta} \times c\,\vec{B}_{\perp} \right] \tag{11.100}$$

$$\widetilde{B}_{||} = B_{||} \qquad \tilde{\vec{B}}_{\perp} = \gamma \left[\vec{B}_{\perp} + \vec{\beta} \times \frac{\vec{E}_{\perp}}{c} \right] \tag{11.101}$$

recovering the results in Griffiths §12.3.2, where the transformation properties of the fields are derived. (These are the equivalent of Equations 12.109 in Griffiths, accounting for the fact that those equations give the field in F in terms of the field in $\widetilde{F}$ and thus have a sign flip on the velocity.) We see that these properties have fallen out of the Lorentz transformation of the $\mathcal{F}$ tensor.

Notice how the expressions for $\vec{E}$ and $\vec{B}$ are analogues of each other, differing only by a sign flip on the $\vec{\beta}\times$ term and factors of c to convert units.

We first saw the mixing of electric and magnetic fields via transformation between inertial frames in the context of Faraday's Law. The field tensor codifies this mixing (with γ's added) and gives it a satisfying rationale by tying it to the general requirement of Lorentz covariance (*i.e.*, special relativity).

One must of course remember that both Faraday's Law and the invariance of the speed of light that lead to special relativity, and more generally the principle that physical laws are independent of inertial reference frame, are empirical observations that we codify via an underlying principle. The principles are profound, which may seem sufficient justification to assume them, but, ultimately, empirical verification of predictions is what justifies those principles. This is the difference between modern empiricism and the ancient Greek approach to science, where the principles were deemed sufficient and formulation and testing of predictions deemed unnecessary!

Fields of a Moving Point Charge

Let's now derive the fields of a moving point charge using Lorentz transformation properties. This is again much (much!) easier than doing it without special relativity, as is done in Griffiths §10.3.2 from the Lienard-Wiechert potentials. The fields in the charge's rest frame are

$$\vec{E}(\vec{r}, t) = \frac{1}{4\pi\epsilon_o} \frac{q}{r^3} \vec{r} \qquad \vec{B}(\vec{r}, t) = 0 \qquad (11.102)$$

We take $\vec{v} = v\,\hat{x}$. Applying the Lorentz transformation to the fields, we obtain

$$\widetilde{E}_x(r^\mu) = \widetilde{E}_{||} = E_{||} = E_x = \frac{1}{4\pi\epsilon_o} \frac{q}{r^3} x \qquad (11.103)$$

$$\widetilde{B}_x(r^\mu) = \widetilde{B}_{||} = B_{||} = B_x = 0 \qquad (11.104)$$

$$\widetilde{\vec{E}}_{yz}(r^\mu) = \widetilde{\vec{E}}_\perp = \gamma\left[\vec{E}_\perp - \vec{\beta} \times c\,\vec{B}_\perp\right] = \frac{1}{4\pi\epsilon_o} \frac{\gamma\,q}{r^3}\,(y\,\hat{y} + z\,\hat{z}) \qquad (11.105)$$

$$\widetilde{\vec{B}}_{yz}(r^\mu) = \widetilde{\vec{B}}_\perp = \gamma\left[\vec{B}_\perp + \vec{\beta} \times \frac{\vec{E}_\perp}{c}\right] = \vec{\beta} \times \frac{\widetilde{\vec{E}}}{c} \qquad (11.106)$$

These fields are still in terms of the F-frame coordinates, so they are not very useful as is.

As with the Lienard-Wiechert potential example, we use the Lorentz transformation of the space-time vector to rewrite r^{μ} in terms of $\widetilde{r}^{\mu}$,

$$c\,t = \gamma\left[c\,\widetilde{t} - \beta\,\widetilde{x}\right] \qquad\qquad x = \gamma\left[-\beta\,c\,\widetilde{t} + \widetilde{x}\right] \qquad (11.107)$$

and $y = \widetilde{y}$ and $z = \widetilde{z}$. Therefore,

$$\widetilde{E}_x = \frac{1}{4\,\pi\,\epsilon_o}\,\frac{\gamma\,q}{\left[\gamma^2\left(\widetilde{x} - v\,\widetilde{t}\right)^2 + \widetilde{y}^2 + \widetilde{z}^2\right]^{3/2}}\,\left(\widetilde{x} - v\,\widetilde{t}\right) \qquad (11.108)$$

$$\widetilde{\vec{E}}_{yz} = \frac{1}{4\,\pi\,\epsilon_o}\,\frac{\gamma\,q}{\left[\gamma^2\left(\widetilde{x} - v\,\widetilde{t}\right)^2 + \widetilde{y}^2 + \widetilde{z}^2\right]^{3/2}}\,\left(\widetilde{y}\,\hat{y} + \widetilde{z}\,\hat{z}\right) \qquad (11.109)$$

which we may write as

$$\widetilde{\vec{E}}(\widetilde{r}^{\mu}) = \frac{1}{4\,\pi\,\epsilon_o}\,\frac{\gamma\,q}{\left[\gamma^2\left(\widetilde{x} - v\,\widetilde{t}\right)^2 + \widetilde{y}^2 + \widetilde{z}^2\right]^{3/2}}\,\vec{R}(\widetilde{r}^{\mu}) \qquad (11.110)$$

$$= \frac{1}{4\,\pi\,\epsilon_o}\,\frac{q}{\gamma^2\left[\left\{\left(\widetilde{x} - v\,\widetilde{t}\right)^2 + \widetilde{y}^2 + \widetilde{z}^2\right\} - \beta^2\left(\widetilde{y}^2 + \widetilde{z}^2\right)\right]^{3/2}}\,\vec{R}(\widetilde{r}^{\mu}) \qquad (11.111)$$

where $\vec{R} = \widetilde{\vec{r}} - \vec{v}\,\widetilde{t}$.

Just as we did for the Lienard-Wiechert potentials, we may generalize this result in terms of current position:

$$\tilde{\vec{E}} = \frac{1}{4\,\pi\,\epsilon_o}\, \frac{q\,\vec{R}(\widetilde{r}^\mu)}{[R(\widetilde{r}^\mu)]^3}\, \frac{1 - \beta^2}{\left[1 - \beta^2\,\sin^2\theta\right]^{3/2}} \tag{11.112}$$

The expression for the magnetic field may be summarized as

$$\tilde{\vec{B}}(\widetilde{r}^\mu) = \vec{\beta} \times \frac{\tilde{\vec{E}}(\widetilde{r}^\mu)}{c} \tag{11.113}$$

We see the phenomenology is different from that for potentials (Equations 11.78 and 11.79): along and behind the direction of motion, the field is weakened by a factor γ^{-2} while, transverse to the direction of motion, the field is enhanced by a factor γ.

Electromagnetic Field Tensor Invariant

We can construct an invariant quantity by contracting $\mathcal{F}$ over both indices, $F_{\mu\nu}\,F^{\mu\nu}$. To see what value this takes on, let's first calculate $F_{\mu\nu}$:

$$F^{\mu\nu} = \begin{bmatrix} 0 & -E_x/c & -E_y/c & -E_z/c \\ E_x/c & 0 & -B_z & B_y \\ E_y/c & B_z & 0 & -B_x \\ E_z/c & -B_y & B_x & 0 \end{bmatrix} \tag{11.114}$$

$$\implies \quad F_{\mu\nu} = \begin{bmatrix} 0 & E_x/c & E_y/c & E_z/c \\ -E_x/c & 0 & -B_z & B_y \\ -E_y/c & B_z & 0 & -B_x \\ -E_z/c & -B_y & B_x & 0 \end{bmatrix} \tag{11.115}$$

Then, $F_{\mu\nu}\,F^{\mu\nu}$ consists of multiplying the two matrices not with standard matrix multiplication but rather element by element and then summing over all elements. That yields the Lorentz invariant:

$$F_{\mu\nu}\,F^{\mu\nu} = -\frac{2}{c^2}\,\vec{E}\cdot\vec{E} + 2\,\vec{B}\cdot\vec{B} \quad \implies \quad \boxed{-\frac{c^2}{2}\,F_{\mu\nu}\,F^{\mu\nu} = E^2 - c^2\,B^2} \tag{11.116}$$

This implies that the field strengths scale together from frame to frame: if one increases or decreases, so does the other. Finding this invariant without first having the Faraday tensor would have been quite difficult!

The Dual Electromagnetic Field Tensor and a Second Field Tensor Invariant

We may define the *dual tensor* $\mathcal{G}$ using the completely antisymmetric four-index Levi-Civita symbol $\epsilon^{\mu\nu\lambda\sigma}$:

$$G^{\mu\nu} = \frac{1}{2}\,\epsilon^{\mu\nu\lambda\sigma}\,F_{\lambda\sigma} \tag{11.117}$$

which yields

$$G^{00} = 0 \qquad\qquad\qquad\qquad G^{jj} = 0 \tag{11.118}$$

$$-G^{0j} = G^{j0} = \frac{1}{2}\,\epsilon^{0j}{}_{\lambda\sigma}\,F^{\lambda\sigma} = \frac{1}{2}\,\epsilon^{0j}{}_{ik}\,F^{ik} \qquad G^{ij} = \frac{1}{2}\,\epsilon^{ij}{}_{\lambda\sigma}\,F^{\lambda\sigma} \tag{11.119}$$

$$= -\frac{1}{2}\,\epsilon_{jik}\,\epsilon_{ik\ell}\,B_\ell \qquad\qquad = \frac{1}{2}\left(\epsilon^{ij}{}_{0k}\,F^{0k} + \epsilon^{ij}{}_{k0}\,F^{k0}\right)$$

$$= \frac{1}{2}\,\epsilon_{jik}\,\epsilon_{\ell ik}\,B_\ell \qquad\qquad\qquad = \epsilon^{ij}{}_{k0}\,F^{k0}$$

$$= \delta_{j\ell}\,B_\ell = B_j \qquad\qquad\qquad\quad = \frac{1}{c}\,\epsilon_{ijk}\,E_k$$

where $\epsilon^{0j}{}_{\lambda\sigma} = \epsilon^{0j}{}_{ik}$ on the left side because $\epsilon^{\mu\nu\lambda\sigma}$ vanishes if any of its indices repeat, so the j, λ, and σ indices must all take on space values 1, 2, 3 and thus all three must be latin indices. Similarly, on the right side, because its first two indices are latin indices, $\epsilon^{ij}{}_{\lambda\sigma}$ is only nonzero if one of λ or σ is zero, which then requires the other to become a latin index so it takes on space values.

We may write $\mathcal{G}$ out component by component:

$$
G^{\mu\nu} = \begin{bmatrix}
0 & -B_x & -B_y & -B_z \\
B_x & 0 & E_z/c & -E_y/c \\
B_y & -E_z/c & 0 & E_x/c \\
B_z & E_y/c & -E_x/c & 0
\end{bmatrix}
\tag{11.120}
$$

which is obtained from $F^{\mu\nu}$ by the replacement $\vec{E}/c \to \vec{B}$ and $\vec{B} \to -\vec{E}/c$.

$\mathcal{G}$ gives us the opportunity to form another invariant. While $G_{\mu\nu}\, G^{\mu\nu} = F_{\mu\nu}\, F^{\mu\nu}$ because the resulting contraction of the Levi-Civita symbol with itself yields an identity operator (or, just make the $\vec{E}$ and $\vec{B}$ replacements in $F_{\mu\nu}\, F^{\mu\nu} = E^2 - c^2\, B^2$ and the result is the same up to a sign), a new invariant is obtained by

$$
F_{\mu\nu}\, G^{\mu\nu} = -\frac{4}{c}\, \vec{E} \cdot \vec{B} \qquad \implies \qquad \boxed{-\frac{c}{4}\, F_{\mu\nu}\, G^{\mu\nu} = \vec{E} \cdot \vec{B}}
\tag{11.121}
$$

This invariant implies two things: 1) if the fields are perpendicular in one frame (or one vanishes), then they remain perpendicular (or one vanishes) in any other frame; 2) if the above invariant is nonzero in some frame, then, since the field strengths increase or decrease together, the magnitude of the cosine of the angle between them changes in the opposite fashion, so the angle between them changes in the same fashion as the fields; e.g., if the field strengths increase, they must become more perpendicular to each other.

Maxwell's Equations

Finally, we can rewrite Maxwell's Equations in a very simple form using the electromagnetic field tensor and its dual. The inhomogeneous equations (the ones with source terms on the right side, Gauss's Law and Ampere's Law) are:

$$\partial_\mu F^{\mu\nu} = \partial_\mu \partial^\mu A^\nu - \partial_\mu \partial^\nu A^\mu = \Box^2 A^\nu - \partial^\nu \partial_\mu A^\mu = \mu_o J^\nu - 0 \qquad (11.122)$$

$$\boxed{\partial_\mu F^{\mu\nu} = \mu_o J^\nu} \qquad (11.123)$$

where the second term vanishes due to the Lorenz gauge condition (Equation 11.71). The homogenous equations are

$$\partial_\mu G^{\mu\nu} = \frac{1}{2} \partial_\mu \epsilon^{\mu\nu\lambda\sigma} (\partial_\lambda A_\sigma - \partial_\sigma A_\lambda) = \partial_\mu \epsilon^{\mu\nu\lambda\sigma} \partial_\lambda A_\sigma = 0 \qquad (11.124)$$

$$\boxed{\partial_\mu G^{\mu\nu} = 0} \qquad (11.125)$$

where the second step is possible because $\epsilon^{\mu\nu\lambda\sigma}$ is antisymmetric under exchange of λ and σ (doing this exchange in the second term yields a copy of the first term) and the third step holds because $\partial_\mu \partial_\lambda$ is symmetric under exchange of these indices. It is interesting that physics statements — Faraday's law and the divergencelessness of $\vec{B}$ — reduce to a mathematical identity thanks to the definitions of $\mathcal{F}$ and $\mathcal{G}$. This is something we have seen before: $\vec{\nabla} \cdot \vec{B} = 0$ can be viewed as a consequence of $\vec{B} = \vec{\nabla} \times \vec{A}$. It is a sign that we have defined $\mathcal{G}$ well!

It is straightforward to see that these equations yield the standard Maxwell Equations:

$$\partial_\mu F^{\mu 0} = \mu_o J^0 \qquad\qquad \partial_\mu F^{\mu i} = \mu_o J^i \qquad (11.126)$$

$$\partial_0 F^{00} + \partial_i F^{i0} = \mu_o \rho c \qquad\qquad \frac{1}{c}\partial_t F^{0i} + \partial_j F^{ji} = \mu_o J^i \qquad (11.127)$$

$$0 + \frac{\partial}{\partial r^i}\frac{E_i}{c} = \frac{1}{c\,\epsilon_o}\rho \qquad -\frac{1}{c^2}\frac{\partial E^i}{\partial t} - \epsilon_{jik}\frac{\partial B_k}{\partial r^j} = \mu_o J^i \qquad (11.128)$$

$$\vec{\nabla}\cdot\vec{E} = \frac{\rho}{\epsilon_o} \qquad\qquad \vec{\nabla}\times\vec{B} = \mu_o \vec{J} + \epsilon_o\,\mu_o\frac{\partial\vec{E}}{\partial t} \qquad (11.129)$$

Since the structure of $\mathcal{G}$ parallels that of $\mathcal{F}$ with the replacement $E^i/c \to B^i$ and $B^i \to -E^i/c$, we may reuse the above arithmetic to see that:

$$\partial_\mu G^{\mu 0} = 0 \qquad\qquad \partial_\mu G^{\mu i} = 0 \qquad (11.130)$$

$$\vec{\nabla}\cdot\vec{B} = 0 \qquad\qquad -\frac{1}{c}\vec{\nabla}\times\vec{E} = c\,\epsilon_o\,\mu_o\frac{\partial\vec{B}}{\partial t} \qquad (11.131)$$

$$\vec{\nabla}\times\vec{E} = -\frac{\partial\vec{B}}{\partial t} \qquad (11.132)$$

Relativistic Dynamics with Electromagnetic Fields

Lorentz Force and Four-Momentum

We can also rewrite the Lorentz Force in a relativistically covariant way. As with the definition of the covariant source density, we use the rest-frame relation and the Lorentz transformation to figure out what the correct form is. In the rest frame of a particle of charge q, we know the nonrelativistic force is

$$\frac{d}{dt}\left(m\frac{dx^i}{dt}\right) = q\,E^i = q\,c\,F^{i0} \tag{11.133}$$

where $m\,dx^i/dt$ is the nonrelativistic momentum (we don't use v^i or p^i to avoid confusion with what will follow). How do we formulate a relativistically covariant version? The quantity being differentiated on the left side is $m\,dx^i/dt$; its natural generalization is, as you know, the covariant momentum $p^\mu = m\,v^\mu$. (This is analogous to the covariant current density, multiplying a scalar with the four-momentum.) On the right side, c is the rest-frame time component of v^μ, so $c\,F^{i0}$ looks like the time component of $v_\nu\,F^{i\nu}$. We are still stuck with a d/dt on the left side, but that becomes relativistically invariant if we replace it with $d/d\tau$ where τ is the proper time of the particle. Thus, we are motivated to write

$$\boxed{\frac{dp^\mu}{d\tau} = q\,F^{\mu\nu}\,v_\nu} \tag{11.134}$$

This expression is now relativistically covariant because both sides of the equation are rank 1 tensors (four-vectors).

Of course, we need to check that this generalization yields the correct Lorentz Force in an arbitrary frame, so let's evaluate it in a frame in which the charged particle is moving. We have for the space components (recall $v_\mu = (\gamma\,c, -\gamma\,\vec{v})$ (-1 on space components because of the lowered index), and writing $\vec{F}\cdot\hat{r}_i$, $\vec{v}\cdot\hat{r}_i$, and $\vec{p}\cdot\hat{r}_i$ to avoid sign ambiguities for space components (note $\vec{F}$ is force while $F^{\mu\nu}$ is the field tensor!):

$$\frac{dp^i}{d\tau} = q\left(F^{i0}\,\gamma\,c - F^{ij}\,\gamma\,\vec{v}\cdot\hat{r}_j\right) \tag{11.135}$$

$$\frac{dt}{d\tau}\frac{dp^i}{dt} = q\left(\frac{E_i}{c}\,\gamma\,c + \epsilon_{ijk}\,B_k\,\gamma\,\vec{v}\cdot\hat{r}_j\right) \tag{11.136}$$

$$\gamma\left(\vec{F}\cdot\hat{r}_i\right) = \gamma\,q\left(E_i + \left(\vec{v}\times\vec{B}\right)_i\right) \tag{11.137}$$

$$\implies \qquad \vec{F} = q\left(\vec{E} + \vec{v}\times\vec{B}\right) \tag{11.138}$$

where $dt/d\tau = \gamma$ (time dilation) and we note the subtlety that, in relativity, the natural generalization of Newton's Second Law is to replace $\vec{F}\cdot\hat{r}_i = (d/dt)(\vec{p}\cdot\hat{r}_i)$ by

$$\vec{F}\cdot\hat{r}_i = \frac{dp^i}{dt} = \frac{d}{dt}\left(\gamma\,m\,\vec{v}\cdot\hat{r}_i\right) \tag{11.139}$$

This generalization ensures the force and momentum can get as large as one wants without $|\vec{v}|$ exceeding c. We note that Griffiths defines $K^\mu = dp^\mu/d\tau$ as the *Minkowski force* because it is a four-vector (transforms by the Lorentz transformation rule); dp^μ/dt is not.

Let's look at the time component that comes along for free in the above. It is

$$\frac{d}{d\tau} \gamma \, m \, c = q \left(F^{00} \, \gamma \, c - F^{0i} \, \gamma \, \vec{v} \cdot \widehat{r}_i \right) \tag{11.140}$$

$$\frac{dt}{d\tau} \frac{d}{dt} \gamma \, m \, c = q \left[(0)(c) - \left(-\frac{E_i}{c} \right) \gamma \, \vec{v} \cdot \widehat{r}_i \right] \tag{11.141}$$

$$\gamma \frac{d}{dt} \gamma \, m \, c = \gamma \, q \, \frac{E_i}{c} \, \vec{v} \cdot \widehat{r}_i \tag{11.142}$$

$$\frac{dU}{dt} = q \, \vec{v} \cdot \vec{E} \tag{11.143}$$

where we have been motivated to define $U = \gamma \, m \, c^2$ as the relativistic generalization of the energy of the particle. This is just conservation of energy: the rate of change of the particle energy is given by the work being done by the electric field (the magnetic field does no work).

We may now see the full meaning of the four-momentum we defined earlier:

$$\overset{\Rightarrow}{p} = m \, \overset{\Rightarrow}{v} \qquad \Longleftrightarrow \qquad p^\mu = m \, v^\mu = \gamma \, m \, (c, \vec{v}) = \left(\frac{U}{c}, \gamma \, \vec{p} \right) \tag{11.144}$$

which is entirely consistent with what we have derived above. It may seem strange that the interpretation of the time component of $\overset{\Rightarrow}{p}$ as the particle energy comes as an afterthought rather than being something we design in, but this makes sense given our interest in covariant quantities: we could not necessarily know ahead of time what quantity should be combined with $\vec{p}$ to make a four-vector. Formulating a relativistically covariant generalization of the Lorentz Force made the meaning of the time component of $\overset{\Rightarrow}{p}$ self-evident.

Lagrangian Formulation of Relativistic Electrodynamics (Optional)

We will show how we can recover all of the above by beginning with a Lagrangian formulation the interaction of the EM field with particles and for the EM field itself.

Lagrangian Formulation for Particle-Field Interaction

In Ph106a, you learned about the appropriate Lagrangian to use for a free particle as well as for a particle interacting with an electromagnetic field, so we review that here. Recall that the Lagrangian for a free particle is

$$L = -\frac{m\,c^2}{\gamma} = -m\,c^2 \sqrt{1 - \frac{|\vec{v}|^2}{c^2}} \tag{11.145}$$

where $|\vec{v}|^2$ is the modulus of the three-velocity. This quantity makes sense because it gives a Lorentz-invariant action:

$$S = \int_{t_1}^{t_2} dt\, L = \int_{t_1}^{t_2} d\tau\, \gamma\, L = -m\,c^2 \int_{t_1}^{t_2} d\tau \tag{11.146}$$

where τ is the proper time, not the volume element! We can check that this Lagrangian recovers the nonrelativistic free particle limit:

$$L \xrightarrow{\beta \ll 1} -m\,c^2 \left(1 - \frac{1}{2}\frac{|\vec{v}|^2}{c^2}\right) = -m\,c^2 + \frac{1}{2}\,m\,|\vec{v}|^2 \tag{11.147}$$

which, aside from an irrelevant constant, is what we expect. Note the importance of the negative sign in the definition of the full relativistic Lagrangian.

We need to consider similarly Lorentz-invariant ways to introduce the interaction with the EM field. In our earlier discussion of the potential form for the Lorentz Force Law, we showed via the construction of a velocity-dependent potential that the appropriate nonrelativistic Lagrangian for this interaction is

$$L_{int} = -q \left(V - \vec{v} \cdot \vec{A} \right) \tag{11.148}$$

This is already a Lorentz-invariant quantity when written in terms of the covariant potential and the four-velocity:

$$L_{int} = -q \, v^\mu A_\mu = -q \left[c \, \frac{V}{c} - \vec{v} \cdot \vec{A} \right] = -q \left(V - \vec{v} \cdot \vec{A} \right) \tag{11.149}$$

We can confirm that it yields a sensible canonical momentum (again using $\vec{v} \cdot \hat{r}_i$ for the components of the three-velocity to avoid confusion with the space components of the four-velocity):

$$\pi^i = \frac{\partial L}{\partial \left(\vec{v} \cdot \hat{r}_i \right)} = -m \, c^2 \frac{1}{2} \, \gamma \left(-\frac{2 \left(\vec{v} \cdot \hat{r}_i \right)}{c^2} \right) - q \left(-A^i \right) = \gamma \, m \left(\vec{v} \cdot \hat{r}_i \right) + q \, A^i \tag{11.150}$$

which is what we expect from the nonrelativistic version $\vec{\pi} = m \vec{v} + q \vec{A}$ along with the relativistic replacement $m \vec{v} \rightarrow \gamma \, m \, \vec{v}$.

The Euler-Lagrange Equations are

$$0 = \frac{d}{dt}\frac{\partial L}{\partial \left(\vec{v} \cdot \hat{r}_i \right)} - \frac{\partial L}{\partial x^i} \tag{11.151}$$

$$= \frac{d}{dt}\left(\gamma\, m\, (\vec{v} \cdot \hat{r}_i) + q\, A^i \right) + q\frac{\partial}{\partial x^i}\left(V - \vec{v} \cdot \vec{A} \right) \tag{11.152}$$

$$= \frac{d}{dt}\left(\gamma\, m\, (\vec{v} \cdot \hat{r}_i) \right) + q\left[\frac{\partial A^i}{\partial t} + \left(\vec{v} \cdot \vec{\nabla} \right) A^i \right] + q\frac{\partial V}{\partial x^i} - q\frac{\partial}{\partial x^i}\left(\vec{v} \cdot \vec{A} \right) \tag{11.153}$$

$$= \frac{d}{dt}\left(\gamma\, m\, (\vec{v} \cdot \hat{r}_i) \right) + q\left[\frac{\partial V}{\partial x^i} + \frac{\partial A^i}{\partial t} \right] - q\left[\vec{v} \times \left(\vec{\nabla} \times \vec{A} \right) \right] \cdot \hat{r}_i \tag{11.154}$$

where we used the convective derivative (Equation 10.24) to expand dA^i/dt and then we used the $BAC - CAB$ rule to reconstruct the triple cross-product in the last equation. Separating the time derivative of the momentum from the other terms, we find

$$\frac{d}{dt}\left(\gamma\, m\, (\vec{v} \cdot \hat{r}_i) \right) = q\left[-\frac{\partial V}{\partial x^i} - \frac{\partial A^i}{\partial t} \right] + q\left[\vec{v} \times \left(\vec{\nabla} \times \vec{A} \right) \right] \cdot \hat{r}_i \tag{11.155}$$

$$= q\left[\vec{E} + \vec{v} \times \vec{B} \right] \cdot \hat{r}_i \tag{11.156}$$

which is exactly the relativistic version of the Lorentz Force Law we derived earlier (Equations 11.138 and 11.139).

It is somewhat unsatisfying that we do not trivially recover the relativistically covariant version of the Lorentz Force Law, Equation 11.134, but that is because the standard Euler-Lagrange Equations are inherently non-covariant — they treat time and space coordinates differently (though the Lagrangian itself can be composed of Lorentz covariant quantities). We will deal with this in the next section.

Did we have other options for L_{int}? The only quantities we have for the EM field from which to construct a Lagrangian are the covariant potential A^μ, the field strength tensor $F^{\mu\nu}$, and its dual $G^{\mu\nu}$, and the only quantity from the particle is the four velocity v^μ. Had we tried to construct L_{int} using either of the field strength tensors instead, we would have obtained a Lagrangian quadratic in v^μ since these tensors have two indices to contract over. If one starts to work out the resulting equations of motion, one finds that they just aren't sensible and cannot yield the nonrelativistic limit. So we are left to use only A^μ and v^μ, yielding the form we already checked above.

Lagrangian Density for the EM Field and Recovery of Maxwell's Equations

We previously introduced the concept of Lagrangian densities in the context of deriving the Schrödinger Equation by treating the quantum mechanical wavefunction ψ as a classical field and extending the Lagrangian formalism to such fields, treating the value of the wavefunction at each point in space and an infinite set of degrees of freedom evolving in time, $\psi(\vec{r}, t)$. We will do the same here for the covariant potential and the EM field tensor.

The obvious question is: what do we write down for the Lagrangian density for the EM field? A good starting principle is that we want the Lagrangian density to be a Lorentz scalar. We know of two Lorentz scalars we can build with the EM field tensor, $F_{\mu\nu}F^{\mu\nu}$ and $F_{\mu\nu}G^{\mu\nu}$. We did note it before, but the latter version is undesirable because $G^{\mu\nu}$ is a *pseudotensor*, meaning that it does not change sign under coordinate inversion. The Lagrangian density itself should be invariant under coordinate inversion, so the first form is but the latter form is not. So we choose the first form.

In order to get Maxwell's Equations with sources, we clearly need to include an interaction term. The generalization of the interaction Lagrangian for a single particle with the EM field, $-q\,v^\mu\,A_\mu$ is $-J^\mu A_\mu$.

With foreknowledge of the appropriate coefficients needed to obtain the correct Maxwell's Equations, we therefore propose for the Lagrangian density

$$\mathcal{L} = -\frac{1}{4\,\mu_o}\,F_{\mu\nu}F^{\mu\nu} - J_\mu A^\mu$$

Next, we must obtain the Euler-Lagrange equations for this Lagrangian. We have to choose the degrees of freedom. While it might seem like they should be the components of $F^{\mu\nu}$, we know these components are highly redundant. We have the mathematical constraint that arises via construction that $\partial_\mu G^{\mu\nu} = 0$ (Equation 11.123). This, along with the presence of A^μ in the interaction Lagrangian density, suggests that we should treat A^μ as our degrees of freedom.

The second question is: what are the relativistically covariant Euler-Lagrange Equations? It turns out the correspondence is (see, *e.g.*, Goldstein, Poole, and Safko)

$$\frac{d}{dt}\left(\frac{\partial L}{\partial \dot{q}_k}\right) - \frac{\partial L}{\partial q_k} = 0 \qquad \longrightarrow \qquad \partial^\mu\left(\frac{\partial \mathcal{L}}{\partial\left(\partial^\mu A^\nu\right)}\right) - \frac{\partial \mathcal{L}}{\partial A^\nu} = 0$$

Given the above two points, it makes sense to rewrite the Lagragian density in a manner that makes the dependence on A^ν explicit and also makes all appearances of ∂^μ and A^ν consistent with the version in the Euler-Lagrange Equations:

$$\mathcal{L} = -\frac{1}{4\,\mu_o}\,g_{\alpha\beta}g_{\gamma\delta}\left(\partial^\beta A^\delta - \partial^\delta A^\beta\right)\left(\partial^\alpha A^\gamma - \partial^\gamma A^\alpha\right) - J_\mu A^\mu$$

Let's calculate the required derivatives:

$$\frac{\partial \mathcal{L}}{\partial \left(\partial^\mu A^\nu\right)} = -\frac{1}{4\,\mu_o}\, g_{\alpha\beta} g_{\gamma\sigma} \left[\delta_\mu^{\ \beta} \delta_\nu^{\ \sigma} F^{\alpha\gamma} - \delta_\mu^{\ \sigma} \delta_\nu^{\ \beta} F^{\alpha\gamma} + \delta_\mu^{\ \alpha} \delta_\nu^{\ \gamma} F^{\beta\sigma} - \delta_\mu^{\ \alpha} \delta_\nu^{\ \gamma} F^{\beta\sigma}\right]$$

$$= -\frac{1}{\mu_o} F_{\mu\nu} \quad \text{by symmetry of } g_{\mu\nu} \text{ and antisymmetry of } F_{\mu\nu}$$

$$\frac{\partial \mathcal{L}}{\partial A^\nu} = -J_\mu$$

Thus, we have

$$\partial^\mu \left(-\frac{1}{\mu_o} F_{\mu\nu}\right) + J_\mu = 0 \qquad \Longrightarrow \qquad \partial_\mu F^{\mu\nu} = \mu_o J^\mu$$

which recovers Equation 11.123! We will not be able to recover Equation 11.125 because it is not dynamical but rather it is a mathematical identity given the construction of $F^{\mu\nu}$ from A^μ.

Relativistic Conservation Laws (Optional)

Maxwell Energy-Momentum Tensor and Conservation of Energy-Momentum

It is natural to generalize the Maxwell Stress Tensor that we defined some time ago.
We define the *Maxwell Energy-Momentum Tensor* $\mathcal{T}$ as

$$T^{\mu\nu} = \frac{1}{\mu_o}\left[F^{\mu}_{\ \lambda}\,F^{\lambda\nu} + \frac{1}{4}\,g^{\mu\nu}\,F_{\lambda\sigma}\,F^{\lambda\sigma}\right] \tag{11.157}$$

The components of $\mathcal{T}$ are (recall, we found the relativistic invariant $F_{\mu\nu}\,F^{\mu\nu}$ before):

$$T^{00} = \frac{1}{\mu_o}\left[F^0_{\ 0}\,F^{00} + g_{ij}\,F^{0j}\,F^{i0} + \frac{1}{4}\,g^{00}\left(-\frac{2}{c^2}\right)\left(E^2 - c^2\,B^2\right)\right] \tag{11.158}$$

$$= \frac{1}{\mu_o}\left[(-1)\left(-\frac{E^2}{c^2}\right) - \frac{1}{2}\left(\frac{E^2}{c^2} - B^2\right)\right] = \frac{\epsilon_o}{2}\left(E^2 + c^2\,B^2\right) = u_{field} \tag{11.159}$$

where u_{field} is the EM field energy density.

The space-space components are

$$T^{ij} = \frac{1}{\mu_o} \left[F^{i\lambda} g_{\lambda\sigma} F^{\sigma j} - \frac{g^{ij}}{2} \left(\frac{E^2}{c^2} - B^2 \right) \right] \tag{11.160}$$

$$= \frac{1}{\mu_o} \left[g_{00} F^{i0} F^{0j} + g_{k\ell} \left(F^{ik} F^{\ell j} \right) + \frac{\delta^{ij}}{2} \left(\frac{E^2}{c^2} - B^2 \right) \right] \tag{11.161}$$

$$= \frac{1}{\mu_o} \left[F^{i0} F^{0j} - \left(F^{ik} F^{kj} \right) + \frac{\delta^{ij}}{2} \left(\frac{E^2}{c^2} - B^2 \right) \right] \tag{11.162}$$

$$= \frac{1}{\mu_o} \left[-\frac{1}{c^2} E_i E_j - \epsilon_{ik\ell} B_\ell \, \epsilon_{kjm} B_m + \frac{\delta^{ij}}{2} \left(\frac{E^2}{c^2} - B^2 \right) \right] \tag{11.163}$$

$$= \epsilon_o \left[-E_i E_j + c^2 \, \epsilon_{i\ell k} B_\ell \, \epsilon_{jmk} B_m + \frac{\delta^{ij}}{2} \left(E^2 - c^2 B^2 \right) \right] \tag{11.164}$$

$$= \epsilon_o \left[-E_i E_j + c^2 \left(\delta_{ij} \delta_{\ell m} - \delta_{im} \delta_{\ell j} \right) B_\ell \, B_m + \frac{\delta^{ij}}{2} \left(E^2 - c^2 B^2 \right) \right] \tag{11.165}$$

$$= \epsilon_o \left[-E_i E_j - c^2 B_i \, B_j + \frac{\delta^{ij}}{2} \left(E^2 + c^2 B^2 \right) \right] = - \left(\underline{\underline{T}} \right)_{ij} \tag{11.166}$$

i.e., the *negative* of the Maxwell Stress Tensor we defined in Equation 8.31. (To be clear, $\mathcal{T}$ is the four-tensor while $\underline{\underline{T}}$ is the three-tensor.) Note that we allow space-space (roman) indices to be contracted without requiring one raised and one lowered index.

The space-time components are

$$T^{0i} = \frac{1}{\mu_o} F^{0\lambda} g_{\lambda\sigma} F^{\sigma i} = \frac{1}{\mu_o} F^{0j} g_{jk} F^{ki} = -\frac{1}{\mu_o} F^{0j} \delta_{jk} F^{ki} = -\frac{1}{\mu_o} F^{0j} F^{ji} \quad (11.167)$$

where we used $F^{00} = 0$ in the first step to reduce λ to j, the diagonality of g in the second step to reduce the sum over σ to k, and finally the fact that $g_{jk} = -\delta_{jk}$ in the third step. Then, we have

$$T^{0i} = -\frac{1}{\mu_o} \left(-\frac{E_j}{c}\right) (-\epsilon_{jik} B_k) = \frac{1}{\mu_o} \frac{1}{c} \epsilon_{ijk} E_j B_k = \frac{1}{c} S_i = c \, p_{field,i} \quad (11.168)$$

where $\vec{S} = \vec{E} \times \vec{B}/\mu_o$ is the Poynting vector. Summarizing,

$$T^{\mu\nu} = \begin{bmatrix} u_{field} & c\,\vec{g} \\ c\,\vec{g} & -\underline{\underline{\mathcal{T}}} \end{bmatrix} \quad (11.169)$$

With the energy-momentum tensor in hand, it is natural to rewrite our energy and linear momentum conservation laws using it. If we take the four-divergence of the energy-momentum tensor, we find

$$\partial_\mu T^{\mu\nu} = \frac{1}{\mu_o}\left[(\partial_\mu F^\mu_{\ \lambda})F^{\lambda\nu} + F^\mu_{\ \lambda}\partial_\mu F^{\lambda\nu} + \frac{1}{4}\partial^\nu\left(F_{\lambda\sigma}F^{\lambda\sigma}\right)\right] \qquad (11.170)$$

We use the inhomogeneous Maxwell Equation, $\partial_\mu F^{\mu\nu} = \mu_o J^\nu$, to rewrite the first term and move it to the left side, and we also rewrite the right side (splitting the first term into two copies and applying the product rule to the second term):

$$\partial_\mu T^{\mu\nu} - J_\lambda F^{\lambda\nu} = \frac{1}{2\mu_o}F_{\mu\lambda}\left[\partial^\mu F^{\lambda\nu} + \partial^\mu F^{\lambda\nu} + \partial^\nu F^{\mu\lambda}\right] \qquad (11.171)$$

We may rewrite the last two terms using the homogeneous Maxwell Equation, which we rewrite as:

$$0 = \partial_\sigma G^{\sigma\lambda} = \frac{1}{2}\partial_\sigma \epsilon^{\sigma\lambda\nu\mu}F_{\nu\mu} = \partial^\lambda F^{\nu\mu} + \partial^\mu F^{\lambda\nu} + \partial^\nu F^{\mu\lambda} \qquad (11.172)$$

$$-\partial^\lambda F^{\nu\mu} = \partial^\mu F^{\lambda\nu} + \partial^\nu F^{\mu\lambda} \qquad (11.173)$$

The last step in the first line is obtained by just writing out all the terms and combining the ones that differ by a single flip of two indices; it may seem strange that four equations ($\lambda = 0$, 1, 2, 3) became 64 equations (λ, μ, $\nu = 0$, 1, 2, 3), but many of them vanish and the others are redundant (see how convenient the dual tensor is!).

Using this in the equation for the divergence of the energy-momentum tensor, and then reordering the indices on the second term on each side (and picking up minus signs), we have

$$\partial_\mu T^{\mu\nu} - J_\lambda F^{\lambda\nu} = \frac{1}{2\,\mu_o} F_{\mu\lambda} \left[\partial^\mu F^{\lambda\nu} - \partial^\lambda F^{\nu\mu}\right] \tag{11.174}$$

$$\partial_\mu T^{\mu\nu} + F^{\nu\lambda} J_\lambda = \frac{1}{2\,\mu_o} F_{\mu\lambda} \left[\partial^\mu F^{\lambda\nu} + \partial^\lambda F^{\mu\nu}\right] \tag{11.175}$$

The quantity in brackets on the right side is now symmetric in μ and λ, while $F_{\mu\lambda}$ is antisymmetric, so the right side vanishes. Moving the field-current term to the right side, we obtain

$$\partial_\mu T^{\mu\nu} = -F^{\nu\lambda} J_\lambda \tag{11.176}$$

If we write out the time and space components of this four-vector equation, we obtain

$$\frac{\partial u_{field}}{\partial t} + \vec{\nabla} \cdot \vec{S} = -\vec{J} \cdot \vec{E} = -\frac{\partial u_{mech}}{\partial t} \tag{11.177}$$

$$\frac{\partial \vec{g}}{\partial t} - \vec{\nabla} \cdot \underline{\underline{T}} = -\left[\rho\vec{E} + \vec{J} \times \vec{B}\right] = -\frac{\partial \vec{p}_{mech}}{\partial t} \tag{11.178}$$

which are Equations 8.15 and 8.40 from our discussion of conservation laws.

Angular Momentum Tensor

The relativistic angular momentum tensor $\mathcal{M}$ has coordinate representation

$$M^{\mu\nu\sigma} = -[T^{\mu\nu}r^\sigma - T^{\mu\sigma}r^\nu] = T^{\mu\sigma}r^\nu - T^{\mu\nu}r^\sigma \qquad (11.179)$$

which is the natural generalization of the three-dimensional angular momentum current density tensor we defined earlier, Equation 8.57; we will see its space-space-space components recover that definition (up to a sign, which is the same sign difference between the relativistic and nonrelativistic versions as we saw for the stress tensor). However, in three dimensions, we could write those as a cross-product. That is not possible in four dimensions. Instead, we use the analogous construction, the antisymmetrized product of $\overset{\Rightarrow}{r}$ and $\mathcal{T}$. This is called the wedge product and is written as

$$\mathcal{M} = -\mathcal{T} \wedge \overset{\Rightarrow}{r} \qquad (11.180)$$

where the $\wedge$ indicates the antisymmetric difference over the index of the space-time position four-vector and the last index of the stress tensor.

Let's work out what the components of this tensor are:

$$M^{000} = 0 \tag{11.181}$$

$$-M^{0i0} = M^{00i} = -\left[T^{00} r^i - T^{0i} r^0 \right] = -u\, r^i + \frac{S^i}{c}\, c\, t = -u\, r^i + c\, g^i\, c\, t \equiv -c^2 N^i \tag{11.182}$$

$$M^{0ij} = -\left[T^{0i} r^j - T^{0j} r^i \right] = -\frac{S^i}{c}\, r^j + \frac{S^j}{c}\, r^i = c\,(\vec{r} \wedge \vec{g})^{ij} \equiv c\, \ell_{field}^{ij} \tag{11.183}$$

$$M^{i00} = 0 \tag{11.184}$$

$$-M^{ij0} = M^{i0j} = -\left[T^{i0} r^j - T^{j0} r^i \right] = -\left[\frac{S^i}{c} r^j - \left(\underline{\underline{T}} \right)^{ij} c\, t \right] = \left(\underline{\underline{T}} \right)^{ij} c\, t - c\, g^i r^j \tag{11.185}$$

$$M^{ijk} = -\left[T^{ij} r^k - T^{ik} r^j \right] = \left(\underline{\underline{T}} \right)^{ij} r^k - \left(\underline{\underline{T}} \right)^{ik} r^j = \left(\underline{\underline{T}} \wedge \vec{r} \right)^{ijk} = \left(-\underline{\underline{\mathcal{M}}} \right)^{ijk} \tag{11.186}$$

where we have defined the relativistic energy three-moment $c^2 \vec{N}$, the three-dimensional angular momentum density tensor $\underline{\underline{\ell}}_{field} = \vec{r} \wedge \vec{g}$, and the three-dimensional torque tensor $\underline{\underline{\mathcal{M}}} = -\underline{\underline{T}} \wedge \vec{r}$. The latter two are the wedge-product generalizations of the three-dimensional versions that involved cross products (Equations 8.67 and 8.57). The interpretation of the space-space-time and space-time-space components will become clear later. We unfortunately cannot write the breakdown in a nice matrix form as we did for $\mathcal{T}$ because $\mathcal{M}$ has three indices.

Let's write down a conservation law for this tensor using $\partial_\mu T^{\mu\nu} = -F^{\nu\lambda} J_\lambda$ and $\partial_\mu r^\nu = \delta_\mu{}^\nu$:

$$\partial_\mu M^{\mu\nu\sigma} = -\partial_\mu T^{\mu\nu} r^\sigma + \partial_\mu T^{\mu\sigma} r^\nu \tag{11.187}$$

$$= \left[\left(F^{\nu\lambda} J_\lambda\right) r^\sigma - \left(F^{\sigma\lambda} J_\lambda\right) r^\nu\right] + \left[-T^{\mu\nu}\delta_\mu{}^\sigma + T^{\mu\sigma}\delta_\mu{}^\nu\right] \tag{11.188}$$

$$= \left[r^\sigma F^{\nu\lambda} - r^\nu F^{\sigma\lambda}\right] J_\lambda + \left[-T^{\sigma\nu} + T^{\nu\sigma}\right] \tag{11.189}$$

$$= \left(\vec{\Rightarrow}{r} \wedge \mathcal{F}\right)^{\sigma\nu\lambda} J_\lambda = \left(-\vec{\Rightarrow}{r} \wedge \mathcal{F}\right)^{\nu\sigma\lambda} J_\lambda \tag{11.190}$$

where, in the penultimate step, we used the symmetry of the stress tensor and, in the last step, we used the asymmetry of the wedge product (the $\nu\sigma$ indices are the wedge-product indices). Let's break this down piece-by-piece.

The time-time component ($\nu = 0$, $\sigma = 0$) is trivial because $M^{\mu00} = 0$ and the right hand side is also antisymmetric in these indices.

The space-time components ($\nu = 0$, $\sigma = i$ and vice versa) imply a relationship between center-of-mass motion and the total linear momentum. We will not explore that here.

Let's consider the space-space components $\mu = i$, $\nu = j$ next. The left-hand side is

$$\partial_\mu M^{\mu ij} = \partial_0 M^{0ij} + \partial_k M^{kij} = \frac{1}{c} \frac{\partial}{\partial t} \left(c\, \ell^{ij}_{field} \right) + \nabla_i \left(-\underline{\underline{\mathcal{M}}} \right)^{ijk} = \left(\frac{\partial}{\partial t} \underline{\underline{\ell}}_{field} - \vec{\nabla} \cdot \underline{\underline{\mathcal{M}}} \right)^{ij} \tag{11.191}$$

The right-hand side is

$$- \left[r^i F^{j\lambda} - r^j F^{i\lambda} \right] J_\lambda = -r^i \frac{E^j}{c} \rho c - r^i \left(-\epsilon_{jkm} B^m \right) \left(-J^k \right) - (i \leftrightarrow j) \tag{11.192}$$

$$= - \left(\vec{r} \wedge \rho \vec{E} \right)^{ij} - \left[r^i \left(\vec{J} \times \vec{B} \right)^j - (i \leftrightarrow j) \right] \tag{11.193}$$

$$= - \left(\vec{r} \wedge \rho \vec{E} \right)^{ij} - \left[\vec{r} \wedge \left(\vec{J} \times \vec{B} \right) \right]^{ij} = -\frac{\partial}{\partial t} \underline{\underline{\ell}}^{ij}_{mech} \tag{11.194}$$

where we have written down the differential version of Equation 8.51, replacing the cross products in that expression with wedge products, and we have also replaced the mechanical angular momentum density cross product $\vec{\ell}_{mech}$ with the wedge product version $\underline{\underline{\ell}}_{mech}$. Thus, we have

$$\frac{\partial}{\partial t} \underline{\underline{\ell}}_{field} - \vec{\nabla} \cdot \underline{\underline{\mathcal{M}}} = -\frac{\partial}{\partial t} \underline{\underline{\ell}}_{mech} \tag{11.195}$$

which is Equation 8.68 showing that the rate of change of the angular momentum density is given by the divergence of the torque tensor.

Section 12
Radiation

Lecture 34:

Radiation I:
Potentials, Fields, and Power Radiated by
an Accelerated Point Charge

Date Revised: 2024/05/23 09:45
Corrected errors in Equations 12.32 and 12.33 regarding
which variables are held fixed while taking partial derivatives
Date Given: 2024/05/21

Potentials, Fields, and Power Radiated by an Accelerated Point Charge

Introduction and Study Guide

Our practical goal is to calculate the power radiated by an accelerated point charge. To get there, we need to calculate the potentials of an accelerated point charge, then the fields, and from the fields the power.

We proceed in a very different manner than Griffiths. In §10.3.1, Griffiths uses the retarded potential formulae (Equation 10.71) to do a direct calculation of the Lienard-Wiechert potential, which is challenging. Then, in §10.3.2, Griffiths brute-force differentiates the potentials to get the fields. Instead, we have used relativity to obtain the potentials of a particle moving at fixed velocity — the Lienard-Wiechert potentials — and its fields in terms of the current position of the particle. We can rewrite the results in terms of the retarded time. Then, we return to the Lienard-Wiechert potentials and incorporate the effect of acceleration through possible time-dependence of the velocity by direct differentiation allowing the velocity to vary, which is much easier now that we have accounted for all other time dependences via the relativity-based derivation of the Lienard-Wiechert potentials and corresponding fields. Finally, from the fields we calculate the radiated power pattern for the general case, specializing to slowly moving particles at the end to obtain the Larmor Formula.

We calculate the power radiated by an accelerated charge before moving on to the general topic of dipole and multipole radiation. Griffiths proceeds in the opposite order for reasons that are not clear, first doing dipole/multipole radiation in Section 11.1 before doing the radiation of a moving point charge in Section 11.2.1.

Our treatment of the calculation of the fields follows, partially, that of M. Cross's Ph106c lecture notes, which do not appear to follow any specific textbook. The treatment of radiation follows that of Heald and Marion Sections 8.7 and 8.8.

We will not cover the topic of radiation reaction (Griffiths 11.2.2 and 11.2.3) — it's interesting, but there is not much to add to what Griffiths says.

Potentials and Fields of a Fixed-Velocity Point Charge in Terms of Current Position

Let's first rewrite the potentials and fields of a moving point charge that we calculated using Lorentz Transformations in a form that will be useful going forward. From Equations 11.78 and 11.79:

$$
\begin{array}{c}
\text{Lienard-Wiechert} \\
\text{scalar potential} \\
\text{using current position}
\end{array}
\quad
V(\vec{r}, t) = \frac{1}{4\pi\epsilon_o} \frac{q}{R(\vec{r}, t)} \frac{1}{\sqrt{1 - \beta^2 \sin^2\theta}}
\tag{12.1}
$$

$$
\begin{array}{c}
\text{Lienard-Wiechert} \\
\text{vector potential} \\
\text{using current position}
\end{array}
\quad
\vec{A}(\vec{r}, t) = \frac{\mu_o}{4\pi} \frac{q\,c\,\vec{\beta}}{R(\vec{r}, t)} \frac{1}{\sqrt{1 - \beta^2 \sin^2\theta}}
\tag{12.2}
$$

It is very important for us to note that, even though the above potentials were derived using Lorentz Transformations — *i.e.*, fixed velocity — they turn out to be the same as what one would get by working from the general forms for the retarded potentials (Equation 10.71). That derivation is done in Griffiths §10.3.1.

There is an intuitive reason for this matchup, which does not hold for the fields. The retarded potential expressions involve the position of the point charge. The process of doing the integral can involve Jacobians, which involves first derivatives and thus might involve the velocity. But there is no way for second derivatives to appear, and thus acceleration cannot be relevant. If acceleration does not matter, then the fixed-velocity derivation using Lorentz Transformations must be valid even in the accelerating case.

And from Equations 11.112 and 11.113:

$$
\begin{array}{c}
\text{Electric field} \\
\text{of moving point charge} \\
\text{using current position}
\end{array}
\qquad
\vec{E}(\vec{r}, t) = \frac{q}{4\pi\epsilon_o} \frac{\vec{R}}{R^3} \frac{1 - \beta^2}{\left[1 - \beta^2 \sin^2\theta\right]^{3/2}}
\tag{12.3}
$$

$$
\begin{array}{c}
\text{Magnetic field} \\
\text{of moving point charge} \\
\text{using current position}
\end{array}
\qquad
\vec{B}(\vec{r}, t) = \frac{\mu_o}{4\pi} \frac{q\,c\,\vec{\beta} \times \vec{R}}{R^3} \frac{1 - \beta^2}{\left[1 - \beta^2 \sin^2\theta\right]^{3/2}}
\tag{12.4}
$$

$$
\begin{array}{c}
\text{Relation between electric and} \\
\text{magnetic fields of moving point charge} \\
\text{using current position}
\end{array}
\qquad
\vec{B} = \frac{1}{c} \vec{\beta} \times \vec{E}
\tag{12.5}
$$

Comments and reminders:

▶ The forms suggest that the potential information at the *field point* (the point at which one wants to know the potentials and fields) acts as if the particle had kept moving at the velocity it had when the potential information left the particle. Effectively, the retarded potential carries information not just about the position of the particle at the retarded time but also about its velocity (in the relation $\vec{A} = V\,\vec{v}/c^2$ and in the angular dependence)! And, of course, it is not affected yet by later time events!

▶ The potentials and fields are forward-backward symmetric in magnitude. We shall see that, when we rewrite using retarded time, they are not!

▶ The potentials are enhanced in the transverse direction by a factor $\gamma = 1/\sqrt{1-\beta^2}$ and take on the static values in the forward and backward directions along the direction of motion.

▶ The fields have enhanced angular and β dependence. The field is enhanced by a factor $1/\sqrt{1-\beta^2}$ (the Lorentz factor of the particle, γ) in the plane transverse to the direction of motion (at $\theta = \pi/2$) and reduced by a factor $1-\beta^2$ along the axis of motion ($\theta = 0$ or $\theta = \pi$).

▶ The relation $\vec{B} \propto \vec{\beta} \times \vec{E}$ along with $\vec{E} \propto \vec{R}$ implies that $\vec{B}$ wraps around $\vec{\beta}$ following the right-hand rule.

Potentials and Fields of a Fixed-Velocity Point Charge in Terms of Retarded Position

When we calculate the fields of an accelerating charge, we find that it is more convenient (or, rather, less inconvenient) to start from expressions for the potentials and fields in terms of the *retarded position*, the position of the particle at the retarded time. We therefore need to relate the current position to the retarded position. We make use of the following figure, with $\vec{w}(t)$ being the particle trajecory and $\vec{\beta} = \vec{w}'/c$ its constant velocity:
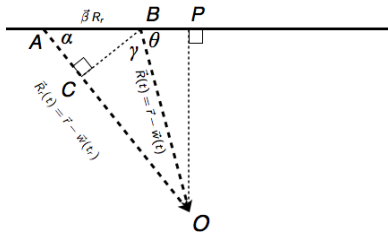
The point O is the position $\vec{r}$ at which we want to calculate the potential, the *field point*. The points A and B are the position of the particle at the retarded time t_r and the current time t. The vectors AO and BO are given by

$$AO: \quad \vec{R}_r(t) = \vec{r} - \vec{w}(t_r) \qquad BO: \quad \vec{R}(t) = \vec{r} - \vec{w}(t) \qquad (12.6)$$

Griffiths uses a boldface script r for our $\vec{R}_r$ and a boldface R for our $\vec{R}$. These are the *retarded relative position* and the *current relative position*.

We can relate γ to α and θ:

$$\gamma = \pi - \left(\theta + \left(\frac{\pi}{2} - \alpha\right)\right) = \frac{\pi}{2} - (\theta - \alpha) \tag{12.7}$$

$$\implies \quad \cos\gamma = \cos\left[\frac{\pi}{2} - (\theta - \alpha)\right] = \sin(\theta - \alpha) = \sin\theta\cos\alpha - \cos\theta\sin\alpha \tag{12.8}$$

Then, from the figure,

$$R_r \sin\alpha = R\sin\theta \tag{12.9}$$

because the right triangles with $\vec{R}_r$ and $\vec{R}$ as hypotenuses share the same vertical side, PO. We also have

$$|AP| = |AB| + |BP| \tag{12.10}$$

$$\implies \quad R_r \cos\alpha = \beta R_r + R\cos\theta \tag{12.11}$$

We use the above two relations to substitute for $\cos\alpha$ and $\sin\alpha$ in the expression for $\cos\gamma$:

$$\cos\gamma = \sin\theta\left[\sin\theta\,\frac{\beta R_r + R\cos\theta}{R_r} - \cos\theta\,\frac{R\sin\theta}{R_r}\right] = \beta\sin\theta \tag{12.12}$$

Thus, we can rewrite the denominator of Equation 12.1:

$$R\sqrt{1 - \beta^2 \sin^2 \theta} = R\sqrt{1 - \cos^2 \gamma} \qquad (12.13)$$

From the figure, we also see

$$R\sqrt{1 - \cos^2 \gamma} = |BO| \sin \gamma = |CO| = |AO| - |AC| \qquad (12.14)$$

$$= R_r - \vec{\beta} R_r \cdot \widehat{R}_r = R_r \left(1 - \vec{\beta} \cdot \widehat{R}_r\right) \qquad (12.15)$$

which is obtained as follows. The vector AO is $\vec{R}_r$, so $|AO| = R_r$. AC is part of a right triangle whose hypotenuse is AB, the distance between the retarded position and the current position. $|AB| = \beta R_r$, the distance the particle moves during the time it takes light to travel the distance from the retarded position to the field point. The angle between AC and AB is θ, the angle between the velocity of the particle and the retarded position vector. So $\cos \theta = \widehat{\beta} \cdot \widehat{R}_r$. Therefore, $|AC| = |AB| \cos \theta = \beta R_r \widehat{\beta} \cdot \widehat{R}_r = \vec{\beta} \cdot R_r \widehat{R}_r$. So

$$R\sqrt{1 - \beta^2 \sin^2 \theta} = R_r \left(1 - \vec{\beta} \cdot \widehat{R}_r\right) \qquad (12.16)$$

Inserting this into Equations 12.1 and 12.2, we obtain

$$
\boxed{
\begin{array}{l}
\text{Lienard-Wiechert} \\
\text{scalar potential} \\
\text{using retarded position}
\end{array}
\qquad
V(\vec{r}, t) = \frac{1}{4\pi\epsilon_o} \, \frac{q}{R_r(\vec{r}, t)} \, \frac{1}{1 - \vec{\beta}(t_r) \cdot \widehat{R}_r(\vec{r}, t)}
}
$$

(12.17)

$$
\boxed{
\begin{array}{l}
\text{Lienard-Wiechert} \\
\text{vector potential} \\
\text{using retarded position}
\end{array}
\qquad
\vec{A}(\vec{r}, t) = \frac{\mu_o}{4\pi} \, \frac{q\, c\, \vec{\beta}(t_r)}{R_r(\vec{r}, t)} \, \frac{1}{1 - \vec{\beta}(t_r) \cdot \widehat{R}_r(\vec{r}, t)}
}
$$
(12.18)

These are versions of the Lienard-Wiechert potentials that use the *retarded position* of
the particle. It has been assumed that the particle has constant velocity between t_r
and t. The $\beta \to 0$ limit is also clear here: the correction term becomes unity.
Comments:

▶ These potentials are what one might expect based on the retarded time, but
with a correction factor related to the angle between the direction of motion
and the position vector between the moving particle and the observer.

▶ While $\vec{\beta} \cdot \widehat{R}_r > 0$ is possible, the correction factor is always nonnegative because
$\beta < 1$.

▶ The potentials are stronger than the static case in the half-space ahead of the
particle and weaker in the half-space behind it, with the static value obtained on
the boundary. ("Static" is meaningless for $\vec{A}$, of course.) This effect can be
understood in terms of a "piling up" effect that is like that of the non-relativistic
Doppler shift. This phenomenon is discussed in more detail in Griffiths §10.3.1.

- ▶ How do we explain the above change from forward-backward-symmetric to asymmetric?

  - ▶ The potential is forward-backward symmetric when we consider the particle at two points (with O ahead or behind the particle) with the same value of *current distance* R.
  - ▶ However, because the particle is moving asymmetrically in the direction $\vec{\beta}$, these two points equidistant from O in R are not equidistant from O in R_r: $R_r > R$ when O is ahead of the particle and $R_r < R$ when O is behind the particle. This difference confounds our forward-backward asymmetry expectations (for a reason we will explain in the comment after the next one).
  - ▶ If we think about two points that are equidistant in R_r, then we recognize that the one for which O is ahead of the particle will have $R < R_r$ and the one for which O is behind the particle will have $R > R_r$. In addition, θ will be closer to $\pi/2$ for the former than for the latter. These effects make the denominator in the "current position" expressions smaller for the former case, thus giving the forward enhancement and backward reduction we expect.
  - ▶ Considering again the two points equidistant in R, we noted above that the point for which O is ahead of the particle has larger R_r: the forward enhancement is canceled by the greater distance R_r, rendering the potential strength equal for the two points.

We can also use the above form to obtain the fields in terms of the retarded position. Recall the relation that we used to convert the Lienard-Wiechert potentials from retarded position to current position:

$$R \sqrt{1 - \beta^2 \sin^2 \theta} = R_r \left(1 - \vec{\beta} \cdot \widehat{R}_r\right) \tag{12.19}$$

Revisiting the geometry, we also have

$$\vec{R} = \vec{R}_r - \vec{\beta} R_r = R_r \left(\widehat{R}_r - \vec{\beta}\right) \tag{12.20}$$

In addition, we note that, because $\vec{E} \propto \vec{R} = R_r(\widehat{R}_r - \vec{\beta})$, then $(\widehat{R}_r - \vec{\beta}) \times \vec{E} = 0$ and therefore

$$\vec{B} = \frac{1}{c} \vec{\beta} \times \vec{E} = \frac{1}{c} \left[\vec{\beta} + \left(\widehat{R}_r - \vec{\beta}\right)\right] \times \vec{E} = \frac{1}{c} \widehat{R}_r \times \vec{E} \tag{12.21}$$

Inserting the above relations into the equations for the fields in terms of the current position, we have

Electric field of fixed
velocity point charge
using retarded position

$$\vec{E}(\vec{r}, t) = \frac{q}{4\pi\epsilon_o} \frac{\left(\vec{R}_r - \vec{\beta}\, R_r\right)}{R_r^3} \frac{(1 - \beta^2)}{\left[1 - \vec{\beta} \cdot \widehat{R}_r\right]^3}$$ (12.22)

Magnetic field of fixed
velocity point charge
using retarded position

$$\vec{B}(\vec{r}, t) = \frac{\mu_o}{4\pi} \frac{q\, c\, \vec{\beta} \times \vec{R}_r}{R_r^3} \frac{(1 - \beta^2)}{\left[1 - \vec{\beta} \cdot \widehat{R}_r\right]^3}$$ (12.23)

Relation between electric and magnetic
fields of fixed velocity point charge
using retarded position

$$\vec{B} = \frac{1}{c} \vec{\beta} \times \vec{E} = \frac{1}{c} \widehat{R}_r \times \vec{E}$$ (12.24)

These match Griffiths Equations 10.72 and 10.73 with zero acceleration ($\vec{a} = 0$). $\vec{B}$ wraps around $\vec{\beta}$ as for the current position fields.

Comments:

- ▶ These fields have a reduction factor $1 - \beta^2$ in the numerator, but the overall reduction/enhancement relative to the static case is more complex.

- ▶ Unlike the current position fields, there is a forward-backward asymmetry with enhancement $(1 - \beta^2)/(1 - \beta)^3$ in the forward direction and reduction $(1 - \beta^2)/(1 + \beta)^3$ in the backward direction.

- ▶ These factors are different from those for the potentials, which had only one power of $1 - \vec{\beta} \cdot \widehat{R}_r$ in the denominator.

- ▶ The electric field does not point along $\vec{R}_r$: the field points along the *current relative position* vector, not the retarded relative position vector

- ▶ Again, information about both the position and velocity seems to be transmitted in the fields.

Potentials and Fields of an Accelerated Point Charge

In calculating the fields from the Lienard-Wiechert potentials, we assumed the charge was moving uniformly — that $\vec{\beta}$ was constant — by passing it through all derivatives. Let's now drop this assumption. Since our prior derivations assumed that $\vec{\beta}$ was the velocity vector at the retarded time, we will replace $\vec{\beta}$ with $\vec{\beta}(t_r)$ in all our prior expressions.

We will find that it is more convenient (or, rather, less inconvenient) to use the retarded position expressions for the Lienard-Wiechert potentials.

We assume the particle has a trajectory $\vec{w}(t)$ and that its instantaneous velocity is $\vec{\beta}(t) = \vec{w}'(t)/c$. We will end up evaluating these and derivatives thereof at t_r.

First, let's consider the partial derivatives of $\vec{\beta}(t_r)$. Evaluating these is difficult because of the somewhat circular dependence of $\vec{w}(t_r)$ and t_r, so we use an unobvious but clever technique to do so. We start with the definition of t_r, $t_r = t - |\vec{r} - \vec{w}(t_r)|/c$, rewriting it as

$$c^2 (t - t_r)^2 = (\vec{r} - \vec{w}(t_r)) \cdot (\vec{r} - \vec{w}(t_r)) \tag{12.25}$$

Let's take the partial derivatives of both sides. First, with respect to t holding $\vec{r}$ fixed, and recalling that $c(t - t_r) = R_r$ and $\vec{\beta} = \vec{w}'/c$,

$$2c^2(t - t_r)\left(1 - \frac{\partial t_r}{\partial t}\bigg|_{\vec{r}}\right) = 2(\vec{r} - \vec{w}(t_r)) \cdot (-\vec{w}'(t_r)) \frac{\partial t_r}{\partial t}\bigg|_{\vec{r}} \tag{12.26}$$

$$-\left[cR_r - \vec{R}_r \cdot c\,\vec{\beta}(t_r)\right] \frac{\partial t_r}{\partial t}\bigg|_{\vec{r}} = -cR_r \tag{12.27}$$

$$\frac{\partial t_r}{\partial t}\bigg|_{\vec{r}} = \frac{1}{1 - \vec{\beta}(t_r) \cdot \widehat{R}_r} \tag{12.28}$$

We repeat the same procedure with $\partial/\partial r_i$, holding t fixed:

$$-2c^2(t - t_r)\frac{\partial t_r}{\partial r_i}\bigg|_t = 2\left[(r_i - w_i(t_r)) + (\vec{r} - \vec{w}(t_r)) \cdot (-\vec{w}'(t_r)) \frac{\partial t_r}{\partial r_i}\bigg|_t\right] \tag{12.29}$$

$$-\left[cR_r - \vec{R}_r \cdot c\,\vec{\beta}(t_r)\right] \frac{\partial t_r}{\partial r_i}\bigg|_t = R_{r,i} \tag{12.30}$$

$$\frac{\partial t_r}{\partial r_i}\bigg|_t = -\frac{1}{c}\frac{R_{r,i}/R_r}{1 - \vec{\beta}(t_r) \cdot \widehat{R}_r} \tag{12.31}$$

From these, we can calculate derivatives of $\vec{\beta}$, defining $\vec{\beta}' = \vec{w}''/c$ as the first derivative of $\vec{\beta}$ or second derivative of $\vec{w}/c$ with respect to its argument:

$$\left.\frac{\partial \beta_i}{\partial t}\right|_{\vec{r},\widehat{R}_r} = \beta_i'(t_r) \left.\frac{\partial t_r}{\partial t}\right|_{\vec{r},\widehat{R}_r} \quad\Longrightarrow\quad \left.\frac{\partial \vec{\beta}}{\partial t}\right|_{\vec{r},\widehat{R}_r} = \vec{\beta}'(t_r) \left.\frac{\partial t_r}{\partial t}\right|_{\vec{r},\widehat{R}_r} \tag{12.32}$$

$$\left.\frac{\partial \beta_i}{\partial r_j}\right|_{t,\widehat{R}_r} = \beta_i'(t_r) \left.\frac{\partial t_r}{\partial r_j}\right|_{t,\widehat{R}_r} \quad\Longrightarrow\quad \left.\frac{\partial \vec{\beta}}{\partial r_j}\right|_{t,\widehat{R}_r} = \vec{\beta}'(t_r) \left.\frac{\partial t_r}{\partial r_j}\right|_{t,\widehat{R}_r} \tag{12.33}$$

where the partial derivatives of t_r are what we just calculated.

Next, we can use these expressions to calculate derivatives of $\vec{\beta} \cdot \widehat{R}_r$, the quantity involving $\vec{\beta}$ that appears in the denominator of the potential expressions. Remember that the prior process of calculating the fields using Lorentz transformation takes all the necessary derivatives of all parts of the potentials, including derivatives of $\widehat{R}_r$, except derivatives of $\vec{\beta}$. Therefore, when evaluating derivatives of $\vec{\beta} \cdot \widehat{R}_r$, we do not need to take derivatives of $\widehat{R}_r$, so we indicate it is held fixed:

$$\left.\frac{\partial}{\partial t}\left(\vec{\beta} \cdot \widehat{R}_r\right)\right|_{\vec{r},\widehat{R}_r} = \left.\frac{\partial}{\partial t}\sum_i \beta_i \frac{R_{r,i}}{R_r}\right|_{\vec{r},\widehat{R}_r} = \sum_i \beta_i' \frac{R_{r,i}}{R_r}\left.\frac{\partial t_r}{\partial t}\right|_{\vec{r},\widehat{R}_r} = \left(\vec{\beta}' \cdot \widehat{R}_r\right)\left.\frac{\partial t_r}{\partial t}\right|_{\vec{r},\widehat{R}_r} \tag{12.34}$$

$$\left.\frac{\partial}{\partial r_j}\left(\vec{\beta} \cdot \widehat{R}_r\right)\right|_{t,\widehat{R}_r} = \left.\frac{\partial}{\partial r_j}\sum_i \beta_i \frac{R_{r,i}}{R_r}\right|_{t,\widehat{R}_r} = \sum_i \beta_i' \frac{R_{r,i}}{R_r}\left.\frac{\partial t_r}{\partial r_i}\right|_{t,\widehat{R}_r} = \left(\vec{\beta}' \cdot \widehat{R}_r\right)\left.\frac{\partial t_r}{\partial r_i}\right|_{t,\widehat{R}_r} \tag{12.35}$$

Next, to calculate $\vec{E}$, we start from Equations 12.17 and 12.18 but now allow $\vec{\beta}$ to be differentiated also. We only show the new terms, understanding that $\vec{\beta}$ and $\vec{\beta}\,'$ are evaluated at t_r:

$$-\frac{\partial V}{\partial r_i} = -\frac{\partial V}{\partial r_i}\bigg|_{\vec{\beta}} - \frac{1}{4\pi\epsilon_o}\frac{q}{R_r}\frac{-1}{\left[1-\vec{\beta}\cdot\widehat{R}_r\right]^2}\left(-\frac{\partial}{\partial r_i}\left[\vec{\beta}\cdot\widehat{R}_r\right]\bigg|_{t,\widehat{R}_r}\right) \tag{12.36}$$

$$= -\frac{\partial V}{\partial r_i}\bigg|_{\vec{\beta}} + \frac{q}{4\pi\epsilon_o}\frac{1}{c}\frac{R_{r,i}}{R_r^2}\frac{1}{\left[1-\vec{\beta}\cdot\widehat{R}_r\right]^3}\vec{\beta}\,'\cdot\widehat{R}_r \tag{12.37}$$

$$\implies \quad -\vec{\nabla}_{\vec{r}}V = -\vec{\nabla}_{\vec{r}}V\bigg|_{\vec{\beta}} + \frac{q}{4\pi\epsilon_o}\frac{1}{c}\frac{\widehat{R}_r}{R_r}\frac{1}{\left[1-\vec{\beta}\cdot\widehat{R}_r\right]^3}\vec{\beta}\,'\cdot\widehat{R}_r \tag{12.38}$$

$$-\frac{\partial\vec{A}}{\partial t} = -\frac{\partial\vec{A}}{\partial t}\bigg|_{\vec{\beta}} - \frac{\mu_o}{4\pi}\frac{q\,c}{R_r}\left[\frac{-\vec{\beta}}{\left[1-\vec{\beta}\cdot\widehat{R}_r\right]^2}\left(-\frac{\partial}{\partial t}\left[\vec{\beta}\cdot\widehat{R}_r\right]\bigg|_{\vec{r},\widehat{R}_r}\right) + \frac{1}{1-\vec{\beta}\cdot\widehat{R}_r}\frac{\partial\vec{\beta}}{\partial t}\bigg|_{\vec{r},\widehat{R}_r}\right] \tag{12.39}$$

$$= -\frac{\partial\vec{A}}{\partial t}\bigg|_{\vec{\beta}} - \frac{\mu_o}{4\pi}\frac{q\,c}{R_r}\frac{1}{\left[1-\vec{\beta}\cdot\widehat{R}_r\right]^3}\left[\vec{\beta}\left(\vec{\beta}\,'\cdot\widehat{R}_r\right) + \left(1-\vec{\beta}\cdot\widehat{R}_r\right)\vec{\beta}\,'\right] \tag{12.40}$$

where we used our formulae from the previous two pages.

We can combine the three terms, recognizing that $\epsilon_o\,\mu_o = 1/c^2$:

$$\vec{E} = \vec{E}\Big|_{\vec{\beta}} + \frac{q}{4\,\pi\,\epsilon_o}\,\frac{1}{c\,R_r}\,\frac{1}{\left[1 - \vec{\beta}\cdot\widehat{R}_r\right]^3}\,\left[\left(\widehat{R}_r - \vec{\beta}\right)\left(\widehat{R}_r\cdot\vec{\beta}'\right) - \left(1 - \vec{\beta}\cdot\widehat{R}_r\right)\vec{\beta}'\right]$$

(12.41)

Finally, using the $BAC - CAB$ rule and adding powers of R_r to turn $\widehat{R}_r$ into $\vec{R}_r$:

$$\boxed{\vec{E}(\vec{r}, t) = \vec{E}(\vec{r}, t)\Big|_{\vec{\beta}} + \frac{q}{4\,\pi\,\epsilon_o}\,\frac{1}{R_r^3}\,\frac{1}{c}\,\frac{\vec{R}_r \times \left[\left(\vec{R}_r - \vec{\beta}\,R_r\right) \times \vec{\beta}'\right]}{\left[1 - \vec{\beta}\cdot\widehat{R}_r\right]^3}}$$

Electric field
of accelerated
point charge
using retarded
position

(12.42)

($\vec{\beta}$ and $\vec{\beta}'$ are evaluated at t_r.) We see that the expression is very similar to the fixed-$\vec{\beta}$ term, Equation 12.22, except that $\left(\vec{R}_r - \vec{\beta}\,R_r\right)\left(1 - \beta^2\right)$ is replaced with the triple cross product. We also see that this term is smaller than the fixed-$\vec{\beta}$ term by a factor $(R_r/c)\,\beta'$, which is sensible: this factor is the ratio of the light travel time to the characteristic *relativistic* acceleration timescale (*i.e.*, where the latter is normalized to c, not to $c\,\beta$, as we discussed in the context of Jefimemko's Equations, Equations 10.106 and 10.107.)

Similarly,

$$B_i = B_i|_{\vec{\beta}} + \sum_{j,k} \epsilon_{ijk} \frac{\partial}{\partial r_j} \left[\frac{\mu_o}{4\pi} \frac{q\,c\,\beta_k(t_r)}{R_r(\vec{r},t)} \frac{1}{1 - \vec{\beta}(t_r) \cdot \hat{R}_r(\vec{r},t)} \right]\Bigg|_{t,\hat{R}_r} \tag{12.43}$$

$$= B_i|_{\vec{\beta}} + \frac{\mu_o}{4\pi} \frac{q\,c}{R_r} \sum_{j,k} \epsilon_{ijk} \left[\frac{\frac{\partial \beta_k}{\partial r_j}\Big|_{t,\hat{R}_r}}{1 - \vec{\beta}\cdot\hat{R}_r} + \frac{\beta_k \frac{\partial}{\partial r_j}\left(\vec{\beta}\cdot\hat{R}_r\right)\Big|_{t,\hat{R}_r}}{\left[1 - \vec{\beta}\cdot\hat{R}_r\right]^2} \right] \tag{12.44}$$

$$= B_i|_{\vec{\beta}} + \frac{\mu_o}{4\pi} \frac{q\,c}{R_r} \sum_{j,k} \epsilon_{ijk} \left[\frac{\beta_k'\left(-\frac{R_{r,j}}{c\,R_r}\right)}{\left[1 - \vec{\beta}\cdot\hat{R}_r\right]^2} + \frac{\beta_k\left(\vec{\beta}'\cdot\hat{R}_r\right)\left(-\frac{R_{r,j}}{c\,R_r}\right)}{\left[1 - \vec{\beta}\cdot\hat{R}_r\right]^3} \right] \tag{12.45}$$

$$\implies \vec{B} = \vec{B}|_{\vec{\beta}} + \frac{\mu_o}{4\pi} \frac{q}{R_r} \left[-\frac{\hat{R}_r \times \vec{\beta}'}{\left[1 - \vec{\beta}\cdot\hat{R}_r\right]^2} + \frac{\vec{\beta}\times\hat{R}_r}{\left[1 - \vec{\beta}\cdot\hat{R}_r\right]^3} \hat{R}_r\cdot\vec{\beta}' \right] \tag{12.46}$$

Therefore (multiplying top and bottom by R_r^2),

$$\vec{B}(\vec{r}, t) = \vec{B}(\vec{r}, t)\Big|_{\vec{\beta}} + \frac{\mu_o}{4\pi} \frac{q}{R_r^3} \frac{\left(R_r - \vec{\beta} \cdot \vec{R}_r\right) \vec{\beta}' \times \vec{R}_r + \left(\vec{R}_r \cdot \vec{\beta}'\right) \vec{\beta} \times \vec{R}_r}{\left[1 - \vec{\beta} \cdot \widehat{R}_r\right]^3} \qquad (12.47)$$

($\vec{\beta}$ and $\vec{\beta}'$ are evaluated at t_r.) As with $\vec{E}$, we see a form similar to the fixed-$\vec{\beta}$ term, now with replacement of $\vec{\beta} \times \vec{R}_r(1 - \beta^2)$ by the complicated expression in the numerator. Like for the electric field, this term is smaller than the fixed-$\vec{\beta}$ term by a factor $(R_r/c)\beta'$. If we compare this to Equation 12.41, the acceleration electric field prior to application of the $BAC - CAB$ rule, we can also conclude

$$\vec{B} = \frac{1}{c}\widehat{R}_r \times \vec{E} \neq \frac{1}{c}\vec{\beta} \times \vec{E} \qquad \begin{array}{l} \text{Magnetic field of accelerated point} \\ \text{charge using retarded position} \end{array} \qquad (12.48)$$

Only one of the relations between $\vec{E}$ and $\vec{B}$ is preserved.

The direction of these acceleration fields is dramatically different from that of the fixed-velocity fields. The latter were similar in direction to the static field, with $\vec{E} \propto \vec{R}$, the current position vector and thus being radial, and $\vec{B}$ wrapping around $\vec{\beta}$ in the usual azimuthal manner. By contrast, the acceleration electric field is perpendicular to $\vec{R}_r$; thus it is *transverse* to the *vector from the retarded position*. We will see below that the remaining direction choice for $\vec{E}$ is set by the acceleration vector $\vec{\beta}\,'$: the electric field is in the plane formed by $\vec{R}_r$ and $\vec{\beta}\,'$. The magnetic field direction is completely determined by $\vec{R}_r$ and $\vec{E}$: it is perpendicular to both, making it also transverse relative to the retarded position (and thus also determined by $\vec{R}_r$ and $\vec{\beta}\,'$).

It is striking how the fixed-velocity fields seem to emanate from the current position of the particle, and thus they carry information about the particle's position and velocity at the retarded time but transfer it forward to the current position, while the acceleration fields clearly emanate from the particle's retarded position and are determined by its acceleration at the retarded time. The acceleration fields "make no assumption" about the acceleration being fixed at times in the future, while the fixed-velocity fields assume fixed velocity into the future.

We can view this differing behavior of fixed-velocity and acceleration fields as a consequence of Galilean relativity (which is also the first postulate of special relativity): for the fixed-velocity field, we are just taking the electrostatic field of a point charge at rest and transforming frames. That transformation is valid at all space and time points if the point charge frame has been moving at fixed velocity for all time, and thus that transformation encodes information about the particle position assuming fixed velocity. It can only depend on current position. By contrast, the acceleration fields have to depend on retarded position due to causality.

Lecture 35:

Radiation II:
Potentials, Fields, and Power Radiated by
an Accelerated Point Charge (cont.)
Bremsstrahlung and Synchrotron Radiation
Lienard's and Larmor's Formulae
General Theory of Radiation

Date Revised: 2024/05/24 09:45
Date Given: 2024/05/25

Poynting Vector Radiated by an Accelerating Point Charge

With the above detailed formulae for the fields due to an accelerated charge, we can make our first study of *radiation*, which consists of the propagating fields created by accelerated charges. We will of course calculate the power radiated using the Poynting vector, $\vec{S} = \vec{E} \times \vec{B}/\mu_0$ (no need for complex conjugations or taking real parts because all fields are real here; also, we don't time average because we are not yet considering sinusoidal behavior). We can quickly see that only the terms involving the acceleration $\vec{a} = c\,\vec{\beta}\,'$ are important at large distances by extracting the dependences of various terms from the full expressions:

$$\left.\vec{E}\right|_{\vec{\beta}}: \qquad \frac{q}{\epsilon_o}\,\frac{1}{\gamma^2}\,\frac{1}{R_r^2} \qquad\qquad \left.\vec{B}\right|_{\vec{\beta}}: \qquad \frac{q}{\epsilon_o}\,\frac{1}{c}\,\frac{\beta}{\gamma^2}\,\frac{1}{R_r^2} \qquad (12.49)$$

$$\left.\vec{E} - \vec{E}\right|_{\vec{\beta}}: \qquad \frac{q}{\epsilon_o}\,\frac{1}{c}\,\vec{\beta}\,'\,\frac{1}{R_r} \qquad\qquad \left.\vec{B} - \vec{B}\right|_{\vec{\beta}}: \qquad \frac{q}{\epsilon_o}\,\frac{1}{c^2}\,(1+\beta)\,\vec{\beta}\,'\,\frac{1}{R_r} \quad (12.50)$$

The $\big|_{\vec{\beta}}$ terms are called the *velocity* terms and the others are called the *acceleration* terms. Only the latter are important at large distances, the so-called *far field*.

We note a number of important facts about the fields:

- ▶ We again recognize that the acceleration terms differ from the velocity terms by a factor $(R_r/c)\,\beta'$, which is the ratio of the light travel time to the relativistic acceleration timescale τ_a. This factor was introduced by the derivatives of $\vec{\beta}$ that were taken to obtain the fields from the potentials.

- ▶ Note, too, that this factor causes the replacement of a factor of R_r in the denominator with the length quantity c/β', which is the distance light travels in an acceleration timescale. This is a key replacement, as it is what makes the acceleration fields dominant at large distances, yielding the $1/R^2$ law for the radiated power!

- ▶ The acceleration electric field is perpendicular to $\widehat{R}_r$ (because it is the result of a cross-product including $\widehat{R}_r$), which points along the line-of-sight from the retarded position of the point charge to the field point. This may be easier to visualize if you imagine looking at the retarded position from the perspective of the field point: $\widehat{R}_r$ is the vector pointing at you along that line-of-sight.

- ▶ The acceleration magnetic field is perpendicular to both this line-of-sight and the acceleration electric field (because the magnetic field is proportional to the cross product of $\widehat{R}_r$ and the electric field).

Let's now calculate the Poynting vector:

$$\vec{S} \approx \frac{1}{\mu_o} \left(\vec{E} - \vec{E} \big|_{\vec{\beta}} \right) \times \left(\vec{B} - \vec{B} \big|_{\vec{\beta}} \right) = \frac{1}{2\,\mu_o} \left(\vec{E} - \vec{E} \big|_{\vec{\beta}} \right) \times \frac{1}{c} \left[\widehat{R}_r \times \left(\vec{E} - \vec{E} \big|_{\vec{\beta}} \right) \right]$$

$$\boxed{\vec{S} \approx \frac{1}{c\,\mu_o}\, \widehat{R}_r \left| \vec{E} - \vec{E} \big|_{\vec{\beta}} \right|^2} \tag{12.51}$$

where the last step used the $BAC - CAB$ rule and $\widehat{R}_r \cdot (\vec{E} - \vec{E}|_{\vec{\beta}}) = 0$. We see that the Poynting vector is along the line-of-sight from the retarded position to the field point.

We'll consider the cases of the acceleration parallel and perpendicular to the velocity separately and then add the two to get the fully relativistic result. We can then either take its non-relativistic limit to get the Larmor Formula, or we can start from a non-relativistic version of the electric field to get the same result.

Acceleration Parallel to Velocity: Bremsstrahlung

Let's first do the parallel acceleration case. A standard application of this case is to calculate *bremsstrahlung* ("braking radiation"), the radiation an electron gives off as it is decelerated by interaction with matter (primarily the Coulomb force from the positive charge of nuclei.) We set $\vec{\beta} \times \vec{\beta}' = 0$. The relevant piece of Equation 12.42 is

$$\vec{E}_{\vec{\beta}'||\vec{\beta}} = \frac{q}{4\pi\epsilon_o} \frac{1}{R_r} \frac{\widehat{R}_r \times \left(\widehat{R}_r \times \frac{1}{c}\vec{\beta}'\right)}{\left[1 - \vec{\beta}\cdot\widehat{R}_r\right]^3} = -\frac{q}{4\pi\epsilon_o} \frac{1}{R_r} \frac{\frac{1}{c}\vec{\beta}'_\perp}{\left[1 - \vec{\beta}\cdot\widehat{R}_r\right]^3} \quad (12.52)$$

$$\text{with} \quad \vec{\beta}'_\perp = \vec{\beta}' - \widehat{R}_r\left(\widehat{R}_r \cdot \vec{\beta}'\right) \qquad \begin{array}{l}\text{projection of acceleration} \\ \text{perpendicular to line-of-sight from}\end{array} \quad (12.53)$$

$$\text{or} \quad \vec{a}_\perp = \vec{a} - \widehat{R}_r\left(\widehat{R}_r \cdot \vec{a}\right) \qquad \text{the retarded position } (\widehat{R}_r) \qquad (12.54)$$

where we used the $BAC - CAB$ rule to evaluate the triple cross product. The result tells us that only the projection of the acceleration perpendicular to the line-of-sight is responsible for the far-field radiation. (Be sure not to confuse this with the case of acceleration perpendicular to the velocity! In fact, in this case, the acceleration is parallel to the velocity.) A bit strange: velocity and acceleration in one direction only yields radiation at angles away from that direction, and the radiation is strongest in the direction perpendicular to the acceleration and velocity. If one thinks about the relationship between the acceleration and the electric field direction, though, it makes perfect sense: the electric field in a given direction is only affected by acceleration parallel to that direction.

Next, if we define $\vec{a}$ to be the z-axis of a spherical coordinate system and θ the polar angle of the field point, then we have $\vec{a} = a\,\widehat{z}$, $|\vec{a}_{\perp}| = a\sin\theta$, and $\vec{\beta}\cdot\widehat{R}_r = \beta\cos\theta$, so

$$\vec{S} = \frac{\mu_o\,q^2}{16\,\pi^2}\,\frac{1}{c\,R_r^2}\,\frac{a^2\sin^2\theta}{[1-\beta\cos\theta]^6}\,\widehat{R}_r \tag{12.55}$$

(We rewrote ϵ_o in terms of c and μ_o.) The energy radiated into a unit area per unit *field point* time interval is $d^2U/dt\,dA = \widehat{R}_r\cdot\vec{S}$.

However, we will also be interested in knowing the energy *lost by the particle* per unit time, so we need to convert the dt in the above to a dt_r by dividing by $\partial t_r/\partial t$, which we calculated in Equation 12.28. The distinction between the two is the same (nonrelativistic!) "piling up" effect that we discussed earlier in the context of calculating the Lienard-Wiechert potential. Therefore (multiplying by R_r^2 to convert from power per unit area to power per unit solid angle):

$$\frac{dP}{d\Omega} = \frac{d^2U}{dt_r\,d\Omega} = \frac{d^2U}{dt\,d\Omega}\left(\left.\frac{\partial t_r}{\partial t}\right|_{\vec{\beta}}\right)^{-1} = R_r^2\,\widehat{R}_r\cdot\vec{S}\left(\left.\frac{\partial t_r}{\partial t}\right|_{\vec{\beta}}\right)^{-1}$$

$$= \frac{\mu_o\,q^2}{16\,\pi^2}\,\frac{1}{c}\,\frac{a^2\sin^2\theta}{[1-\beta\cos\theta]^6}\left(1-\vec{\beta}\cdot\widehat{R}_r\right) \tag{12.56}$$

$$\boxed{\frac{dP}{d\Omega} = \frac{\mu_o\,q^2}{16\,\pi^2}\,\frac{a^2}{c}\,\frac{\sin^2\theta}{[1-\beta\cos\theta]^5}} \quad \begin{array}{c}\text{power per solid angle radiated by a}\\\text{point charge accelerated}\\\text{parallel to direction of motion}\end{array} \tag{12.57}$$

The following figure is a polar plot that illustrates the shape of this function for a particle moving from left to right. When $\beta \ll 1$, one obtains a distribution symmetric between the forward and the reverse direction. As β increases, the denominator begins to take effect and the radiation is strongly directed into the forward hemisphere (though the radiation along the direction of motion continues to vanish), with the peak intensity found on a cone.
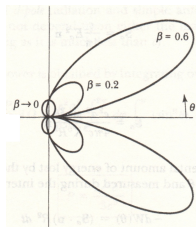
© 2012 Heald and Marion, *Classical Electromagnetic Radiation*

Bremsstrahlung is present in a variety of environments. In astrophysics, it is called "free-free emission" and arises in hot, ionized plasmas where the electrons experience acceleration as they Coulomb scatter with ions. It is the dominant emission mechanism of the hot gas in galaxy clusters (where the bremsstrahlung appears at X-ray energies) and of hot ionized plasma in our galaxy (where the lower temperature gas yields bremsstrahlung at radio frequencies, up to 100–200 GHz). Bremsstrahlung is also a dominant mechanism for energy loss by high-energy electrons passing through matter, such as electrons produced by radioactive beta decay or by pair-production by gamma rays.

One can integrate the above over all angles to obtain the total radiated power

$$
P = \frac{\mu_o \, q^2}{6 \, \pi} \, \frac{a^2}{c \, (1 - \beta^2)^3} = \frac{\mu_o \, q^2}{6 \, \pi} \, \frac{a^2}{c} \, \gamma^6 \qquad
\begin{array}{c}
\text{total power radiated by a} \\
\text{point charge accelerated} \\
\text{parallel to direction of motion}
\end{array} \qquad (12.58)
$$

where $\gamma = (1 - \beta^2)^{-1/2}$ is the usual relativistic Lorentz factor. Note the strong dependence on γ!

Acceleration Perpendicular to Velocity: Synchrotron Radiation

Now, let's do the perpendicular acceleration case. Unfortunately, in this case we have to keep both terms in the accelerated electric field, Equation 12.42, and the expression is not simple. If we set $\vec{\beta} = \beta \, \widehat{z}$ and $\vec{a} = a \, \widehat{x}$, then one can work out the vector algebra (see, *e.g.*, Heald and Marion §8.8) to show

$$\frac{dP}{d\Omega} = \frac{\mu_o \, q^2}{16 \, \pi^2} \, \frac{a^2}{c} \, \frac{(1 - \beta \, \cos\theta)^2 - \left(1 - \beta^2\right) \sin^2\theta \, \cos^2\phi}{[1 - \beta \, \cos\theta]^5} \qquad \begin{array}{c} \text{power per solid angle} \\ \text{radiated by a} \\ \text{point charge} \\ \text{moving along } \widehat{z} \text{ and} \\ \text{accelerated along } \widehat{x} \end{array} \qquad (12.59)$$

θ is the polar angle relative to the particle's direction of motion (*i.e.*, angle away from $\widehat{z}$ in this case), while ϕ is the azimuthal angle around the particle's direction of motion, with $\phi = 0$ being along the acceleration direction ($\widehat{x}$ in this case).

The classic example of synchrotron radiation is a charged particle executing circular motion due to the Lorentz force in a uniform magnetic field, for which the velocity and acceleration are perpendicular. The resulting radiation pattern (valid for any case of perpendicular velocity and acceleration, not just circular motion) is shown below. As $\beta \to 1$, the radiation becomes "beamed" in the direction of motion. In the $\beta \to 0$ limit, the radiation pattern is a donut whose axis is along $\vec{a}$, so the radiation in the forward and backward directions and in the perpendicular direction along $\widehat{y}$ have the same intensity.
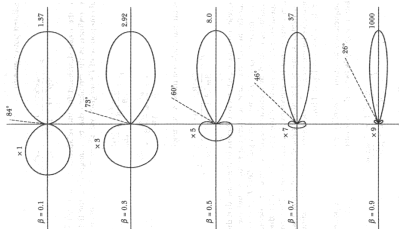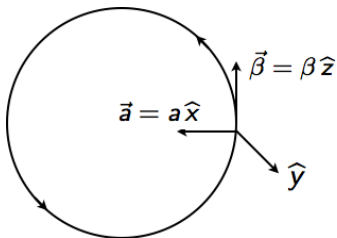
© 2012 Heald and Marion, *Classical Electromagnetic Radiation*

The numbers at the top give the Lorentz factor, the dashed lines indicate the angle at which the radiated power vanishes, and the $\times$ factors indicate the enhancement of the "backward" lobe for visualization (which, in fact, moves to the forward hemisphere as $\beta \to 1$).

Synchrotron radiation is also present in a variety of environments. Synchrotron light sources put electrons in a circular ring, producing intense light in the direction tangent to the circle. Such light sources produce photons in energy from tens of eV (hard UV) to tens of keV (hard X-ray), mostly for studies of materials. Synchrotron radiation is used as a beam monitor for high-energy physics electron storage rings (LEP and BaBar in the 1990s–2000s). Synchrotron radiation is ubiquitous in astrophysical environments. It is the dominant mechanism for radio emission from our galaxy at tens of GHz and lower frequencies, arising from electrons spiraling in the magnetic fields of supernova remnants. It is the dominant source of emission in pulsars and magnetars, spinning neutron stars with enormous magnetic fields. It is also the dominant mechanism in the jets of accelerated particles emitted by the supermassive black holes that power active galactic nuclei: the magnetic field winds up and threads the jets emitted from the poles of the rotating black hole, and the particles accelerated in the jet execute circular motion around the field lines, emitting synchrotron radiation. Synchrotron radiation is generally highly polarized because the field line geometry creates a preferred direction.

The total radiated power is (weaker but still strong dependence on γ)

$$P = \frac{\mu_o \, q^2}{6 \, \pi} \frac{a^2}{c \, (1 - \beta^2)^2} = \frac{\mu_o \, q^2}{6 \, \pi} \frac{a^2}{c} \gamma^4 \qquad \begin{array}{c} \text{total power radiated by a} \\ \text{point charge accelerated} \\ \text{perpendicular to} \\ \text{direction of motion} \end{array} \qquad (12.60)$$

Arbitrary Acceleration Direction

By extrapolation from these two cases and returning to Equation 12.42, we can write the general formula for the power per unit solid angle:

$$\frac{dP}{d\Omega} = \frac{\mu_o\,q^2}{16\,\pi^2}\,\frac{a^2}{c}\,\frac{\left|\widehat{R}_r \times \left[\left(\widehat{R}_r - \vec{\beta}\right) \times \widehat{a}\right]\right|^2}{[1 - \beta\cos\theta]^5}$$

power per solid angle
radiated by a
point charge
with velocity $\vec{\beta}$ and
acceleration $\vec{a}$

(12.61)

Combining the two total power formulae, which only differ by a factor of γ^2, yields the total power radiated for an arbitrary angle between velocity and acceleration:

$$P = \frac{\mu_o\,q^2}{6\,\pi}\,\frac{a^2}{c}\,\frac{1 - \left|\vec{\beta} \times \widehat{a}\right|^2}{(1 - \beta^2)^3} = \frac{\mu_o\,q^2}{6\,\pi}\,\frac{a^2}{c}\,\gamma^6\left(1 - \left|\vec{\beta} \times \widehat{a}\right|^2\right)$$

total power radiated
by an accelerated
point charge
Lienard's Formula

(12.62)

Non-relativistic Limit

The nonrelativistic or "slowly moving charge" limit is obtained by letting $\beta \to 0$ but allowing $\beta' \neq 0$ in Equation 12.42. The electric field simplifies to

$$\vec{E}(\vec{r}, t) = \vec{E}(\vec{r}, t)\Big|_{\vec{\beta}} + \frac{q}{4\pi\epsilon_o} \frac{\vec{R}_r \times \left[\vec{R}_r \times \vec{\beta}'\right]}{c\,R_r^3} \qquad \begin{array}{c}\text{Electric field} \\ \text{of accelerated} \\ \text{point charge in} \\ \text{non-relativistic limit}\end{array} \qquad (12.63)$$

where the first, fixed-velocity term still falls off more quickly than the second, acceleration term in this limit.

We will calculate the radiated power from Equation 12.51. If we again define $\vec{a}$ to be the z-axis of a spherical coordinate system and θ the polar angle of the field point so that $\vec{a} = a\,\hat{z}$, $\vec{R}_r = \hat{r}$ ($= \vec{R}$ because $\beta \to 0$), then we have

$$\vec{R}_r \times \left[\vec{R}_r \times \vec{\beta}'\right] = \frac{a\,r^2}{c}\hat{r} \times [\hat{r} \times \hat{z}] = -\frac{a\,r^2}{c}\hat{r} \times \hat{\phi}\sin\theta = \frac{a\,r^2}{c}\hat{\theta}\sin\theta \qquad (12.64)$$

Therefore, canceling out factors of $r^2 = R_r^2$,

$$\left|\vec{E} - \vec{E}\Big|_{\vec{\beta}}\right|^2 = \left(\frac{q}{4\pi\epsilon_o}\right)^2 \frac{a^2\sin^2\theta}{c^4\,r^2} \qquad (12.65)$$

Therefore,

$$\vec{S} = \frac{1}{c\mu_o} \widehat{R}_r \left| \vec{E} - \vec{E} \right|_{\vec{\beta}}^2 = \widehat{r} \, \frac{\mu_o \, q^2}{16\,\pi^2} \, \frac{a^2 \sin^2 \theta}{c \, r^2} \tag{12.66}$$

(We rewrote ϵ_o in terms of c and μ_o and used $\widehat{R} = \widehat{r}$ and $R_r = r$.) The energy radiated into a unit area per unit time interval is $d^2 U/dt \, dA = \widehat{r} \cdot \vec{S}$, so the radiated power per unit solid angle and total power radiated then become (the observer-emitter time correction $\partial t/\partial t_r|_{\vec{\beta}} = 1$ in this limit):

$$\boxed{\frac{dP}{d\Omega} = \frac{\mu_o \, q^2}{16\,\pi^2} \, \frac{a^2 \sin^2 \theta}{c} \qquad P = \frac{\mu_o \, q^2}{6\,\pi} \, \frac{a^2}{c} \qquad \begin{array}{c} \text{power per solid angle and total} \\ \text{power radiated by a slowly moving} \\ \text{point charge with acceleration } \vec{a} \\ \textit{Larmor's Formula} \end{array}} \tag{12.67}$$

The angular dependence is simply $\sin^2 \theta$.

We could have obtained this limit from the fully relativistic Equation 12.61 by letting $\beta \to 0$ and $\beta' \neq 0$ as we did here. We would also see that the angular radiation pattern of the parallel and perpendicular acceleration cases (bremsstrahlung and synchrotron) both converge to the above simple, symmetric $\sin^2 \theta$ dependence because the direction of motion no longer breaks the symmetry.

General Theory of Radiation

Introduction and Study Guide

As usual, we don't follow Griffiths closely. We immediately go to the full radiation calculation via the technique of Fourier Transforms, skipping his example of electric dipole radiation. This makes it easier to interpret the radiation field in terms of spherical outgoing waves.

We note three length scales that will be important in the discussion and on whose relative sizes we will base various approximations:

- ▶ d: the length scale of the source distribution
- ▶ r: the distance from the source distribution to the field point, the point where we want to know the potentials, fields, and radiated power
- ▶ λ: the wavelength of the emitted radiation

Fourier Transforms

We are going to use Fourier techniques to simplify the derivations in the case of radiation from an arbitrary source distribution, so we need to define the formalism.

We use the concept of orthonormal functions that we developed in connection to separation of variables. We state without proof (see ACM 95/100) that any function of time can be written in the form (with $\omega = 2\pi f$):

$$g(t) = \int_{-\infty}^{\infty} df \, \widetilde{g}(f) \, e^{-i\omega t} \qquad\qquad \widetilde{g}(f) = \int_{-\infty}^{\infty} dt \, g(t) \, e^{i\omega t} \qquad (12.68)$$

For the special case of the delta function,

$$\delta(t) = \int_{-\infty}^{\infty} df \, e^{-i\omega t} \qquad\qquad \delta(f) = \int_{-\infty}^{\infty} dt \, e^{i\omega t} = 2\pi\,\delta(\omega) \qquad (12.69)$$

The function $\widetilde{g}(f)$ is the *Fourier Transform* of the function $g(t)$. It has the units of g divided by frequency. For $g(t)$ to which no boundary conditions have been applied (typical BC would be to assume periodicity with a particular frequency f_0 or to assume Dirichlet or Neumann BC at two times t_a and t_b), there is no restriction on the values of f — any value must be allowed. (In real life, the rate at which you can sample $g(t)$ sets an upper limit on the values of f for which information is available — the so-called Nyquist criterion — but we won't worry about real life here...).

Note that our Fourier Transform sign convention for the argument of the exponential is the opposite of that used in electrical engineering (which gives the impedance of inductors and capacitors as $i\,\omega\,L$ and $1/i\,\omega\,C$). As long as one is self-consistent, one can choose whichever convention one likes. But be careful when comparing to other texts.

The value of the Fourier Transform lies in the linearity of electrodynamics. Specifically, define

$$\rho(\vec{r}, t) = \rho_0(\vec{r})\, e^{-i\,\omega\,t} \qquad \vec{J}(\vec{r}, t) = \vec{J}_0(\vec{r})\, e^{-i\,\omega\,t} \qquad (12.70)$$

and suppose the resulting fields are

$$\vec{E}(\vec{r}, t) = \vec{\mathcal{E}}[\rho_0, \vec{J}_0](\vec{r}, f)\, e^{-i\,\omega\,t} \qquad \vec{B}(\vec{r}, t) = \vec{\mathcal{B}}[\rho_0, \vec{J}_0](\vec{r}, f)\, e^{-i\,\omega\,t} \qquad (12.71)$$

where $\vec{\mathcal{E}}[\ ](\vec{r}, f)$ and $\vec{\mathcal{B}}[\ ](\vec{r}, f)$ indicate functional dependence: put in the spatial functions $\rho_0(\vec{r})$, $\vec{J}_0(\vec{r})$ and the frequency f and what you get out are the spatial functions $\vec{\mathcal{E}}(\vec{r}, f)$, $\vec{\mathcal{B}}(\vec{r}, f)$. The linearity of Maxwell's Equations assures us that the harmonic time dependence is carried through from the sources to the fields. More importantly, linearity assures us that, if we have a source distribution with arbitrary time dependence that can be broken down in terms of components with harmonic time dependence like that shown above, then we can calculate the fields for each component using $\vec{\mathcal{E}}[\ ](\vec{r}, f)$ and $\vec{\mathcal{B}}[\ ](\vec{r}, f)$ and then sum them up to get the total field. The Fourier Transform is the tool for doing all this.

Let's be more specific about the procedure. In all generality, define

$$\widetilde{\rho}(\vec{r}, f) = \int_{-\infty}^{\infty} dt\, \rho(\vec{r}, t)\, e^{i\,\omega\,t} \qquad \widetilde{\vec{J}}(\vec{r}, f) = \int_{-\infty}^{\infty} dt\, \vec{J}(\vec{r}, t)\, e^{i\,\omega\,t} \qquad (12.72)$$

The motivation for these expressions is that multiplication by the $e^{i\,\omega\,t}$ factor followed by the integration would pick out the $\rho_0(\vec{r})$ and $\vec{J_0}(\vec{r})$ with the $e^{-i\,\omega\,t}$ time dependence shown on the previous page. Then, from the above solution that determines fields from sources, we know

$$\widetilde{\vec{E}}(\vec{r}, f) = \vec{\mathcal{E}}[\widetilde{\rho}(\vec{r}, f), \widetilde{\vec{J}}(\vec{r}, f)] \qquad \widetilde{\vec{B}}(\vec{r}, f) = \vec{\mathcal{B}}[\widetilde{\rho}(\vec{r}, f), \widetilde{\vec{J}}(\vec{r}, f)] \qquad (12.73)$$

and then, by linearity,

$$\vec{E}(\vec{r}, t) = \int_{-\infty}^{\infty} df\, \widetilde{\vec{E}}(\vec{r}, f)\, e^{-i\,\omega\,t} \qquad \vec{B}(\vec{r}, t) = \int_{-\infty}^{\infty} df\, \widetilde{\vec{B}}(\vec{r}, f)\, e^{-i\,\omega\,t} \qquad (12.74)$$

Therefore, we only need determine the functions $\vec{\mathcal{E}}[\ ](\vec{r}, f)$ and $\vec{\mathcal{B}}[\ ](\vec{r}, f)$ by determining the fields for a harmonic time dependence $e^{-i\,\omega\,t}$ of the sources and then we can use Fourier Transforms to calculate the fields for arbitrary source time dependence.

Linearity also holds, surprisingly, for quadratic quantities like energy and radiation if one time averages. Specifically, let's calculate the time-averaged Poynting vector:

$$\left\langle \vec{S}(\vec{r}, t) \right\rangle = \frac{1}{2\,\mu_o} \left\langle \mathcal{R}\left(\vec{E}^*(\vec{r}, t) \times \vec{B}(\vec{r}, t) \right) \right\rangle \tag{12.75}$$

$$= \frac{1}{2\,\mu_o} \left\langle \mathcal{R}\left(\int_{-\infty}^{\infty} df_1 \int_{-\infty}^{\infty} df_2 \, e^{i(\omega_1 - \omega_2)t} \, \tilde{\vec{E}}^*(\vec{r}, f_1) \times \tilde{\vec{B}}(\vec{r}, f_2) \right) \right\rangle \tag{12.76}$$

$$= \frac{1}{2\,\mu_o} \mathcal{R}\left(\int_{-\infty}^{\infty} df_1 \int_{-\infty}^{\infty} df_2 \, \left\langle e^{i(\omega_1 - \omega_2)t} \right\rangle \tilde{\vec{E}}^*(\vec{r}, f_1) \times \tilde{\vec{B}}(\vec{r}, f_2) \right) \tag{12.77}$$

Now,

$$\left\langle e^{i\,\omega\,t} \right\rangle = \lim_{T \to \infty} \frac{1}{T} \int_{-T/2}^{T/2} dt \, e^{i\,\omega\,t} = \lim_{T \to \infty} \frac{\int_{-T/2}^{T/2} dt \, e^{i\,\omega\,t}}{\int_{-T/2}^{T/2} dt} = \frac{\delta(f)}{\delta(f = 0)} \tag{12.78}$$

So,

$$\left\langle \vec{S}(\vec{r}, t) \right\rangle = \frac{1}{2\,\mu_o} \, \mathcal{R} \left(\int_{-\infty}^{\infty} df_1 \int_{-\infty}^{\infty} df_2 \, \frac{\delta(f_1 - f_2)}{\delta(f = 0)} \, \tilde{\vec{E}}^*(\vec{r}, f_1) \times \tilde{\vec{B}}^*(\vec{r}, f_2) \right) \quad (12.79)$$

$$= \frac{1}{\delta(f = 0)} \, \frac{1}{2\,\mu_o} \, \mathcal{R} \left(\int_{-\infty}^{\infty} df_1 \, \tilde{\vec{E}}^*(\vec{r}, f_1) \times \tilde{\vec{B}}^*(\vec{r}, f_1) \right) \quad (12.80)$$

There is a funny normalizing factor that corrects for the fact that the Fourier Transforms of the fields have the units of field divided by frequency; since there is only one integral over frequency left, the $\delta(f = 0)$ provides another unit of frequency in the numerator as needed. Recall, $\delta(f = 0) = \lim_{T \to \infty} T$ is present because of the time-averaging. It will always be canceled by a similar factor in the numerator, eliminating what appears to be division by ∞. (We will not actually be calculating power from the Fourier Transforms directly, so this will turn out not to be an issue below.) More importantly, the above tells us that the *time-averaged* Poynting vector is obtained by summing up the contributions from each frequency in a linear way: the power at different frequencies just adds up, so we can calculate the power for a given frequency and then do sums to get total power. The cross-terms drop away in the time average.

A final point to make is that time derivatives become powers of ω for Fourier transforms. This is seen by differentiating the expression for $g(t)$:

$$\frac{dg}{dt} = \frac{d}{dt} \int_{-\infty}^{\infty} df \, \tilde{g}(f) \, e^{-i\,\omega\,t} = \int_{-\infty}^{\infty} df \, \tilde{g}(f) \, \frac{d}{dt} \, e^{-i\,\omega\,t} = \int_{-\infty}^{\infty} df \, (-i\,\omega) \, \tilde{g}(f) \, e^{-i\,\omega\,t} \quad (12.81)$$

So, in what we do below, a factor of ω will be rewritten as a time derivative.

Radiation from an Arbitrary Source Distribution

Let's now consider the radiation from an arbitrary configuration of time-varying charges and currents. It will be no surprise that we need to apply techniques similar to those used for calculating the potentials and fields of electric and magnetic multipole configurations. As motivated by our discussion of Fourier Transforms, we start with source distributions having simple harmonic time dependence

$$\rho(\vec{r}, t) = \rho_0(\vec{r})\, e^{-i\,\omega\, t} \qquad \vec{J}(\vec{r}, t) = \vec{J}_0(\vec{r})\, e^{-i\,\omega\, t} \tag{12.82}$$

We assume that ρ_0 and $\vec{J}_0$ are zero outside some volume $\mathcal{V}$ near the origin. The retarded scalar and vector potentials are (Equations 10.71; recall, $t_r = t - |\vec{r} - \vec{r}'|/c$)

$$V(\vec{r}, t) = \frac{1}{4\,\pi\,\epsilon_o} \int_{\mathcal{V}} d\tau'\, \frac{\rho\,(\vec{r}, t_r)}{|\vec{r} - \vec{r}'|} \qquad \vec{A}(\vec{r}, t) = \frac{\mu_o}{4\,\pi} \int_{\mathcal{V}} d\tau'\, \frac{\vec{J}(\vec{r}, t_r)}{|\vec{r} - \vec{r}'|} \tag{12.83}$$

When $e^{-i\,\omega\, t}$ is evaluated at $t = t_r$, we get

$$e^{-i\,\omega\, t_r} = e^{-i\,\omega\, t} e^{i\,k\,|\vec{r} - \vec{r}'|} \qquad \text{with} \qquad k = \frac{\omega}{c} \tag{12.84}$$

Let's rewrite the retarded potentials using this information. With the assumption of harmonic time dependence and no sources outside $\mathcal{V}$ ($\rho = 0$, $\vec{J} = 0$ outside $\mathcal{V}$), Ampere's Law tells us that, outside $\mathcal{V}$,

$$\epsilon_o \, \mu_o \, \frac{\partial \vec{E}}{\partial t} = \vec{\nabla} \times \vec{B} \qquad \Longrightarrow \qquad \vec{E} = c^2 \, \frac{i}{\omega} \, \vec{\nabla} \times \vec{B} \tag{12.85}$$

Therefore, we only need to calculate $\vec{A}$. It becomes

$$\vec{A}(\vec{r}, t) = \frac{\mu_o}{4\pi} \, e^{-i\omega t} \int_{\mathcal{V}} d\tau' \, \vec{J_0}(\vec{r}') \, \frac{e^{i \, k \, |\vec{r} - \vec{r}'|}}{|\vec{r} - \vec{r}'|} \tag{12.86}$$

Now, for large distances, $|\vec{r}| \gg |\vec{r}'|$ for $\vec{r}'$ inside $\mathcal{V}$, so we can Taylor expand $|\vec{r} - \vec{r}'|$. (In our electric and magnetic multipole expansions, we used Equation 3.149 to expand $|\vec{r} - \vec{r}'|^{-1}$, but here we want to expand $|\vec{r} - \vec{r}'|$ in the argument of the exponential too.). This expansion is:

$$|\vec{r} - \vec{r}'| = \sqrt{r^2 - 2\,\vec{r} \cdot \vec{r}' + (r')^2} = r\sqrt{1 - 2\,\widehat{r} \cdot \frac{\vec{r}'}{r} + \left(\frac{r'}{r}\right)^2} \tag{12.87}$$

$$= r - \widehat{r} \cdot \widehat{r}' \, r' + \mathcal{O}\left(\frac{d^2}{r}\right) \tag{12.88}$$

where d is the characteristic size of the source distribution.

Putting these two together, we have

$$\frac{e^{i\,k\,|\vec{r}-\vec{r}'|}}{|\vec{r}-\vec{r}'|} \approx \frac{1}{r}\left[1 + \hat{r}\cdot\hat{r}'\,\frac{r'}{r} + \mathcal{O}\left(\frac{d^2}{r^2}\right)\right]\exp\left[i\,k\left(r - \hat{r}\cdot\hat{r}'\,r' + \mathcal{O}\left(\frac{d^2}{r}\right)\right)\right] \quad (12.89)$$

$$= \frac{e^{i\,k\,r}}{r}\left[1 + \hat{r}\cdot\hat{r}'\,\frac{r'}{r} + \mathcal{O}\left(\frac{d^2}{r^2}\right)\right]\exp\left(-i\,k\,r'\,\hat{r}\cdot\hat{r}' + i\,k\,\mathcal{O}\left(\frac{d^2}{r}\right)\right) \quad (12.90)$$

Now that the dependences are clear, let's figure out what we need to keep. The first factor is independent of r' and multiplies the whole expression, so no issue there. In the second factor, there are power law dependences on r' and r, and there are no direct power-law dependences elsewhere, so we may look at this factor alone. In the limit $d \ll r$, we may keep the first term and discard the second and remaining terms because they falls off as higher powers of r.

Now, looking at the exponential, it has a completely imaginary argument, which causes the phase of the exponent to vary. The ratio of the second term to the first term of this phase factor is $(kd^2/r)/(kd) = d/r$, so it varies much less quickly than the first term in the limit $d \ll r$, so the first one will dominate the phase variation of the argument of the exponential in this limit. However, we also want to require that the second term in the argument is not large in an absolute sense: if it is of order unity, then it causes fast enough variations in the phase of the argument of the exponential that it will result in cancellations and make that term vanish. We thus require $k\,d^2/r \ll 2\,\pi$, or $d^2/\lambda \ll r$: this is the so-called "far-field" approximation, with d^2/λ being the "far-field distance."

Making these approximations and inserting into our expression for the vector potential:

$$
\boxed{
\begin{array}{l}
\text{radiation from arbitrary} \\
\quad \text{source distribution} \\
\quad \text{vector potential} \\
\quad d \ll r,\ d^2/\lambda \ll r
\end{array}
\qquad
\vec{A}(\vec{r}, t) = \frac{\mu_o}{4\pi} \frac{e^{i(kr - \omega t)}}{r} \int_{\mathcal{V}} d\tau'\, \vec{J}_0(\vec{r}')\, e^{-ikr'\,\hat{r}\cdot\hat{r}'}
}
\qquad (12.91)
$$

It may seem strange that we dropped the $\hat{r} \cdot \hat{r}'\, r'$ term from the expansion of $|\vec{r} - \vec{r}'|^{-1}$, the one that yielded the electric dipole and magnetic dipole potentials. The reason we don't need it here is because of the exponential: its variation over the scale of the source distribution (note that we did not make the approximation $k\,d = 2\pi\,d/\lambda \ll 1$!) prevents the cancellation that occurred for the monopole terms in the static case and necessitated keeping the $\hat{r} \cdot \hat{r}'\, r'$ term.

This expression makes very explicit the physical picture that the harmonic dependence of the source current drives a spherical outgoing wave: the $e^{i(kr - \omega t)}/r$ factor.

From the above, we can calculate the fields, time-averaged Poynting vector, radiation pattern, and total power radiated (for harmonic time dependence!):

$$
\boxed{\vec{B} = \vec{\nabla} \times \vec{A} \qquad \vec{E} = c^2\, \frac{i}{\omega}\, \vec{\nabla} \times \vec{B}}
$$

$$
\boxed{\left\langle \vec{S} \right\rangle = \frac{1}{2\,\mu_o}\, \mathcal{R}\left(\left\langle \vec{E}^* \times \vec{B} \right\rangle \right) \qquad \left\langle \frac{dP}{d\Omega} \right\rangle = r^2\, \hat{r} \cdot \left\langle \vec{S} \right\rangle \qquad \left\langle P \right\rangle = \int d\Omega\, \frac{dP}{d\Omega}}
$$

If we make the further approximation $d \ll \lambda$, then the exponential can be Taylor-expanded to obtain

radiation from
arbitrary
source distribution
vector potential
$d \ll r,\ d \ll \lambda$

$$\vec{A}(\vec{r},t) = \frac{\mu_o}{4\pi} \frac{e^{i(kr-\omega t)}}{r} \sum_{m=0}^{\infty} \frac{(-ik)^m}{m!} \int_{\mathcal{V}} d\tau'\ \vec{J}_0(\vec{r}')\ (r'\,\hat{r} \cdot \hat{r}')^m$$

(12.92)

We have dropped the $d^2/\lambda \ll r$ requirement because it is implied by the other two requirements. This is now the multipole expansion for radiation: successive terms probe finer and finer structure of the source distribution. We see next how these various terms give electric dipole, magnetic dipole, and electric quadrupole radiation.

The different multipole terms are not an analogue of the multipole terms we calculated for the electric scalar and magnetic vector potentials (Equations 3.226 and 5.105): in those cases, we had $[(r')^m/r^{m+1}]\,P_m(\hat{r} \cdot \hat{r}')$ in the integrand, here we have $[(k\,r')^m/r]\,(\hat{r} \cdot \hat{r}')^m$ in the integrand: same units, different formula. In the radiation case, all the multipole terms fall off as $1/r$, but they depend on different moments of the current distribution. In contrast, for the static potential multipoles, the higher terms fell off with increasing powers of r (and depended on higher moments of the charge and current distributions, but calculated in a different manner).

Lecture 36:

Radiation III:
Electric and Magnetic Dipole Radiation

Applications of Radiation I:
Classical Scattering

Date Revised: 2024/05/28 05:00
Date Given: 2024/05/28

Electric Dipole Radiation

We take the $m = 0$ term, which is

$$\vec{A}(\vec{r}, t) = \frac{\mu_o}{4\pi} \frac{e^{i(kr - \omega t)}}{r} \int_{\mathcal{V}} d\tau' \, \vec{J_0}(\vec{r}') \tag{12.93}$$

We saw this expression in our study of magnetic multipoles, Equations 5.103 and 5.104, where it vanished because of continuity and the steady-state assumption $\vec{\nabla} \cdot \vec{J} = -\partial \rho / \partial t = 0$. Here, we do not make that assumption, so

$$J_{0,i} \, e^{-i\omega t} = \vec{\nabla} \cdot \left(r_i \, \vec{J_0} \, e^{-i\omega t} \right) - r_i \, \vec{\nabla} \cdot \left(\vec{J_0} \, e^{-i\omega t} \right) = \vec{\nabla} \cdot \left(r_i \, \vec{J_0} \, e^{-i\omega t} \right) + r_i \frac{\partial}{\partial t} \left(\rho_0 \, e^{-i\omega t} \right)$$

$$= \vec{\nabla} \cdot \left(r_i \, \vec{J_0} \, e^{-i\omega t} \right) - i\omega \, r_i \, \rho_0 \, e^{-i\omega t} \tag{12.94}$$

When we do the integral, the first term vanishes: it can be turned into a surface integral of $r_i \vec{J}$ at the boundary of $\mathcal{V}$ and, since the sources are contained in $\mathcal{V}$, there can be no current flowing through the boundary of $\mathcal{V}$. That leaves the second term, so the vector potential becomes

$$\vec{A}(\vec{r}, t) = \frac{\mu_o}{4\pi} \frac{e^{i(kr - \omega t)}}{r} \int_{\mathcal{V}} d\tau' \, (-i\omega) \, \rho_0(\vec{r}') \, \vec{r}' = -i \frac{\mu_o}{4\pi} \frac{\omega \, \vec{p_0}}{r} \, e^{i(kr - \omega t)} \tag{12.95}$$

where we have used the definition of the dipole moment, Equation 3.229. We see that $\vec{A}$ is in the direction of the current (recall, $\vec{J_p} = d\vec{P}/dt$), and, in the static limit ($\omega \to 0$), the expression vanishes, both as we expect.

To calculate the magnetic field, we need to take the curl. The curl will act on both the $1/r$ and the phase factor in the exponent. But the action of the curl on $1/r$ will yield $1/r^2$, which we can drop if we also assume $k\,r \gg 1$ or $\lambda \ll r$. So we calculate

$$\vec{\nabla} \times \left(\vec{p}_0\, e^{i\,k\,r} \right) = -\vec{p}_0 \times \vec{\nabla} e^{i\,k\,r} = -\vec{p}_0 \times e^{i\,k\,r} \vec{\nabla}(i\,k\,r) = -\vec{p}_0 \times e^{i\,k\,r}\, i\,k\,\widehat{r} \quad (12.96)$$

Therefore

$$\vec{B}(\vec{r}, t) = \frac{\mu_o}{4\,\pi} \frac{\omega^2\,\widehat{r} \times \vec{p}_0}{c\,r} e^{i\,(k\,r - \omega\,t)} \qquad d \ll \lambda \ll r \qquad (12.97)$$

where the factor of ω/c came from k. We obtain the electric field from $\vec{E} = c^2(i/\omega)\,\vec{\nabla} \times \vec{B} = c^2\,(-k/\omega)\,\widehat{r} \times \vec{B} = -c\,\widehat{r} \times \vec{B}$, where we used Equation 12.85 and a similar technique for the calculation of the curl (involving more terms, but again one drops all terms of order $1/r^2$). This yields (using $\mu_o = 1/(c^2\,\epsilon_o)$)

$$\vec{E}(\vec{r}, t) = -\frac{1}{4\,\pi\,\epsilon_o} \frac{\omega^2\,\widehat{r} \times (\widehat{r} \times \vec{p}_0)}{c^2 r} e^{i\,(k\,r - \omega\,t)} \qquad d \ll \lambda \ll r \qquad (12.98)$$

The geometry of the fields is as follows. The magnetic field is normal to both the line-of-sight to the dipole at the origin ($\widehat{r}$) and to the dipole's direction $\vec{p}_0$. So, for a dipole along $\widehat{z}$, $\vec{B}$ oscillates in the $\widehat{\phi}$ direction. The electric field is normal to the line-of-sight and to $\vec{B}$, so it oscillates in the $\widehat{\theta}$ direction. Both $\widehat{\theta}$ and $\widehat{\phi}$ are in the plane normal to $\widehat{r}$. As expected, $\vec{E}$ is in the direction of the dipole, *i.e.*, of the acceleration of the charges involved.

Let's now generalize these using our discussion of Fourier Transforms. Let's assume an arbitrary time dependence of the charge density $\rho(\vec{r}, t)$ and the dipole moment $\vec{p}(t)$. The former has a Fourier decomposition (Equation 12.72), which we can use to Fourier decompose the latter:

$$\widetilde{\rho}(\vec{r}, f) = \int_{-\infty}^{\infty} dt\, \rho(\vec{r}, t)\, e^{i\,\omega\, t} \quad \Longleftrightarrow \quad \rho(\vec{r}, t) = \int_{-\infty}^{\infty} df\, \widetilde{\rho}(\vec{r}, f)\, e^{-i\,\omega\, t}$$

$$\Longrightarrow \quad \vec{p}(\vec{r}, t) = \int_{\mathcal{V}} d\tau'\, \rho(\vec{r}', t)\, \vec{r}' = \int_{\mathcal{V}} d\tau'\, \vec{r}' \int_{-\infty}^{\infty} df\, \widetilde{\rho}(\vec{r}', f)\, e^{-i\,\omega\, t} \qquad (12.99)$$

$$= \int_{-\infty}^{\infty} df\, \widetilde{\vec{p}}(f)\, e^{-i\,\omega\, t} \quad \text{with} \quad \widetilde{\vec{p}}(f) = \int_{\mathcal{V}} d\tau'\, \widetilde{\rho}(\vec{r}', f)\, \vec{r}' \qquad (12.100)$$

Now, the quantity $\widetilde{\vec{p}}(f)\, e^{-i\,\omega\, t}$ is just like $\vec{p}_0\, e^{-i\,\omega\, t}$ since both $\vec{p}_0$ and $\widetilde{\vec{p}}(f)$ have no spatial dependence. Therefore, our expressions for $\vec{A}$, $\vec{B}$, and $\vec{E}$ for harmonic dependence apply to it. Since Maxwell's Equations are linear, we can just sum up the contribution to the fields from each frequency f using Fourier Transforms.

That is, we may write

$$\tilde{\vec{A}}(\vec{r}, f)\, e^{-i\,\omega\,t} = -i\,\frac{\mu_o}{4\,\pi}\,\frac{\omega\,\tilde{\vec{p}}(f)}{r}\,e^{i\,(k\,r - \omega\,t)} \qquad (12.101)$$

where we have factored the $e^{-i\,\omega\,t}$ dependence out of the vector potential and added
the f argument to indicate that each $\tilde{\vec{p}}(f)$ generates its own $\tilde{\vec{A}}(\vec{r}, f)$. Now, sum up
over frequency components using the Fourier Transform:

$$\vec{A}(\vec{r}, t) = \int_{-\infty}^{\infty} df\, \tilde{\vec{A}}(\vec{r}, f)\, e^{-i\,\omega\,t} = \int_{-\infty}^{\infty} df\, \left[-i\,\frac{\mu_o}{4\,\pi}\,\frac{\omega\,\tilde{\vec{p}}(f)}{r}\,e^{i\,k\,r} \right] e^{-i\,\omega\,t} \qquad (12.102)$$

Next, recall how time derivatives are related to factors of $-i\,\omega$ in Fourier Transforms,
Equation 12.81, and also recall Equation 12.84, which tells us that
$e^{i\,(k\,r - \omega\,t)} = e^{-i\,\omega\,t_r}$ (recall, $r \gg r'$ is assumed!). These let us rewrite the above as

$$\vec{A}(\vec{r}, t) = \frac{\mu_o}{4\,\pi}\,\frac{1}{r} \int_{-\infty}^{\infty} df\, \left[-i\,\omega\,\tilde{\vec{p}}(f) \right] e^{-i\,\omega\,t_r} = \frac{\mu_o}{4\,\pi}\,\frac{\dot{\vec{p}}(t_r)}{r} \qquad (12.103)$$

where $\dot{\vec{p}}$ indicates the derivative of $\vec{p}$ with respect to its argument. Note that we were
able to trivially convert from $-i\,\omega$ to d/dt even though the time is evaluated at t_r
because the approximation $r \gg r'$ simplifies the t-dependence in the argument of the
complex exponential (i.e., $k|\vec{r} - \vec{r}'| \approx k\,r$).

One can do similar Fourier analyses for $\vec{B}$ and $\vec{E}$. One could instead take the curl and time derivative of the above expression, using again the $r \gg r'$ approximation to make these derivatives easy to take. The Fourier analysis path seems more physical because it makes a bit clearer how the $\hat{r} \times$ factors below arise.

The result is the full generalization of our harmonic time dependence expressions:

$$
\boxed{
\begin{array}{c}
\text{electric dipole radiation} \\
\text{vector potential} \\
d \ll \lambda \ll r
\end{array}
\qquad
\vec{A}(\vec{r}, t) = \frac{\mu_o}{4\pi} \frac{\dot{\vec{p}}(t_r)}{r}
}
\tag{12.104}
$$

$$
\boxed{
\begin{array}{c}
\text{electric dipole radiation} \\
\text{magnetic field} \\
d \ll \lambda \ll r
\end{array}
\qquad
\vec{B}(\vec{r}, t) = -\frac{\mu_o}{4\pi} \frac{\hat{r} \times \ddot{\vec{p}}(t_r)}{c\, r}
}
\tag{12.105}
$$

$$
\boxed{
\begin{array}{c}
\text{electric dipole radiation} \\
\text{electric field} \\
d \ll \lambda \ll r
\end{array}
\qquad
\vec{E}(\vec{r}, t) = \frac{1}{4\pi\epsilon_o} \frac{\hat{r} \times \left(\hat{r} \times \ddot{\vec{p}}(t_r) \right)}{c^2 r}
}
\tag{12.106}
$$

Note that the dipole moment derivatives are evaluated at t_r! These equations match Equations 11.54, 11.56, and 11.57 of Griffiths.

Let's define a coordinate system with $\ddot{\vec{p}} \propto \widehat{z}$ and θ the spherical coordinate polar angle, so then

$$\widehat{r} \times \widehat{z} = -\widehat{\phi} \sin \theta \qquad \widehat{r} \times (\widehat{r} \times \widehat{z}) = \widehat{r} \times \left(-\widehat{\phi} \sin \theta\right) = \widehat{\theta} \sin \theta \qquad (12.107)$$

Thus, for an arbitrary time dependence, we obtain (using $dP/d\Omega = r^2 \, \widehat{r} \cdot \vec{S}$, no complex conjugations or real parts necessary because these formulae do not assume harmonic time dependence, and $\vec{S} \propto \vec{E} \times \vec{B} \propto \widehat{\theta} \times -\widehat{\phi} = \widehat{r}$):

$$\boxed{\begin{array}{ccc} \text{electric dipole} \\ \text{radiation pattern} & \dfrac{dP}{d\Omega} = \dfrac{\mu_o}{16\,\pi^2} \, \dfrac{\ddot{p}^2 \sin^2 \theta}{c} & P = \dfrac{\mu_o}{6\,\pi} \, \dfrac{\ddot{p}^2}{c} \\ d \ll \lambda \ll r \end{array}} \qquad (12.108)$$

Notice the similarity to the Larmor formula! Had we calculated the Larmor formula for two point charges forming a dipole, going through the fields to see how the factors of $q\,d$ add coherently, we would have obtained the above result. (The dipole approximation automatically incorporates $\beta \to 0$.) If we assume harmonic time dependence and time-average (also now taking the necessary complex conjugations and real parts if we use complex notation), we obtain:

$$\boxed{\begin{array}{ccc} \text{electric dipole} \\ \text{radiation pattern} \\ \text{for harmonic} & \left\langle \dfrac{dP}{d\Omega} \right\rangle = \dfrac{\mu_o}{32\,\pi^2} \, \dfrac{p_0^2 \, \omega^4 \sin^2 \theta}{c} & \langle P \rangle = \dfrac{\mu_o}{12\,\pi} \, \dfrac{p_0^2 \, \omega^4}{c} \\ \text{time dependence} \\ d \ll \lambda \ll r \end{array}} \qquad (12.109)$$

Magnetic Dipole and Electric Quadrupole Radiation

Next, let's look at the $m = 1$ term:

$$\vec{A}(\vec{r}, t) = \frac{\mu_o}{4\pi} \frac{e^{i(k\, r - \omega\, t)}}{r} (-i\, k) \int_{\mathcal{V}} d\tau'\, \vec{J}_0(\vec{r}')\, \widehat{r} \cdot \widehat{r}'\, r' \qquad (12.110)$$

This term is a factor of order $k\, d = 2\pi\, d/\lambda$ smaller than the electric dipole term. We also calculated this term in the process of deriving the vector potential due to a magnetic dipole, starting at Equation 5.106. Again, we now have to allow $\vec{\nabla} \cdot \vec{J} = -\partial\rho/\partial t \neq 0$. This yields (many steps left out!)

$$\vec{J}_0(\vec{r}')\, \widehat{r} \cdot \widehat{r}'\, r' = \frac{1}{2} \left[\left(\vec{r}' \times \vec{J}_0 \right) \times \widehat{r} \right] + \frac{1}{2}\, r' \left[\left(\widehat{r} \cdot \widehat{r}' \right) \vec{J}_0 + \left(\widehat{r} \cdot \vec{J}_0 \right) \widehat{r}' \right] \qquad (12.111)$$

The second two terms involve quantities similar to those we dealt with for the electric dipole term. We can integrate these terms by parts to obtain expressions involving $\vec{\nabla} \cdot \vec{J}_0$ (and therefore bringing in one more power of r' to cancel the dimensions of $\vec{\nabla}$), which can be evaluated using continuity, $\vec{\nabla} \cdot \vec{J}_0 e^{-i\,\omega\, t} = -\partial \left(\rho_0 e^{-i\,\omega\, t} \right)/\partial t = i\,\omega\,\rho_0 e^{-i\,\omega\, t}$. They therefore result in terms containing $i\,\omega\,\rho$ and two powers of r'. These are electric quadrupole terms. They are of the same order of magnitude as the terms we will keep, but they are complicated and so we will drop them for this study.

The first term recalls the definition of the magnetic dipole moment (Equation 5.119)

$$\vec{m}_0 = \int_{\mathcal{V}} d\tau' \, \frac{1}{2} \, \vec{r}' \times \vec{J}_0(\vec{r}') \tag{12.112}$$

Using this definition, we may rewrite the vector potential as

$$\vec{A}(\vec{r}, t) = i \, \frac{\mu_o}{4\pi} \, \frac{\omega \, \hat{r} \times \vec{m}_0}{c \, r} \, e^{i(k\,r - \omega\,t)} \qquad d \ll \lambda \ll r \tag{12.113}$$

This expression is smaller than the analogous expression for electric dipole radiation (Equation 12.95) by a factor $(m_0/c)/p$, which we can see is $k\,d$ as expected:

$$\frac{1}{p} \, \frac{m_0}{c} = \frac{1}{d^3 \rho \, d} \, \frac{d^3 J \, d}{c} = \frac{1}{\rho \, d} \, \frac{\omega \, d \, \rho \, d}{c} = k \, d \tag{12.114}$$

where $J \propto \omega \, d \, \rho$ follows from our evaluation of $\vec{J}$ in the electric dipole radiation case. The fields are easily derived using the same procedure as for the electric dipole term to evaluate the curls, yielding

$$\vec{B}(\vec{r}, t) = -\frac{\mu_o}{4\pi} \, \frac{\omega^2 \, \hat{r} \times (\hat{r} \times \vec{m}_0)}{c^2 r} \, e^{i(k\,r - \omega\,t)} \qquad d \ll \lambda \ll r \tag{12.115}$$

For the electric field, we can use the harmonic relation $\vec{E} = -c\,\hat{r} \times \vec{B}$ (from Equation 12.85, $\vec{E} = c^2(i/\omega)\vec{\nabla} \times \vec{B}$) to find

$$\vec{E}(\vec{r}, t) = \frac{1}{4\pi\epsilon_o} \frac{\omega^2\,\hat{r} \times (\hat{r} \times (\hat{r} \times \vec{m}_0))}{c^3 r}\, e^{i(k\,r - \omega\,t)} \tag{12.116}$$

In the electric dipole case, we had $\vec{B} \propto (1/c)\,\hat{r} \times \vec{E} \propto \hat{r} \times (\hat{r} \times (\hat{r} \times \vec{p}))$, but we also derived $\vec{B} \propto \hat{r} \times \vec{p}$. We thus suspect that the quadruple vector product in $\vec{E}$ for magnetic dipole radiation reduces to $\hat{r} \times \vec{m}_0$. Let's prove that explicitly using the $BAC - CAB$ rule for an arbitrary vector $\vec{a}$:

$$\hat{r} \times [\hat{r} \times (\hat{r} \times \vec{a})] = \hat{r}\,[\hat{r} \cdot (\hat{r} \times \vec{a})] - (\hat{r} \times \vec{a})\,(\hat{r} \cdot \hat{r}) = -\hat{r} \times \vec{a} \tag{12.117}$$

Thus, we may rewrite $\vec{E}$ in the magnetic dipole case as

$$\vec{E}(\vec{r}, t) = -\frac{1}{4\pi\epsilon_o} \frac{\omega^2\,\hat{r} \times \vec{m}_0}{c^3 r}\, e^{i(k\,r - \omega\,t)} \qquad d \ll \lambda \ll r \tag{12.118}$$

Griffiths derives the magnetic dipole fields for the special case $\vec{m}_0 \propto \hat{z}$. We saw for the ideal electric dipole that $\hat{r} \times (\hat{r} \times \hat{z}) = \hat{\theta}\,\sin\theta$ and $\hat{r} \times (\hat{r} \times (\hat{r} \times \hat{z})) = \hat{\phi}\,\sin\theta$. Applying that here yields $\vec{B} \propto -\hat{\theta}\,\sin\theta$ and $\vec{E} \propto \hat{\phi}\,\sin\theta$, thus matching Griffiths Equations 11.36 and 11.37.

As we did with the electric dipole radiation field, we can generalize these expressions using Fourier Transforms, yielding

$$
\boxed{
\begin{array}{c}
\text{magnetic dipole radiation} \\
\text{vector potential} \\
d \ll \lambda \ll r
\end{array}
\qquad
\vec{A}(\vec{r}, t) = -i\,\frac{\mu_o}{4\,\pi}\,\frac{\hat{r} \times \dot{\vec{m}}(t_r)}{c\,r}
}
\tag{12.119}
$$

$$
\boxed{
\begin{array}{c}
\text{magnetic dipole radiation} \\
\text{magnetic field} \\
d \ll \lambda \ll r
\end{array}
\qquad
\vec{B}(\vec{r}, t) = \frac{\mu_o}{4\,\pi}\,\frac{\hat{r} \times \left(\hat{r} \times \ddot{\vec{m}}(t_r) \right)}{c^2 r}
}
\tag{12.120}
$$

$$
\boxed{
\begin{array}{c}
\text{magnetic dipole radiation} \\
\text{electric field} \\
d \ll \lambda \ll r
\end{array}
\qquad
\vec{E}(\vec{r}, t) = \frac{1}{4\,\pi\,\epsilon_o}\,\frac{\hat{r} \times \ddot{\vec{m}}(t_r)}{c^3 r}
}
\tag{12.121}
$$

where again all the dipole moments and derivatives thereof are evaluated at the retarded time. The vector potential is perpendicular to both the vector rate of change of the dipole and the position vector, the electric field is perpendicular to the second derivative of the dipole and the position vector, and the magnetic field is perpendicular to the second derivative of the dipole and the electric field. There is no analogue of these generic results in Griffiths, who only considers the special case of $\vec{m} \propto \hat{z}$ and harmonic time dependence, as we derived above. For the special case of a dipole $\vec{m} \propto \hat{z}$ that also has $\dot{\vec{m}} \propto \hat{z}$ and $\ddot{\vec{m}} \propto \hat{z}$, the vector potential and electric field are along $\hat{\phi}$ and the magnetic field is along $\hat{\theta}$. Note how $\vec{A}$ and $\vec{E}$ have the same direction as the infinitesimal current and the acceleration of the charges, respectively.

To determine the power pattern, we can use the calculations we did for the electric dipole case because the dependences are similar: up to signs and normalization, $\vec{B}$ for the electric case matches $\vec{E}$ for the magnetic case and $\vec{E}$ for the electric case matches $\vec{B}$ for the magnetic case. In both cases, $\vec{E} \times \vec{B} \propto +\hat{r}$ as is necessary for outgoing radiation, so we can dispense with the signs. The normalization is set by the replacement $p \rightarrow m/c$. So, we have (again, $dP/d\Omega = r^2\,\hat{r} \cdot \vec{S}$, all fields real)

> magnetic dipole
> radiation pattern
> $d \ll \lambda \ll r$
> $$\frac{dP}{d\Omega} = \frac{\mu_o}{16\,\pi^2}\,\frac{\dddot{m}^2 \sin^2\theta}{c^3} \qquad P = \frac{\mu_o}{6\,\pi}\,\frac{\dddot{m}^2}{c^3}$$
> (12.122)

Again, we see an analogue to the Larmor formula. If we assume harmonic time dependence and time-average (taking complex conjugations and real parts as necessary for complex notation), we obtain:

> magnetic dipole
> radiation pattern
> for harmonic
> time dependence
> $d \ll \lambda \ll r$
> $$\left\langle \frac{dP}{d\Omega} \right\rangle = \frac{\mu_o}{32\,\pi^2}\,\frac{m_0^2\,\omega^4 \sin^2\theta}{c^3} \qquad \langle P \rangle = \frac{\mu_o}{12\,\pi}\,\frac{m_0^2\,\omega^4}{c^3}$$
> (12.123)

Given the replacement $p \rightarrow m/c$, the magnetic dipole radiation power is down by a factor of $[(m/c)/p]^2 = (k\,d)^2 = (2\,\pi\,d/\lambda)^2$ relative to electric dipole radiation.

Postscript

Griffiths is unfortunately lacking in examples on the topic of radiation. The reason for this is that, with the above formalism in hand, the basic examples consist largely of calculating electric and magnetic dipole moments and then plugging into the above formulae. Rather than do a lot of this plug-and-chug, it makes sense to move on to classical scattering theory and antennas, which provide more meaty and meaningful applications of the above.

If it seems like the above discussion was a bit tortured — the term-by-term manipulation of Equation 12.92, the by-hand splitting of the $m = 1$ term into magnetic dipole and electric quadrupole radiation, the reappearance of the same electric and magnetic field dependences — that impression is correct. There is a more unified way of treating the multipole expansion, which is given in Jackson Chapter 16, but which we do not have time to present.

Section 13
Applications of Radiation

Classical Scattering Theory

What is Scattering?

When we considered wave propagation in a medium having a permittivity and a permeability, we assumed the medium was completely uniform, with polarization and magnetization volume density that reflect none of the atomic nature of the material or any variation in polarization or magnetization density with position. This also meant the medium was infinite in extent, with no boundaries except planar boundaries infinite in transverse extent. When we calculated the polarization density using the simple dispersion model, we again assumed that this model gave us a smooth polarization volume density.

Scattering is, by contrast, an effect that happens when an EM wave propagates through a *nonuniform* medium. Now the polarization and magnetization volume density can be slightly or even substantially position dependent. At one extreme, we have a single polarizable particle in vacuum. At the other extreme, we have a gas or a dense medium with a fluctuating density. In all cases, we treat each nonuniformity as its own dipole that radiates in response to an EM wave. When more than one scatterer is present, we assume they are randomly positioned relative to one another so there is no coherence between the radiated fields of individual scatterers.

Scattering to Large Distances of Polarized Light by a Small Object, $d \ll \lambda \ll r$

Consider a linearly polarized EM wave

$$\vec{E}(\vec{r}, t) = \vec{E}_0 \, e^{i(\vec{k} \cdot \vec{r} - \omega t)} \tag{13.1}$$

incident on a single particle of size $d \ll \lambda$ ($k \, d \ll 2 \pi$) and of frequency-dependent polarizability $\alpha(\omega)$. Then the dipole moment of the particle is (in complex notation)

$$\vec{p} = \alpha(\omega) \, \vec{E} = \alpha(\omega) \, \vec{E}_0 \, e^{-i \omega t} \tag{13.2}$$

(Recall we derived models for $\alpha(\omega)$ in Section 5.) This is an oscillating, accelerating electric dipole with $\vec{p}$, $\dot{\vec{p}}$, and $\ddot{\vec{p}}$ in the same direction as the incoming polarization vector. We can calculate its radiated power pattern using Equation 12.109, which adds the assumption $d \ll \lambda \ll r$ (scatterer much smaller than wavelength, which is smaller than the distance to field point):

$$\left\langle \frac{dP}{d\Omega} \right\rangle = \frac{\mu_o}{32 \, \pi^2} \frac{\alpha^2 \, |\vec{E}_0|^2 \, \omega^4 \sin^2 \theta}{c} \qquad \langle P \rangle = \frac{\mu_o}{12 \, \pi} \frac{\alpha^2 \, |\vec{E}_0|^2 \, \omega^4}{c} \tag{13.3}$$

where θ is the polar angle relative to the polarization vector of the incoming wave. Note that the emitted power pattern has no dependence on ϕ, the azimuthal angle around the polarization vector, implying the scattered light does not care about the incoming wave direction, only its polarization.

We define the particle's *differential* and *total scattering cross section* as

$$\frac{d\sigma}{d\Omega} = \frac{\left\langle \frac{dP}{d\Omega} \right\rangle}{\text{incident power flux}} \qquad \sigma = \frac{\langle P \rangle}{\text{incident power flux}} \qquad (13.4)$$

With the incident power flux $I = \left\langle |\vec{S}| \right\rangle = c\,\epsilon_o\,|\vec{E}_0|^2/2$, we obtain

$$\boxed{\begin{array}{l} \text{linearly polarized} \\ \text{cross section} \\ d \ll \lambda \ll r \end{array} \qquad \frac{d\sigma_{pol}}{d\Omega} = \left(\frac{\alpha}{4\,\pi\,\epsilon_o}\right)^2 \frac{\omega^4}{c^4} \sin^2\theta \quad \sigma_{pol} = \frac{8\,\pi}{3}\left(\frac{\alpha}{4\,\pi\,\epsilon_o}\right)^2 \frac{\omega^4}{c^4}} \qquad (13.5)$$

Because $\alpha/4\,\pi\,\epsilon_o$ carries units of volume, σ carries units of area and $d\sigma/d\Omega$ units of area/solid angle. The reason for this is that we assume an incoming plane wave of infinite transverse extent, so we cannot calculate the fraction of its total power scattered: that would vanish because the power is infinite. Rather, we calculate the power scattered given an incoming power per unit area, so we have to multiply by an area to get the right output units.

Scattering to Large Distances of Unpolarized Light by a Small Object, $d \ll \lambda \ll r$

It is in general useful to calculate the corresponding results for unpolarized incoming light: light with equal amounts of the two complementary linear polarizations and no specific phase relationship between those two polarizations. (A specific phase relationship would just result in a different polarization.) This excites oscillating dipole moments simultaneously in the two directions corresponding to these two polarizations. Since the two polarization vectors are orthogonal, we need to write their scattering cross-sections in the same coordinate system in order to sum their contributions. So far, we have written the scattering cross-section in terms of angles relative to the polarization vector (the oscillating dipole moment). Instead, let's write it in terms of angles relative to the incoming wave's propagation vector $\widehat{k}$ and the viewing direction (the outgoing scattered wave direction) $\widehat{r}$.

Consider a coordinate system whose z-axis is the incident wave direction $\widehat{k}$ and whose xz plane (the *scattering plane*) is defined by $\widehat{k}$ and the scattered direction (viewing direction) $\widehat{r}_{sc}$. Let θ_{sc} be the polar angle of $\widehat{r}_{sc}$ relative to $\widehat{k}$, in the scattering plane. The incoming light has components polarized parallel to and perpendicular to the scattering plane. Since we will consider unpolarized light, we know that the result will have no dependence on the azimuthal angle around $\widehat{k}$, which is why we can choose $\widehat{r}_{sc}$ in the xz plane.
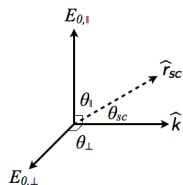
For the polarization component perpendicular to the scattering plane, the scattering plane makes an angle $\pi/2$ with the polarization vector. The angle θ_{sc} is the ϕ angle in a coordinate system whose z-axis is this polarization vector. Thus, for any θ_{sc}, the angle θ in the differential cross section formula is $\theta_\perp = \pi/2$ in the diagram. Assuming the incoming light is unpolarized implies half of the incoming power is in this polarization component, yielding

$$\frac{d\sigma_\perp}{d\Omega} = \frac{1}{2} \left. \frac{d\sigma}{d\Omega} \right|_{\theta = \pi/2} = \frac{1}{2} \left(\frac{\alpha}{4\pi\epsilon_o} \right)^2 \frac{\omega^4}{c^4} \tag{13.6}$$

For the polarization component parallel to the scattering plane, the scattering plane lies in the xz-plane of the coordinate system whose z-axis is this polarization vector and whose $\pm x$-axis is the incident direction $\hat{k}$. Thus, the polar angle in the scattering coordinate system and the polar angle in the incident polarization coordinate system are complementary, so $\sin\theta_{||} = \cos\theta_{sc}$ and therefore

$$\frac{d\sigma_{||}}{d\Omega} = \frac{1}{2} \left. \frac{d\sigma}{d\Omega} \right|_{\theta = \pi/2 - \theta_{sc}} = \frac{1}{2} \left(\frac{\alpha}{4\pi\epsilon_o} \right)^2 \frac{\omega^4}{c^4} \cos^2\theta_{sc} \tag{13.7}$$

We want to combine these cross sections to get the total cross section. Do we need to worry about constructive or destructive addition of the scattered waves in the two polarizations? No, a fact we can see both conceptually and mechanically.

We can see this conceptually as follows. We first recognize that the two incoming polarizations excite dipoles parallel to the incoming field directions and therefore perpendicular to each other:

$$\ddot{\vec{p}}_{||} \propto \vec{E}_{in,||} \qquad \ddot{\vec{p}}_{\perp} \propto \vec{E}_{in,\perp} \qquad \vec{E}_{in,||} \perp \vec{E}_{in,\perp} \implies \ddot{\vec{p}}_{||} \perp \ddot{\vec{p}}_{\perp} \qquad (13.8)$$

Then, we recall that the electric field of the radiated wave is in the direction of the projection of $\ddot{\vec{p}}$ transverse to the line-of-sight direction ($\widehat{r}_{sc}$). Therefore, the radiated electric fields for the two incoming polarizations must be perpendicular to each other (and to the line of sight). We know that fields in orthogonal polarizations do not interfere with each other and their powers just add, hence we just add the cross sections, accounting for the fact that half the power is in each polarization.

We can see this mechanically by calculating the Poynting vector of the summed wave:

$$\mu_o \, \vec{S}_{tot} = \left(\vec{E}_{rad,||} + \vec{E}_{rad,\perp} \right)^* \times \left(\vec{B}_{rad,||} + \vec{B}_{rad,\perp} \right)^* \tag{13.9}$$

The fields of the scattered (radiated) wave satisfy (Equations 12.105 and 12.106)

$$\vec{E}_{rad} \propto \hat{r} \times \left(\hat{r} \times \ddot{\vec{p}} \right) \propto \hat{r} \left(\hat{r} \cdot \vec{E}_{in} \right) - \vec{E}_{in} \qquad \vec{B}_{rad} \propto -\hat{r} \times \ddot{\vec{p}} \propto \hat{r} \times \vec{E}_{in} \tag{13.10}$$

Therefore, the cross terms in the Poynting vector are (applying the $BAC - CAB$ rule many times):

$$\vec{E}_{rad,\perp}^* \times \vec{B}_{rad,||} \propto \left[\hat{r} \left(\hat{r} \cdot \vec{E}_{in,||} \right) - \vec{E}_{in,||} \right] \times \left[\hat{r} \times \vec{E}_{in,\perp} \right] \tag{13.11}$$

$$= \left(\hat{r} \cdot \vec{E}_{in,||} \right) \left[\hat{r} \times \left(\hat{r} \times \vec{E}_{in,\perp} \right) \right] - \vec{E}_{in,||} \times \left(\hat{r} \times \vec{E}_{in,\perp} \right) \tag{13.12}$$

$$= \left(\hat{r} \cdot \vec{E}_{in,||} \right) \left[\hat{r} \left(\hat{r} \cdot \vec{E}_{in,\perp} \right) - \vec{E}_{in,\perp} \right] \tag{13.13}$$

$$- \left[\hat{r} \left(\vec{E}_{in,||} \cdot \vec{E}_{in,\perp} \right) - \vec{E}_{in,\perp} \left(\hat{r} \cdot \vec{E}_{in,||} \right) \right] \tag{13.14}$$

The third term vanishes because the polarizations are perpendicular to one another. The second term cancels the fourth term. The first term vanishes because, given the geometry, $\hat{r}$ is in the plane perpendicular to $\vec{E}_{in,\perp}$. So the whole cross term vanishes.

The other cross term is obtained by exchanging $||$ and $\perp$ and so it also vanishes.

Since the power radiated in the two polarizations just sum, we may add the two cross sections together (incoherent sum). We also integrate over all angles to get the total scattering cross section:

$$\begin{array}{ll} \text{unpolarized} \\ \text{cross} \\ \text{section} \\ d \ll \lambda \ll r \end{array} \quad \frac{d\sigma_{unpol}}{d\Omega} = \left(\frac{\alpha}{4\,\pi\,\epsilon_o} \right)^2 \frac{\omega^4}{c^4} \frac{1 + \cos^2 \theta_{sc}}{2} \quad \sigma_{unpol} = \frac{8\,\pi}{3} \left(\frac{\alpha}{4\,\pi\,\epsilon_o} \right)^2 \frac{\omega^4}{c^4} \quad (13.15)$$

The total scattering cross section is the same as the polarized case because the polarized cross section has no dependence on polarization angle.

Application: Rayleigh Scattering, or Why the Sky is Blue

Recall our discussion of the permittivity and index of refraction of dispersive materials and how it depends on frequency. In particular, it is fairly independent of frequency in regions well away from any known bound states. For the atmosphere, the relevant states are either vibrational and rotational states of molecules or electronic states of the constituent atoms. The atmosphere is dominated by N_2 and O_2, which are symmetric molecules with no net dipole moment; hence, in our approximation, there is no coupling of light to their rotational and vibrational states. The electronic states of N, O, and the third major component, Ar, are above optical frequencies. Therefore, the permittivity becomes constant over the optical part of the spectrum. We may infer from this that the polarizability α becomes constant. (Given the above, it actually vanishes for N_2 and O_2 and the scattering is entirely by the individual atoms, not molecules.)

Thus, the frequency dependence of the scattering cross section comes entirely from the ω^4 term. This strong dependence on frequency implies that the blue portion of the solar spectrum scatters much more than the red portion: seven times comparing 400 nm to 650 nm.

The above strong dependence leads to a number of interesting phenomena:

▶ When we look away from the sun, the light we are seeing is primarily scattered light, which is dominantly blue.

▶ When we look at the sun at sunrise or sunset, when the path length through the atmosphere is large, the blue light has been scattered out of the line-of-sight and we see a red sunrise or sunset.

▶ Because of the angular dependence in in Equation 13.15, when we look in a direction normal to the sun's rays (e.g., straight up when the sun is low in the sky, but not so low that all the blue light has been lost), we can only see scattered light in the polarization perpendicular to the scattering plane. (The polarization of sunlight parallel to the scattering plane has its polarization vector pointed in the same direction as our line-of-sight $\hat{r}_{sc}$, so we get no scattered light in that polarization.) So, if we look in this direction with a polarizer, we can make the sky appear bluer by selecting the scattered polarization or less blue by selecting the unscattered polarization: the light in the unscattered polarization consists only of light reflected from our surroundings (which is not preferentially blue).

Application: Thomson Scattering

A simple situation in which to apply our theory of scattering quantitatively is the case of Thomson scattering, which is scattering of an electromagnetic wave off of a free electron. Recall that we calculated the relation between a bound, damped electron's dipole moment and the incident electric field to be (Equation 9.191)

$$\widetilde{p}(t) = e\,\widetilde{x}(t) = \frac{q^2/m}{\omega_0^2 - \omega^2 - i\,\gamma\,\omega}\,\widetilde{E}_0\,e^{-i\,\omega\,t} \tag{13.16}$$

If we take the limit $\omega \gg \omega_0$ and $\omega \gg \sqrt{\gamma\,\omega}$ (the plasma limit we defined in Section 5), we obtain

$$\alpha_{free} = \frac{\widetilde{p}_{free}(t)}{\widetilde{E}_0\,e^{-i\,\omega\,t}} = -\frac{e^2}{m\,\omega^2} \tag{13.17}$$

In this limit, the polarizability is determined entirely by the electron's inertia.

Plugging in to our total scattering cross section equation, Equation 13.15, and rearranging (canceling powers of ω, moving c^2 around), we obtain the differential and total scattering cross section of a free electron:

$$
\boxed{
\begin{array}{l}
\text{free-electron} \\
\text{(Thomson)} \\
\text{cross section}
\end{array}
\quad
\frac{d\sigma_{Thomson}}{d\Omega} = \left(\frac{e^2}{4\,\pi\,\epsilon_o\,m_e\,c^2} \right)^2 \frac{1 + \cos^2\theta_{sc}}{2}
\quad
\sigma_{Thomson} = \frac{8\,\pi}{3} \left(\frac{e^2}{4\,\pi\,\epsilon_o\,m_e\,c^2} \right)^2
}
$$

$$(13.18)$$

The total cross section is called the *Thomson cross section* and is ubiquitous because it sets the scale of scattering of EM waves off of electrons, even when quantum mechanical effects are taken into account. Notice that it is independent of frequency because the ω^{-2} dependence of α cancels the ω^4 dependence of dipole radiation. Though, recall we have assumed $(p/e) \sim d \ll \lambda \ll r$. This assumption fails at high enough frequency.

The quantity in parentheses is called the classical electron radius

$$
r_{classical} = \frac{e^2}{4\,\pi\,\epsilon_o\,m_e\,c^2}
\tag{13.19}
$$

because it is the radius one obtains by equating the rest mass energy of the electron, $m_e\,c^2$, to the Coulomb energy required to assemble of sphere of uniform charge density of total charge e and radius $r_{classical}$. On the one hand, it is remarkable that this radius gives $\sigma_{Thomson}$ up to a factor of $8/3$; on the other hand, it is the only length scale available, so perhaps we should not be surprised.

Lecture 37:

Applications of Radiation II:
Classical Scattering
Antennas

Date Revised: 2024/05/30 07:00
Date Given: 2024/05/30

Application: Small ($d \ll \lambda$) Dielectric Sphere Scattering

Another simple application is to consider scattering off of a dielectric sphere of radius $d \ll \lambda$ and permittivity ϵ. The polarization of such a sphere was calculated in Griffiths Example 4.7, which we quoted in Equation 4.71, with:

$$\vec{p} = 4\pi\,\epsilon_o\,d^3\,\frac{\epsilon - \epsilon_o}{\epsilon + 2\,\epsilon_o}\,\vec{E}_0 \tag{13.20}$$

α is easily read off the above expression, yielding

> dielectric sphere scattering cross section
> $$d \ll \lambda \ll r$$

$$\frac{d\sigma_{diel.\,sph.}}{d\Omega} = \left(\frac{\epsilon - \epsilon_o}{\epsilon + 2\,\epsilon_o}\right)^2 (k\,d)^4\,d^2\,\frac{1 + \cos^2\theta_{sc}}{2} \tag{13.21}$$

$$\sigma_{diel.\,sph.} = \frac{8\pi}{3}\left(\frac{\epsilon - \epsilon_o}{\epsilon + 2\,\epsilon_o}\right)^2 (k\,d)^4\,(\pi\,d^2) \tag{13.22}$$

where $(k\,d)^4$ comes from $(\omega/c)^4\,d^4$. This is also termed "Rayleigh scattering" because of the similar frequency dependence. ("Mie scattering," important for dust, pollen, smoke, and water droplets in the atmosphere, is the case $d \not\ll \lambda$ and is not treated here because our multipole expansion fails.)

Application: Scattering in a Gas

For a gas, because the atoms are in random positions, we may apply our scattering theory to the medium as a whole. Let's determine the rate at which incoming power is attenuated due to scattering to off-axis angles. The decay rate of the transmitted power per unit area F is given by

$$F(z) = F(0)\, e^{-\gamma z} \qquad \gamma = n\,\sigma \qquad (13.23)$$

where n is the density of atoms or molecules and σ is the scattering cross section per atom or molecule. To find σ, we need the polarizability, α. Recalling the definition of the dielectric constant ϵ_r, we have

$$\vec{P} = n\,\vec{p} = n\,\alpha\,\vec{E} \quad \text{and} \quad \vec{P} = \chi_e\,\epsilon_o\,\vec{E} = (\epsilon - \epsilon_o)\,\vec{E} \implies \alpha = \frac{\epsilon - \epsilon_o}{n} \qquad (13.24)$$

Inserting this into our expression for σ_{unpol}, Equation 13.15, we obtain

$$\gamma_{dilute} = n\,\sigma = n\,\frac{8\,\pi}{3}\left(\frac{\epsilon - \epsilon_o}{4\,\pi\,\epsilon_o\,n}\right)^2 \frac{\omega^4}{c^4} = \frac{8\,\pi^3}{3}\left(\frac{\epsilon - \epsilon_o}{\epsilon_o}\right)^2 \frac{1}{n\,\lambda^4} \qquad (13.25)$$

This attenuation constant is applicable to the scattering of sunlight because α is frequency-independent, again explaining why the sun becomes redder as it sets: the redder wavelengths have a substantially smaller attenuation constant (larger attenuation length) than the bluer wavelengths.

Application: Scattering in a Dense Medium

In a dense medium (*e.g.*, a liquid), individual scatterers can no longer be treated as totally independent. Instead, we treat the medium as consisting of cells of size d such that $d \ll \lambda$ while also $d \gg \xi$ where ξ is the correlation length of fluctuations in the medium's density. The number of scatterers per unit volume is $1/d^3$ because that is the number of cells per unit volume. The "polarizability per scatterer", which is α in the case of a single scatterer, is given by $\alpha\,\delta N_d$ where δN_d is the rms fluctuation in the number of scatterers in cells of size d. Thus, the attenuation length becomes (using $\alpha = (\epsilon - \epsilon_o)/n$ from the previous page)

$$\gamma_{dense} = (\text{\# of cells per unit volume}) \times (\text{cross section per cell}) \qquad (13.26)$$

$$= \frac{1}{d^3}\,\frac{8\,\pi}{3}\left(\frac{\alpha\,\delta N_d}{4\,\pi\,\epsilon_o}\right)^2 \frac{\omega^4}{c^4} = \frac{8\,\pi^3}{3}\left(\frac{\epsilon - \epsilon_o}{\epsilon_o}\right)^2 \frac{1}{n\,\lambda^4}\,\frac{\langle \delta N_d^2 \rangle}{n\,d^3} \qquad (13.27)$$

In an ideal gas (not what we are dealing with), we should make the cell size match the volume per atom because the correlation length vanishes for an ideal gas. Thus, the cell volume d^3 is $1/n$ and the rms fluctuation per cell is 1 because that is the rms for a Poisson distribution with mean value of 1. With these values, we recover the prior dilute expression.

More generally, one can derive from thermodynamics that

$$\frac{\langle \delta N_d^2 \rangle}{n\,d^3} = n\,k_B\,T\,\beta_T \qquad \text{with} \qquad \beta_T = -\frac{1}{V}\left(\frac{\partial P}{\partial V}\right)_T \tag{13.28}$$

where β_T is the *isothermal compressibility* (basically, how hard it is stuff more atoms or molecules into a given volume). With this, we have

$$\gamma = \frac{8\,\pi^3}{3}\left(\frac{\epsilon - \epsilon_o}{\epsilon_o}\right)^2 \frac{1}{n\,\lambda^4}\,k_B\,T\,\beta_T \tag{13.29}$$

This kind of Rayleigh scattering is quite important for light propagation in liquids; scattering, rather than absorption, is the primary reason for light attenuation in liquids.

One perhaps surprising overlap of particle physics and thermodynamics is the fact that, in almost any detector that involves the propagation of Cerenkov or scintillation light in a liquid — *e.g.*, water Cerenkov and liquid scintillator detectors of neutrinos such as in the SuperKamiokande, Sudbury Neutrino Observatory, MiniBooNE, MicroBoone, Daya Bay, LZ, XENONnT, and DUNE experiments (among many others), and liquid noble detectors based on scintillation light used for these purposes as well as dark matter searches and coherent neutrino-nucleus scattering measurements — Rayleigh scattering is the dominant mechanism for attenuation and must be modeled well to accurately simulate the detector behavior.

Also, interestingly, β_T can diverge at the liquid-gas critical point, giving rise to extremely large scattering called *critical opalescence*, which provides a useful way of measuring the critical point of a particular material.

Antennas

Introduction and Study Guide

Antennas are our second major application of the theory of radiation we have developed. This material is not covered in Griffiths, it is largely from Heald and Marion §9.4, §9.5, and §9.7.

A $d \ll \lambda$ Electric Dipole Antenna

Let us consider an electric dipole given by a line current along the z-axis $I(t) = I_0 e^{-i\omega t}$ that has length d. This is the current that flows in the case of an oscillating perfect electric dipole, with $I(t)d = \dot{p} = q_0 \dot{d}$. Plugging this into the electric dipole radiation power formulae, 12.108, and assuming harmonic time dependence and $\vec{p} \propto \hat{z}$, we obtain

$$\left\langle \frac{dP}{d\Omega} \right\rangle = \frac{\mu_o}{32\pi^2} \frac{\omega^2 I_0^2 d^2 \sin^2\theta}{c} \qquad \langle P \rangle = \frac{\mu_o}{12\pi} \frac{\omega^2 I_0^2 d^2}{c} = \frac{(k\,d)^2}{6\pi} \langle I^2 \rangle Z_0 \qquad (13.30)$$

with the free-space impedance $Z_0 = \sqrt{\mu_o/\epsilon_o}$ and $\langle I^2 \rangle = I_0^2/2$ the mean-square current.

We see the first hint of a concept of an antenna's *radiation resistance*, which gives the relation between mean-square current and the radiated power. We will see that, for an antenna radiating into free space, the maximal efficiency for radiated power is obtained when $\langle P \rangle / \langle I^2 \rangle = Z_0$. This short line antenna does not do a good job of radiating into free space because of the prefactor $(kd)^2/6\pi \ll 1$. We immediately see that, to match free-space well, an antenna must have a size of order the wavelength it is radiating. The approximations made for the multipole expansion and electric and magnetic dipole radiation fail and we must go back to the more general expressions.

General Setup for $d \ll \lambda$ Antennas

Given the above, let us recall a number of results from our general discussion of radiating systems. The form for the vector potential under the assumption $d \ll r$ and $d^2/\lambda \ll r$, but not assuming $d \ll \lambda$, Equation 12.91, is

$$\vec{A}(\vec{r}, t) = \frac{\mu_o}{4\pi} \frac{e^{i(kr - \omega t)}}{r} \int_{\mathcal{V}} d\tau' \, \vec{J}_0(\vec{r}') \, e^{-i k r' \hat{r} \cdot \hat{r}'}$$

We did not derive expressions for the fields in this general case, so we do so here. We recall that, with the above assumed harmonic time dependence, the application of the curl to get the magnetic and electric fields yields $i k \hat{r} \times$ factors from the argument of the exponential (neglecting the terms of higher order in r that come from differentiating the $1/r$ dependence). Therefore,

$$
\boxed{
\begin{array}{c}
\text{antenna} \\
\text{magnetic} \\
\text{field}
\end{array}
\qquad
\vec{B}(\vec{r}, t) = i \, \frac{\mu_o}{4\pi} \frac{\omega}{c} \frac{e^{i(kr - \omega t)}}{r} \, \hat{r} \times \int_{\mathcal{V}} d\tau' \, \vec{J}_0(\vec{r}') \, e^{-i k r' \hat{r} \cdot \hat{r}'}
}
\qquad (13.31)
$$

$$
\boxed{
\begin{array}{c}
\text{antenna} \\
\text{electric} \\
\text{field}
\end{array}
\qquad
\vec{E}(\vec{r}, t) = \frac{i \, c^2}{\omega} \vec{\nabla} \times \vec{B} = -c \, \hat{r} \times \vec{B}
}
\qquad (13.32)
$$

We will consider antennas that can be treated as line currents, so $d\tau' \, \vec{J}_0(\vec{r}') \to I \, d\vec{\ell}'$.

Antenna Gain

An antenna is fundamentally a device for either radiating power or receiving power with a desired pattern. Antennas are frequently intended to radiate to or receive from a particular direction. This can be quantified using the idea of *antenna gain*:

$$\text{antenna gain} \qquad G(\widehat{r}) = \frac{\left\langle \frac{dP}{d\Omega}(\widehat{r}) \right\rangle}{\langle P/4\pi \rangle} \qquad (13.33)$$

which is just the ratio of the power per unit solid angle into a particular direction divided by the total power per unit solid angle into all directions. The antenna radiates more power in directions for which this number is large. If an antenna is fully omnidirectional, then $G(\widehat{r}) = 1$ in all directions. We shall see that this same number also quantifies how directional an antenna is in reception mode.

Gain is frequently quoted in decibels (dB), with

$$G(\widehat{r}) \text{ [dB]} = 10 \log_{10} G(\widehat{r}) \qquad (13.34)$$

Thus, an omnidirectional antenna has 0 dB gain in all directions while other antennas have positive dB gain in some directions (the directions they are good at transmitting to) and negative dB gain in other directions.

Antenna Impedance

Since the antenna may not be small compared to λ, the mean-square current $\langle I^2 \rangle$ may not be the same everywhere in the antenna. For specificity, we choose to evaluate $\langle I^2 \rangle$ at the point where the antenna is fed by wires or a waveguide, the so-called "gap" drive point. We then define the *antenna impedance* and *radiation resistance*:

$$\boxed{\begin{array}{cc} \text{antenna} \\ \text{impedance} \end{array} \qquad Z_{ant} = \frac{V_g}{I_g}} \qquad \boxed{\begin{array}{cc} \text{radiation} \\ \text{resistance} \end{array} \qquad R_{rad} = \frac{\langle P \rangle}{\langle I_g^2 \rangle}} \qquad (13.35)$$

For an ideal antenna, meaning one that radiates all the power it receives from a transmission line, $Z_{ant} = R_{rad}$ (we use R instead of Z for the radiation component because it is always real by definition). For non-ideal antennas, there may be additional resistive or reactive components.

Of particular importance is that Z_{ant} determines the relation between power flowing in on a transmission line and power received by the antenna (which can then be radiated). If $Z_{ant} = Z_{line}$, then the antenna receives all power from the transmission line and reflects none. If this is not true, as may be the case because transmission lines have a set of standard impedances, then a transformer can be used to impedance match such a transmission line to an antenna. We discussed earlier how appropriate lengths of transmission line could be used as transformers. These can work quite well for antennas intended for single-frequency use. Broadband impedance matching — matching over a large fractional bandwidth $\Delta\nu/\nu$ — is more difficult.

Center-Driven Line Antennas

A line antenna is an antenna with a line current in the z-direction,
$\vec{J}_0(\vec{r}\,') \, d\tau' = I(\vec{r}\,') \, d\ell \, \hat{z}$. In general, solving for the current and vector potential even approximately is nontrivial. Jackson §9.4.B demonstrates that, for a thin-wire antenna that is perfectly conducting and with wire radius $a \ll \lambda$, d and that has azimuthal symmetry about the antenna's long direction d, the vector potential is exactly sinusoidally dependent on the distance along the antenna. From this, he concludes the current is approximately sinusoidal. We will assume this behavior in the following.

We assume a configuration as illustrated below, where the antenna consists of a long wire of length d in the z-direction, symmetrically placed about the origin, with a gap of negligible extent at $z = 0$. The two pieces of the antenna are connected to the two electrodes of a transmission line or two portions of a waveguide that have a potential difference.

© 1999 Jackson, *Classical Electrodynamics*
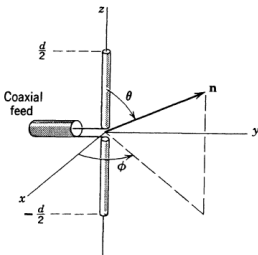
The current in the wire is assumed to follow

$$I(z) = I_0 \sin\left[k\left(\frac{d}{2} - |z|\right)\right] \tag{13.36}$$

This form is sinusoidal as desired and also satisfies the boundary condition that the current vanish at the end. The *derivative* is not continuous at $z = 0$ except for certain values of d. The current at the gap drive point is $I_{g,0} = I_0 \sin\frac{k\,d}{2}$. Given the current distribution, we may calculate the magnetic field, noting that $\widehat{r} \times d\ell\,\widehat{z} = -\widehat{\phi}\,\sin\theta$:

$$\vec{B}(\vec{r}, t) = -i\,\widehat{\phi}\,\frac{\mu_o}{4\pi}\,\frac{\omega}{c}\,I_0\,\frac{e^{i(k\,r - \omega\,t)}}{r}\,\sin\theta\int_{-d/2}^{d/2} dz'\,e^{-i\,k\,z'\cos\theta}\,\sin\left[k\left(\frac{d}{2} - |z'|\right)\right] \tag{13.37}$$

Let us rewrite this as the product of a function of λ and a function of d/λ and θ:

$$\boxed{\begin{array}{l}\text{center-driven}\\ \text{line antenna}\\ \text{magnetic field}\end{array}\qquad \vec{B}(\vec{r}, t) = -i\,\widehat{\phi}\,\frac{\mu_o}{4\pi}\,\frac{\omega}{c}\,I_0\,\frac{\lambda}{2}\,\frac{e^{i(k\,r - \omega\,t)}}{r}\,f\left(\frac{d}{\lambda}, \theta\right)} \tag{13.38}$$

$$\boxed{\begin{array}{l}\text{center-driven}\\ \text{line antenna}\\ \text{magnetic field}\\ \text{pattern}\end{array}\qquad f\left(\frac{d}{\lambda}, \theta\right) = \frac{\sin\theta}{\pi}\int_{-\pi\,d/\lambda}^{\pi\,d/\lambda} d\zeta\,e^{-i\,\zeta\cos\theta}\,\sin\left(\frac{\pi\,d}{\lambda} - |\zeta|\right)} \tag{13.39}$$

Comparing with our calculation of the radiation field due to a line current electric dipole, we see that if we make the correspondence,

$$I_0 \, d \leftrightarrow I_0 \frac{\lambda}{2} \qquad\qquad \sin\theta \leftrightarrow f\left(\frac{d}{\lambda}, \theta\right) \tag{13.40}$$

and note that the relation between $\vec{E}$ and $\vec{B}$ is the same as for the line current electric dipole, then the expressions calculated for the line current electric dipole can be translated to this case, giving

$$\left\langle \frac{dP}{d\Omega} \right\rangle = \frac{\mu_o}{32\,\pi^2} \, \frac{\omega^2 I_0^2 \frac{\lambda^2}{4}}{c} \left| f\left(\frac{d}{\lambda}, \theta\right) \right|^2 = \frac{Z_0}{32} \, I_0^2 \left| f\left(\frac{d}{\lambda}, \theta\right) \right|^2 \tag{13.41}$$

The gain and radiation resistance are

$$\boxed{\begin{array}{ll} \text{antenna} \\ \text{gain} \\ \text{pattern} \end{array} \quad G(\theta) = \frac{\left| f\left(\frac{d}{\lambda}, \theta\right) \right|^2}{\left\langle |f|^2 \right\rangle} \quad \text{with} \quad \left\langle |f|^2 \right\rangle = \frac{1}{4\,\pi} \int d\Omega \left| f\left(\frac{d}{\lambda}, \theta\right) \right|^2} \tag{13.42}$$

$$\boxed{\begin{array}{ll} \text{antenna} \\ \text{radiation} \\ \text{resistance} \end{array} \quad R_{rad} = \frac{\langle P \rangle}{\langle I_g^2 \rangle} = \frac{\frac{Z_0}{32} I_0^2 \, 4\,\pi \left\langle |f|^2 \right\rangle}{\frac{1}{2} I_{g,0}^2} = \frac{\pi}{4} \, \frac{1}{\sin^2 \frac{k\,d}{2}} \left\langle |f|^2 \right\rangle Z_0} \tag{13.43}$$

For the sinusoidal current we have assumed, we can evaluate the field pattern for arbitrary d. This is done by using the trigonometric product identities

$$\cos A \sin B = \frac{1}{2}\left[\sin\left(A+B\right) - \sin\left(A-B\right)\right] \tag{13.44}$$

$$\sin A \sin B = \frac{1}{2}\left[\cos\left(A-B\right) - \cos\left(A+B\right)\right] \tag{13.45}$$

and also using the Euler formula for $e^{-i k z' \cos\theta}$. This yields

$$\boxed{\begin{array}{c}\text{center-driven} \\ \text{line antenna} \\ \text{radiation} \\ \text{field pattern}\end{array} \quad f\left(\frac{d}{\lambda}, \theta\right) = \frac{2}{\pi}\,\frac{\cos\left(\frac{k\,d}{2}\,\cos\theta\right) - \cos\frac{k\,d}{2}}{\sin\theta}} \tag{13.46}$$

We can calculate these quantities for some specific antenna lengths. A plot comparing the radiation patterns of our idealized electric dipole and half-wave and full-wave antennas is given at the end.

▶ *half-wave antenna, $d = \lambda/2$*

The quantity $kd/2 = \pi/2$, so the current, field pattern, $\langle |f|^2 \rangle$, and radiation resistance are

$$\mathrm{I}(z) = \mathrm{I}_0 \sin\left(\frac{\pi}{2} - k|z'|\right) = \mathrm{I}_0 \cos kz \tag{13.47}$$

$$f\left(\frac{d}{\lambda} = \frac{1}{2}, \theta\right) = \frac{2 \cos\left(\frac{\pi}{2} \cos\theta\right)}{\pi \sin\theta} \qquad \langle |f|^2 \rangle = 0.247 \tag{13.48}$$

$$G_{max} = \frac{|f(d/\lambda = 1/2, \theta = \pi/2)|^2}{\langle |f|^2 \rangle} = \frac{(2/\pi)^2}{0.247} = 1.64 = 2.15 \text{ dB} \tag{13.49}$$

$$\mathrm{I}_{g,0} = \mathrm{I}_0 \qquad R_{rad} \approx 73\,\Omega \tag{13.50}$$

This type of antenna is therefore well-matched to $75\,\Omega$ coaxial cable. The antenna radiates an azimuthally uniform pattern but has highest gain in the $\theta = \pi/2$ plane. It is not a great match to free space because R_{rad} is quite different from $Z_0 = 377$ Ω.

▶ *full-wave antenna, $d = \lambda$*

The current is

$$I(z) = I_0 \sin\left(\pi - k|z'|\right) = I_0 \sin k|z'| = I_0 \left|\sin kz'\right| \tag{13.51}$$

The current at the drive point formally vanishes and the antenna's radiation resistance is formally infinite: basically, no matter what voltage the transmission line has on it, it cannot drive a current at the antenna drive point. Clearly, not a very good antenna to drive with a transmission line! The corrections we ignored in making the assumption of perfectly sinusoidal current will enable $I_g \neq 0$. However, we can still calculate the radiation pattern using our formula for f with $k\, d/2 = \pi$, giving

$$f\left(\frac{d}{\lambda} = 1, \theta\right) = \frac{2}{\pi} \frac{\cos\left(\pi \cos\theta\right) - \cos\pi}{\sin\theta} = \frac{4}{\pi} \frac{\cos^2\left(\frac{\pi}{2}\cos\theta\right)}{\sin\theta} \tag{13.52}$$

$$\left\langle |f|^2 \right\rangle = 0.672 \tag{13.53}$$

$$G_{max} = \frac{|f(d/\lambda = 1, \theta = \pi/2)|^2}{\left\langle |f|^2 \right\rangle} = \frac{(4/\pi)^2}{0.672} = 2.41 = 3.82 \text{ dB} \tag{13.54}$$

The full-wave antenna has a more peaked radiation pattern than the half-wave antenna, though it is harder to drive.

▶ *folded half-wave antenna*

The idea here is to fold a full-wave antenna at $z = \pm\lambda/4$ and connect the two ends together back at the center. This joining changes the boundary condition so that the joined point should be a point of peak current and the fold points are now current zeros. If s is the distance along the antenna from the drive point, so that $s = \pm\lambda/2$ sits at $z = 0$, the current becomes $I(s) = I_0 \cos ks$ instead of $I(s) = I_0 |\sin ks|$. The current is negative at $s = \pm\lambda/2 = d/2$, but, because the antenna has been folded so this piece of the antenna is pointed in the opposite direction as the unfolded piece, this current that is negative relative to s is positive relative to z. The antenna now looks like two half-wave antennas right next to each other. The radiation pattern is the same as a half-wave antenna, but the effective current I_0 for a given gap drive current $I_{g,0}$ is enhanced by a factor of two relative to the half-wave antenna, so the radiation resistance increases by a factor of $(I_0/I_{g,0})^2 = 4$:

$$I_0 = 2\, I_{g,0} \qquad R_{rad} = 4 \times 73\,\Omega = 292\,\Omega \qquad (13.55)$$

This higher radiation resistance is a better match to free space, which has $Z_0 = 377\,\Omega$ as noted before. Flat-pair or twisted-pair transmission line, which have impedances of about $300\,\Omega$, are best for feeding this antenna.

▶ *Arbitrary Length*

Given on the following slide is a plot of the antenna impedance (Equation 13.35) as a function of length in units of λ (a "Smith chart"). The length is marked with tickmarks on the trajectory. The impedance is a complex number whose resistive (real) part is the horizontal axis and whose reactive (imaginary) part is the vertical axis. The real part is dominated by the radiation resistance (Equation 13.35), which is always real by definition. The reactive component simply indicates that the current and voltage at the drive point (which are supplied by the feeding transmission line) are out of phase, with the sign of the phase defining whether the antenna has a capacitive or inductive reactance.

The right side of the trajectory, at $d/\lambda = 1$, would go off to infinity for an ideal full-wave antenna; nonidealities keep it finite. At $d/\lambda = 3/2$, the antenna looks similar to a half-wave antenna but with a somewhat higher impedance.

One can use a plot like this to pick the transmission line that best impedance matches the antenna one wants to use, the choice of which is usually driven by the radiation pattern or the available space. If there is an impedance mismatch, one might use a transmission line transformer to transform the impedance or lumped element capacitors or inductors to "tune out" (cancel) the reactive part of the antenna impedance.
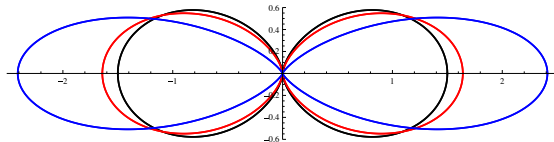
Polar plot of gain patterns (power units) of ideal electric dipole (black), half-wave antenna (red), and full-wave antenna (blue). The greater gain of the full-wave antenna is due to its $\cos^4$ dependence (vs. $\cos^2$ for the half-wave antenna). Figure courtesy of M. Cross.
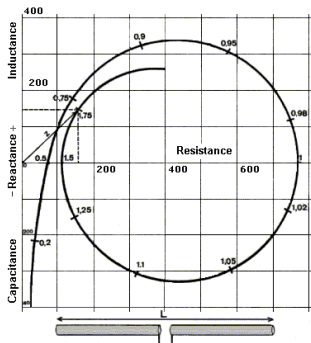
Antenna impedance as a function of antenna length in units of λ. Tickmarks indicate d/λ. The axes are in Ω and give the complex Z_{ant}, which is real for integer multiples of $\lambda/2$, implying $Z_{ant} = R_{rad}$ at these lengths.
NOTE: This plot follows the engineering sign convention. Our sign convention is the opposite, so the curve should be mirrored about the horizontal axis. $d < \lambda/2$ will still be the capacitive portion and $\lambda/2 < d < \lambda$ the inductive section (and so on for $d > \lambda$).

http://www.astrosurf.com/luxorion/qsl-swr.htm

Antenna Reciprocity Theorem

One can prove a theorem that the power *received* by an antenna from a plane wave (due to the current excited in it by the incoming wave's electric field) in direction $\widehat{k}$ is

$$P_r(\widehat{k}) = \frac{\lambda^2}{4\,\pi}\, G(-\widehat{k}) \times \text{incident flux} \tag{13.56}$$

The quantity $(\lambda^2/4\,\pi)\, G(-\widehat{k})$ is the *effective area of the antenna* in the direction $\widehat{k}$ (sort of like the cross section for scattering we discussed earlier.) By its definition, $\int d\Omega\, G(\widehat{r}) = 4\,\pi$, so the above also tells us that the total effective area of the antenna integrated over all possible incoming angles is λ^2. This is an important theorem in antenna theory because it provides a normalizing factor for measuring antenna efficiency: if you surround an antenna with a blackbody radiator, you know the incident flux. The antenna should receive a power λ^2 times the flux incident from the blackbody. You can then measure the power exiting the antenna onto a transmission line and take the ratio of observed to expected power to determine the antenna's overall efficiency. This could also be used, for example, to determine the antenna's radiation resistance.

The proof of this theorem is interesting because it involves the *Lorentz Reciprocity Theorem* and can be found in Drabowich *et al.*, *Modern Antennas*, §4.1.

Driven Arrays of Line Antennas

Suppose one has an array of N antennas, each displaced from the origin by $\vec{\Delta}_j$, $j = 1$ to N, all fed identically with current, in phase. Let $\vec{E}_0$ be the radiation field pattern of an individual antenna situated at the origin. Then the radiation pattern of the array is easily calculated by recognizing that each antenna's contribution to the total electric field is the same up to the phase offsets due to the $e^{-i k r' \hat{r} \cdot \hat{r}'}$ factor arising from the displacement in $\vec{r}'$ between antennas:

$$\vec{E} = \sum_{j=1}^{N} \frac{i \, c^2}{\omega} \, \vec{\nabla} \times \left[\vec{\nabla} \times \vec{A}_j(\vec{r}, t) \right] \tag{13.57}$$

$$= \sum_{j=1}^{N} \frac{i \, c^2}{\omega} \, \vec{\nabla} \times \left[\vec{\nabla} \times \frac{\mu_o}{4 \pi} \, \frac{e^{i(k r - \omega t)}}{r} \int_{\mathcal{C}_j} d\vec{\ell}' \, I(\vec{r}') \, e^{-i k \hat{r} \cdot \vec{r}'} \right] \tag{13.58}$$

$$= \sum_{j=1}^{N} \frac{i \, c^2}{\omega} \, \vec{\nabla} \times \left[\vec{\nabla} \times \frac{\mu_o}{4 \pi} \, \frac{e^{i(k r - \omega t)}}{r} \int_{\mathcal{C}_0} d\vec{\ell}' \, I(\vec{r}') \, e^{-i k \hat{r} \cdot (\vec{\Delta}_j + \vec{r}')} \right] \tag{13.59}$$

$$= \vec{E}_0 \sum_{j=1}^{N} e^{-i k \hat{r} \cdot \vec{\Delta}_j} \tag{13.60}$$

where we factored $e^{-i k \hat{r} \cdot \vec{\Delta}_j}$ out of the integral and recognized $\vec{E}_0$ in what remains.

You hopefully recognize the sum over phase factors of this type as the same kind of sum present in the calculation of the interference pattern due to an array of slits in a screen. Let us consider an analogue of that, an array of N center-fed antennas oriented with their currents flowing along the z-axis and displaced from the origin with uniform spacing Δ. Let the first antenna be at the origin, and let $N\Delta$, the extent of the array, satisfy $N\Delta \ll r$. The radiation field pattern is

$$\vec{E} = \vec{E}_0 \sum_{j=1}^{N} e^{-i\,k\,\hat{r}\cdot\vec{\Delta}_j} = \vec{E}_0 \sum_{j=1}^{N} e^{-i\,k\,\hat{r}\cdot(j-1)\,\vec{\Delta}} = \vec{E}_0 \sum_{j=1}^{N} e^{-i\,k\,\hat{r}\cdot\vec{\Delta}\,(j-1)} \qquad (13.61)$$

$$= \vec{E}_0 \frac{1 - e^{-i\,k\,N\,\hat{r}\cdot\vec{\Delta}}}{1 - e^{-i\,k\,\hat{r}\cdot\vec{\Delta}}} \qquad (13.62)$$

where we have used the standard result $\sum_{j=1}^{N} r^{j-1} = (1 - r^N)/(1 - r)$ for geometric series. Thus, if $f_0(d/\lambda, \theta)$ is the radiation pattern corresponding to $\vec{E}_0$, we have

$$\left| f\left(\frac{d}{\lambda}, \theta\right) \right|^2 = \left| f_0\left(\frac{d}{\lambda}, \theta\right) \right|^2 \left| \frac{1 - e^{-i\,k\,N\,\hat{r}\cdot\vec{\Delta}}}{1 - e^{-i\,k\,\hat{r}\cdot\vec{\Delta}}} \right|^2 \qquad (13.63)$$

$$\boxed{\begin{array}{c} \text{driven-array} \\ \text{antenna} \\ \text{radiation pattern} \end{array} \quad \left| f\left(\frac{d}{\lambda}, \theta\right) \right|^2 = \left| f_0\left(\frac{d}{\lambda}, \theta\right) \right|^2 \frac{\sin^2 \frac{N\,k\,\hat{r}\cdot\vec{\Delta}}{2}}{\sin^2 \frac{k\,\hat{r}\cdot\vec{\Delta}}{2}}} \qquad (13.64)$$

The factor at the end is the same factor one finds in the calculation of the interference pattern from an array of slits. It has the following properties:

$$\text{peak position at} \qquad k\,\hat{r}\cdot\vec{\Delta} = 2\,n\,\pi \qquad\qquad (13.65)$$

$$\text{peak height} \qquad N^2 \qquad\qquad (13.66)$$

$$\text{peak width} \qquad |k\,\hat{r}\cdot\vec{\Delta} - 2\,n\,\pi| \approx 2\,\sqrt{6}/N \qquad\qquad (13.67)$$

$$\text{secondary peaks at} \qquad k\,\hat{r}\cdot\vec{\Delta} = m\,\pi/N \qquad (m \neq 2\,n\,N\,\pi) \qquad (13.68)$$

$$\text{secondary peak height} \qquad \mathcal{O}(1) \qquad\qquad (13.69)$$

These results for the peak height and width are obtained by Taylor expanding the sines near $k\,\hat{r}\cdot\vec{\Delta} = 2\,n\,\pi$ to first and third order, while the secondary maxima are found by requiring $k\,\hat{r}\cdot\vec{\Delta} = m\,\pi/N \neq 2\,n\,\pi$.

This structure is easier to understand if we set $\vec{\Delta} = \Delta\,\hat{y}$, which yields $k\,\hat{r}\cdot\vec{\Delta} = k\,\Delta\sin\theta\,\sin\phi$. If we additionally assume $\Delta < \lambda$, this ensures there are at most two primary peaks at $\phi = 0, \pi$ because $k\,\Delta\sin\theta\,\sin\phi < 2\,\pi$ for all (θ, ϕ). The figure below shows the gain in the xy-plane ($\theta = \pi/2$) for $\Delta = \lambda/2$, $N = 5$. Off this plane, there are no primary peaks, only secondary peaks. Figure courtesy of M. Cross.
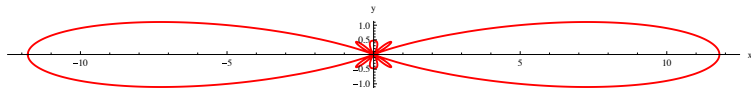
Driven arrays of antennas are pervasive. Here are some examples:

- ▶ The radar in the nose cone of most airplanes is a driven array antenna. The phase of the current driving the individual elements is varied, which is the equivalent of changing the orientation of $\vec{\Delta}$ so that the emitted (and received) radiation beam can be pointed away from the direction of motion of the plane without steering a physical antenna.

- ▶ Much of radio astronomy is done with driven antenna arrays operated in reception mode. The signals from the individual antennas are routed to a central "correlator." In principle, one should sum the signals from all the antennas. But the interesting behavior comes from the cross-terms in the sum, so, instead, all independent pairs of voltage signals (proportional to electric field) are multiplied, which gives the cross-terms. The advantage of the array approach is the very narrowly peaked gain function thus obtained: the width is roughly the same as that of a single radio telescope whose diameter is the extent of the array. These are called "radio interferometers." The Jansky Very Large Array in New Mexico and the Atacama Large Millimeter Array in Chile are examples for which the individual antennas are themselves radio telescopes. At longer wavelengths, a number of projects use arrays of simple antennas to obtain very large collecting areas (appreciable fractions of a square kilometer, including the Long Wavelength Array in the Owens Valley operated by Caltech) that view a large fraction of the sky at one time and for which different resolution functions can be obtained by digitizing and then doing the multiplications in a computer. There is also the Event Horizon Telescope, which combines telescopes spread over the entire planet and has imaged the event horizon of the black holes at the centers of M87 and our galaxy.

Passive Arrays of Line Antennas

Many television or radio antennas are arrays of linear antennas with one driven element and many passive elements. These passive elements are driven by radiation received from the driven element, and then they radiate with a definite phase relative to the driven antenna. The combined radiation pattern of the driven and passive antennas can be designed to be peaked in a particular direction. As we will see below, this forward peaking of the emitted radiation pattern also ensures that the antenna is a good receiver in that direction.

Here, we consider a *Yagi-Uda* antenna, shown below.
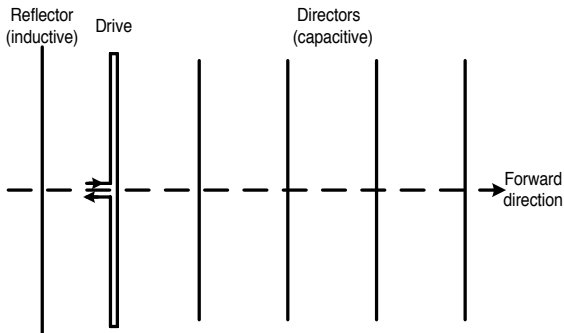
The important features of this design are as follows:

- There is one driven antenna, which is a folded half-wave antenna (for reasons discussed earlier).

- There is one antenna placed behind the the drive antenna that is intended as a back-reflector: it is spaced $\lambda/4$ behind the driven antenna so that the wave that reflects from it is in phase in the forward direction (remember the extra π phase shift due to a reflection from a conductor). This is the equivalent of saying that the wave that is radiated by the back-reflector, which is excited by the driven antenna, is in-phase with the driven antenna's wave in the forward direction and out of phase in the backward direction.

- There are multiple "director" antennas placed ahead of the driven antenna. They are intended to constructively interfere with the driven antenna in the forward direction and destructively in the backward direction. If there were just one, it would be placed $\lambda/2$ in front of the driven antenna (so the reflected wave in the reverse direction is out of phase). With multiple directors, the optimal spacing turns out to be about $\lambda/3$.

► A more sophisticated discussion implies that the lengths of the passive antennas are also important. The spacing argument is an oversimplification because it assumes the back-reflector acts like an infinite conducting plane. Really, we should think about the antenna impedance and ask what needs to be done to cancel the $-i\,k\,\hat{r}\cdot\vec{\Delta}$ phase factor that arises from the spacing between the driven antenna and the back-reflector or directors. This issue did not arise for the phased array because, using identical antennas, there are no path length differences for the direction of maximum constructive interference, the direction perpendicular to the spacing $\vec{\Delta}$ between elements. But here we want constructive interference in the direction in which there *are* path length differences, and so we need to cancel out the corresponding phase factors using the antenna impedances.

Recall the figure showing the antenna impedance as a function of size: antennas that have $d < \lambda/2$ are capacitive and those with $\lambda/2 < d < \lambda$ are inductive. In the forward direction, the phase factor $-i\,k\,\hat{r}\cdot\vec{\Delta}$ is negative imaginary. With our $-i\,\omega\,t$ time dependence, a capacitive element has a positive imaginary component to its impedance (positive phase factor), so the directors should be capacitive to cancel this negative imaginary phase and yield constructive interference in the forward direction (and, conversely, destructive interference in the reverse direction). So, the directors should have $d_d < \lambda/2$. Conversely, the back-reflector should be inductive and thus have $\lambda/2 < d_r < d$. Usually, d_d and d_r are not much different from d.

Diffraction

Introduction and Study Guide

Diffraction is the third major application of the theory of radiation we have developed. This material is not covered in Griffiths; it is largely from Heald and Marion Chapter 12 and from Jackson Section 10.5. We will study scalar diffraction theory only (vector diffraction theory gets rather complicated; it is covered in Jackson Section 10.6) and focus on the Fraunhofer limit ($d \ll \lambda \ll r$).

We may use a scalar rather than vector approach because we will focus primarily on using the wave equation for EM waves, in which $\vec{E}$ and $\vec{B}$ decouple and in which each component of $\vec{E}$ and $\vec{B}$ individually satisfies the wave equation, and also because we will not be calculating radiated fields from currents but rather directly from received fields. This point will become clearer as we proceed.

We will therefore assume a scalar field with harmonic time dependence $\psi(\vec{r})\,e^{-i\,\omega\,t}$ that satisfies the wave equation.

$$\left(\nabla^2 - \frac{1}{c^2}\,\frac{\partial^2}{\partial t^2}\right) \psi(\vec{r})\,e^{-i\,\omega\,t} = 0 \qquad (13.70)$$

Because of the harmonic time dependence, the time derivative part simplifies and the wave equation becomes the Helmholtz equation

$$(\nabla^2 + k^2)\,\psi(\vec{r}) = 0 \qquad \text{with} \qquad k = \omega/c \qquad (13.71)$$

The Helmholtz-Kirchoff Integral

The first step we must take is to figure out, given the EM fields on some bounding surface of a region due to sources external to the region, how to calculate the fields that propagate to points inside that region. See the picture below at left, where the sources are marked with × symbols, the bounding surface is shown in gray, and the point at which we want to calculate the field is P.
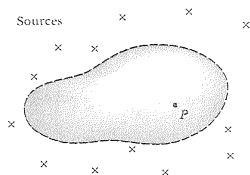
© 2012 Heald and Marion, *Classical Electro-*

magnetic Radiation
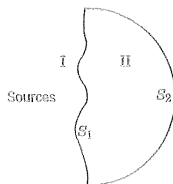
© 1999 Jackson, *Classical Electrodynamics*
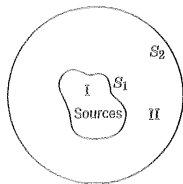
If one wants to think instead about the classical diffraction geometry, think of the boundary as an opaque screen with a hole in it with fields incident on the hole from sources on one side of the hole. We want to calculate the fields at points P on the other side of the screen. The figures above at the right could describe these geometries, with the sources in region I and the point P in region II.

Let us suppose that, along with having the scalar wave function ψ that satisfies the Helmholtz equation, we also have the Green Function for the equation, *i.e.*, the function $G(\vec{r}, \vec{r}')$ that satisfies the Helmholtz equation with a unit point source:

$$\left(\nabla_{\vec{r}}^2 + k^2\right) G(\vec{r}, \vec{r}') = -\delta(\vec{r} - \vec{r}') \qquad (13.72)$$

The sign on the right side for the source comes from the inhomogeneous wave equation in Lorenz gauge, Equation 10.21, where it is due to the choice of sign convention for the d'Alembertian (negative sign in front of the spatial derivatives).

Next, applying Green's Theorem (Equation 3.13) with $\psi(\vec{r}')$ and $G(\vec{r}, \vec{r}')$, we have

$$-\oint_{\mathcal{S}} da' \left[\psi(\vec{r}') \, \hat{n}(\vec{r}') \cdot \vec{\nabla}_{\vec{r}'} \, G(\vec{r}, \vec{r}') - G(\vec{r}, \vec{r}') \, \hat{n}(\vec{r}') \cdot \vec{\nabla}_{\vec{r}'} \, \psi(\vec{r}') \right]$$

$$= \int_{\mathcal{V}(\mathcal{S})} d\tau' \left[\psi(\vec{r}') \nabla_{\vec{r}'}^2 \, G(\vec{r}, \vec{r}') - G(\vec{r}, \vec{r}') \nabla_{\vec{r}'}^2 \, \psi(\vec{r}') \right]$$

$$= -\psi(\vec{r}) \qquad \text{for } \vec{r} \in \mathcal{V}, \text{ 0 otherwise} \qquad (13.73)$$

where the left side has acquired a negative sign because we define $\hat{n}$ to the be *inward* rather than outward normal on the surface $\mathcal{S}$. We have used the Helmholtz equation to evaluate the $\nabla_{\vec{r}'}^2$ terms on the right-hand side, and we have also assumed $G(\vec{r}, \vec{r}')$ is symmetric in its arguments as usual. The delta function in the Helmholtz Equation for G gives the one nonzero term on that side. When evaluating $\nabla_{\vec{r}'}^2 \, \psi(\vec{r}')$, there is no source term because the sources are outside $\mathcal{V}$. We have thus obtained what we want, an equation for $\psi(\vec{r} \in \mathcal{V})$ in terms of ψ and its normal gradient on the boundary. We need the Green Function for the specific geometry, of course.

We then make two approximating assumptions:

▶ For any geometry of interest, it is acceptable to use the Green function for the Helmholtz equation for a free-space geometry, $G(\vec{r}, \vec{r}') = \exp(i\,k\,R)/R$ with $\vec{R} = \vec{r} - \vec{r}'$, independent of the specific geometry. We first need to evaluate the action on G of $\vec{\nabla}_{\vec{r}'}$,

$$\vec{\nabla}_{\vec{r}'}\, G(\vec{r}, \vec{r}') = \vec{\nabla}_{\vec{r}'} \frac{e^{i\,k\,R}}{R} = \left(i\,k - \frac{1}{R}\right)\left(-\widehat{R}\right)\frac{e^{i\,k\,R}}{R} \qquad (13.74)$$

where we have $-\widehat{R}$ instead of $\widehat{R}$ because the gradient is with respect to $\vec{r}'$, not $\vec{r}$. Inserting this into the equation for $\psi(\vec{r})$:

$$\psi(\vec{r}) = -\frac{1}{4\,\pi} \oint_{\mathcal{S}} da'\, \frac{e^{i\,k\,R}}{R}\, \widehat{n}(\vec{r}') \cdot \left[\vec{\nabla}_{\vec{r}'}\psi(\vec{r}') + \left(i\,k - \frac{1}{R}\right)\widehat{R}\,\psi(\vec{r}')\right] \qquad (13.75)$$

▶ We may break the surface $\mathcal{S}$ into surfaces $\mathcal{S}_1$ and $\mathcal{S}_2$ where $\mathcal{S}_1$ consists of a screen with apertures for which we wish to calculate the diffractive effects of a wave incident from outside of $\mathcal{S}$ and where $\mathcal{S}_2$ consists of the remainder of $\mathcal{S}$, taken off to infinity. At such large distances,

$$\psi \to f(\theta, \phi)\, \frac{e^{ikr}}{r} \qquad \widehat{r} \cdot \vec{\nabla}\psi \to \left(i\,k - \frac{1}{r}\right)\psi \qquad (13.76)$$

At these distances, $r' \gg r$, so $\vec{R} \to -\vec{r}'$, $R \to r'$, and $\hat{n} \to -\hat{r}'$ and the integrand becomes

$$\frac{e^{ikr}}{r} \left(-\hat{r}'\right) \cdot \left[\left(i\,k - \frac{1}{r'}\right) \psi\,\hat{r}' + \left(i\,k - \frac{1}{r'}\right)\left(-\hat{r}'\right)\psi(\vec{r}')\right] = 0 \qquad (13.77)$$

That is, the $\mathcal{S}_2$ integral vanishes.

We thus arrive at the Helmholtz-Kirchoff Integral,

$$\psi(\vec{r}) = -\frac{1}{4\pi} \int_{\mathcal{S}_1} da' \, \frac{e^{i\,k\,R}}{R} \, \hat{n}(\vec{r}') \cdot \left[\vec{\nabla}_{\vec{r}'}\psi(\vec{r}') + \left(i\,k - \frac{1}{R}\right)\hat{R}\,\psi(\vec{r}')\right] \qquad (13.78)$$

where $\mathcal{S}_1$ is now just the diffracting screen. (The utility of this for scalar fields is reflected by the fact that it was first derived by Helmholtz in 1859 for acoustic waves. Kirchoff extended it to EM waves in 1882.)

To do this integral, we need to know both $\psi(\vec{r})$ and $\hat{n} \cdot \vec{\nabla}\psi(\vec{r})$ on the screen $\mathcal{S}_1$, which already hints at a problem: by analogy to Laplace's equation in electrostatics, we should not need to know both. Nevertheless, soldiering on, the Kirchoff approximation consists of assuming these functions vanish everywhere but in the apertures in the screen and that their values there are given by the values they would have in the absence of the screen. This is not an unreasonable approximation if we assume the screen is a perfect absorber.

The problem with this approximation is that one can show that, if both ψ and its normal gradient vanish on any finite surface, then ψ vanishes everywhere. (This is not obvious.) This is both inconsistent with the second assumption and it does not recover the values in the opening (*i.e.*, one cannot use the formula for a point $\vec{r}$ in the apertures in $\mathcal{S}_1$). We will see below, nevertheless, that the approximation is useful.

The formally correct method would have been to not make the first approximation about the Green function being the free-space Green function. One should specify either Dirichlet or Neumann boundary conditions (knowing ψ or its normal gradient on $\mathcal{S}_1$) and find the appropriate Green function as we did in electrostatics, requiring it satisfy the appropriate condition:

$$\text{Dirichlet:} \quad G_D(\vec{r} \in \mathcal{S}, \mathcal{V}, \vec{r}' \in \mathcal{S}) = 0 \tag{13.79}$$

$$\text{Neumann:} \quad \hat{n} \cdot \vec{\nabla}_{\vec{r}'} G_N(\vec{r} \in \mathcal{S}, \mathcal{V}, \vec{r}' \in \mathcal{S}) = 0 \tag{13.80}$$

One then gets the appropriate Generalized Helmholtz-Kirchoff Integral from Equation 13.73

$$\text{Dirichlet:} \quad \psi(\vec{r}) = \int_{\mathcal{S}_1} da' \, \psi(\vec{r}') \, \hat{n}(\vec{r}') \cdot \vec{\nabla}_{\vec{r}'} G(\vec{r}, \vec{r}') \tag{13.81}$$

$$\text{Neumann:} \quad \psi(\vec{r}) = -\int_{\mathcal{S}_1} da' \, G(\vec{r}, \vec{r}') \, \hat{n}(\vec{r}') \cdot \vec{\nabla}_{\vec{r}'} \psi(\vec{r}') \tag{13.82}$$

It turns out (again, not obvious) these expressions are consistent with the input approximation: in the Dirichlet case, that ψ vanishes on $\mathcal{S}_1$ except in the apertures, where it equals the incident wave; and in the Neumann case, that the normal gradient of ψ vanishes on $\mathcal{S}_1$ except in the apertures, where it equals the normal gradient of the incident wave.

So, to do the calculation fully correctly, one needs to know the Green Function for the Helmholtz equation in the particular geometry. In simple cases, it is easy to find.

Plane Screen Case

Let us consider a specific, common case, where $\mathcal{S}_1$ is a plane with an aperture or apertures in it. If we place the plane at $z = 0$, one can show using the Method of Images that the Green Functions are

$$\text{Dirichlet:} \qquad G_D(\vec{r}, \vec{r}') = \frac{1}{4\pi} \left(\frac{e^{ikR}}{R} - \frac{e^{ikR'}}{R'} \right) \qquad (13.83)$$

$$\text{Neumann:} \qquad G_N(\vec{r}, \vec{r}') = \frac{1}{4\pi} \left(\frac{e^{ikR}}{R} + \frac{e^{ikR'}}{R'} \right) \qquad (13.84)$$

$$\text{with} \qquad \vec{r}' = x'\hat{x} + y'\hat{y} + z'\hat{z} \qquad \vec{r}'' = x'\hat{x} + y'\hat{y} - z'\hat{z} \qquad (13.85)$$

$$\text{and} \qquad \vec{R} = \vec{r} - \vec{r}' \qquad \vec{R}' = \vec{r} - \vec{r}'' \qquad (13.86)$$

The resulting Generalized Helmholtz-Kirchoff Integrals are then

$$\text{Dirichlet:} \qquad \psi(\vec{r}) = -\frac{1}{2\pi} \int_{\mathcal{S}_1} da' \, \frac{e^{ikR}}{R} \, \hat{n}(\vec{r}') \cdot \hat{R} \left(ik - \frac{1}{R} \right) \psi(\vec{r}') \qquad (13.87)$$

$$\text{Neumann:} \qquad \psi(\vec{r}) = -\frac{1}{2\pi} \int_{\mathcal{S}_1} da' \, \frac{e^{ikR}}{R} \, \hat{n}(\vec{r}') \cdot \vec{\nabla}_{\vec{r}'} \psi(\vec{r}') \qquad (13.88)$$

which are two times the first or second terms of the Helmholtz-Kirchoff integral.

These forms look quite different from each other. Let's consider a simple case, though, to see how they correspond. We assume the incident wave is sourced by a point source many wavelengths from the screen so the incident wave is a spherically symmetric expanding wave, $\psi(\vec{r} \in S_1) = \psi_0 \exp(i\,k\,|\vec{r} - \vec{r}_{src}|)/|\vec{r} - \vec{r}_{src}|$. We assume the observation point $\vec{r} \in \mathcal{V}$ is also many wavelengths from the screen so that $k\,R = 2\,\pi\,R/\lambda \gg 1$ and we can ignore the additional $1/R$ terms in the integrands. Under these assumptions, all three expressions (Helmholz-Kirchoff, Equation 13.78, and the two generalized forms, Equations 13.87 and 13.88) can be written in the common form (taking the origin of $\vec{r}'$ to be the center of the apertures in the screen)

$$\psi(\vec{r}) = -\frac{i\,k}{2\,\pi}\,\psi_0 \int_{apertures} da' \, \frac{e^{i\,k\,|\vec{r}' - \vec{r}_{src}|}}{|\vec{r}' - \vec{r}_{src}|} \, \frac{e^{i\,k\,|\vec{r} - \vec{r}'|}}{|\vec{r} - \vec{r}'|} \, \mathcal{O}(\theta_{src}, \theta_{obs}) \qquad (13.89)$$

where $\mathcal{O}(\theta_{src}, \theta_{obs}) = \begin{cases} \cos\theta_{obs} \equiv \hat{z} \cdot \hat{r}' & \text{Dirichlet} \\ \cos\theta_{src} \equiv -\hat{z} \cdot \hat{r}_{src} & \text{Neumann} \\ \frac{1}{2}\left(\cos\theta_{obs} + \cos\theta_{src}\right) = \frac{1}{2}\left(\hat{z} \cdot \hat{r}' - \hat{z} \cdot \hat{r}_{src}\right) & \text{Kirchoff} \end{cases}$

is the *obliquity factor*. The $\hat{z}$ comes from $\hat{n}(\vec{r}')$: we get $\hat{z} \cdot \hat{r}'$ from $\hat{n}(\vec{r}') \cdot \hat{R}$ and we get $-\hat{z} \cdot \hat{r}'_{src}$ from $\hat{n}(\vec{r}') \cdot \vec{\nabla}_{\vec{r}'} \psi(\vec{r}')$. The "Kirchoff" version is termed the Fresnel-Kirchoff diffraction integral. The other two versions do not have defined names.

By way of explanation: The first factor in the integrand simply calculates the value of the field in the aperture. The denominator allows for variation of the intensity over the aperture, while the numerator allows for the variation of the phase over the aperture. The second factor then propagates the field from the aperture to the observation point in similar fashion, accounting for both intensity and phase variation as the variable of integration varies over the aperture.

In the limit of a source very far away from the screen, so that $d \ll r$ (a further approximation in addition to our assumption $r_{src} \gg \lambda$, $r \gg \lambda$), the angles θ_{src} and θ_{obs} do not vary much over the integral. The three cases thus just have different normalizations, so we see they tie together somewhat. We do not have time here to address the issue of the differing normalizations.

Incoming Plane Wave Limit

Given that similar expressions are present as for our discussion of radiation, one can imagine taking similar limits. Let's first do this for the source field. The first step would be to assume that the source is in the far-field, where, if d is the size scale of the aperture, we assume $d \ll r_{src}$ and $d^2/\lambda \ll r_{src}$, as we did in deriving Equation 12.91. This would yield

$$\psi(\vec{r}) = -\frac{i\,k}{2\,\pi} \frac{e^{i\,k\,r_{src}}}{r_{src}} \psi_0 \int_{apertures} da' \, e^{i\,k\,r'\,\hat{r}_{src}\cdot\hat{r}'} \frac{e^{i\,k\,|\vec{r}-\vec{r}'|}}{|\vec{r}-\vec{r}'|} \mathcal{O}(\theta_{src}, \theta_{obs}) \qquad (13.90)$$

If we further assume $d \ll \lambda$, we can Taylor expand the exponential in the same way as we did to obtain the multipole expansion for radiation, Equation 3.226, yielding

$$\psi(\vec{r}) = -\frac{i\,k}{2\,\pi} \frac{e^{i\,k\,r_{src}}}{r_{src}} \sum_{m=0}^{\infty} \frac{(-i\,k)^m}{m!} \psi_0 \int_{apertures} da' \, (r'\,\hat{r}_{src}\cdot\hat{r}')^m \frac{e^{i\,k\,|\vec{r}-\vec{r}'|}}{|\vec{r}-\vec{r}'|} \mathcal{O}(\theta_{src}, \theta_{obs})$$
$$(13.91)$$

The most extreme approximation would be to assume the source is along z axis and so far away that the phase variation over the aperture is negligible — i.e., the incoming wave looks like a plane wave with $\hat{k} = \hat{z}$, in which case one keeps only the $m = 0$ term:

$$\psi(\vec{r}) = -\frac{i\,k}{2\,\pi}\,\frac{e^{i\,k\,r_{src}}}{r_{src}}\,\psi_0 \int_{apertures} da'\,\frac{e^{i\,k\,|\vec{r}-\vec{r}'|}}{|\vec{r} - \vec{r}'|}\,\mathcal{O}(\theta_{src} = 0, \theta_{obs}) \tag{13.92}$$

This is equivalent to assuming the incoming wave is a plane wave with $\hat{k} = \hat{z}$ and hence we can term this the *incoming plane wave limit*.

Fraunhofer, Fresnel, and the Fraunhofer Limit

We will work with the Fresnel-Kirchoff diffraction integral in the incoming plane wave limit. This is not general, but will illustrate the important aspects of diffraction.

We may consider a similar set of approximations when evaluating the terms that involve the observation point $\vec{r}$. We first make the same approximation as earlier, $d \ll r$ (note that we have already assumed $\lambda \ll r$ in deriving the Fresnel-Kirchoff diffraction integral). We will not yet make the far-field approximation, $d^2/\lambda \ll r$. We *will* make the *paraxial approximation*, which is that θ_{obs} is close to zero so $\cos \theta_{obs} \approx 1$ *and* $\hat{r} \cdot \hat{r}' \approx 0$ (*i.e.*, $\hat{r}$ is almost perpendicular to the plane in which $\hat{r}'$ lives, the aperture plane). This yields (similar to the case of radiation but keeping one more term in the exponential for now)

$$\psi(\vec{r}) = -\frac{i\,k}{2\,\pi} \frac{e^{i\,k\,r_{src}}}{r_{src}} \frac{e^{i\,k\,r}}{r} \psi_0 \int_{apertures} da' \, e^{-i\,k\,r'\,\hat{r}\cdot\hat{r}'} \, e^{i\,k\,(r')^2/r} \qquad (13.93)$$

The paraxial approximation was used to ignore the term of order kd^2/r that includes $\hat{r} \cdot \hat{r}'$ (which comes from the expansion of $\hat{r} - \hat{r}'$ to that order) and also to reduce the obliquity factor to unity.

The Fraunhofer limit consists of ignoring the factor of order d^2/r in the argument of the exponential; this is equivalent to taking the far-field limit, $r \gg d^2/\lambda$. When we do not take that limit, the analysis is called Fresnel diffraction and the analysis gets very involved. If we do indeed make the far-field approximation, the expression reduces to

$$\psi(\vec{r}) = -\frac{i\,k}{2\,\pi} \frac{e^{i\,k\,r_{src}}}{r_{src}} \frac{e^{i\,k\,r}}{r} \psi_0 \int_{apertures} da' \exp\left[-i\,k\left(x'\,\theta_{obs,x} + y'\,\theta_{obs,y}\right)\right] \quad (13.94)$$

where we have used the fact that the component of $\hat{r}$ along $\hat{z}$ is orthogonal to $\hat{r}'$, so we only retain the components of $\hat{r}$ in the transverse plane. In the paraxial limit, those components are x/r and y/r, involving the transverse coordinates of $\vec{r}$. These quantities are the sines of the x and y projections of θ_{obs}, which, in the paraxial limit, can be taken to simply be the angles themselves. The above is thus the Fraunhofer diffraction integral that you have seen in prior courses.