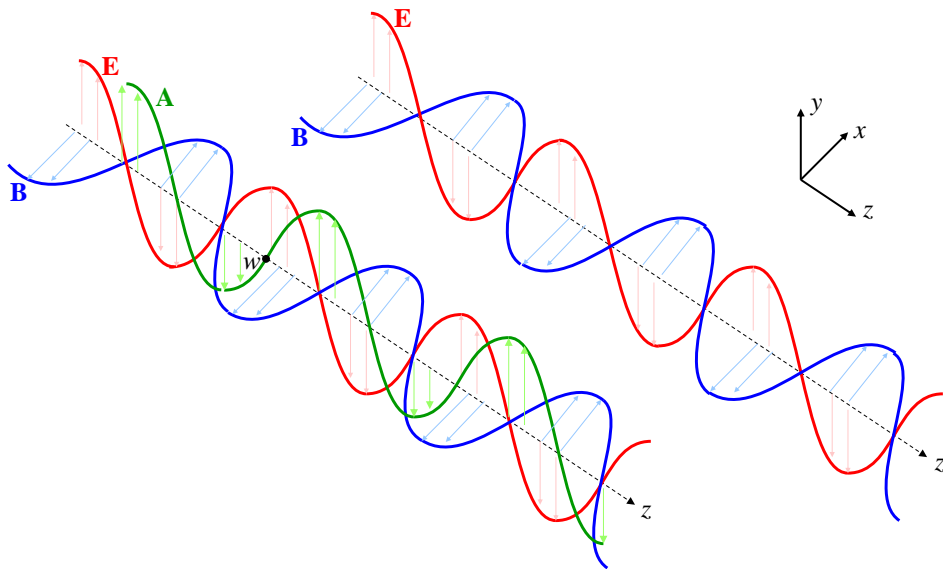# Funky Electromagnetic Concepts

The Anti-Textbook*

A Work in Progress. See elmichelsen.physics.ucsd.edu/ for the latest versions of the Funky Series.
Please send me comments.

Eric L. Michelsen

"But Mr. Faraday, of what use is all this?" - unknown woman

"Madam, of what use is a newborn baby?" - Michael Faraday

"With electromagnetism, as with babies, it's all a matter of potential."
- Bill Nye, the Science Guy

* Physical, conceptual, geometric, and pictorial physics that didn't fit in your textbook.

"Remarkably clear and easy to reproduce." -- physics graduate student.

"Finally, a systematic approach to boundary value problems." -- physics graduate student.

2006 values from NIST. For more physical constants, see http://physics.nist.gov/cuu/Constants/ .

??Needs updating for the 2019 revamp of SI.

Speed of light in vacuum	$c = 299\ 792\ 458$ m s^{-1} (exact)
Boltzmann constant	$k = 1.380\ 6504(24)$ x 10^{-23} J K^{-1}
Stefan-Boltzmann constant Relative standard uncertainty	$\sigma = 5.670\ 400(40)$ x 10^{-8} W m^{-2} K^{-4} ± 7.0 x 10^{-6}
Avogadro constant Relative standard uncertainty	$N_A, L = 6.022\ 141\ 79(30)$ x 10^{23} mol^{-1} ± 5.0 x 10^{-8}
Molar gas constant	$R = 8.314\ 472(15)$ J mol^{-1} K^{-1}
calorie	4.184 J (exact)
Electron mass	$m_e = 9.109\ 382\ 15(45)$ x 10^{-31} kg
Proton mass	$m_p = 1.672\ 621\ 637(83)$ x 10^{-27} kg
Proton/electron mass ratio	$m_p/m_e = 1836.152\ 672\ 47(80)$
Elementary charge	$e = 1.602\ 176\ 487(40)$ x 10^{-19} C
Electron g-factor	$g_e = -2.002\ 319\ 304\ 3622(15)$
Proton g-factor	$g_p = 5.585\ 694\ 713(46)$
Neutron g-factor	$g_N = -3.826\ 085\ 45(90)$
Muon mass	$m_\mu = 1.883\ 531\ 30(11)$ x 10^{-28} kg
Inverse fine structure constant	$\alpha^{-1} = 137.035\ 999\ 679(94)$
Planck constant	$h = 6.626\ 068\ 96(33)$ x 10^{-34} J s
Planck constant over 2π	$\hbar = 1.054\ 571\ 628(53)$ x 10^{-34} J s
Bohr radius	$a_0 = 0.529\ 177\ 208\ 59(36)$ x 10^{-10} m
Bohr magneton	$\mu_B = 927.400\ 915(23)$ x 10^{-26} J T^{-1}

Other values:

Jansky (Jy), flux and spectral density 10^{-26} W/m^2/Hz $= 10^{-23}$ erg/s/cm^2/Hz

Contents

1   Introduction

Why Funky?

The purpose of the "Funky" series of documents is to help develop an accurate physical, conceptual, geometric, and pictorial understanding of important physics topics. We focus on areas that don't seem to be covered well in most texts. The Funky series attempts to clarify those neglected concepts, and others that seem likely to be challenging and unexpected (funky?). The Funky documents are intended for serious students of physics; they are not "popularizations" or oversimplifications.

Physics includes math, and we're not shy about it, but we also don't hide behind it.

> Without a conceptual understanding, math is gibberish.

This work is one of several aimed at graduate and advanced-undergraduate physics students. Go to http://physics.ucsd.edu/~emichels for the latest versions of the Funky Series, and for contact information. We're looking for feedback, so please let us know what you think.

How to Use This Document

> This work is not a text book.

There are plenty of those, and they cover most of the topics quite well. This work is meant to be used *with* a standard text, to help emphasize those things that are most confusing for new students. When standard presentations don't make sense, come here.

> If you don't understand something, read it again *once*, then keep reading.
> Don't get stuck on one thing. Often, the following discussion will clarify things.

You should read all of this introduction to familiarize yourself with the notation and contents. After that, this work is meant to be read in the order that most suits you. Each section stands largely alone, though the sections are ordered logically. Simpler material generally appears before more advanced topics. You may read it from beginning to end, or skip around to whatever topic is most interesting.

The index is not yet developed, so go to the web page on the front cover, and text-search in this document.

What's Wrong With Existing Electromagnetic Expositions?

They're not precise enough with their definitions. Usually, when there appears to be an obvious contradiction, it is a confusion of definitions. Many widely used references have terribly unclear definitions, and one purpose of these notes is to help resolve them. Also, many texts are not visual or graphical enough. They rely way too much on algebra or advanced math, and not enough on insight.

My Story

The Funky series of notes is the result of my going to graduate school in physics after 20 years out of school. Although I had been an engineer all that time, most of my work involved software and design architectures that are far removed from fundamental science and mathematics. I expected to be a little rusty, but I found that the rust ran deeper than I realized.

There are many things I wish I had understood better while taking my classes (first at San Diego State University, then getting my PhD at University of California, San Diego). The Funky series is my attempt to help other students acquire a deeper understanding of physics.

12/21/2021 17:12   Page 5 of 108

Thank You

I owe a big thank you to many professors at both SDSU and UCSD, for their generosity, even when I wasn't a real student: Dr. Herbert Shore, Dr. Peter Salamon, Dr. Arlette Baljon, Dr. Andrew Cooksy, Dr. George Fuller, Dr. Tom O'Neil, Dr. Terry Hwa, and others.

Notation

[Square brackets] in text indicates asides that can be skipped without loss of continuity. They are included to help make connections with other areas of physics.

arg A for a complex number A, arg A is the angle of A in the complex plane; i.e., $A = |A|e^{i(\arg A)}$.

[Interesting points that you may skip are "asides," shown in square brackets, or smaller font and narrowed margins. Notes to myself may also be included as asides.]

> Common misconceptions are sometimes written in dark red dashed-line boxes.

> Mnemonics (memory aids) or other tips are given in green boxes.

2 Circuits

Lumped Element Circuits

??

Circuits Reference Desk

Which end of a resistor is positive? An inductor? A capacitor? A diode? A battery? It all comes down to two simple conventions: the passive convention, and the active convention. We start with the resistor, then proceed to the more complicated devices.

> These principles extend directly to AC analysis, using phasors and complex impedance.

$$v = iR \qquad i = C\frac{dv}{dt} \qquad v = L\frac{di}{dt} \qquad i = I_0\left(e^{vq/kT} - 1\right) \qquad \begin{array}{l} v = \text{constant} \\ i = \text{arbitrary} \end{array}$$

passive convention   active convention
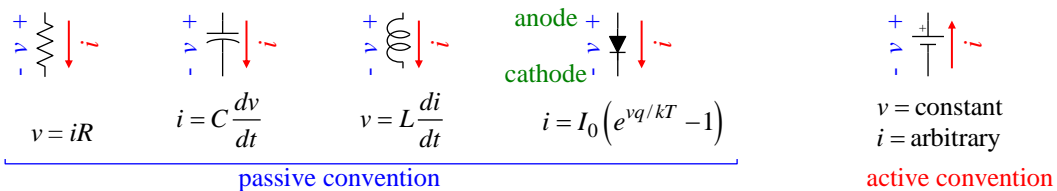
Figure 2.1   Reference polarities for all components, except sources, follow the passive convention. The reference polarity is reversed for the battery (a source) compared to all other components; it uses the active convention.

> For any circuit element, conventions *define* a **reference polarity** for the voltage, and also a **reference direction** for the current.
> We must write our *I-V* equations consistently with those choices.

For a resistor: the current always flows from + to -, so we *choose* reference directions consistent with that (diagram above, left). This allows us to write Ohm's law without minus signs: *v = iR*. All other passive devices follow this same convention. Therefore, when the power *P = vi* is positive, the device is *consuming* power from the circuit.

Sources, both voltage and current sources, always use the active convention: reference current flows *out of* the reference positive voltage. Therefore, when the power *P = vi* is positive, the source is *supplying* power to the circuit.

For an arbitrary circuit, we may not know ahead of time which end of a resistor will end up being +. For example:

Figure 2.2 Two valid choices of reference directions for the same circuit.

We don't know, without being given numbers and doing the circuit analysis, whether the middle resistor's current will flow up or down. No problem: we just choose an *arbitrary* polarity (this defines both voltage and current, since their relationship is fixed by the conventions: reference current flows from reference + to reference –). We do the circuit analysis *assuming* this polarity. Note that:

> Ohm's law applies for *both positive and negative* voltages and currents.

If we find that the voltage (and therefore current) is negative, it just means that the current is really flowing opposite to our reference choice. Note, however, that resistors always consume power from the circuit:

$$P = VI \qquad \text{is always positive: the resistor } \textit{consumes} \text{ power.}$$

For a capacitor: Things are a tad messier, because current does *not* always flow from + to –. A capacitor stores energy in its electric field. If we increase the voltage on the capacitor from zero, the capacitor is drawing energy from the rest of the circuit (it is charging). In this case, it is qualitatively similar to a resistor, and its current flows from + to –. Therefore, to be consistent with resistors, we *define* our reference directions for this case: reference current flows from reference + to reference – (just like a resistor). But wait! The current through a capacitor is *not* related to the polarity of voltage across it; the current is related to the *rate of change* of voltage:

$$i = C\frac{dv}{dt} \qquad \text{units: } C/s = (C/V)(V/s).$$

If the voltage is negative, but increasing (becoming less negative), the current through the capacitor is still positive. Thus the *I-V* equation above is always valid: charging or discharging. When a capacitor is discharging, either its voltage is + and its current is –, or its voltage is – and its current is +. Either way, the capacitor is delivering energy to the circuit, and temporarily acts more like a battery than a resistor. But we can't change our reference directions on the circuit based on the charging/discharging state of the capacitor from moment to moment. Also, either way, the capacitor's *power* "consumed" is

$$P = VI \qquad where \quad \text{negative power means it } \textit{supplies} \text{ energy to the circuit.}$$

For an inductor: things are similar, but we exchange "voltage" and "current", and replace "electric" with "magnetic." Again, the current does *not* always flow from + to –. An inductor stores energy in its magnetic field. If we increase the current from zero, the inductor is drawing energy from the rest of the circuit (loosely, it is "charging"). In this case, it is qualitatively similar to a resistor, and its current flows from + to –. Again, to be consistent with resistors, we *define* our reference directions for this case: reference current flows from reference + to reference – (just like a resistor). But wait! The voltage across an inductor is *not* related to direction of the current through it; the voltage is related to the *rate of change* of the current. Therefore, if the current is negative, but increasing (becoming less negative), the voltage across the inductor is still positive. This allows us to write a single *I-V* equation for both cases:

$$v = L\frac{di}{dt} \qquad \text{units: } V = (\textit{flux-linkages}/\textit{ampere})(\textit{amperes}/s)$$

When an inductor is "discharging," the current is decreasing, and the inductor is supplying energy to the circuit. Now it is acting more like a battery than a resistor. But we made our reference choice for the inductor, and we must stick with it. The above *I-V* equation is valid at all times. Again,

$$P = VI \qquad where \quad \text{negative power means it } \textit{supplies} \text{ energy to the circuit.}$$

Note that when the inductor current is increasing (either becoming more positive *or* less negative), v is positive. When the current is decreasing (either becoming less positive *or* more negative), v is negative, which means the + reference terminal is really at negative voltage with respect to the – terminal. In all cases, the above equations are correct. We achieved that consistency by defining a single reference polarity.

For a diode: Resistors, capacitors, and inductors are all "symmetric" or "unpolarized" devices: you can reverse the two leads with no effect. Diodes, in contrast, are polarized: one lead is the "anode"; the other is the "cathode." You must connect them properly. The reference voltage is defined always with + on the anode (and therefore – on the cathode); reference current flows from + to –. Diodes always consume power (like resistors do). Consistently with that, these conventions require that

$$P = VI \qquad \text{is always positive: the diode } \textit{consumes} \text{ power.}$$

For a battery: Batteries usually *supply* energy to the circuit, so we define them as having *positive* power when they do so (the opposite of all other devices here). This requires the *opposite* reference directions:

> Batteries use the opposite convention from other devices:
> reference current flows through the battery from reference – to reference +.

However, it is possible to force current "backwards" through a battery, and then it will *consume* energy from the circuit (as demanded by the fundamental definitions of voltage and current). This is how we recharge a rechargeable battery. Thus:

large V,   small V, $P = VI$ *where*   positive power $\Rightarrow$ battery *supplies* energy
forward   reverse   to the circuit;
current,   current,   negative power $\Rightarrow$ battery *consumes* energy
positive   negative   from the circuit.
power   power
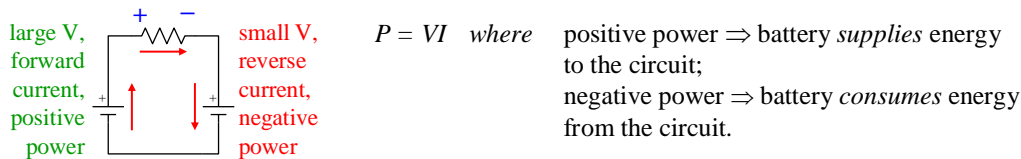
For a transformer: Transformers are a new breed, because they have 4 terminals, rather than two:

$$v_1 = M \frac{di_2}{dt} \qquad\qquad v_2 = M \frac{di_1}{dt}$$

$$v_2 = N v_1 \qquad\qquad\qquad i_2 = \frac{1}{N} i_1$$
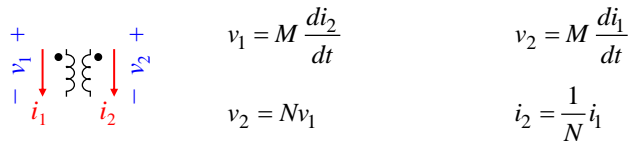
Figure 2.3 An ideal transformer relates the voltages and currents on both sides.

The dots on the two windings define the reference directions and polarities, as shown in the diagram above. Essentially, the reference directions are the same as for an inductor, except that v_1 depends on i_2, and v_2 depends on i_1. This implies that when v_1 is positive on the dot-side, v_2 is also positive on the dot-side.

In an ideal transformer, all the magnetic flux, Φ, passes through every turn of both the primary and secondary windings. From Faraday's law, $V = -d\Phi/dt$, applied to each turn of wire:

$$v_1 = -N_1 \frac{d\Phi}{dt}, \qquad v_2 = -N_2 \frac{d\Phi}{dt} \qquad \Rightarrow \qquad v_2 = \frac{N_2}{N_1} v_1 = N v_1 \qquad where \quad N \equiv \frac{N_2}{N_1} .$$

Also, an *ideal* transformer has a highly magnetizeable core such that the flux, Φ, requires virtually no current to create it. This means that the primary and secondary currents must cancel each other, leaving nearly zero MMF (magneto-motive force). Therefore:

$$N_1 i_1 = N_2 i_2 \qquad or \qquad i_2 = \frac{1}{N} i_1 \qquad where \quad N \equiv \frac{N_2}{N_1} \text{ as before} .$$

> The secondary voltage varies as the turns ratio, N.
> The secondary current varies as the *inverse* of the turns ratio.

This implies that the power into the primary equals the power out of the secondary. Unlike an inductor, an ideal transformer does not store energy. Instead, a transformer transfers energy from one side of the transformer to the other. The direction of energy transfer depends on the circuits to which the two transformer sides are connected. Details are beyond the scope of this section.

Brief Note on Phasor Analysis

This section assumes you are familiar with phasor analysis, also known as Fourier mode analysis (described later). The idea of phasor analysis is to consider one pure sinusoidal frequency at a time. Since the circuit response is linear with the excitation, a sum of sinusoidal excitations results in a response equal to the sum of the individual sinusoidal responses.

For capacitors, we have (using exp(+$i\omega t$) time dependence, which is standard for circuits):

$$i(t) = C\frac{dv}{dt} \qquad \Rightarrow \qquad \tilde{i} = i\omega C\tilde{v} \qquad where \quad \tilde{i}, \tilde{v} \text{ are the phasors for current and voltage .}$$

The phase of the derivative is positive, which means the derivative *leads* the original function. In this case, the current leads the voltage, or equivalently, the voltage lags the current (Figure 2.4 left).
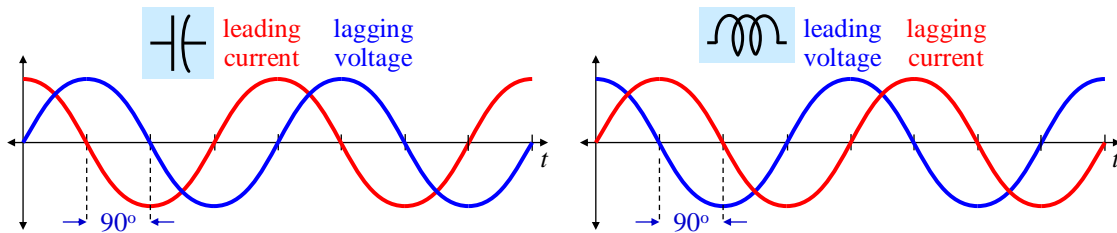
Figure 2.4 (Left) In a capacitor, voltage lags current. (Right) In an inductor, current lags voltage.

For inductors, things are reversed:

$$v(t) = L\frac{di}{dt} \qquad \Rightarrow \qquad \tilde{v} = i\omega L\tilde{i} \qquad where \quad \tilde{i}, \tilde{v} \text{ are the phasors for current and voltage .}$$

The voltage leads the current, or the current lags the voltage (above right).

> You can remember the lead/lag relationships for circuit elements mnemonically as follows:
> Capacitors oppose a change in voltage, so the voltage lags the current.
> Inductors oppose a change in current, so the current lags the voltage.

Similarly:

> You can remember the lead/lag relationships for derivatives as follows:
> Derivatives show where you're going before you get there, so they *lead* the function.
> Integrals sum up where you've been, and so *lag* the function.

3 Electromagnetic Fundamentals

Fundamental Definitions

Charge Is a Fundamental Electrical Quantity

Physically, there are 4 fundamental macroscopic quantities in the universe: length, time, mass, and charge (L, T, M, Q). This is a conceptual truth that does *not* depend on how we choose our units of measure.

The SI system defines two arbitrary units: time and mass. Time is defined in terms of atomic cesium radiation. Mass is defined in terms of the prototype kilogram stored in Paris. From there, SI defines the speed of light, a constant of nature, with an exact value. From time and c, the meter is defined by the distance light travels in vacuum in a specific time. From time, mass, and distance, the SI system defines force. As noted below, μ_0 and ε_0 are defined exact values. Then from force, time, and μ_0, SI defines the ampere, and then the coulomb.

For example, in the MKSA (aka SI) system, the unit of charge is the coulomb. However, as a matter of laboratory convenience, the SI system defines the *ampere* first, and then defines the coulomb as the amount of charge flowing past a point in 1 second, with a current of 1 ampere. This is simply because it's easier to measure magnetic forces in the lab than electrostatic forces. This does not change the theoretical reality that *charge* is one of the 4 fundamental measurable physical quantities. The ampere is a compound quantity involving charge and time. Particles have charge; they don't have currents.

There are at least three CGS systems of units, the most popular being **gaussian units**. In contrast to SI, CGS unit systems define the *charge* as the first electrical quantity, and current in terms of charge. Therefore, in CGS systems, charge is a compound unit of $M^{1/2}L^{3/2}/T$.

> The variety of human chosen definitions for units has no bearing on the nature of physics,
> or the fact that there are 4 fundamental, macroscopic, measurable quantities:
> length, time, mass, and charge.

The Coulomb Force Constant is Defined Exactly

This section is superseded by the 2019 revamp of SI. Updates TBD. ??Recall that c and μ_0 are defined exactly, which means so are ε_0 and the force constant k_e:

$$c \equiv 299\,792\,458 \text{ m/s} \qquad \mu_0 \equiv 4\pi \times 10^{-7} \qquad \Rightarrow \qquad \varepsilon_0 = \frac{1}{c^2 \mu_0} \text{ is exact, and so is } k_e :$$

$$k_e = \frac{1}{4\pi\varepsilon_0} = \frac{c^2 \mu_0}{4\pi} = c^2 \times 10^{-7} = 8\,987\,551\,787.3681764 \text{ Nm}^2/\text{C}^2 \approx 9.0 \times 10^9 \text{ Nm}^2/\text{C}^2 .$$

Units of Measure: SI vs. Gaussian

We choose not to capitalize "gaussian."

When converting *equations* between SI and Gaussian units, three simple observations will help tremendously in remembering the conversions. Three formulas are easy to remember in both units, since they are so common: (1) electric potential, (2) the speed of light, and (3) the Lorentz magnetic force. These provide all the information needed for most conversions. By comparing these equations in each set of units, we can see what substitutions are needed to take almost any equation from one system to the other. We show here conversion from SI to Gaussian, because undergraduates usually learn in SI, then have to switch to Gaussian as graduates:

Notes	SI		Gauss		SI $\rightarrow$ Gauss
(1)	$\phi(r) = q / 4\pi\varepsilon_0 r$	$\rightarrow$	$\phi(r) = q / r$	$\Rightarrow$	$\varepsilon_0 \rightarrow 1/4\pi$.
(2)	$c^2 \varepsilon_0 \mu_0 = 1$	$\rightarrow$	$\varepsilon_0 \mu_0 = 1$	$\Rightarrow$	$\mu_0 \rightarrow 4\pi / c^2$.
(3)	$\mathbf{F} = q\mathbf{v} \times \mathbf{B}$	$\rightarrow$	$\mathbf{F} = q\dfrac{\mathbf{v}}{c} \times \mathbf{B}$	$\Rightarrow$	$\mathbf{B} \rightarrow \mathbf{B}/c$.

(1) This does *not* mean that ε_0 is $1/4\pi$ in Gaussian units; in fact, $\varepsilon_0 = 1$ in Gaussian units. It means that when including the *whole package* of conversions to take an *equation* from SI to Gaussian, the final result is that ε_0 turns into $1/4\pi$. Gaussian units have no concept of ε_0 or μ_0.

(2) We used #1 above ($\varepsilon_0 \rightarrow 1/4\pi$) to eliminate ε_0.

(3) This means that the B-field in Gaussian units is *inflated* by the factor c, so when using it, you have to *deflate* it to make it equivalent to the SI **B**. For example, that's why a propagating wave has:

(SI) $E = cB$,   but   (gaussian) $E = B$.

Example: Biot-Savart Law:

(SI) $d\mathbf{B} = \dfrac{\mu_0}{4\pi} \dfrac{I d l \times \hat{\mathbf{r}}}{r^2}$ $\rightarrow$ (gaussian) $\dfrac{d\mathbf{B}}{c} = \dfrac{4\pi}{4\pi c^2} \dfrac{I d l \times \hat{\mathbf{r}}}{r^2}$ $\Rightarrow$ $d\mathbf{B} = \dfrac{1}{c} \dfrac{I d l \times \hat{\mathbf{r}}}{r^2}$.

Most (but not all) graduate EM courses seem to use gaussian units. There's almost a cult following of them. For some problems, gaussian units are better, but for many problems, SI is better. Theoreticians seem to prefer gaussian.

More examples: Maxwell's Equations (ME) in gaussian units, in vacuum:

$$\nabla \cdot \mathbf{E} = 4\pi\rho, \qquad \nabla \cdot \mathbf{B} = 0, \qquad \nabla \times \mathbf{E} = -\dfrac{1}{c}\dfrac{\partial \mathbf{B}}{\partial t}, \quad \nabla \times \mathbf{B} = \dfrac{4\pi}{c}\mathbf{J} + \dfrac{1}{c}\dfrac{\partial \mathbf{E}}{\partial t} \qquad (\text{gaussian, vacuum}).$$

Note that in gaussian units, the two source distributions $\rho(\mathbf{r})$ and $\mathbf{J}(\mathbf{r})$ get 4π factors. Since they combine into a 4-vector $j^\mu \equiv (c\rho, \mathbf{J})$, they must have the same pre-factor. In matter, the electric dipole per unit volume, **P**, and the magnetic dipole per unit volume **M**, are also sources, so they also get 4π factors:

$$\mathbf{E} = \mathbf{D} - 4\pi\mathbf{P}, \qquad\qquad \mathbf{B} = \mathbf{H} + 4\pi\mathbf{M} \qquad\qquad (\text{gaussian}).$$

We can find ME from the three simple SI $\rightarrow$ gaussian conversion rules. Gaussian units have $\varepsilon_0 = \mu_0 = 1$, so in vacuum, they don't usually use **D** or **H**; they stick with **E** and **B**. In materials, they still need **D** and **H**, so the lack of **D** and **H** is only a supposed advantage for gaussian units in vacuum. We give here the conversion only for the simpler case of vacuum.

For Gauss' laws, we start with $\mathbf{D} = \varepsilon_0 \mathbf{E} \rightarrow \mathbf{E}/4\pi$, and the second equation is unchanged:

$$\nabla \cdot \mathbf{E} = 4\pi\rho, \qquad \nabla \cdot \mathbf{B} = 0 \qquad (\text{gaussian, vacuum}).$$

For Faraday's Law, $\mathbf{B} \rightarrow \mathbf{B}/c$:

$$\nabla \times \mathbf{E} = -\dfrac{1}{c}\dfrac{\partial \mathbf{B}}{\partial t} \qquad (\text{gaussian}).$$

Two ways to convert Ampere-Maxwell: first, by direct substitution. $\mathbf{H} = \mathbf{B}/\mu_0 \rightarrow (\mathbf{B}/c)c^2/4\pi = \mathbf{B}(c/4\pi)$, and again $\mathbf{D} = \varepsilon_0 \mathbf{E} \rightarrow \mathbf{E}/4\pi$:

$$\nabla \times \mathbf{B} = \dfrac{4\pi}{c}\mathbf{J} + \dfrac{1}{c}\dfrac{\partial \mathbf{E}}{\partial t} \qquad (\text{gaussian, vacuum}).$$

Alternatively, it may be easier to first put the SI equation in terms of $\mathbf{B}$, $\mathbf{E}$, and c, by multiplying by μ_0, using $\mathbf{D} = \varepsilon_0\mathbf{E}$, and $\mu_0\varepsilon_0 = 1/c^2$, yielding $\nabla \times \mathbf{B} = \mu_0\mathbf{J} + \dfrac{1}{c^2}\dfrac{\partial \mathbf{E}}{\partial t}$. Then $\mathbf{B} \rightarrow \mathbf{B}/c$ and $\mu_0 \rightarrow 4\pi/c^2$ gives the above.

More subtle is the Poynting vector:

$$\mathbf{S} = \mathbf{E} \times \mathbf{H} \text{ (SI)} \quad \rightarrow \quad \mathbf{S} = \frac{c}{4\pi}\mathbf{E} \times \mathbf{H} \text{ (gaussian)}.$$

As above, to take $\mathbf{H}_{\text{SI}}$ to $\mathbf{H}_{\text{G}}$ we replace $\mathbf{H}_{\text{SI}}$ with $\mathbf{B}/\mu_0$, and use rules 2 and 3:

$$\mathbf{S} = \mathbf{E} \times \frac{\mathbf{B}}{\mu_0} \text{ (SI)} \quad \rightarrow \quad \mathbf{S} = \mathbf{E} \times \frac{\mathbf{B}}{c}\frac{c^2}{4\pi} = \frac{c}{4\pi}\mathbf{E} \times \mathbf{B} \text{ (gaussian)}.$$

These two examples allow a shortcut rule: $\mathbf{H}_{\text{SI}} \rightarrow (c/4\pi)\mathbf{H}_{\text{G}}$.

About the Size of Things

Classical electromagnetics requires three size scales: microscopic (atomic), mesoscopic (continuum approximated average-atomic), and macroscopic (equipment). Many references define explicitly only macroscopic and microscopic scales, but this is not tenable, and so may cause confusion. We give three examples below requiring explicit recognition of the mesoscopic scale.

The first scale is microscopic: the level of atoms, molecules, and fundamental particles such as electrons and protons. A hydrogen atom is about 10^{-10} m in diameter, and most inorganic molecules are no bigger than 10^{-9} m across. (Polymers and biomolecules can be much bigger.) Therefore, when discussing the microscopic structure of fields, even a crude approximation requires resolving 10^{-11} m or less. A proton is about 10^{-15} m, and in nonrelativistic conditions, an electron is typically much *bigger*, about the size of the atom, 10^{-10} m.

Next up, the mesoscopic scale is for "continuous" charge and current distributions, $\rho(\mathbf{r})$ and $\mathbf{J}(\mathbf{r})$, and "infinitesimal" volumes $d\tau$ [Jac p20m]. It must be large compared to microscopic, so we can average over molecules to make (with high accuracy) a continuum. The mesoscopic scale must also be much smaller than the feature size of any apparatus in question. We may take it as roughly 100 molecules long, about 10^{-7} m, or 0.1 micron. This allows for 10^6 molecules per unit of "infinitesimal" mesoscopic volume. ([Cullwick *Electromagnetism and Relativity*, 1959 p??] uses 1000 molecules, but underestimates the size of each, to reach the same scale of 10^{-7} m.)

Finally, the macroscopic scale must resolve the smallest feature size of any apparatus in question. We may take it as roughly 100 mesoscopic "sizes" long, or about 10^{-5} m, or 10 microns. This allows 10^6 mesoscopic volumes per macroscopic volume. Calculations with features smaller than this scale cannot blithely assume the continuum approximation; they must justify their approximations explicitly.

As an example, the mesoscopic scale is required to even have a concept of smooth current $i(t)$, or densities $\rho(\mathbf{r})$ and $\mathbf{J}(\mathbf{r})$. Charge is discrete, and current is the flow of discrete microscopic charges. We can well approximate these as continuous at the mesoscopic scale.

Similarly, polarization of dielectrics and permeable materials requires the mesoscopic scale for the polarization per unit volume, $\mathbf{P}(\mathbf{r})$ and $\mathbf{M}(\mathbf{r})$, to be well-approximated as a pure dipole. The mesoscopic dipole in a volume $d\tau$ comprises millions of distinct molecular dipoles.

Finally, as a practical calculational example: What is the force per unit area on each of two identical large charged plates (Figure 3.1a)?
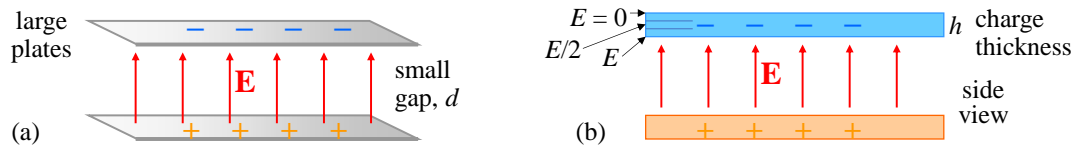
Figure 3.1 (a) Electric field between two charged plates. (b) View of charge distribution on a mesoscopic scale, illustrating its thickness, though h may be much less than d.

A naive calculation would say:

$$E = \sigma / \varepsilon, \qquad F / a = \sigma E \qquad \text{(wrong)} .$$

This is a factor of two too big. The problem lies in the fact that the thickness of the "surface" charge cannot be taken to be truly zero. Any realistic surface charge has some thickness h at the mesoscopic scale (Figure 3.1b). The charge is distributed through a finite height in the plates. We can think of the charge as a series of thin (mesoscopic) layers. Then in the top plate, E decreases "smoothly" as we proceed upward through the plate, until $E = 0$. The average E pulling on the top charges is $E/2$, so the force on the plate is:

$$F / a = \sigma \langle E \rangle = \sigma E / 2 .$$

This is also an example a common and useful idealization (zero surface charge thickness) breaking down, and giving an incorrect answer.

Is the Electric Field Real?

Is the electric field real? Or is it just a tool to help model reality? This question has been asked over the history of our understanding of EM fields, and the answer has evolved over time. For the electric field:

First, people noticed that certain particles pushed or pulled on each other. To model that, they invented the idea of "charge", and Coulomb's force law. At this point, there is no need for any fields, since electric forces are all two-particle interactions.

In more complicated situations, particularly when there is a fixed set of charges, called the "source" charges, it was convenient to pre-compute the effect of these source charges on hypothetical "test" charges. This pre-computation worked because of the linearity of the Coulomb force. The result was a vector function of space, and was given the name "electric field." At this point, the electric field is purely a mathematical convenience for describing point-charge interactions, with no physical significance. [Note that all forces, by the definition of a vector, must add.]

Later, experiments showed that if energy is to be conserved, then it is convenient to say that this "electric field" actually stores energy. Now the electric field starts to seem like a physical reality.

Still later, experiments showed that a time varying electric field could produce waves which propagate very far away from any source charges. Essentially, the electric field (together with the magnetic field) was viewed as self-sustaining. This makes fields seem quite real indeed, as they are now far separated from their sources, and seem to exist independently of them. This view is somewhat unsatisfactory, though, because the ultimate source of all EM radiation is still accelerating charges.

However, this time-dependent theory of EM fields also leads to the experimental fact that the effect of a charge on a distant charge is delayed by a propagation time. What "thing" is propagating? If fields are real entities, then the answer is clear.

Furthermore, when a particle radiates, it recoils immediately, in a direction opposite to the radiation. This implies a transfer of momentum from the particle to the radiation, which is immediate, and independent of when that radiation might later be absorbed. Thus, the energy and momentum of EM radiation appears to be quite "real," and the EM field's energy and momentum exist independently of the particles.

The final evidence favoring fields as "real" is that they contribute to the center of mass, now called **center of energy**, of a system of particles. This goes beyond conservation of energy, which counts only the total energy, because now the *location* of the energy in space matters. To satisfy the condition that the center of energy remain fixed for an isolated system, we must associate an energy density with the electric and magnetic fields, such that:

$$\text{Energy density:} \quad \rho_{electric} \propto E^2, \quad \rho_{magnetic} \propto B^2 .$$

The currently accepted classical (non-quantum) model is that fields are real things: they carry both energy and momentum, and we can say *where* the energy and momentum is located, e.g., the energy density varies as the square of the field.

Quantum mechanics is fully consistent with this classical picture, as photons (quantized EM fields) are treated fully and equally with matter as particles: they carry both energy and momentum.

The complete picture is somewhat more complicated due to the details of magnetic fields, and their interaction with electric fields. However, the above reasoning and conclusions remain valid.

Just For Reference: Faraday's Law

What is the meaning of the minus sign in Faraday's law? Faraday's law relates the induced voltage in a circuit surrounding a magnetic field to the rate of change of that field:

$$V = -\frac{d\Phi}{dt} .$$

The minus sign can *only* be interpreted with respect to a standard set of reference directions and polarities. As with reference directions in electric circuits (see *Lumped Element Circuits* elsewhere), a **reference direction** (of a current, B-field, or E-field) is the direction which is called positive in the equations. A **reference polarity** (of a voltage) is the polarity which is called positive in the equations. The actual polarity may be the same as our reference choice, or it may be opposite. We may not know until we solve some equations. But there is no problem either way, because if the actual polarity is opposite our reference choice, it will simply have a negative value.
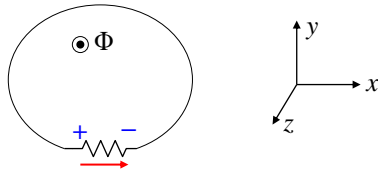

Figure 3.2 Reference directions for flux Φ, current, and reference polarity for voltage.

Now, Faraday's law: the reference direction for current is that which would produce the reference direction for Φ, and Φ is in the same direction as **B**. In the diagram above, **B** and Φ are positive in the *z*-direction (out of the page). The reference direction for the current is counter-clockwise, which produces (by the right hand rule) a B-field out of the page. We can lump the resistance of the loop into a single equivalent resistor. The reference polarity for the voltage must be consistent with the reference direction for the current, and this forces the choice shown. We now have a consistent set of reference directions and polarity for all four of the flux, B-field, current, and voltage. Positive voltage makes positive B-field.

This defines the meaning of the minus sign in Faraday's law. Lenz' law says that if the B-field changes, the voltage induced will try to drive a current that produces a B-field which opposes the change. This is the minus sign: decreasing B (negative $d\Phi/dt$) causes *positive* voltage, which boosts B. Increasing B (positive $d\Phi/dt$) causes *negative* voltage, which reduces B.

The minus sign emphasizes that if the induced voltage reinforced the change in B-field, that would induce more voltage, further changing the B-field, which induces more voltage, further changing the B-field, in a never ending death spiral of infinite current and B-field.

Making Friends With Maxwell

Maxwell's Equations (ME) describe the behavior of the E and B fields in the general case, including any nearby charges and currents. Outside materials, Maxwell's Equations are universal; they apply at all times, under all conditions: static, steady state, accelerating, relativistic. They also apply to dielectric and permeable materials, provided those materials are linear and not moving.

Ultimately, all fields result from charge and current sources, but in some situations, the charges are too far away to have any significant influence. We then use ME to study the fields alone, in the absence of charges, such as propagating waves far from their sources. The electromagnetic static and velocity fields of charges drop off as $1/r^2$, but the *radiation* fields drop off like $1/r$. Therefore, far from the sources, the static/velocity fields are negligible. Note that the radiation *power* drops like $1/r^2$, because $\mathbf{S} = \mathbf{E} \times \mathbf{H}$, and **E** and **H** each fall as $1/r$; power must drop as $1/r^2$ to conserve energy.

Here are three features of ME, and tips for remembering the equations:

- Gauss' Law, with electric but no magnetic charge.
- Faraday's Law with a minus sign to keep fields and energy finite.
- Ampere-Maxwell law, with a plus sign (opposite Faraday's Law), again to keep fields and energy finite.

We start with SI units, which are designed to make calculations of energy and power simpler, and this also makes ME simpler than in gaussian units. Therefore, we find it easier to understand and remember ME in SI units. Then for completeness, we mention gaussian units, but the concepts are independent of units.

There are a few concepts that help to remember ME. The first is that there exist electric charges (monopoles), but no magnetic charges. Two of the four ME are just Gauss' law:

$$\nabla \cdot \mathbf{D} = \rho, \qquad \nabla \cdot \mathbf{B} = 0 \qquad (\text{SI}),$$

the latter because the magnetic "charge" is always zero, since they don't exist. How do we remember it's **D** and **B**, and not **E** or **H**? Units help: **D** and ρ are "coulombs per something," so they have to go together in a simple equation. For the second equation, **B** is the physical field, which is always a result of currents and/or fundamental magnetic dipoles, which are divergenceless sources. Any superposition of divergenceless fields is also divergenceless.

The third ME is Faraday's law that a changing magnetic field produces an electric field:

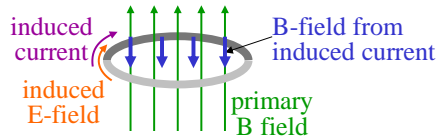$$\nabla \times \mathbf{E} = -\frac{\partial \mathbf{B}}{\partial t} \qquad (\text{SI}).$$

Figure 3.3 Faraday's Law example: An increasing primary B-field induces an E-field, which drives a current, which produces a secondary (induced) B-field that opposes the *change* (here, the increase) in B.

Picture a current loop with a B-field inside (Figure 3.3). E is induced around the loop. The negative sign is Lenz' Law: the induced E field must drive current to oppose the *change* in the B-field. If the induced E *enhanced* the change in B, that would induce more E, and then more B, in an infinite spiral of boundless energy from nothing. Note the electric and magnetic fields are on separate sides of the equation.

The fourth and last equation is the Ampere-Maxwell Law. It is the most complicated because magnetic fields are "curly" from both current sources *and* a changing E-field. The magnetic and electric terms are again on separate sides of the equation:

$$\nabla \times \mathbf{H} = \mathbf{J} + \frac{\partial \mathbf{D}}{\partial t} \qquad (\text{SI}).$$

Here again, units help: **H** is simple C/s per something, **J** is coulombs-something, and so is **D**. No **E** or **B**. For the plus sign, remember that it's $+\partial \mathbf{D}/\partial t$ by thinking: E induces B, and B induces E. When including Faraday's Law above, if both $\partial \mathbf{D}/\partial t$ and $\partial \mathbf{B}/\partial t$ had the same sign, a changing E could create a changing B that *enhances* the changing E. Again, we'd have a runaway spiral of infinite fields from nothing. Therefore the $\partial/\partial t$ terms in Faraday's Law and the Ampere-Maxwell Law must be opposite. You may recall that the $\partial/\partial t$ terms having opposite signs in ME gave us the wave-equation for EM fields with the proper signs.

For completeness, we present ME in gaussian units, which has no effect on the concepts above.

$$\nabla \cdot \mathbf{E} = 4\pi\rho, \qquad \nabla \cdot \mathbf{B} = 0, \qquad \nabla \times \mathbf{E} = -\frac{1}{c}\frac{\partial \mathbf{B}}{\partial t}, \quad \nabla \times \mathbf{B} = \frac{4\pi}{c}\mathbf{J} + \frac{1}{c}\frac{\partial \mathbf{E}}{\partial t} \qquad (\text{gaussian, vacuum}).$$

We can find these from the three simple SI $\rightarrow$ gaussian conversion rules explained elsewhere. Note that in gaussian units, the two source distributions $\rho(\mathbf{r})$ and $\mathbf{J}(\mathbf{r})$ get 4π factors. Since they combine into a 4-vector $j^\mu \equiv (c\rho, \mathbf{J})$, they must have the same pre-factor.

Stunning Phasors and Fourier Space

Phasors are a convenient way to mathematically represent oscillations. Phasors and Fourier space are used heavily in EM waves and propagation, as well as classical mechanics, quantum mechanics, and any other physics that involves oscillations. Phasors are also used extensively in engineering. Virtually all physics curricula use phasors, though many do not explain the concept, or use the word "phasor." As a

result, many physics students are limited by not clearly understanding phasors. A full understanding of phasors includes understanding ratios of phasors (e.g., impedance), and simple extensions of other concepts, such as a complex propagation vector, polarization vector, complex permittivity, and complex index of refraction. Phasors also provide a simple but sturdy foundation on which to build more advanced concepts: e.g., a Fourier transform is a phasor-valued function of frequency; a quantum wave-function can be considered a phasor-valued function of space.

Working with phasors is also called working in **Fourier space** or in **Fourier modes**. A phasor can also be called a **complex amplitude** [Gri E&M p??]. (Quantum mechanics uses the term "complex amplitude" for a complex number whose relationship to a sinusoid is somewhat abstract.)

This section requires that you understand complex numbers in both polar and rectangular form, and the basic calculus of complex functions of real variables.

We start by noting that any real-valued sinusoid is fully characterized by 3 numbers: amplitude, phase, and frequency:

$$c(t) = C\cos(\omega t + \theta), \qquad C \text{ real .}$$

> A **phasor** is a complex number that characterizes the amplitude and phase of a sinusoid.

A phasor says nothing about its frequency ω. You must know the frequency from some other condition. Combining two phasors by simple addition only makes sense if they refer to sinusoids of the same frequency. However, phasors of different frequencies are often combined by inserting the time dependence explicitly before combining (quantum mechanics does this routinely).

A phasor A (a complex number) corresponds to the sinusoid above as follows (in the engineering time convention):

$$a(t) = \underbrace{|A|}_{C} \cos\left(\omega t + \underbrace{\arg A}_{\theta}\right) = \text{Re}\left\{Ae^{+i\omega t}\right\}, \qquad where \quad \arg A \equiv \text{the complex angle of } A \text{ .}$$

The sign of the times: For engineers (including electromagnetics) and classical physics [M&T, Tay], and for AC circuit analysis, the time dependence is $e^{+i\omega t}$. For most physicists in quantum mechanics and electromagnetics, the time dependence is $e^{-i\omega t}$. You can remember this mnemonically by thinking that for wave physicists, time goes backwards. The polarity of the exponent is purely conventional, and has no physical significance. This work shows pictures for both time conventions.

The magnitude of the phasor is exactly the (real) amplitude of the sinusoid, and the complex angle of the phasor is exactly the phase of the sinusoid, i.e. the angle of the cosine at $t = 0$. The geometric interpretation of a phasor is that of a rotating stick, which casts a shadow on the horizontal axis. The length of the shadow is a sinusoidal function of time, with amplitude |A|, and starting at an angle arg A:
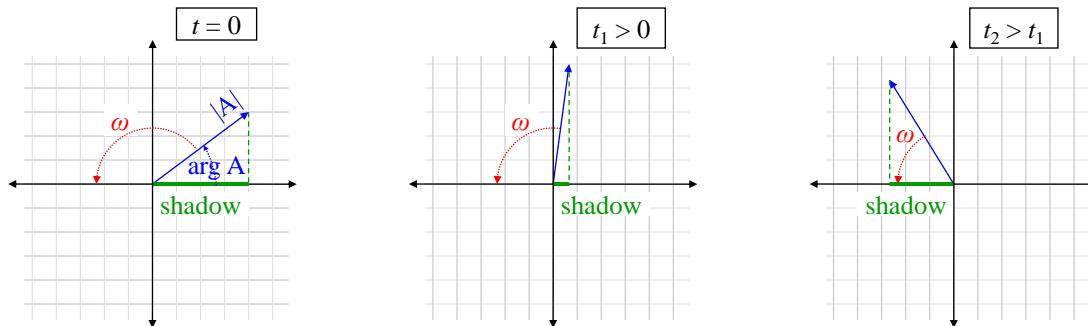
Figure 3.4   NB: For engineers, the stick rotates counter-clockwise, as shown, per $e^{+i\omega t}$.. For wave physicists, the stick usually rotates clockwise (opposite to that shown), per $e^{-i\omega t}$.

We can also view the rotation at frequency ω as complex multiplication by $e^{+i\omega t}$ (engineering), or $e^{-i\omega t}$ (physics). Recall that multiplication by a unit-magnitude complex number simply rotates another complex

number in the complex plane. Now imagine the unit-magnitude angle is not fixed, but changes linearly with time, i.e. multiply not by a fixed complex angle φ, but by an increasing angle ωt. 'ω' is the angular frequency, in rad/s. When we multiply some complex number $r\,e^{i\theta}$ by $e^{i\omega t}$, we get a complex function of time that rotates continuously around the origin in the complex plane. The magnitude of the result is fixed, because $|e^{i\omega t}| = 1$ at all times. But the angle of the result increases with time, at the rate ω.
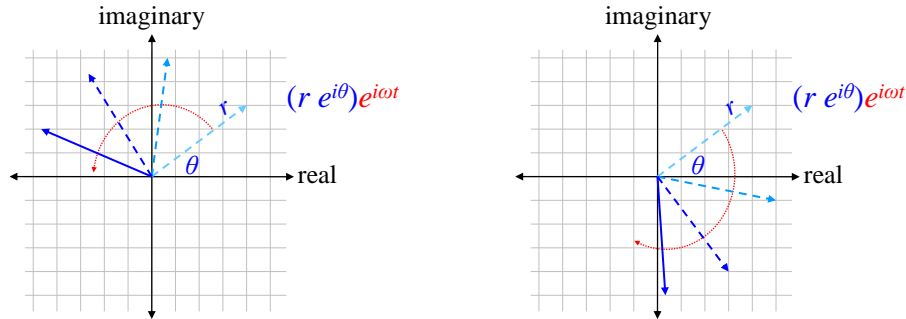
Figure 3.5 Physics: (Left) Rotation in time at negative frequency $\omega < 0$. (Right) Rotation in time at frequency $\omega > 0$. **Engineering**: (Left) Rotation in time at frequency $\omega > 0$. (Right) Rotation in time at negative frequency $\omega < 0$.

The angular frequency is not constrained to be positive; it can just as well be negative. Rotation by a negative frequency rotates in the clock-wise direction, rather than counter clockwise. Hence, both positive and negative frequencies occur in complex rotations, and in signal analysis.

Recall that any linear combination of sinusoids, of arbitrary amplitude and phase (but identical frequency, of course), is another sinusoid (at that frequency). The beauty of phasors is that the phasor representing the sum of two sinusoids is simply the (complex) sum of the phasors representing the original sinusoids (addends). The graphical demonstration of this is both simple, and illuminating:
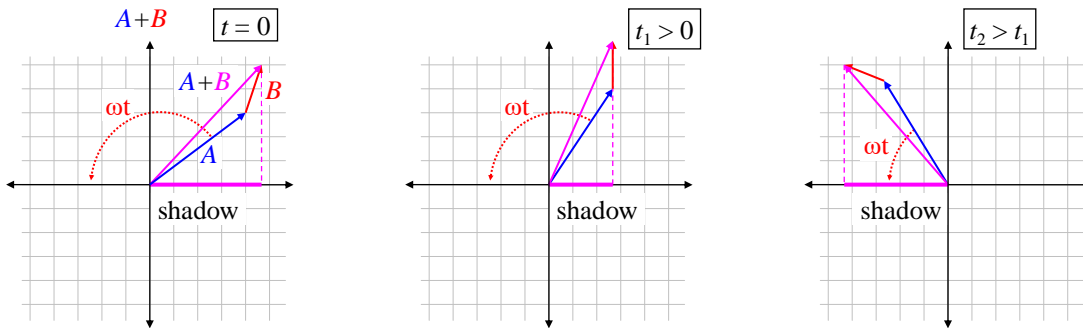
Figure 3.6 Demonstration that the sum of any 2 sinusoids is a sinusoid, and its phasor is the sum of the constituent phasors. Note that all 3 vectors rotate together. The sum of the shadows is the shadow of the vector sum. NB: For physicists, the sticks usually rotate clockwise per $e^{-i\omega t}$.

Note that the phasor for $\cos\omega t$ is 1; the phasor for $\sin\omega t$ is i (physics, or $-j$ for engineering). Thus, we can represent the sine wave $a(t)$ by its **in-phase** (cosine) and **quadrature** (sine) parts (in physics notation $e^{-i\omega t}$):

Let phasor $\qquad A = A_r + iA_i$.

Then $\qquad a(t) = \mathrm{Re}\left\{Ae^{-i\omega t}\right\} = \mathrm{Re}\left\{\left(A_r + iA_i\right)\left(\cos\omega t - i\sin\omega t\right)\right\} = A_r\cos\omega t + A_i\sin\omega t$.

We see that the real and imaginary parts of the phasor A are exactly the in-phase and quadrature components of the sine wave.

> We use complex numbers to represent sinusoids because the arithmetic (and some calculus) of complex numbers (2D vectors) is the same as the arithmetic of adding sinusoids.

12/21/2021 17:12   Page 19 of 108

Direction of the wave-vector: For traveling waves, we can see that the wave-vector points in the direction of propagation, by considering a point of constant phase on the wave. For constant phase, we must have

$$\mathbf{k} \cdot \mathbf{x} - \omega t = const \qquad \Rightarrow \qquad \mathbf{k} \text{ points in direction of propagation ,}$$

because as t increases, so must $\mathbf{k} \cdot \mathbf{x}$, and therefore the point of constant phase ($\mathbf{x} \rightarrow \mathbf{x} + d\mathbf{x}$) must move in the same direction as $\mathbf{k}$ points.

Even in the engineering convention, with time evolution given by $\exp(+i\omega t)$, the wave-vector $\mathbf{k}$ still points in the direction of propagation. The condition for propagation is $\omega t - \mathbf{k} \cdot \mathbf{x} = const$, which can be brought into the physics-convention form by absorbing a minus sign into the constant.

Phasor Calculus

We can easily see that phasors convert differential equations to algebraic equations. This is expected, because phasors are a method of Fourier analysis, which is well known for converting differential to algebraic equations. Let's take the first and second derivatives of a cosine, in both the old-fashioned real-valued way, and this new-fangled phasor way:

We can take time derivatives of phasors by noting that the time derivative of the real part of a complex function equals the real part of the time derivative:

Let $\qquad z(t) \equiv z_r(t) + iz_i(t).$

Then $\qquad \dfrac{d}{dt} \operatorname{Re}\{z(t)\} = \operatorname{Re}\left\{\dfrac{d}{dt} z(t)\right\} \qquad\qquad$ because

$$\operatorname{Re}\left\{\dfrac{d}{dt} z(t)\right\} = \operatorname{Re}\left\{\dfrac{d}{dt} z_r(t) + i\dfrac{d}{dt} z_i(t)\right\} = \dfrac{d}{dt} z_r(t) = \dfrac{d}{dt} \operatorname{Re}\{z(t)\}.$$

Then $\qquad \dfrac{d}{dt} \operatorname{Re}\{Ae^{-i\omega t}\} = \operatorname{Re}\left\{\dfrac{d}{dt} Ae^{-i\omega t}\right\} = \operatorname{Re}\{-i\omega Ae^{-i\omega t}\}$

In phasor notation: $\qquad \dfrac{d}{dt} A = -i\omega A$

For example (in the engineering convention):

$$\cos(\omega t) \qquad A = 1 \qquad \leftrightarrow \qquad \operatorname{Re}\{Ae^{i\omega t}\} = \operatorname{Re}\{1(\cos \omega t + i\sin \omega t)\} = \cos \omega t$$

$$\dfrac{d}{dt} \cos \omega t = -\omega \sin \omega t$$

$$\dfrac{d}{dt} A = i\omega A \qquad \leftrightarrow \qquad \operatorname{Re}\{i\omega Ae^{i\omega t}\} = \operatorname{Re}\{i\omega(\cos \omega t + i\sin \omega t)\}$$

$$= \operatorname{Re}\{i\omega \cos \omega t - \omega \sin \omega t\} = -\omega \sin \omega t$$

$$\dfrac{d^2}{dt^2} \cos \omega t = -\omega^2 \cos \omega t$$

$$\dfrac{d^2}{dt^2} A = (i\omega)^2 A = -\omega^2 A \quad \leftrightarrow \operatorname{Re}\{-\omega^2 Ae^{i\omega t}\} = \operatorname{Re}\{-\omega^2(\cos \omega t + i\sin \omega t)\} = -\omega^2 \cos \omega t$$

> We can also have phasors defining both space and time sinusoidal variations.
> These can be used for traveling waves.

Then the phasor carries the amplitude and phase of the traveling wave, but not its wave-vector $\mathbf{k}$ (spatial frequency) or temporal frequency ω. We must be given $\mathbf{k}$ and ω separately. In 1-dimension, the spatial derivative works like this:

Let $\qquad z(t,x) \equiv z_r(t,x) + iz_i(t,x)$.

Then $\qquad \dfrac{d}{dx} \text{Re}\{z(t,x)\} = \text{Re}\left\{\dfrac{d}{dx} z(t,x)\right\} \qquad$ similar to above with $\dfrac{d}{dt}$,

and $\qquad \dfrac{d}{dx} \text{Re}\left\{Ae^{ikx-i\omega t}\right\} = \text{Re}\left\{\dfrac{d}{dx} Ae^{ikx-i\omega t}\right\} = \text{Re}\left\{ikAe^{ikx-i\omega t}\right\}$

In phasor notation: $\qquad \dfrac{d}{dx} A = ikA$.

In higher dimension space, we replace $\partial/\partial x$ with ∇:

Let $\qquad z(t,\mathbf{r}) \equiv z_r(t,\mathbf{r}) + iz_i(t,\mathbf{r})$.

Then $\qquad \nabla \text{Re}\{z(t,\mathbf{r})\} = \text{Re}\{\nabla z(t,\mathbf{r})\} \qquad$ similar to above with $\dfrac{d}{dx}$.

Then $\qquad \nabla \text{Re}\left\{Ae^{i\mathbf{k}\cdot\mathbf{r}-i\omega t}\right\} = \text{Re}\left\{\nabla Ae^{i\mathbf{k}\cdot\mathbf{r}-i\omega t}\right\} = \text{Re}\left\{i\mathbf{k}Ae^{i\mathbf{k}\cdot\mathbf{r}-i\omega t}\right\}$

In phasor notation: $\qquad \nabla A = i\mathbf{k}A$.

Time Averages

It is often useful to compute the time average of the product of 2 sinusoids. E.g., the time average of the Poynting vector gives the effective radiation power density in W/m^2 (or power per unit area). The time average of two sinusoids does not depend on the absolute phase of either; it depends only on the relative phase of the two. Specifically,

Let $\qquad A$ be a phasor: $\qquad A = A_r + iA_i \quad \leftrightarrow \quad a(t) = \text{Re}\left\{|A|e^{-i\omega t+\arg A}\right\}$

$\qquad\qquad\qquad B$ be a phasor: $\qquad B = B_r + iB_i \quad \leftrightarrow \quad b(t) = \text{Re}\left\{|B|e^{-i\omega t+\arg B}\right\}$

Then: $\qquad \langle a(t)b(t)\rangle_t = \dfrac{1}{2}\text{Re}\{AB^*\}$. $\qquad\qquad$ How so?

One way to see this is to decompose the sinusoids into the cosine (real) and sine (imaginary) components:

$a(t) = A_r \cos\omega t + A_i \sin\omega t \qquad\qquad b(t) = B_r \cos\omega t + B_i \sin\omega t$

$\langle a(t)b(t)\rangle_t = \left\langle \left(A_r \cos\omega t + A_i \sin\omega t\right)\left(B_r \cos\omega t + B_i \sin\omega t\right)\right\rangle_t$

$\qquad\qquad = \left\langle A_r B_r \cos^2\omega t + A_r B_i \cos\omega t \sin\omega t + A_i B_r \cos\omega t \sin\omega t + A_i B_i \sin^2\omega t\right\rangle_t$

$\qquad\qquad = \left\langle A_r B_r \cos^2\omega t\right\rangle_t + \left\langle A_r B_i \cancel{\cos\omega t \sin\omega t}\right\rangle_t + \left\langle A_i B_r \cancel{\cos\omega t \sin\omega t}\right\rangle_t + \left\langle A_i B_i \sin^2\omega t\right\rangle_t$

Only the cos-cos and sin-sin terms contribute to the time average, because the time average of cos-sin is zero (they are orthogonal functions; sin is odd, cos is even). Therefore,

Use $\qquad \left\langle \cos^2\right\rangle_t = \left\langle \sin^2\right\rangle_t = \dfrac{1}{2}$

$\langle a(t)b(t)\rangle_t = \left\langle A_r B_r \cos^2\omega t\right\rangle_t + \left\langle A_i B_i \sin^2\omega t\right\rangle_t = \dfrac{1}{2}A_r B_r + \dfrac{1}{2}A_i B_i = \dfrac{1}{2}\text{Re}\{AB^*\}$

[Notice that $\text{Re}\{AB^*\}$ is analogous to the dot-product of two spatial vectors: it is the product of parallel components of A and B, i.e. $|A||B|\cos\theta$, where $\theta = \arg B - \arg A$. Only the parallel components contribute to the time average.]
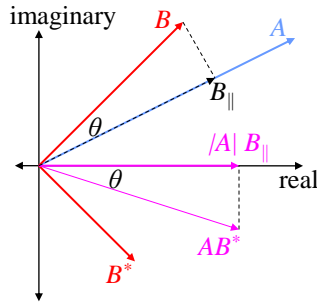
Figure 3.7 Phasors showing time-average power: only the parallel components contribute to the average.

If you'd rather grind through the integration, we can demonstrate the time average formula that way. We start with the sinusoids as real-valued cosines, and later switch back to the complex formula (after the integration):

$$a(t) = |A|\cos(\omega t + \arg A) = \text{Re}\left\{ Ae^{-i\omega t + \arg A} \right\}$$

$$b(t) = |B|\cos(\omega t + \arg B) = \text{Re}\left\{ Be^{-i\omega t + \arg B} \right\}$$

$$\langle a(t)b(t) \rangle_t = \int_{period} dt \, |A|\cos(\omega t + \arg A)|B|\cos(\omega t + \arg B)$$

$$= |AB| \int_{period} dt \, \cos(\omega t + \arg A)\cos(\omega t + \arg B)$$

$$= |AB| \int_{period} dt \, \cos(\omega t + (\arg A - \arg B))\cos(\omega t)$$

The last equation is because the integral is over a full period, so shifting the starting and ending point by a fixed time interval, or angle, doesn't change its value. This demonstrates that the time average depends only on the phase *difference* between A and B, and not their absolute phases. Finally,

$$\text{Use} \quad \cos(c+d) = \cos c \cos d - \sin c \sin d \quad where \quad c = \omega t, \; d = (\arg A - \arg B)$$

$$\langle a(t)b(t) \rangle_t = |AB| \int_{period} dt \left[\cos \omega t \cos(\arg A - \arg B) - \sin \omega t \sin(\arg A - \arg B) \right]\cos(\omega t)$$

$$= \left[|AB|\cos(\arg A - \arg B) \int_{period} dt \, \cos^2(\omega t) \right]$$

$$- \left[|AB|\sin(\arg A - \arg B) \underbrace{\int_{period} dt \, \sin \omega t \cos(\omega t)} \right]$$

$$= \frac{1}{2}|AB|\cos(\arg A - \arg B) = \frac{1}{2}\text{Re}\left\{ |AB|e^{i(\arg A - \arg B)} \right\} = \frac{1}{2}\text{Re}\left\{ |A|e^{i\arg A}|B|e^{-i\arg B} \right\}$$

$$= \frac{1}{2}\text{Re}\{AB*\} \quad QED$$

Vector Potentials I Have Known

Many E&M questions are well-suited to using the magnetic vector potential, **A**(**r**), for the magnetic fields. Given a current distribution, we can (in principle) find the vector potential from:

$$\mathbf{A}(\mathbf{r}) = \frac{\mu_0}{4\pi} \int \frac{\mathbf{J}(\mathbf{r}\,')}{|\mathbf{r} - \mathbf{r}\,'|} \, d^3 r\,' \qquad\qquad \text{[Jac 5.32 p181]} \,.$$

However, many problems start with a given magnetic field (rather than current distribution), but our analysis tools use the vector potential. How can we find the vector potential for a given magnetic field? There is no explicit "inverse curl," but usually it is fairly easy to do by inspecting the formula for curl in the coordinates of interest. For example, for a constant B-field, $\mathbf{B} = \nabla \times \mathbf{A} = B_z \hat{\mathbf{z}}$ and rectangular coordinates (chosen from other symmetries of the problem), we look at the formula for curl in those coordinates:

$$\nabla \times \mathbf{A} = \left(\frac{\partial A_z}{\partial y} - \frac{\partial A_y}{\partial z} \right) \hat{\mathbf{x}} + \left(\frac{\partial A_x}{\partial z} - \frac{\partial A_z}{\partial x} \right) \hat{\mathbf{y}} + \left(\frac{\partial A_y}{\partial x} - \frac{\partial A_x}{\partial y} \right) \hat{\mathbf{z}} \,.$$

We are given only a B_z component, so we look at the coefficient of $\hat{\mathbf{z}}$ above. We see that we can get it from either a $\partial A_y / \partial x$, or a $\partial A_x / \partial y$ term. Both terms are simple, so we choose one, and set:

$$\frac{\partial A_y}{\partial x} = B_z \qquad\qquad \Rightarrow \qquad\qquad A_y(x) = \int_0^x dx \, B_z = B_z x \,.$$

If other aspects of the problem favored an A_x instead, we could choose the other term in the curl. Note that the curl operator is linear, so if there are more components of $\mathbf{B}$, we can build them up from $\mathbf{A}$ terms, and sum them for the final $\mathbf{A}$.

Suppose we have a constant B_z in cylindrical coordinates,

$$\nabla \times \mathbf{A} = \left(\frac{1}{r} \frac{\partial A_z}{\partial \phi} - \frac{\partial A_\phi}{\partial z} \right) \hat{\mathbf{r}} + \left(\frac{\partial A_r}{\partial z} - \frac{\partial A_z}{\partial r} \right) \hat{\boldsymbol{\phi}} + \frac{1}{r} \left(\frac{\partial}{\partial r} \left(r A_\phi \right) - \frac{\partial A_r}{\partial \phi} \right) \hat{\mathbf{z}}$$

Again we see two terms to choose from: $\dfrac{1}{r} \dfrac{\partial}{\partial r} \left(r A_\phi \right)$ and $-\dfrac{1}{r} \dfrac{\partial A_r}{\partial \phi}$. If the problem has axial symmetry, then the 2nd term must be zero. Also, the 2nd term cannot have a constant derivative, and be single valued at $\phi = 0 = 2\pi$. So we choose the first term, and set

$$\frac{1}{r} \frac{\partial}{\partial r} \left(r A_\phi \right) = B_z$$

$$\frac{\partial}{\partial r} \left(r A_\phi \right) = r B_z$$

$$r A_\phi = \int r B_z \, dr = \frac{B_z}{2} r^2 \qquad \Rightarrow \qquad A_\phi = \frac{B_z}{2} r$$

Spherical coordinates don't come up all that often in magnetics problems, and are a little more complicated, but we can use the same method.

$$\nabla \times \mathbf{A} = \frac{1}{r \sin\theta} \left(\frac{\partial}{\partial \theta} \left(\sin\theta \, A_\phi \right) - \frac{\partial A_\theta}{\partial \phi} \right) \hat{\mathbf{r}} + \left(\frac{1}{r \sin\theta} \frac{\partial A_r}{\partial \phi} - \frac{1}{r} \frac{\partial}{\partial r} r A_\phi \right) \hat{\boldsymbol{\theta}} + \frac{1}{r} \left(\frac{\partial}{\partial r} \left(r A_\theta \right) - \frac{\partial A_r}{\partial \theta} \right) \hat{\boldsymbol{\phi}}$$

For a constant B_z, we must first decompose $\hat{\mathbf{z}}$ into its r and θ components:

$$\hat{\mathbf{z}} = \cos\theta\,\hat{\mathbf{r}} - \sin\theta\,\hat{\boldsymbol{\theta}} \Rightarrow \quad \begin{cases} \dfrac{1}{r\sin\theta}\left(\dfrac{\partial}{\partial\theta}\left(\sin\theta A_\phi\right) - \dfrac{\partial A_\theta}{\partial\phi}\right) = \cos\theta \\[2mm] \dfrac{1}{r\sin\theta}\dfrac{\partial A_r}{\partial\phi} - \dfrac{1}{r}\dfrac{\partial}{\partial r} rA_\phi = -\sin\theta \end{cases}$$

Choose: $A_\theta = 0 \quad \Rightarrow \qquad \dfrac{1}{r\sin\theta}\dfrac{\partial}{\partial\theta}\left(\sin\theta A_\phi\right) = \cos\theta$

$$\sin\theta A_\phi = r\int d\theta\,\cos\theta\sin\theta = \frac{r}{2}\sin^2\theta \qquad \Rightarrow \qquad A_\phi = \frac{r}{2}\sin\theta$$

Plug into the 2nd equation from $\hat{\mathbf{z}}$:

$$\frac{1}{r\sin\theta}\frac{\partial A_r}{\partial\phi} - \frac{1}{r}\frac{\partial}{\partial r} r\left(\frac{r}{2}\sin\theta\right) = -\sin\theta$$

$$\frac{1}{r\sin\theta}\frac{\partial A_r}{\partial\phi} - \sin\theta = -\sin\theta, \qquad \frac{\partial A_r}{\partial\phi} = 0 \Rightarrow \qquad A_r = const, choose \qquad A_r = 0$$

Image-ination

The methods of images are often described very tersely. We offer here a more careful discussion, allowing a motivated development for dielectric interfaces. We use gaussian units.

Introduction to images: Consider first the simplest case, a point source charge near a perfect conductor, which can be well-approximated in practice (Figure 3.8a, b). What are the fields everywhere? Conductors have mobile charges that move to cancel any fields inside them, so inside the conductor, the answer is $\mathbf{E} = \mathbf{0_v}$. There remains only the outside region, which is influenced by the source charge and by the induced mobile charge on the conductor surface. It is possible to solve this by simultaneously solving for the induced sheet charge density $\sigma(x, y)$ on the conductor surface, and the electric field. Then for each observation point, you'd have to integrate over the induced charge density. This is highly inconvenient. The method of images converts this into an easier, but equivalent, question.
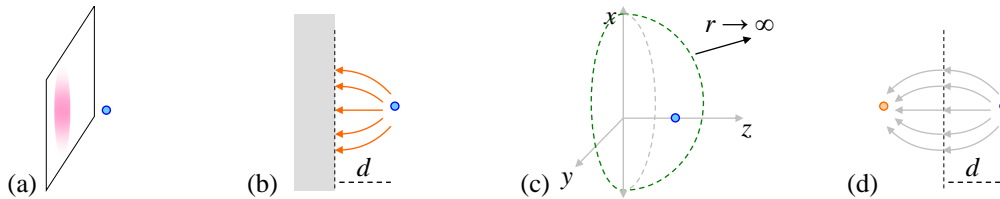
Figure 3.8 (a) A positive point charge near a conductor attracts mobile negative charge in the conductor. (b) Side view. (c) The containing surface of the region of interest is the x-y plane and a hemisphere going out to infinity. (d) System with same boundary conditions and charge density in the region of interest.

The uniqueness theorem says that the electric potentials and fields in any volume are determined by the values on the surface of the volume (either $V(\mathbf{r})$ or $\mathbf{E}(\mathbf{r})$), and the charge density within the volume. If we can find a system with the same V (or $\mathbf{E}$) on the surface, and point charge inside, it will also have the same potential and fields inside the surface. Our original system has a point charge at $(0, 0, d)$. Being a conductor, its potential is constant in the x-y plane. We *choose* to call this $V = 0$.

It is easy to *guess* that it has potential $V = 0$ on its hemispherical surface at infinity (Figure 3.8c), but guessing is unreliable. We must demonstrate it. First, recall that from far enough away, any region of charge looks essentially like a point charge. View the source charge and induced charge from far away, and they look like a net point charge. If this charge is nonzero, it will either attract or expel charge from the conductor until it is zero. In our conductor, the field from the source charge extends to infinity, and therefore so does the surface charge density $\sigma(x, y)$. Nonetheless, we can still take a limit in the x-y plane toward infinity, where the total charge approaches zero, and the potential then also approaches zero. Therefore, the potential at infinity (anywhere on the hemisphere) is also zero.

If we can find another system with the same point charge and surface of zero potential, regardless of its construction *outside* the region, it will have the same fields *inside* the region as ours. There is no general method for finding such a system. However, Figure 3.8b reminds us of a dipole field in the region (Figure 3.8d). For this dipole, the potential at infinity can be taken as 0, and therefore also in the x-y plane. Hence, it is equivalent, *only* in the region of interest ($z > 0$), to our original system. From the image system, it is trivial to solve for the potential inside the system. We leave the details to the reader.

Images for dielectric interface: A harder problem is the case of a point charge embedded in a dielectric, and near a dielectric interface. Figure 3.9a illustrates the question, where we chose $\varepsilon_2 > \varepsilon_1$ for illustration, but all our results are completely general. What are the fields everywhere? Again the method of images simplifies the problem significantly. Qualitatively, the E-field vectors bend outward at the interface, because $\mathbf{E}_{\parallel 2} = \mathbf{E}_{\parallel 1}$, but in the z-direction, the higher dielectric constant extinguishes more of the E-field: $E\perp_2 = (\varepsilon_1 / \varepsilon_2)E\perp_1$. Note that in contrast to the conducting slab, the potential $V(x, y)$ is nontrivial, and unknown. That's why this problem is harder.
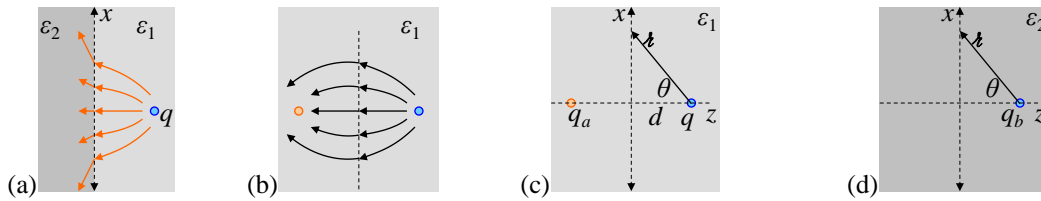

Figure 3.9 (a) Side view of dielectric interface, $\varepsilon_2 > \varepsilon_1$, with point source charge. (b) All-ε_1 medium, with its image charge. (c) Geometry of the all-ε_1 system. (d) All-ε_2 medium, with its image charge.

Consider first the right-side region, $z > 0$. We can only hope to find a simpler but equivalent system with no interface, and the same boundary values. This suggests Figure 3.9b. We have removed ε_2, made the whole system a slab of ε_1, and added a point charge q_a (much like the conductor case where we removed the conductor and replaced it with a point charge). Note that only some of the field lines terminate on the image charge, because the field lines are spreading out at the boundary (as in the original system). This tells us $|q_a| < |q|$, which we'll use as a check later.

To produce the same boundary fields in the x-y plane (and therefore along the x-axis), the image charge must have the same $\hbar(x, y)$ dependence as the source charge (Figure 3.9c), so it must be in the mirrored location, a distance d from the x-y plane. (The entire system is axially symmetric about the z-axis, so $\sigma(x, y) = \sigma(r)$, with $r \equiv (x^2 + y^2)^{1/2}$.)

From the original problem, we have two boundary conditions (constraints) that must be satisfied along the x-axis:

$$E_{x2} = E_{x1}$$

$$\varepsilon_2 E_{z2} = \varepsilon_1 E_{z1} \qquad \text{or} \qquad D_{z2} = D_{z1} \qquad \text{or} \qquad \varepsilon_2 \left.\frac{\partial V}{\partial z}\right|_{z=0^-} = \varepsilon_1 \left.\frac{\partial V}{\partial z}\right|_{z=0^+}$$

We notice that one constraint depends on the ε's, and the other doesn't. This guarantees that the two constraints will yield independent equations. Figure 3.10a shows the vacuum equivalent system: the dielectrics have been removed and replaced by a sheet charge at the interface. Near the sheet charge, it has no vertical component of $\mathbf{E}$; therefore $\mathbf{E}_{\parallel}$ is entirely from the source charge. Then the two matching conditions yield:

$$E_x(x) = \frac{q}{\varepsilon_1 \hbar^2}\sin\theta + \frac{q_a}{\varepsilon_1 \hbar^2}\sin\theta, \qquad D_z(x) = -\frac{q}{\hbar^2}\cos\theta + \frac{q_a}{\hbar^2}\cos\theta \qquad \text{(gaussian)} \tag{3.1}$$

$$\text{where} \quad \hbar, x, \theta \text{ are redundant parameters related by } \sin\theta = x/\hbar, \ \tan\theta = x/d .$$

Since E_x and D_z are unknown, we have insufficient information to solve for q_a. (Also, we haven't used ε_2, so we can't yet have a solution.) Thus, a single image charge cannot be enough to solve the original problem. But we got two independent equations, and only three unknowns. There is no other suitable

place to put another image charge in Figure 3.9c, and we must include ε_2 somehow, so our only choice is to consider the left-side region $z < 0$, as well.

Figure 3.9d shows the left-side region extended over all space, with ε_2 everywhere. There is no source charge here, only an image charge q_b, which must be outside the left-side. It must have the same $\lambda(x)$ dependence as the source, so it must be at the same location the source was. The fields at the boundary are:

$$E_x(x) = \frac{q_b}{\varepsilon_2 \lambda^2} \sin \theta, \qquad\qquad\qquad D_z(x) = -\frac{q_b}{\lambda^2} \cos \theta. \qquad\qquad (3.2)$$

We now have four equations and four unknowns: q_a, q_b, E_x, and D_z. We must find q_a and q_b such that the fields E_x and D_z in all configurations, Figure 3.9a, c, and d, are the same *along the x axis*. Therefore, we eliminate E_x and D_z from eqs. (3.1) and (3.2), yielding two independent equations for q_a and q_b. We then use q_a, q, and Figure 3.9c to find the fields in the right-side; we use q_b and Figure 3.9d to find the fields in the left-side.

$$E_x = \frac{q_b}{\varepsilon_2 \lambda^2} \sin \theta = \frac{q}{\varepsilon_1 \lambda^2} \sin \theta + \frac{q_a}{\varepsilon_1 \lambda^2} \sin \theta \qquad \Rightarrow \qquad \frac{q_b}{\varepsilon_2} = \frac{q + q_a}{\varepsilon_1}$$

$$D_z = -\frac{q_b}{\lambda^2} \cos \theta = -\frac{q}{\lambda^2} \cos \theta + \frac{q_a}{\lambda^2} \cos \theta \qquad \Rightarrow \qquad q_b = q - q_a \qquad\qquad (3.3)$$

[We can alternatively get the first equation above from the potential on the x-axis:

$$V(x) = \frac{q + q_a}{\varepsilon_1 \lambda} = \frac{q_b}{\varepsilon_2 \lambda}.]$$

Solving:

$$\frac{q - q_a}{\varepsilon_2} = \frac{q + q_a}{\varepsilon_1}, \qquad\qquad q(\varepsilon_1 - \varepsilon_2) = q_a(\varepsilon_2 + \varepsilon_1), \qquad \boxed{q_a = \frac{\varepsilon_1 - \varepsilon_2}{\varepsilon_1 + \varepsilon_2} q}$$

$$\boxed{q_b = q - q_a = \frac{2\varepsilon_2}{\varepsilon_1 + \varepsilon_2} q}$$

(This agrees with [Gri 2013 p196], [Jac 1999 p154-6], and is valid in both gaussian and SI units.)

We check that for $\varepsilon_2 > \varepsilon_1$, and $q > 0$, $q_a < 0$, and its magnitude is less than q, as noted earlier. Furthermore, the case of the conducting slab is the limit for $\varepsilon_1 = 1$, and $\varepsilon_2 \to \infty$; this agrees, with $q_a = -q$ (and q_b is not relevant). Finally, it may not be obvious, but q_a and q_b must also depend only on the ratio $\varepsilon_1/\varepsilon_2$, and not on either one independently. This is because each dielectric constant reduces the electric field from a source charge (compared to vacuum) by a multiplicative factor of ε_r. Dividing top and bottom of the above equations by ε_2 shows that they are, indeed, a function only of $\varepsilon_1/\varepsilon_2$. This is also evident already from the two RHS equations of (3.3).

But hold your horses! We have found a self-consistent pair of image systems in Figure 3.9c and d, but we have not shown that they are equivalent to the *original* system. Recall we eliminated the original system E_x and D_z to solve for q_a and q_b. To verify this method of images, we must prove the image systems are equivalent to the original system, Figure 3.9a. So we must now verify that these image charges have the same boundary values as the original system. Then the method of images still makes it easier to compute the fields everywhere *inside* the boundary.
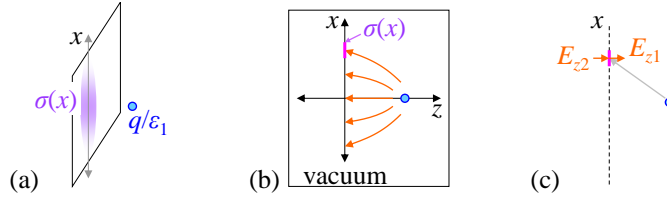
Figure 3.10 (a) Perspective view of the vacuum-equivalent sheet charge at the dielectric interface. (b) Side view of the equivalent system in vacuum, with sheet charge density $\sigma(x)$. (c) Reference E-fields for $\sigma(x)$.

So our goal now is to find the fields of the real system at the interface, as a function of the givens, q, ε_1, and ε_2. The dielectric interface acquires the equivalent of a sheet charge density, $\sigma(x, y)$, which is radially symmetric with respect to the z-axis, defined in Figure 3.9c. So without loss of generality, we confine ourselves to the x-axis. The fields in the real system are identical to those in a vacuum system which has the sheet charge as an additional source (Figure 3.10a), and the source charge reduced by ε_1: $q_{vac} = q/\varepsilon_1$. At any point x on the interface, we have three unknowns: E_{z1}, E_{z2}, and σ, and we need to find E_{z1}.

We can eliminate E_{z2} with $E_{z2}(x) = \dfrac{\varepsilon_1}{\varepsilon_2} E_{z1}(x)$. It is the sheet charge that implements the interface matching conditions, so from Gauss' Law and Figure 3.10c:

$$E_{z1} = E_{z2} + \sigma = \frac{\varepsilon_1}{\varepsilon_2} E_{z1} + \sigma \qquad \Rightarrow \qquad \sigma(x) = \left(1 - \frac{\varepsilon_1}{\varepsilon_2}\right) E_{z1}(x) .$$

We now have two unknowns: E_{z1} and σ, and the one equation above. A second equation comes from E_{z1} and the source charge. E_{z1} has two contributions: the source charge, and the sheet charge:

$$E_{z1} = \frac{-q}{\varepsilon_1 \hbar^2} + \frac{\sigma}{2} \qquad \Rightarrow \qquad \sigma = \left(\frac{\varepsilon_2 - \varepsilon_1}{\varepsilon_2}\right)\left(\frac{-q}{\varepsilon_1 \hbar^2} + \frac{\sigma}{2}\right). \tag{3.4}$$

We first solve for $\sigma(x)$:

$$\sigma\left(1 - \frac{\varepsilon_2 - \varepsilon_1}{2\varepsilon_2}\right) = \left(\frac{\varepsilon_2 - \varepsilon_1}{\varepsilon_2}\right)\frac{-q}{\varepsilon_1 \hbar^2}$$

$$\sigma\left(\frac{\varepsilon_2 + \varepsilon_1}{2\varepsilon_2}\right) = \left(\frac{\varepsilon_2 - \varepsilon_1}{\varepsilon_2}\right)\frac{-q}{\varepsilon_1 \hbar^2} \quad .$$

$$\sigma = 2\left(\frac{\varepsilon_2 - \varepsilon_1}{\varepsilon_2 + \varepsilon_1}\right)\frac{-q}{\varepsilon_1 \hbar^2}$$

We now substitute σ into the first equation of (3.4) to find E_{z1} entirely in terms of the givens, q, ε_1, and ε_2:

$$E_{z1}(x, y) = \frac{-q}{\varepsilon_1 \hbar^2}\left(1 + \frac{\varepsilon_2 - \varepsilon_1}{\varepsilon_2 + \varepsilon_1}\right) = \frac{-q}{\varepsilon_1 \hbar^2}\left(\frac{2\varepsilon_2}{\varepsilon_2 + \varepsilon_1}\right) \qquad where \quad \hbar(x, y) \equiv \sqrt{x^2 + y^2 + d^2} .$$

To establish the validity of our method of images, we confirm that this field equals the field produced by our image systems. We have already shown the two sides of the image systems agree with each other, so we need only verify one side to complete the proof. From Figure 3.9b and our earlier results, the image system gives:

Image system: $$E_{z1}(x, y) = \frac{-q + q_a}{\varepsilon_1 \hbar^2} = \frac{-q}{\varepsilon_1 \hbar^2}\left(1 - \frac{\varepsilon_1 - \varepsilon_2}{\varepsilon_1 + \varepsilon_2}\right) = \frac{-q}{\varepsilon_1 \hbar^2}\left(\frac{2\varepsilon_2}{\varepsilon_1 + \varepsilon_2}\right).$$

The image system agrees with the real system at the interface, and both are zero at infinity of a right-side hemisphere (Figure 3.8c). Therefore, the image system gives the same fields as the real system.

Solving Laplace's Equation (Boundary Value Problems)

Although these are generally covered in standard texts, we take a different approach from any we have seen. Note that boundary value problems (BVPs) are essential before moving on to waveguides. We describe solutions to Laplace's equation in eigenfunction-like terms. This has a huge advantage when moving on to waveguides, which are just big eigenfunction boundary value problems. It also is familiar to most students, because eigenfunctions appear everywhere. It can even serve as a simple introduction to eigenfunctions.

We consider both 2-D and 3-D BVPs, in all common coordinate systems: rectangular, polar, cylindrical, and spherical. All problems start from Laplace's equation:

$$\nabla^2 \Phi = 0 ,$$

plus some boundary conditions on Φ. In all common coordinate systems, the nontrivial solutions arise from **separations of variables**. Separation of variables hopes that we can write the solution as a product of functions, each of a single coordinate, such as:

$$\Phi(x, y, z) = X(x)Y(y)Z(z), \qquad \Phi(r, \phi, z) = R(r)Q(\phi)Z(z), \qquad or \qquad \Phi(r, \phi, \theta) = R(r)Q(\phi)P(\theta) .$$

We restrict ourselves to real-valued functions and parameters, since they are of most interest to E&M. [However, quantum mechanics uses complex functions which are simple linear combinations of the real functions we use here.] Finally, note that Laplace's equation is **linear**, so that the weighted sum (**linear superposition**) of any number of solutions is also a solution. We use this property extensively in what follows.

We cover the various cases in order of difficulty. We do not consider orthogonal function expansions for matching boundary conditions; see regular texts [refs??].

Two-D Laplace Solutions

Recall that the 2-D Laplacian operator is:

$$\nabla^2 = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2}, \qquad or \qquad \nabla^2 = \frac{1}{r}\frac{\partial}{\partial r}\left(r\frac{\partial}{\partial r}\right) + \frac{1}{r^2}\frac{\partial^2}{\partial \phi^2} .$$

Two-D Laplace Solutions in Rectangular Coordinates

We start with the simplest solution to $\nabla^2 \Phi = 0$, which is completely omitted from Jackson [Jac ch 2]:

$$\Phi(x, y) = Ax + By + C .$$

By inspection, this clearly satisfies Laplace's equation. Both terms of the Laplacian operator are zero. It turns out, though, that this solution doesn't come up very often in E&M problems, because it doesn't satisfy most physical boundary conditions. Nonetheless, it is a perfectly valid solution, and should be considered in some cases. [It comes up in some 1D quantum mechanical scattering problems.]

More interestingly, we consider separation of variables, which requires:

$$\Phi(x, y) = X(x)Y(y) .$$

Consider the two terms of the Laplacian in rectangular coordinates, i.e., the partial derivatives w.r.t. x and y:

$$\frac{\partial^2 \Phi}{\partial^2 x} = \left(\frac{\partial^2 X}{\partial^2 x}\right)Y(y), \qquad \frac{\partial^2 \Phi}{\partial^2 y} = X(x)\frac{\partial^2 Y}{\partial^2 y} \Rightarrow \qquad \left(\frac{\partial^2 X}{\partial^2 x}\right)Y(y) + X(x)\frac{\partial^2 Y}{\partial^2 y} = 0 .$$

If we could find $X(x)$ and $Y(y)$ with these properties, we'd be set:

$$\frac{\partial^2 X}{\partial x^2} = kX(x), \quad\quad and \quad\quad \frac{\partial^2 Y}{\partial y^2} = -kY(y), \quad then$$

$$\nabla^2 \Phi = kX(x)Y(y) - kX(x)Y(y) = \left(k - k\right)X(x)Y(y) = 0.$$

This implies that $X(x)$ and $Y(y)$ are eigenfunctions of the 2nd derivative operator. In rectangular coordinates, we can easily find such a solution. Recall that $sinh(x)$ and $cosh(x)$ (and equivalently $exp(x)$ and $exp(-x)$) satisfy this, and their eigenvalues are always positive. Also, $sin(x)$ and $cos(x)$ satisfy this, and their eigenvalues are always negative. The solutions are therefore:

$$\begin{cases} X(x) = Ae^{\alpha x} + Be^{-\alpha x}, & \dfrac{\partial^2 X}{\partial x^2} = \alpha^2 \left(Ae^{\alpha x} + Be^{-\alpha x}\right) = \alpha^2 X(x), & and \\[4mm] Y(y) = C\sin\alpha y + D\cos\alpha y, & \dfrac{\partial^2 Y}{\partial y^2} = -\alpha^2 \left(C\sin\alpha y + D\cos\alpha y\right) = -\alpha^2 Y(y) \end{cases}$$

OR

$$\begin{cases} X(x) = C\sin\alpha x + D\cos\alpha x, & \dfrac{\partial^2 X}{\partial x^2} = -\alpha^2 \left(C\sin\alpha x + D\cos\alpha x\right) = -\alpha^2 X(x), & and \\[4mm] Y(y) = Ae^{\alpha y} + Be^{-\alpha y}, & \dfrac{\partial^2 Y}{\partial y^2} = \alpha^2 \left(Ae^{\alpha y} + Be^{-\alpha y}\right) = \alpha^2 Y(y) \end{cases}$$

$$\Phi(x, y) = X(x)Y(y)$$

> As always, the boundary conditions determine the coefficients A, B, C, and D.

Note that

$$\sinh(x) = \frac{1}{2}e^x - \frac{1}{2}e^{-x}, \quad\quad \cosh(x) = \frac{1}{2}e^x + \frac{1}{2}e^{-x}, \quad\quad and$$

$$A\sin(x) + B\cos(x) = E\cos\left(x + \beta\right), \quad where \quad E = \sqrt{A^2 + B^2}, \quad \beta = \tan_{4q}^{-1}\left(B/A\right).$$

$sinh(x)$ and $cosh(x)$ and $exp(x)$ and $exp(-x)$ are linear combinations of each other, and therefore equally good solutions. However, $sinh(x)$ and $cosh(x)$ are more often convenient in E&M problems because $sinh$ has a zero, and $cosh$ has a zero derivative, which are useful for matching common boundary conditions. $exp(\pm x)$ have no zeros, and no zero derivatives. Also, $E\, cos(x + \beta)$ is equivalent to $A\, sin(x) + B\, cos(x)$. So we can conveniently write:

$$\begin{cases} X(x) = A\sinh(\alpha x) + B\cosh(\alpha x), & \dfrac{\partial^2 X}{\partial x^2} = \alpha^2 \left(A\sinh(\alpha x) + B\cosh(\alpha x)\right) = \alpha^2 X(x), & and \\[4mm] Y(y) = E\cos(\alpha y + \beta), & \dfrac{\partial^2 Y}{\partial y^2} = -\alpha^2 E\cos(\alpha y + \beta) = -\alpha^2 Y(y) \end{cases}$$

OR

$$\begin{cases} X(x) = E\cos(\alpha x + \beta), & \dfrac{\partial^2 X}{\partial x^2} = -\alpha^2 E\cos(\alpha x + \beta) = -\alpha^2 X(x), & and \\[4mm] Y(y) = A\sinh(\alpha y) + B\cosh(\alpha y), & \dfrac{\partial^2 Y}{\partial y^2} = \alpha^2 \left(A\sinh(\alpha y) + B\cosh(\alpha y)\right) = \alpha^2 Y(y) \end{cases}$$

$$\Phi(x, y) = X(x)Y(y).$$

> As always, the boundary conditions determine A, B, E, and β in this alternate form.

More on boundary conditions later.

Two-D Laplace Solutions in Polar Coordinates

In polar coordinates, (r, ϕ), we might try the same trick for separation of variables: eigenfunctions of the two Laplacian operator terms, with eigenvalues that sum to zero:

$$\nabla^2 = \frac{1}{r}\frac{\partial}{\partial r}\left(r\frac{\partial}{\partial r}\right) + \frac{1}{r^2}\frac{\partial^2}{\partial \phi^2} \qquad Let \qquad \Phi(r,\phi) = R(r)Q(\phi). \qquad Solve$$

$$\frac{1}{r}\frac{\partial}{\partial r}\left(r\frac{\partial}{\partial r}\right)R(r) = kR(r), \qquad\qquad \frac{1}{r^2}\frac{\partial^2}{\partial \phi^2}Q(\phi) = -kQ(\phi)$$

We see immediately that this won't work, because of the $1/r^2$ factor in front of the $Q(\phi)$ term, and $Q(\phi)$ has no r in it, by definition. But we don't really need *eigenfunctions* of the Laplacian terms; we just need a function $f(r, \phi)$ such that:

$$\frac{1}{r}\frac{\partial}{\partial r}\left(r\frac{\partial}{\partial r}\right)R(r) = f(r)R(r), \qquad\qquad and$$

$$\frac{1}{r^2}\frac{\partial^2}{\partial \phi^2}Q(\phi) = -f(r)Q(\phi). \qquad\qquad Then:$$

$$\nabla^2\Phi = f(r)R(r)Q(\phi) - R(r)f(r)Q(\phi) = 0$$

Given that the 2ⁿᵈ term of the Laplacian has $1/r^2$ before it, the simplest choice for $f(r)$ that we could hope for is:

Let $\qquad f(r) = \dfrac{k}{r^2}.$

Then $\qquad R(r) = Ar^\nu + Br^{-\nu} \qquad\qquad \Rightarrow \qquad \dfrac{1}{r}\dfrac{\partial}{\partial r}\left(r\dfrac{\partial}{\partial r}\right)R(r) = \dfrac{1}{r}\dfrac{\partial}{\partial r}r\left(\nu Ar^{\nu-1} - \nu Br^{-\nu-1}\right)$

$$= \frac{1}{r}\frac{\partial}{\partial r}\left(\nu Ar^\nu - \nu Br^{-\nu}\right) = \frac{1}{r}\left(\nu^2 Ar^{\nu-1} + \nu^2 Br^{-\nu-1}\right) = \boxed{\frac{\nu^2}{r^2}R(r)}$$

and $\qquad Q(\phi) = C\sin\nu\phi + D\cos\nu\phi \;\Rightarrow\; \dfrac{1}{r^2}\dfrac{\partial^2}{\partial\phi^2}Q(\phi) = \dfrac{1}{r^2}\left(-\nu^2 C\sin\nu\phi - \nu^2 D\cos\nu\phi\right) = \boxed{-\dfrac{\nu^2}{r^2}Q(\phi)}$

So $\qquad f(r) = \dfrac{\nu^2}{r^2},$

and we have found our solution to Laplace's equation in 2-D polar coordinates. [It happens now that $Q(\phi)$ *is* an eigenfunction of $\partial^2/\partial\phi^2$, but $R(r)$ is not an eigenfunction of anything.] Note that, depending on the boundary conditions, ν may or may not be an integer. [Jac sec 2.11 p75] covers this well. In particular, if ϕ goes completely around the origin, then ν must be an integer, because it must be that $Q[\nu(2\pi+\phi)] = Q(\nu\phi)$, so that Q is single-valued at every point. The solutions are thus

$$R(r) = Ar^\nu + Br^{-\nu}, \qquad Q(\phi) = C\sin\nu\phi + D\cos\nu\phi, \qquad \Phi(r,\phi) = R(r)Q(\phi).$$

Note that if the domain of Φ includes the origin, then $R(r) = Ar^\nu$, because $Br^{-\nu}$ blows up there.

There is one more case: $\nu = 0$. If this is allowed by boundary conditions and the domain of Φ, then the two Laplacian terms are both zero. There are separate solutions for this case:

$$\frac{1}{r}\frac{\partial}{\partial r}\left(r\frac{\partial}{\partial r}\right)R(r) = 0, \qquad\qquad \frac{1}{r^2}\frac{\partial^2}{\partial \phi^2}Q(\phi) = 0$$

$$R(r) = B\ln\frac{r}{A} \qquad \text{(sometimes written } R(r) = K + B\ln r\text{)} \qquad \Rightarrow$$

$$\frac{1}{r}\frac{\partial}{\partial r}\left(r\frac{\partial}{\partial r}\right)R(r) = \frac{1}{r}\frac{\partial}{\partial r}r\left(\frac{B}{r}\right) = \frac{1}{r}\frac{\partial}{\partial r}B = 0$$

$$\text{and} \qquad Q(\phi) = C + D\phi \qquad \Rightarrow \qquad \frac{1}{r^2}\frac{\partial^2}{\partial \phi^2}Q(\phi) = 0$$

$$\Phi(r,\phi) = R(r)Q(\phi) = \left(B\ln\frac{r}{A}\right)(C + D\phi)$$

Note that this possibility exists only if the domain of Φ excludes the origin, because ($\ln r$) blows up there. Also note that if the domain of Φ allows ϕ to go all the way around the origin, then $D = 0$, so that $Q(\phi)$ is single valued, and hence $Q(\phi) = 1$ (constant).

In this case of the domain surrounding but excluding the origin, we get the most general solution by combining the ($v = 0$, $D = 0$) case with the other integer v [Jac 2.71 p77]:

$$\Phi(r,\phi) = B_0\ln\frac{r}{A_0} + \sum_{v=1}^{\infty}\left[\left(A_v r^v + B_v r^{-v}\right)\left(C_v\sin v\phi + D_v\cos v\phi\right)\right].$$

Three-D Laplace Solutions

Three-D Laplace Solutions in Rectangular Coordinates

First, we extend the almost-trivial solution, which satisfies Laplace's equation by inspection:

$$\nabla^2 = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2}, \qquad\qquad \Phi(x,y,z) = Ax + By + Cz + D.$$

All 3 Laplacian terms are zero. Again, it doesn't come up much, but some references don't even mention it.

For less trivial solutions, we extend the separation of variables approach from the 2D rectangular coordinate case:

$$\Phi(x,y,z) = X(x)Y(y)Z(z).$$

We choose X(x), Y(y), and Z(z) to be eigenfunctions of their respective operators:

$$\frac{\partial^2 X}{\partial x^2} = aX(x) \qquad \Rightarrow \qquad \frac{\partial^2}{\partial x^2}\Phi(x,y,z) = a\Phi(x,y,z), \qquad\qquad \text{and similarly:}$$

$$\frac{\partial^2 Y}{\partial y^2} = bY(y), \qquad \frac{\partial^2 Z}{\partial z^2} = cZ(z) \qquad \Rightarrow \qquad \nabla^2\Phi = \left(a + b + c\right)\Phi = 0.$$

For $\Phi(x, y, z)$ to be a solution, it must have $a + b + c = 0$.

If any coordinate, say z, can be separated into the form $Cz + D$, its eigenvalue is zero, and we revert to the 2D solutions for the remaining two coordinates. In the 2D case, we had only two terms, so the eigenvalues had to be negatives of each other, so that they would sum to zero, and satisfy Laplace's equation. Recall that the eigenvalues of $sinh(x)$ and $cosh(x)$ are always positive, and the eigenvalues of $sin(x)$ and $cos(x)$ are always negative. This meant that one coordinate had to be $A\ sinh(\cdot) + B\ cosh(\cdot)$ (or equivalently $exp(\pm\ \cdot)$), and the other had to be $C\ sin(\cdot) + D\ cos(\cdot)$.

In the 3-D case, there are more possibilities, but all we require is that the 3 eigenvalues sum to zero:

$$a + b + c = 0 \ .$$

This means at least one function must be $sinh(\cdot)$ or $cosh(\cdot)$ (or equivalently $exp(\pm \cdot)$), at least one function must be sin or cos, and the last function can be either, subject to the constraint that the eigenvalues of all 3 functions sum to zero. Some example solutions are:

$$X(x) = \sin 2x, \quad Y(y) = \cos 3y, \quad Z(z) = \sinh\left(\sqrt{13}z\right) \qquad \left(-2^2 - 3^2 + 13 = 0\right)$$

$$X(x) = \cosh(3x), \ Y(y) = \sinh(4y), \ Z(z) = \cos(5z) \qquad \left(3^2 + 4^2 - 5^2 = 0\right)$$

$$X(x) = e^{5x}, \qquad Y(y) = e^{-12y}, \qquad Z(z) = 2\sin(13z) + 7\cos(13z) \qquad \left(5^2 + 12^2 - 13^2 = 0\right)$$

Three-D Laplace Solutions in Cylindrical Coordinates

In cylindrical coordinates, if the z coordinate can be separated into the form $Cz + D$, its eigenvalue is zero, and we revert to the 2D solutions for polar coordinates. This is rare.

Barring that, we now must introduce a slight variation on our separation of variables theme: the Bessel functions. We start with the standard approach of separation of variables:

$$\nabla^2 = \frac{1}{r}\frac{\partial}{\partial r}\left(r\frac{\partial}{\partial r}\right) + \frac{1}{r^2}\frac{\partial^2}{\partial \phi^2} + \frac{\partial^2}{\partial z^2}, \qquad \Phi(r, \phi, z) = R(r)Q(\phi)Z(z) \ . \qquad (3.5)$$

The sum of the above 3 Laplacian terms must be zero. From our 2D work, we have a $Q(\phi) = sin$ or $cos(v\phi)$ solution that can offset an $R(r)/r^2$ term, and a $Z(z)$ solution that can offset either a positive or negative constant term [$Z(z) = sinh$ or $cosh(kz)$ or $Z(z) = sin$ or $cos(kz)$]:

$$\frac{1}{r^2}\frac{\partial^2}{\partial \phi^2}Q(\phi) = -\frac{v^2}{r^2}Q(\phi),$$

$$\frac{\partial^2 Z}{\partial z^2}Z(z) = k^2 Z(z) \ . \qquad (3.6)$$

Wouldn't it be nice if we had an $R(r)$ that would produce the negative of the sum of those $Q(\phi)$ and $Z(z)$ terms? That is:

$$\frac{1}{r}\frac{\partial}{\partial r}\left(r\frac{\partial}{\partial r}\right)R(r) = \left(-k^2 + \frac{v^2}{r^2}\right)R(r) \ . \qquad (3.7)$$

This equation is singular at $r = 0$, which is often a physically valid radius. The singularity is easily removed by defining $R(r = 0)$ as the limit:

$$R(0) \equiv \lim_{r \to 0} R(r) \ .$$

Case 1: We call the eigenvalue of $Z(z)$ "k^2", and it is positive. $Z(z) = sinh$ or $cosh(kz)$.

We don't have any pre-existing functions that satisfy the $R(r)$ equation, so we do the obvious: we make one! Using power series solutions and some fancy math (that we don't need to know here), mathematicians define a family of functions which satisfy the above $R(r)$ equation for $k^2 > 0$, and call them **Bessel functions**. Every value of v gets its own Bessel function, named $J_v(kr)$. Recall that v can be any real number, so there is an uncountably infinite set of Bessel functions. Bessel functions have lots of magic properties (that we don't need to know here). We don't need a separate Bessel function for each value of k, because a simple change of variable eliminates k ($w = kr$) [I need to clarify this??]:

Let $w = kr, \ r = \dfrac{w}{k}$.   Then $\dfrac{\partial}{\partial r} \ \rightarrow \ k\dfrac{\partial}{\partial w}$,   and

$$\frac{1}{r}\frac{\partial}{\partial r}\left(r\frac{\partial}{\partial r}\right)R(r) = \left(-k^2 + \frac{v^2}{r^2}\right)R(r) \ \rightarrow \ \frac{k}{w}k\frac{\partial}{\partial w}\left(\frac{w}{k}k\frac{\partial}{\partial w}\right)R\left(\frac{w}{k}\right) = \left(-k^2 + \frac{k^2 v^2}{x^2}\right)R\left(\frac{w}{k}\right)$$

Define $J_v(w) \equiv R\left(\dfrac{w}{k}\right)$.   Then

$$\frac{1}{w}\frac{\partial}{\partial w}\left(w\frac{\partial}{\partial w}\right)J_v(w) = \left(-1 + \frac{v^2}{w^2}\right)J_v(w)$$

Then $R\left(\dfrac{w}{k}\right) = J_v(w) \quad \Rightarrow \quad R(r) = J_v(kr) \quad \Rightarrow$

$$\Phi(r,\phi,z) = J_v(kr)\big(C\sin v\phi + D\cos v\phi\big)\big(E\sinh kz + F\cosh kz\big)$$

Notice that since the eigenvalue of $Z(z) > 0$ ($k^2 > 0$), $Z(z) = \sinh(kz)$ or $\cosh(kz)$. (We'll get to the $Z(z) = \sin$ or cos case in a minute.)

But wait a minute, something is missing! We have 2 linearly independent solutions for $Q(\phi)$, two linearly independent solutions for $Z(z)$, but only one solution for $R(r)$. Basic differential equation theory tells us we need 2 solutions for *each* of the 3 variables. Well it turns out, for non-integer v, we have two Bessel functions: J_v and J_{-v}. So the solution to Laplace's equation is really

$$R(r) = AJ_v(kr) + BJ_{-v}(kr)$$

$$\Phi(r,\phi,z) = \big(AJ_v(kr) + BJ_{-v}(kr)\big)\big(C\sin v\phi + D\cos v\phi\big)\big(E\sinh kz + F\cosh kz\big), \quad v = \text{non-integer}$$

Finally, since Laplace's equation is linear, the linear combination of any two solutions is also a solution:

$$\boxed{\Phi(r,\phi,z) = \sum_v \ \big(A_v J_v(kr) + B_v J_{-v}(kr)\big)\big(C_v \sin v\phi + D_v \cos v\phi\big)\big(E_v \sinh kz + F_v \cosh kz\big),}$$
$$v = \text{non-integer}$$

Note that each value of v has its own set of A, B, C, D, E, and F. Also, the values of v over which to sum are determined by the problem from the allowed range of ϕ and the boundary conditions.

However, a very common case is v = integer (solution valid for ϕ all around the axis). In that case, J_v is a multiple of J_{-v}, and the two solutions are not independent. So we do more math, and find a variant of the Bessel function which satisfies the equation, and *is* linearly independent. Such a function is a **Neumann function** (pronounced noy'-mon), also called a **Bessel function of the second kind**. Our solution is then:

$$R(r) = AJ_v(kr) + BN_v(kr), \qquad N_v(kr) \equiv \text{Neumann function}, \qquad v = \text{integer}$$

$$\boxed{\Phi(r,\phi,z) = \sum_{v=0}^{\infty} \ \big(A_v J_v(kr) + B_v N_v(kr)\big)\big(C_v \sin v\phi + D_v \cos v\phi\big)\big(E_v \sinh kz + F_v \cosh kz\big)}$$

Note that each value of v has its own set of A, B, C, D, E, and F.

An important property of the Bessel and Neumann functions is:

> Bessel functions are regular at the origin, and Neumann functions blow up there.

This is analogous to the 2D polar case, where a domain that includes the origin must exclude the r^{-n} solution, which blows up. In 3D, a cylinder that extends to $r = 0$ must exclude the Neumann function solution, which blows up at $r = 0$.

Case 2: For the case where $Z(z) = sin$ or $cos(kz)$, its eigenvalue is negative, and we have:

$$\frac{1}{r^2}\frac{\partial^2}{\partial\phi^2}Q(\phi) = -\frac{\nu^2}{r^2}Q(\phi),$$

$$\frac{\partial^2 Z}{\partial z^2}Z(z) = -k^2 Z(z)$$

$$\frac{1}{r}\frac{\partial}{\partial r}\left(r\frac{\partial}{\partial r}\right)R(r) = \left(+k^2 + \frac{\nu^2}{r^2}\right)R(r)$$

This is a different differential equation for $R(r)$, and the Bessel functions don't solve it. No problem; we just make up a new set of functions for this case, do the fancy math/power series solutions, and call them **modified Bessel functions** (not to be confused with Bessel functions of the 2nd or 3rd kinds). We use the same change of variable ($w = kr$), and call the solutions I_ν and K_ν [Jac ??]:

$$\frac{1}{r}\frac{\partial}{\partial r}\left(r\frac{\partial}{\partial r}\right)R(r) = \left(+k^2 + \frac{\nu^2}{r^2}\right)R(r) \qquad \Rightarrow \qquad R(r) = AI_\nu(kr) + BK_\nu(kr), \qquad \nu \text{ integer}$$

.

As before, I_ν is regular at the origin, and K_ν blows up there. [For ν non-integer, I assume we don't use K_ν, and instead use I_ν and $I_{-\nu}$, as in the $Z(z) = sinh$ or $cosh$ case. Need reference??]

Three-D Laplace Solutions in Spherical Coordinates with Axial Symmetry

The Laplacian in spherical coordinates is:

$$\nabla^2 = \frac{1}{r^2}\frac{\partial}{\partial r}\left(r^2\frac{\partial}{\partial r}\right) + \frac{1}{r^2\sin\theta}\frac{\partial}{\partial\theta}\left(\sin\theta\frac{\partial}{\partial\theta}\right) + \frac{1}{r^2\sin^2\theta}\frac{\partial^2}{\partial\phi^2},$$

$$\text{Note:} \quad \frac{1}{r^2}\frac{\partial}{\partial r}\left(r^2\frac{\partial}{\partial r}\right) \equiv \frac{1}{r}\frac{\partial^2}{\partial r^2}(r\bullet) \equiv \frac{2}{r}\frac{\partial}{\partial r} + \frac{\partial^2}{\partial r^2}$$

We start with a common special case of spherical coordinates: axial symmetry. In this case, the whole system is unchanged by any rotation about the z-axis. All of our results are thus independent of ϕ. We now seek a 2D solution $\Phi(r, \theta)$, embedded in a 3D physical space. Applying the simplified Laplacian, and separation of variables, we seek something similar to the 2D polar case:

$$\nabla^2 = \frac{1}{r^2}\frac{\partial}{\partial r}\left(r^2\frac{\partial}{\partial r}\right) + \frac{1}{r^2\sin\theta}\frac{\partial}{\partial\theta}\left(\sin\theta\frac{\partial}{\partial\theta}\right) \qquad \text{(with axial symmetry)}$$

$$\Phi(r,\theta,\phi) = \Phi(r,\theta) = R(r)P(\theta)$$

As with the 2D polar case, the two operators must reduce to $\pm k/r^2$ for $R(r)$ and $P(\theta)$:

$$\frac{1}{r^2\sin\theta}\frac{\partial}{\partial\theta}\left(\sin\theta\frac{\partial}{\partial\theta}\right)P(\theta) = -\frac{k}{r^2}P(\theta)$$

$$\frac{1}{r^2}\frac{\partial}{\partial r}\left(r^2\frac{\partial}{\partial r}\right)R(r) = \frac{k}{r^2}R(r)$$

The difference from the 2D polar case is that the differential operators for the two Laplacian terms are different. We don't have any standard functions that provide the desired results, so (as with cylindrical coordinates), we invent them.

First, to find $P(\theta)$: Fancy math → **Legendre polynomials**, $P_l(cos\ \theta)$. These are a countably infinite set of polynomials, labeled with their (integer) order l, which when applied to $cos\ \theta$, satisfy the above equation [Arf 11.37 p565]:

$$\frac{1}{r^2 \sin\theta}\frac{\partial}{\partial\theta}\left(\sin\theta\frac{\partial}{\partial\theta}\right)P_l(\cos\theta) = -\frac{l(l+1)}{r^2}P_l(\cos\theta)\,.$$

The first few Legendre polynomials, and the corresponding $P(\theta)$, are:

$$P_0(x)=1 \qquad P_1(x)=x \qquad P_2(x)=\frac{1}{2}\left(3x^2-1\right) \qquad P_3(x)=\frac{1}{2}\left(5x^3-3x\right)$$

$$P_4(x)=\frac{1}{8}\left(35x^4-30x^2+3\right), \qquad \text{or}$$

$$P_1(\cos\theta)=\cos\theta \qquad P_2(\cos\theta)=\frac{1}{2}\left(3\cos^2\theta-1\right) \qquad P_3(\cos\theta)=\frac{1}{2}\left(5\cos^3\theta-3\cos\theta\right)$$

Now, to find $R(r)$: Consider

$$R(r)=Ar^l+Br^{-(l+1)} \qquad \Rightarrow \qquad \frac{1}{r^2}\frac{\partial}{\partial r}\left(r^2\frac{\partial}{\partial r}\right)R(r)=\frac{1}{r^2}\frac{\partial}{\partial r}\left[r^2\left(Alr^{l-1}-(l+1)Br^{-(l+2)}\right)\right]$$

$$=\frac{1}{r^2}\frac{\partial}{\partial r}\left(Alr^{l+1}-(l+1)Br^{-l}\right)=\frac{1}{r^2}\left(Al(l+1)r^l+l(l+1)Br^{-(l+1)}\right)=\frac{l(l+1)}{r^2}R(r)$$

Finally, since Laplace's equation is linear, the linear combination of any two solutions is also a solution:

$$\boxed{\Phi(r,\theta,\phi)=\Phi(r,\theta)=R(r)P(\theta)=\sum_{l=0}^{\infty}\left(A_l r^l+B_l r^{-(l+1)}\right)P_l\left(\cos\theta\right)} \qquad \text{(with axial symmetry)}$$

Note that each value of l has its own set of A and B.

Three-D Laplace Solutions in Spherical Coordinates without Axial Symmetry

$$\nabla^2=\frac{1}{r^2}\frac{\partial}{\partial r}\left(r^2\frac{\partial}{\partial r}\right)+\frac{1}{r^2\sin\theta}\frac{\partial}{\partial\theta}\left(\sin\theta\frac{\partial}{\partial\theta}\right)+\frac{1}{r^2\sin^2\theta}\frac{\partial^2}{\partial\phi^2}, \qquad Note:\ \frac{1}{r^2}\frac{\partial}{\partial r}\left(r^2\frac{\partial}{\partial r}\right)\equiv\frac{1}{r}\frac{\partial^2}{\partial r^2}\left(r\bullet\right)$$

$$\Phi(r,\theta,\phi)=R(r)P(\theta)Q(\phi)$$

Without axial symmetry, we need a real function of ϕ. Since the prefactor of $\partial^2/\partial\phi^2$ is $1/(r^2\ sin^2\ \theta)$, which has no ϕ in it, we simply need the eigenfunction of $\partial^2/\partial\phi^2$, which is, as always, sin or $cos(m\phi)$, with eigenvalue $-m^2$:

$$Q(\phi)=C\sin m\phi+D\cos m\phi \qquad \Rightarrow \qquad \frac{\partial^2}{\partial\phi^2}Q(\phi)=-m^2Q(\phi)\,.$$

The Legendre polynomials no longer work for θ, because now we need to satisfy:

$$\frac{1}{r^2\sin\theta}\frac{\partial}{\partial\theta}\left(\sin\theta\frac{\partial}{\partial\theta}\right)P(\theta)=\left(-\frac{l(l+1)}{r^2}+m^2\right)P(\theta)\,,$$

to cancel the new $-m^2$ term from the ϕ part above. For integer m, more fancy math creates a new set of functions, the **associated Legendre functions**, $P_l^m(cos\ \theta)$, which satisfy the above equation for given l and m:

$$\frac{1}{r^2 \sin\theta} \frac{\partial}{\partial\theta}\left(\sin\theta \frac{\partial}{\partial\theta}\right) P_l^m(\cos\theta) = \left(-\frac{l(l+1)}{r^2} + m^2\right) P_l^m(\cos\theta)$$

In other words, combining the θ and ϕ dependence, we see that (excluding r dependence):

$$\nabla^2 P_l^m(\cos\theta)Q(\phi) = \left(-\frac{l(l+1)}{r^2} + \cancel{m^2}\right) P_l^m(\cos\theta)Q(\phi) - \cancel{m^2 P_l^m(\cos\theta)Q(\phi)}$$

$$= -\frac{l(l+1)}{r^2} P_l^m(\cos\theta)Q(\phi)$$

This is exactly the same equation we had for the axial symmetry case, so $R(r)$ is not affected. Note that m is restricted to $-l \le m \le l$. Also, when $m = 0$, the associated Legendre function is simply the original (non-associated) Legendre polynomial of order l. [Associated Legendre functions are only polynomials for even l; for odd l, they have a square root in them.]

> The loss of axial symmetry introduced a ϕ dependence,
> which was cancelled by new functions of θ.
> The radial function was not involved, and the radial function cannot "see" any dependence on ϕ.

Now $R(r)$: compare to the axially symmetric case (above): we added here a $-m^2$ eigenvalue for ϕ, but cancelled all of it in the new $P_l^m(\cos\theta)$. Therefore, the radial function doesn't know anything about ϕ, and our axially symmetric $R(r)$ still works as before:

$$R(r) = Ar^l + Br^{-(l+1)}.$$

Finally, since Laplace's equation is linear, the linear combination of any set of solutions is also a solution:

$$\Phi(r, \theta, \phi) = R(r)P(\theta)Q(\phi) = \sum_{l=0}^{\infty}\left(A_l r^l + B_l r^{-(l+1)}\right) \sum_{m=-l}^{l} P_l^m(\cos\theta)\left(C_{lm}\sin m\phi + D_{lm}\cos m\phi\right)$$

Note that each value of l has its own set of A_l, and B_l, and each (l, m) pair has its own C_{lm}, and D_{lm}.

Boundary Conditions Determine Solutions

Given all our choices in the infinite set of solutions, how do we use boundary conditions to find the unique solution that they demand?

In rectangular coordinates of 2 or 3 dimensions, we have to choose between sin/cos, sinh/cosh, and e^{+x}/e^{-x} for each coordinate. The boundary conditions determine the simplest form for Laplace solutions. It helps to consider the following facts about these functions:

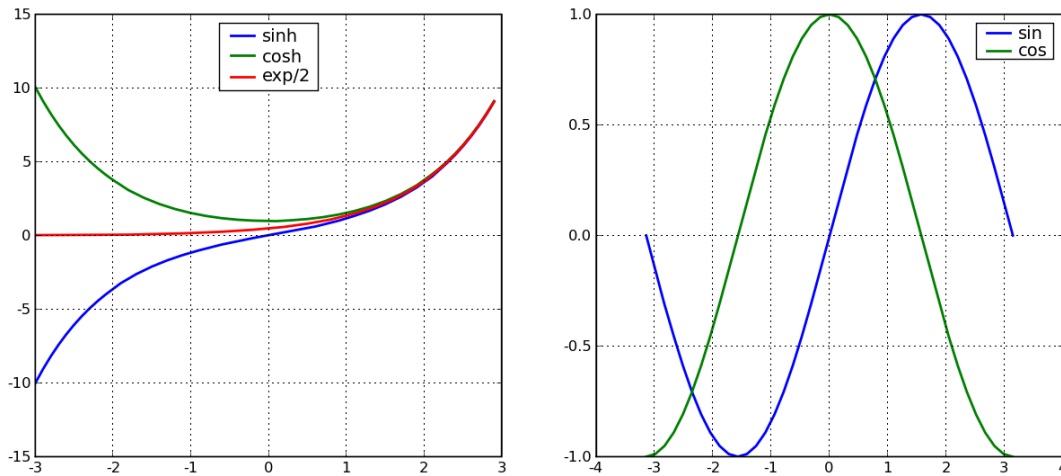| Function	Range	Zeros at	Derivative	Derivative range
sin	–1 to 1	$n\pi$	cos	–1 to 1
cos	–1 to 1	$(n+1/2)\pi$	–sin	–1 to 1
sinh	$-\infty$ to $+\infty$	0	cosh	≥ 1
cosh	1 to ∞	none	sinh	$-\infty$ to ∞
exp(+x)	> 0 to ∞	none	exp(+x)	> 0 to ∞
exp(–x)	> 0 to ∞	none	–exp(–x)	$-\infty$ to < 0

Figure 3.11 (Left) sinh(x) (blue) is odd, and has one zero. cosh(x) (green) is even, and has no zeros. $e^x/2$ (red) has no symmetry and no zeros. (Right) sin(x) (blue) is odd. cos(x) (green) is even.

Recall that in rectangular coordinates, at least one coordinate must be sin/cos, and another other must be sinh/cosh or e^{+x}/e^{-x}. The following two cases often allow a simple choice for the coordinate which is sin/cos:

(1)   Given $\Phi(x{=}0, y) = 0$ and $\Phi(x{=}b, y) = 0$

We try $X(x) = A \sin(\pi x / b)$, because that is the only function above with zeros at $x = 0$ and $x = b$. Similarly for BCs in y, and in z.

(2)   Given $\partial/\partial x\, \Phi(x{=}0, y) = 0$ and $\partial/\partial x\, \Phi(x{=}b, y) = 0$

We try $X(x) = A \cos(\pi x / b)$, because that is the only function above with zero derivatives at $x = 0$ and $x = b$. Similarly for BCs in y, and in 3D, for z.

In 3D, we might find two coordinates which are both sin/cos. Note, though, that other BCs can also be solved with sin/cos, so the above choices are just two of many possible conditions.

Now having a good idea which coordinate(s) is/are sin/cos, we use the other BCs with the table above to choose the simplest function for the other coordinate(s). These functions must be sinh/cosh or e^{+x}/e^{-x}, because they have the positive eigenvalues needed to complement the negative ones from the sin/cos functions already identified in the solution. Some BCs allow us to write the solution as a single function, instead of a linear combination. For example:

(3)   Given $\Phi(x, y{=}0) = 0$ and $\Phi(x, y{=}b) = L$

We try $Y(y) = A \sinh(ky)$, because that is the only function above with zero at $y{=}0$ and non-zero at $y{=}b$. Similarly for x, and z in 3D.

(4)   Given $\Phi(x, y{=}0) = K$ and $\Phi(x, y{=}b) = L$

We try $Y(y) = A \cosh(ky)$ or $A \exp(ky)$, because those are the functions above non-zero at $y{=}0$ and at $y{=}b$.

(5)   Given $\partial/\partial y\, \Phi(x, y{=}0) = 0$ and $\Phi(x, y{=}b) = L$

We try $Y(y) = A \cosh(ky)$, because that is the only function above with zero derivative at $y{=}0$, and non-zero value at $y{=}b$.

(6)   Given $\partial/\partial y\, \Phi(x, y{=}0) = K$ and $\partial/\partial y\, \Phi(x, y{=}b) = L$

We try $Y(y) = A \sinh(ky)$ or $A \exp(ky)$, because those are the functions above with non-zero derivatives at $y=0$, and at $y=b$.

(7)   Given symmetric BCs, such as $\partial/\partial y \; \Phi(x, y = -a) = K$ and $\partial/\partial y \; \Phi(x, y = +a) = K$

We try $Y(y) = A \cosh(ky)$, because it is the only symmetric function above with non-zero derivatives at $y = \pm a$.

(8)   Given anti-symmetric BCs, such as $\partial/\partial y \; \Phi(x, y = -a) = -K$ and $\partial/\partial y \; \Phi(x, y = +a) = +K$

We try $Y(y) = A \sinh(ky)$, because it is the only anti-symmetric function above with non-zero derivatives at $y = \pm a$.

If the BCs do not allow any of the above simple solutions, then the solution must be a linear combination of these functions, with coefficients that must be determined by solving simultaneous equations, in the usual way.

Respecting Orthogonality

TBS: orthogonality wrt to linear operators. Orthogonality of interest of Bessel functions.

Multipoles: Dipoles and Quadrupoles

Suppose we have a small blob of charges, and we look at it from a distance. We are at the origin:
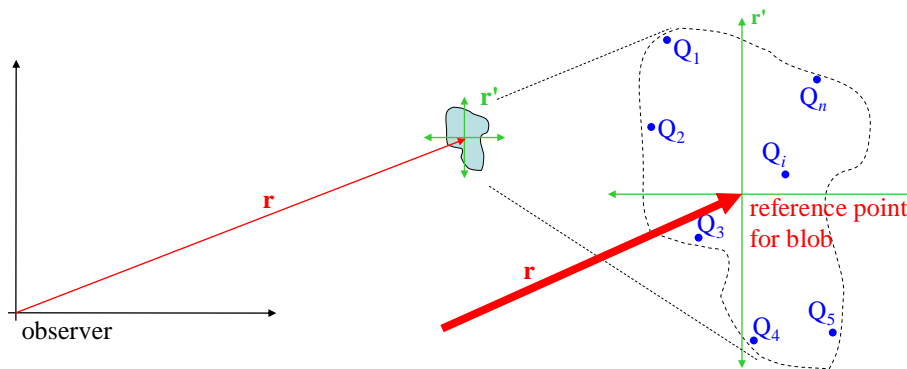

Figure 3.12 A distribution of charges, admired from afar.

We wish to know the potential we feel (here at the origin) from this blob of charges. From that, we can compute the E-field. We suppose that our observation distance is large compared to the size of the blob, as shown, so we can use the **far field approximation**. This is a valuable and widely used approximation. We assume the blob consists of n point charges, with total charge:

$$Q_{total} = \sum_{i=1}^{n} Q_i \; .$$

To zeroth order, the potential is simply:

$$\Phi = \frac{Q_{total}}{r} \qquad where \quad r \equiv |\mathbf{r}| .$$

If all the point charges were exactly at $\mathbf{r}$, this would be the exact potential. Now consider the effect of moving one of the point charges, Q_i, from the point $\mathbf{r}$ by some small amount, $\mathbf{r}'_i$. The potential at the origin due to that charge changes by

$$\Delta\Phi_{Q_i} = \frac{Q_i}{|\mathbf{r} + \mathbf{r}'|} - \frac{Q_i}{r} = Q_i \left(\frac{1}{|\mathbf{r} + \mathbf{r}'|} - \frac{1}{r} \right) = Q_i \left(\Delta \frac{1}{r} \right).$$

That is, the change in potential (seen by us at the origin) equals the charge times the change in $1/r$. The function $1/r$ is a scalar field: it assigns a number to every point in space. Therefore, we can use its gradient to approximate small changes in $1/r$:

$$\Delta \frac{1}{r} \approx \nabla \left(\frac{1}{r} \right) \cdot \mathbf{r}' \quad \Rightarrow \quad \Delta \Phi_{Q_i} \approx Q_i \nabla \left(\frac{1}{r} \right) \cdot \mathbf{r}'.$$

We can find the gradient direction with a simple geometric argument: the gradient points in the direction of greatest rate of increase.

direction of fastest
increase in **r**

r
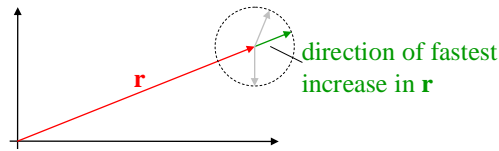
The greatest rate of increase of $1/r$ is opposite to **r** itself: r increases fastest along itself. Therefore, $1/r$ *decreases* fastest along **r**, and $1/r$ *increases* fastest opposite to **r**: The magnitude of the gradient then follows from a simple derivative:

$$\frac{d}{dr} \frac{1}{r} = -\frac{1}{r^2} \quad \Rightarrow \quad \nabla \left(\frac{1}{r} \right) = \frac{-1}{r^2} \hat{\mathbf{r}}.$$

We use this to approximate the change in potential at the origin due to moving Q_i a small distance $\mathbf{r}'$:

$$\Delta \Phi_{Q_i} = Q_i \nabla \left(\frac{1}{r} \right) \cdot \mathbf{r}' = Q_i \frac{-1}{r^2} \hat{\mathbf{r}} \cdot \mathbf{r}'.$$

Now let's move each of the charges by its own small displacement $\mathbf{r}'_i$. By the principle of superposition, we just add up the changes due to moving each charge:

$$\Delta \Phi_{total} = \sum_{i=1}^{n} Q_i \frac{-1}{r^2} \hat{\mathbf{r}} \cdot \mathbf{r}'_i = \frac{-1}{r^2} \hat{\mathbf{r}} \cdot \left(\sum_{i=1}^{n} Q_i \mathbf{r}'_i \right) = \frac{-1}{r^2} \hat{\mathbf{r}} \cdot \mathbf{p}, \qquad where \quad \mathbf{p} \equiv \sum_{i=1}^{n} Q_i \mathbf{r}'_i.$$

Note that **p** is independent of where the blob is located, i.e. independent of the relative position of observer and blob. It is solely a function of the blob, i.e. the distribution of charges. **p** is called the **dipole moment** of the charge distribution. We see that, to first order, the potential due to a net-neutral blob (I mean, charge distribution) falls off as $1/r^2$, and is zero along a line through the blob perpendicular to **p**.

Note that often we consider the blob to be at the origin, and compute the field at some position **r**. This simply reverses the sign of **r** above, to yield the more common formula:

$$\Delta \Phi_{total} = \frac{1}{r^2} \hat{\mathbf{r}} \cdot \mathbf{p} \quad \Rightarrow \quad \Phi \approx \frac{Q_{total}}{r} + \frac{1}{r^2} \hat{\mathbf{r}} \cdot \mathbf{p} + ...$$

A subtle point: the dipole moment of the blob *may* depend on the reference point within the blob from which the **r**' are measured. However, if the total charge is zero, the dipole moment is then independent of the reference point within the blob. Why??

Note that the dipole moment does not completely characterize the charge distribution, nor is it necessarily the most significant component of such a characterization. It is simply the $1/r^2$ component of the potential due to the charge distribution in the far field approximation.

Quadrupoles

The quadrupole moment extends the above approximation to 2nd order. In short, we could say, "Just extend the dipole to the 3-variable 2nd order Taylor expansion of $1/r$," and claim we're done. But that's too arcane, so let's see what it really means.

First, let's recall the meaning of the Taylor expansion of a simple function, $f(x)$, about x_0. To first order, we approximate the first derivative as constant. Hence:

$$\Delta f \approx f'(x_0)\Delta x \quad \Rightarrow \quad f(x + \Delta x) \approx f(x_0) + f'(x_0)\Delta x.$$

To 2nd order, we approximate the 2nd derivative as constant, and therefore the first derivative changes *linearly* with x:

$$\Delta f' \approx f''(x_0)\Delta x \quad \Rightarrow \quad f'(x + \Delta x) \approx f'(x_0) + f''(x_0)\Delta x.$$

We can then find the *average* first derivative f' over the interval Δx:

$$f'_{avg} = f'(x_0) + \frac{1}{2}\Delta f' = f'(x_0) + \frac{1}{2}f''(x_0)\Delta x.$$

This average f' incorporates 2nd order effects. Now use this average f' in the first order estimate for $f(x + \Delta x)$:

$$\Delta f \approx f'_{avg}(x_0)\Delta x = \left(f'(x_0) + \frac{1}{2}f''(x_0)\Delta x \right)\Delta x = f'(x_0)\Delta x + \frac{1}{2}f''(x_0)\left(\Delta x\right)^2$$

$$\Rightarrow \quad f(x + \Delta x) \approx f(x_0) + f'(x_0)\Delta x + \frac{1}{2}f''(x_0)\left(\Delta x\right)^2$$

Dipole moments are the 1st order Taylor coefficients in the expansion of potential due to a blob of charges. Quadrupole moments are simply the 2nd order Taylor coefficients in the same expansion:

$$\Phi(\mathbf{r}) = \frac{Q_{total}}{r} + \frac{\mathbf{p} \cdot \hat{\mathbf{r}}}{r^2} + \frac{quadrupole\ term}{r^3} + \dots\ .$$

We follow a procedure similar to the 2nd order Taylor expansion of $f(x)$ above, but with two additional complications: (1) our function $1/r$ is a function of 3-D space, so its first derivative (gradient) is a vector, not a number; and (2) the 2nd derivative of $1/r$ is the gradient of a gradient, which is the gradient of a vector, which is a rank-2 tensor. We take these two issues in turn. As is traditional, we'll stick to Cartesian coordinates.

[Aside: purists might object to the following explanation, but our goal here is to describe electromagnetics, not differential geometry.]

First, the gradient of a scalar field at a point is a set of 3 *numbers* (a vector), which tell us how the scalar field varies as we move in the x, y, and z directions:

$$\nabla f \equiv \left(\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}, \frac{\partial f}{\partial z} \right).$$

We usually write the gradient as a vector, but for now, you can think of it as just a set of 3 numbers. We use the 3 numbers to approximate changes in f given small changes in position, dx, dy, dz:

$$\Delta f \approx \frac{\partial f}{\partial x}dx, \qquad \Delta f \approx \frac{\partial f}{\partial y}dy, \qquad \Delta f \approx \frac{\partial f}{\partial z}dz.$$

Further, for differentially small changes dx, dy, dz, the Δf are independent and can be summed, to get the familiar total derivative rule:

$$\Delta f \approx \frac{\partial f}{\partial x}dx + \frac{\partial f}{\partial y}dy + \frac{\partial f}{\partial z}dz.$$

This is what we used to compute the dipole moment. It is a first order approximation, because we take the gradient to be *constant* over the intervals dx, dy, dz. To make the 2nd order approximation, we take the 2nd derivative to be constant, thus making the gradient (first derivative) vary *linearly* over the intervals dx, dy, dz.

So now we ask, what is the gradient of a vector field, $\mathbf{g}(\mathbf{r})$? At a given point, it is simply a set of 3 *vectors* that tell us how the vector field varies as we move in the x, y, and z directions:

$$\nabla \mathbf{g} \equiv \left(\frac{\partial \mathbf{g}}{\partial x}, \frac{\partial \mathbf{g}}{\partial y}, \frac{\partial \mathbf{g}}{\partial z} \right)$$

Figure 3.13 Variation of a 2D vector field over small intervals.

Note that all 3 component of $\mathbf{g}$ may vary even if we move only in, say, x. We can approximate the change in the vector $\mathbf{g}$ over small displacements, just like we did for the scalar field:

$$\Delta \mathbf{g} \approx \frac{\partial \mathbf{g}}{\partial x} dx, \qquad \Delta \mathbf{g} \approx \frac{\partial \mathbf{g}}{\partial y} dy, \qquad \Delta \mathbf{g} \approx \frac{\partial \mathbf{g}}{\partial z} dz, \qquad \text{and} \qquad \Delta \mathbf{g} \approx \frac{\partial \mathbf{g}}{\partial x} dx + \frac{\partial \mathbf{g}}{\partial y} dy + \frac{\partial \mathbf{g}}{\partial z} dz .$$

$\nabla \mathbf{g}$ is a set of 3 vectors, each with 3 components. We write $\mathbf{g}$ as a 3 x 3 matrix of 3 column vectors:

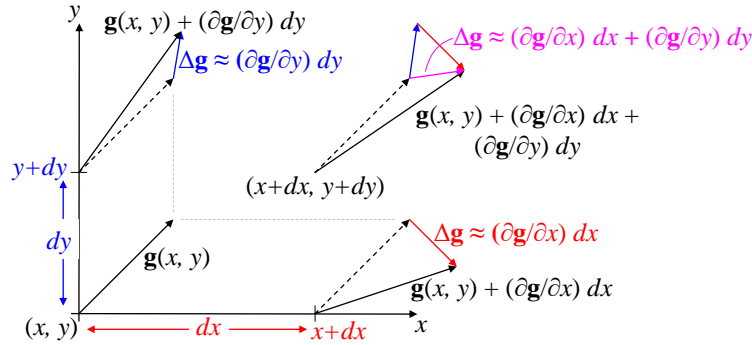$$\nabla \mathbf{g} \equiv \left(\frac{\partial \mathbf{g}}{\partial x}, \frac{\partial \mathbf{g}}{\partial y}, \frac{\partial \mathbf{g}}{\partial z} \right) = \begin{bmatrix} g_{xx} & g_{yx} & g_{zx} \\ g_{xy} & g_{yy} & g_{zy} \\ g_{xz} & g_{yz} & g_{zz} \end{bmatrix}, \qquad where \qquad \begin{pmatrix} g_{xx} \\ g_{xy} \\ g_{xz} \end{pmatrix} \equiv \frac{\partial \mathbf{g}}{\partial x}, \qquad \begin{pmatrix} g_{yx} \\ g_{yy} \\ g_{yz} \end{pmatrix} \equiv \frac{\partial \mathbf{g}}{\partial y}, \qquad \begin{pmatrix} g_{zx} \\ g_{zy} \\ g_{zz} \end{pmatrix} \equiv \frac{\partial \mathbf{g}}{\partial z} .$$

This is a rank-2 tensor. It can operate on a displacement vector to approximate the change in a vector field over that displacement:

$$\Delta \mathbf{g} \approx \frac{\partial \mathbf{g}}{\partial x} dx + \frac{\partial \mathbf{g}}{\partial y} dy + \frac{\partial \mathbf{g}}{\partial z} dz \equiv \nabla \mathbf{g}(d\mathbf{r}) = \begin{bmatrix} g_{xx} & g_{yx} & g_{zx} \\ g_{xy} & g_{yy} & g_{zy} \\ g_{xz} & g_{yz} & g_{zz} \end{bmatrix} \begin{pmatrix} dx \\ dy \\ dz \end{pmatrix} .$$

Note that a tensor is *linear* in its vector argument: $\nabla \mathbf{g}(k\, d\mathbf{r}_1 + d\mathbf{r}_2) = k \nabla \mathbf{g}(d\mathbf{r}_1) + \nabla \mathbf{g}(d\mathbf{r}_2)$. [This implies that if we transform coordinates (e.g., rotation), the columns and then rows of the tensor will transform like vectors; but we don't need that right now.]

Now let's compute the gradient of the gradient of a scalar function $f(x, y, z)$:

$$\nabla f \equiv \left(\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}, \frac{\partial f}{\partial z} \right) = \begin{pmatrix} \dfrac{\partial f}{\partial x} \\[2mm] \dfrac{\partial f}{\partial y} \\[2mm] \dfrac{\partial f}{\partial z} \end{pmatrix}$$

$$\nabla \nabla f = \left[\frac{\partial}{\partial x} \begin{pmatrix} \dfrac{\partial f}{\partial x} \\[2mm] \dfrac{\partial f}{\partial y} \\[2mm] \dfrac{\partial f}{\partial z} \end{pmatrix}, \frac{\partial}{\partial y} \begin{pmatrix} \dfrac{\partial f}{\partial x} \\[2mm] \dfrac{\partial f}{\partial y} \\[2mm] \dfrac{\partial f}{\partial z} \end{pmatrix}, \frac{\partial}{\partial z} \begin{pmatrix} \dfrac{\partial f}{\partial x} \\[2mm] \dfrac{\partial f}{\partial y} \\[2mm] \dfrac{\partial f}{\partial z} \end{pmatrix} \right] = \begin{bmatrix} \dfrac{\partial^2 f}{\partial x^2} & \dfrac{\partial^2 f}{\partial y \partial x} & \dfrac{\partial^2 f}{\partial z \partial x} \\[3mm] \dfrac{\partial^2 f}{\partial x \partial y} & \dfrac{\partial^2 f}{\partial y^2} & \dfrac{\partial^2 f}{\partial z \partial y} \\[3mm] \dfrac{\partial^2 f}{\partial x \partial z} & \dfrac{\partial^2 f}{\partial y \partial z} & \dfrac{\partial^2 f}{\partial z^2} \end{bmatrix}$$

It's just the matrix of all second derivatives. [It is called the **Hessian** matrix.]

Back to quadrupoles: We are the observer at the origin. We want the 2nd order change in potential seen by us due to displacing a single charge Q_i from the blob by a small amount $\mathbf{r'} = (x', y', z')$. We follow the same procedure as above for the 2nd order Taylor expansion of $f(x)$. First, we compute the gradient of the gradient of $1/r$:

$$\nabla \nabla \left(\frac{1}{r} \right) = \nabla \left[\frac{-1}{r^2} \hat{\mathbf{r}} \right] = \nabla \left[\frac{-1}{r^3} \mathbf{r} \right] = \nabla \left[-\left(x^2 + y^2 + z^2 \right)^{-3/2} \left(x\hat{\mathbf{x}} + y\hat{\mathbf{y}} + z\hat{\mathbf{z}} \right) \right]. \; Use \; \nabla \mathbf{g} \equiv \left(\frac{\partial \mathbf{g}}{\partial x}, \frac{\partial \mathbf{g}}{\partial y}, \frac{\partial \mathbf{g}}{\partial z} \right)$$

$$\frac{\partial}{\partial x} \left[-\left(x^2 + y^2 + z^2 \right)^{-3/2} \left(x\hat{\mathbf{x}} + y\hat{\mathbf{y}} + z\hat{\mathbf{z}} \right) \right]$$

$$= \left[\frac{3}{\cancel{2}} \left(x^2 + y^2 + z^2 \right)^{-5/2} \cancel{2} x^2 - \left(x^2 + y^2 + z^2 \right)^{-3/2} \right] \hat{\mathbf{x}} + \frac{3}{\cancel{2}} \left(x^2 + y^2 + z^2 \right)^{-5/2} \cancel{2} x \left(y\hat{\mathbf{y}} + z\hat{\mathbf{z}} \right)$$

$$= \left[\frac{3x^2}{r^5} - \frac{1}{r^3} \right] \hat{\mathbf{x}} + \frac{3xy}{r^5} \hat{\mathbf{y}} + \frac{3xz}{r^5} \hat{\mathbf{z}}$$

Similarly: $\quad \dfrac{\partial}{\partial y} [...] = \dfrac{3yx}{r^5} \hat{\mathbf{x}} + \left[\dfrac{3y^2}{r^5} - \dfrac{1}{r^3} \right] \hat{\mathbf{y}} + \dfrac{3yz}{r^5} \hat{\mathbf{z}}, \quad$ and $\quad \dfrac{\partial}{\partial z} [...] = \dfrac{3zx}{r^5} \hat{\mathbf{x}} + \dfrac{3zy}{r^5} \hat{\mathbf{y}} + \left[\dfrac{3z^2}{r^5} - \dfrac{1}{r^3} \right] \hat{\mathbf{z}}$

$$\nabla \nabla \left(\frac{1}{r} \right) = \frac{3}{r^5} \begin{bmatrix} x^2 & xy & xz \\ yx & y^2 & yz \\ zx & zy & z^2 \end{bmatrix} - \frac{1}{r^3} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} = \frac{3}{r^5} \begin{bmatrix} x^2 - \dfrac{1}{3} r^2 & xy & xz \\[3mm] yx & y^2 - \dfrac{1}{3} r^2 & yz \\[3mm] zx & zy & z^2 - \dfrac{1}{3} r^2 \end{bmatrix}$$

Now find the average gradient over the displacement:

$$\left[\nabla\left(\frac{1}{r}\right)\right]_{avg} \approx \frac{1}{2}\left[\nabla\nabla\left(\frac{1}{r}\right)\right]\begin{pmatrix} x' \\ y' \\ z' \end{pmatrix} = \frac{3}{2r^5}\begin{bmatrix} x^2 - \frac{1}{3}r^2 & xy & xz \\ yx & y^2 - \frac{1}{3}r^2 & yz \\ zx & zy & z^2 - \frac{1}{3}r^2 \end{bmatrix}\begin{pmatrix} x' \\ y' \\ z' \end{pmatrix}$$

$$= \frac{3}{2r^5}\begin{pmatrix} \left(x^2 - \frac{1}{3}r^2\right)x' + xyy' + xzz' \\ yxx' + \left(y^2 - \frac{1}{3}r^2\right)y' + yzz' \\ zxx' + zyy' + \left(z^2 - \frac{1}{3}r^2\right)z' \end{pmatrix}$$

$$\Delta\Phi \approx Q_i\left[\nabla\left(\frac{1}{r}\right)\right]_{avg} \cdot \mathbf{r}' = Q_i\frac{3}{2r^5}\left[\left(x^2 - \frac{1}{3}r^2\right)(x')^2 + xyy'x' + xzz'x'\right.$$

$$\left. + yxx'y' + \left(y^2 - \frac{1}{3}r^2\right)(y')^2 + yzz'y' + zxx'z' + zyy'z' + \left(z^2 - \frac{1}{3}r^2\right)(z')^2\right]$$

There's a trick here that's kind of complicated that I don't have time to write down yet, but it turns out we can drop the $(1/3)r^2$ terms in the first matrix by adding $-(1/3)r'^2$ terms to the quadrupole tensor. Briefly, it's because grad-grad-$1/r$ is traceless, so we can add any multiple of the identity matrix to the quadrupole tensor $D_{\alpha\beta}$, and it won't affect $\Delta\Phi$. So we add $-(1/3)r'^2[I]$ to make $D_{\alpha\beta}$ traceless. This means we can now add any multiple of the identity matrix to grad-grad-$1/r$, and it won't affect $\Delta\Phi$. So we add $(r^2/3)\mathbf{I}_3$, to eliminate those ugly terms.

Now notice that all the terms for $\Delta\Phi$ separate cleanly into **r** factors and **r**' factors. We also use the principle of superposition, to include the effect of all charges in the blob. This leaves us with:

$$\Delta\Phi_{Q_i} \approx Q_i\frac{3}{2r^5}\begin{bmatrix} x^2 & xy & xz \\ yx & y^2 & yz \\ zx & zy & z^2 \end{bmatrix}\begin{pmatrix} \text{element by element} \\ \text{multiplied and} \\ \text{summed with} \end{pmatrix}\begin{bmatrix} x'^2 - \frac{1}{3}r'^2 & x'y' & x'z' \\ y'x' & y'^2 - \frac{1}{3}r'^2 & y'z' \\ z'x' & z'y' & z'^2 - \frac{1}{3}r'^2 \end{bmatrix}.$$

So we define:

$$D_{\alpha\beta} \equiv \sum_{i=1}^{n} Q_i\begin{bmatrix} x_i'^2 - \frac{1}{3}r_i'^2 & x_i'y_i' & x_i'z_i' \\ y_i'x_i' & y_i'^2 - \frac{1}{3}r_i'^2 & y_i'z_i' \\ z_i'x_i' & z_i'y_i' & z_i'^2 - \frac{1}{3}r_i'^2 \end{bmatrix},$$

and then:

$$\Delta\Phi \approx \frac{3}{2r^5}\sum_{\alpha,\beta=1}^{3} D_{\alpha\beta}r_\alpha r_\beta = \frac{3}{2r^3}\sum_{\alpha,\beta=1}^{3} D_{\alpha\beta}\hat{r}_\alpha\hat{r}_\beta, \qquad where \quad \hat{r}_\alpha \equiv \alpha^{th} \text{ component of } \hat{\mathbf{r}}.$$

$D_{\alpha\beta}$ is the **quadrupole tensor**, which can simply be thought of as the matrix of numbers needed to compute the 2nd order change in Φ due to a blob of charges. Notice that the change in Φ varies as $1/r^3$.

[Notice that the quadrupole tensor is a rank-2 tensor, which acts directly on another rank-2 tensor, to produce a scalar: $\Delta\Phi$. The quadrupole tensor acts on the tensor $\mathbf{r} \otimes \mathbf{r}$, which is the tensor product of the displacement vector with itself. We have thus seen two ways rank-2 tensors can act: some can act on a vector to produce a vector, and others can act on another rank-2 tensor to produce a scalar. It is not always sensible for a single rank-2 tensor to act both ways. See *Funky Mathematical Physics Concepts* for details.]

[Aside: The above last few steps would be a lot easier if we recognize that:

$$(x,\, y,\, z)\begin{bmatrix} a & b & c \\ d & e & f \\ g & h & i \end{bmatrix}\begin{pmatrix} x \\ y \\ z \end{pmatrix} = (x,\, y,\, z) \otimes \begin{pmatrix} x \\ y \\ z \end{pmatrix}\begin{pmatrix} \text{element by element} \\ \text{multiplied and} \\ \text{summed with} \end{pmatrix}\begin{bmatrix} a & b & c \\ d & e & f \\ g & h & i \end{bmatrix}$$

$$= \begin{bmatrix} x^2 & xy & xz \\ yx & y^2 & yz \\ zx & zy & z^2 \end{bmatrix}\begin{pmatrix} \text{element by element} \\ \text{multiplied and} \\ \text{summed with} \end{pmatrix}\begin{bmatrix} a & b & c \\ d & e & f \\ g & h & i \end{bmatrix} \equiv \begin{bmatrix} x^2 & xy & xz \\ yx & y^2 & yz \\ zx & zy & z^2 \end{bmatrix} \bullet \begin{bmatrix} a & b & c \\ d & e & f \\ g & h & i \end{bmatrix},$$

where we've defined a matrix dot-product, producing a scalar.]

4 Electromagnetic Propagation and Radiation

Propagation In a Vacuum

Maxwell's equations imply that *all* vacuum EM fields, including inside waveguides, satisfy the wave equations (in Gaussian units) [Jac sec. 8.2]:

$$\nabla^2 \mathbf{E} = \frac{1}{c^2}\frac{\partial^2}{\partial t^2}\mathbf{E}, \qquad \text{i.e., each component of } \mathbf{E} \text{ satisfies the wave equation:}$$

$$\nabla^2 E_x = \frac{1}{c^2}\frac{\partial^2}{\partial t^2}E_x, \qquad \nabla^2 E_y = \frac{1}{c^2}\frac{\partial^2}{\partial t^2}E_y, \qquad and \qquad \nabla^2 E_z = \frac{1}{c^2}\frac{\partial^2}{\partial t^2}E_z.$$

Also for **B**: $\qquad \nabla^2 \mathbf{B} = \frac{1}{c^2}\frac{\partial^2}{\partial t^2}\mathbf{B}$

$$\Rightarrow \qquad \nabla^2 B_x = \frac{1}{c^2}\frac{\partial^2}{\partial t^2}B_x, \qquad \nabla^2 B_y = \frac{1}{c^2}\frac{\partial^2}{\partial t^2}B_y, \qquad and \qquad \nabla^2 B_z = \frac{1}{c^2}\frac{\partial^2}{\partial t^2}B_z$$

> In free space (no waveguide), the E & B fields are *in phase*.

> Many people think that in free space, E & B are out of phase,
> and imagine the energy to be shifted back and forth between the E & B fields, but that is *not true*.

Absorbing an EM wave is like being hit with a stream of warm beans. The energy density at the E & B nodes (zeros) is zero. The energy density is maximum at the peaks (anti-nodes).
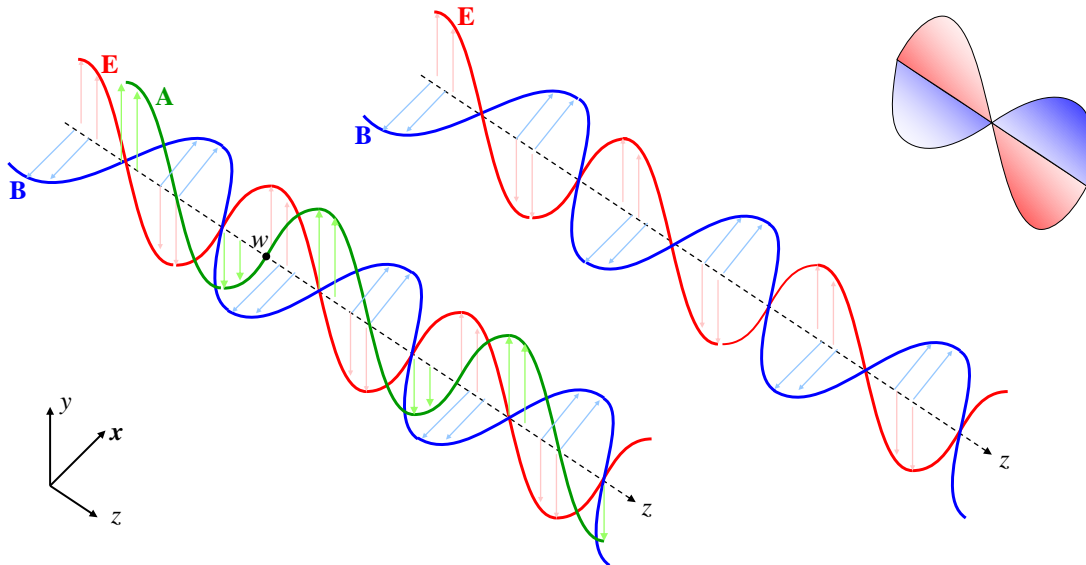
Figure 4.1 Linearly polarized EM wave in free space, propagating in the *z* direction. On the right, only the **E** and **B** fields are shown, for simplicity. On the left, the **A** field is included. **A** is in the plane of **E**, but leads **E** in time and space.

In the radiation gauge, the vector potential **A** for a linearly polarized EM wave is also a linearly polarized transverse wave, in the plane of the **E** field. **A** leads **E** in time (**A** peaks 90° before **E**), and lags in space (**A** peaks 90° after **E**). Consider the point *w* in space in the diagram above:

$$\mathbf{A} = A_y\hat{\mathbf{y}}, \qquad \mathbf{A}(w) = 0, \qquad \frac{\partial A_y(w)}{\partial t} < 0, \qquad \frac{\partial A_y(w)}{\partial z} > 0 \,.$$

E is maximum at *w*, where **A** is 0 and *decreasing* in time:

$$\mathbf{E} = -\frac{1}{c}\frac{\partial \mathbf{A}}{\partial t} \qquad \text{(gaussian)}.$$

At the same point, **B** is (negative, in this case) maximum at *w*, where **A** is 0 and *increasing* in space:

$$\mathbf{B} = \nabla \times \mathbf{A} = \left(\frac{\partial A_z}{\partial y} - \frac{\partial A_y}{\partial z} \right)\hat{\mathbf{x}} = -\frac{\partial A_y}{\partial z}\hat{\mathbf{x}}$$

However, the **A** field *decreases* in magnitude with *increasing* frequency, for the same power density, because **E** & **B** derive from **A** with derivative operators, which introduce factors of *ω* for **E**, or *k* for **B**:

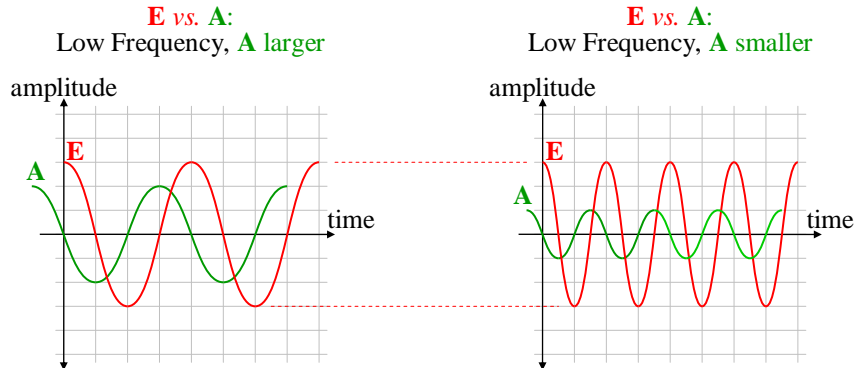**Figure 4.2** Same magnitude **E** field and its associated **A** field at two different frequencies.

Polarization Vector

This section assumes you understand phasors, and the decomposition of a sinusoid into "in-phase" and "quadrature" parts (see "Phasors," above). We start with an overview, and then fill in the details.

Overview: A polarization vector in general applies to any vector (or other multi-component object) which travels in space, and varies sinusoidally in time and space. It gives the relative amplitude and phase of each component of the vector. By convention for EM waves, the polarization vector, **ε**, gives the relative amplitudes and phases of the **E**-field components (E_x, E_y, E_z) for the wave.

In applications where the **A** field is most relevant, we write a polarization vector for the **A**-field, instead. In this case, **ε** is *gauge–dependent*. In some gauges, it has unexpected components, including a "longitudinal" component, which points *along* the propagation direction. This is a gauge artifact, and the polarization vectors for **E** and **B**-fields (in vacuum) always have no component along the propagation direction.

In principle, one could write a polarization vector for a **B**-field, but it's not usually done since it is just a right-hand rotation of the **E**-field polarization vector about the propagation direction by 90°.

[Note that for gravity waves, where the propagating field is the 4 × 4 metric tensor perturbation field, $h_{\mu\nu}$, the polarization "object" is a 4 × 4 polarization tensor.]

For a propagating EM wave, the polarization vector **ε** gives two things:

1. the E-field direction (**polarization**) of the EM wave at any time;

2. the fraction of the total EM wave intensity (power density) carried by each component of polarization.

The polarization vector **ε** is a set of 3 phasors, that describe the sinusoidal oscillations of each component (E_x, E_y, E_z) of the E-field. So we can immediately write the form of the E-field at a point as a function of time (in several different notations):

$$\mathbf{E} \propto \mathrm{Re}\left\{\boldsymbol{\varepsilon} e^{-i\omega t}\right\} = \mathrm{Re}\left\{\left(\varepsilon_x \hat{\mathbf{x}} + \varepsilon_y \hat{\mathbf{y}} + \varepsilon_z \hat{\mathbf{z}}\right)e^{-i\omega t}\right\} = \mathrm{Re}\left\{\begin{bmatrix} \varepsilon_x \\ \varepsilon_y \\ \varepsilon_z \end{bmatrix} e^{-i\omega t}\right\} = \begin{bmatrix} \mathrm{Re}\left\{\varepsilon_x e^{-i\omega t}\right\} \\ \mathrm{Re}\left\{\varepsilon_y e^{-i\omega t}\right\} \\ \mathrm{Re}\left\{\varepsilon_z e^{-i\omega t}\right\} \end{bmatrix}.$$

It is possible to represent any polarization (linear or elliptical) in any propagation direction with a 3D complex vector. Since (ideal) EM waves are transverse, the E-field, and thus the polarization vector, are perpendicular to the wave-vector $\mathbf{k}$, i.e., $\boldsymbol{\varepsilon} \cdot \mathbf{k} = 0$.
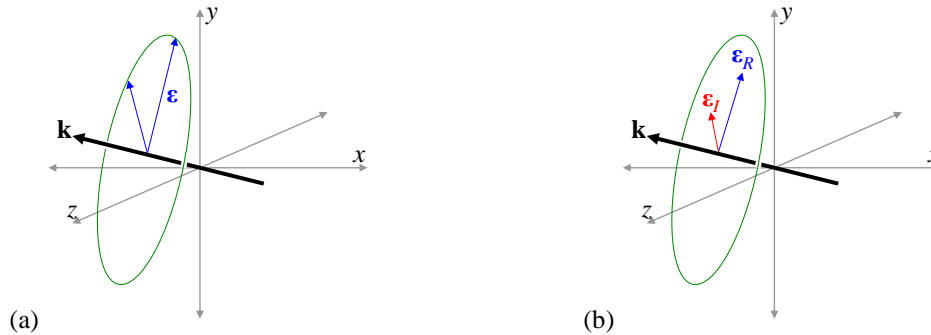

(a)   (b)

Figure 4.3 (Left) Wave vector $\mathbf{k}$ in an arbitrary direction, and some possible real polarization vectors. (Right) Wave vector $\mathbf{k}$, and the real (blue) and imaginary (red) parts of a complex polarization vector.

To keep the x, y, and z intensities normalized with respect to total intensity, the polarization vector is usually normalized to unit magnitude:

$$\boldsymbol{\varepsilon} = \left(\varepsilon_x, \varepsilon_y, \varepsilon_z\right) = \varepsilon_x \hat{\mathbf{x}} + \varepsilon_y \hat{\mathbf{y}} + \varepsilon_z \hat{\mathbf{z}}, \quad where \quad \varepsilon_x, \varepsilon_y, \varepsilon_z \text{ are phasors (complex numbers), and}$$

$$\left|\varepsilon_x\right|^2 + \left|\varepsilon_y\right|^2 + \left|\varepsilon_z\right|^2 = 1.$$

Note, however, there is a subtle distinction between the amplitudes of spatially perpendicular components, and the instantaneous maximum magnitude of the electric field, E_{max}. For a linearly polarized wave, say halfway between the x and y-directions, $E_{max} = (\sqrt{2})E_x$, and the time averaged intensity $<I> \sim E_{max}^2/2$. However, for the same *amplitude* of perpendicular electric field components, but phase shifted to make circular polarization, $E_{max} = E_x = E_y$, and $<I> \sim E_{max}^2$. Note that the *intensity* is the same for both waves: perpendicular component intensities always add; it is E_{max} which is different.

> For linearly polarized waves, $<I> \sim E_{max}^2 / 2$, whereas for circular waves, $<I> \sim E_{max}^2$.

The real part of the polarization vector is the "in-phase" E-field parts. The imaginary part is the time-quadrature E-field parts, i.e. is 90° out of phase with the real part. That is, the real part gives the direction of $\mathbf{E}$ at $t = 0$, and the imaginary part gives the direction of $\mathbf{E}$ a quarter-cycle later:

$$\mathrm{Re}\left\{\boldsymbol{\varepsilon}\right\} \parallel \mathbf{E}(t = 0), \quad\quad\quad \mathrm{Im}\left\{\boldsymbol{\varepsilon}\right\} \parallel \mathbf{E}(t = T/4).$$

Any polarization vector for a propagating wave in a given direction $\mathbf{k}$ can be written as a linear combination of two basis polarization vectors for that $\mathbf{k}$.
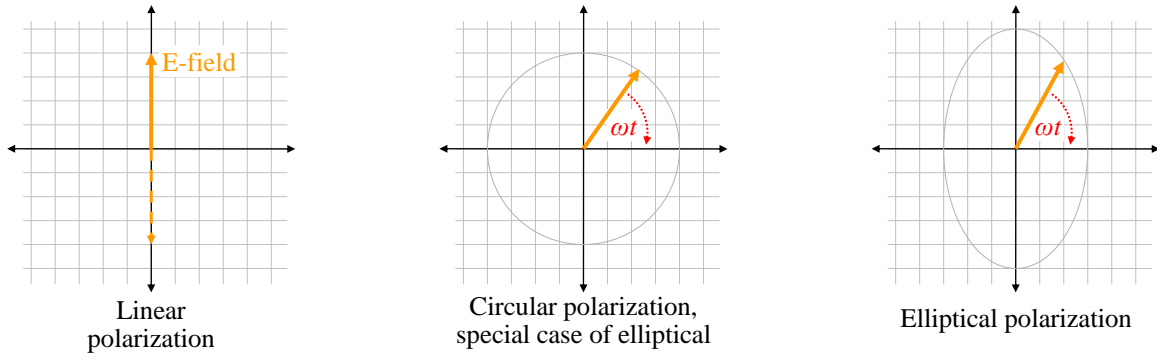
Figure 4.4 EM wave propagating out of the page: sample polarizations.

Non-plane waves can be thought of as an infinite set of tiny plane wave fronts propagating in different directions; therefore, non-plane waves have a separate polarization vector at each point. I.e., the polarization vector is a function of space, $\boldsymbol{\varepsilon}(\mathbf{r})$. Since the reference time for the sinusoidal phase is arbitrary, we conventionally take the first non-zero component of $\boldsymbol{\varepsilon}$ to be real and positive, which sets the reference time, and define the subsequent components relative to that.

Details: We restrict our attention to a single point in space, through which an EM wave is propagating. Since we are at an infinitesimal region of space, any shape wave front may be viewed as a plane wave. For a simple, plane-polarized wave, the time-average intensity P (W/m^2) and E-field amplitude E are related by (*the ε below are permittivity, not polarization vector*):

$$(SI) \qquad P = \frac{1}{2}\frac{E^2}{Z} \qquad \Rightarrow \qquad E = \sqrt{2ZP}, \qquad Z \equiv \sqrt{\frac{\mu}{\varepsilon}} \qquad \text{[Jac 7.13 p298]}$$

$$(gaussian) \qquad P = \frac{c}{4\pi}E^2 \qquad \Rightarrow \qquad E = \sqrt{\frac{4\pi}{c}P} \qquad \text{[L\&L p120]}$$

$$\text{In general:} \qquad E = m\sqrt{P}, \qquad m = \sqrt{2Z} \ \ (SI) \quad \text{or} \quad m = \sqrt{\frac{4\pi}{c}} \ \ (gaussian)$$

Consider an EM wave traveling in an arbitrary direction, and with arbitrary polarization. At any given time, the wave has an E-field with components in the x, y, and z directions:

$$\mathbf{E}(t) = E_x(t)\hat{\mathbf{x}} + E_y(t)\hat{\mathbf{y}} + E_z(t)\hat{\mathbf{z}} .$$

Each component is a sinusoid of frequency $\omega = c|\mathbf{k}|$, and so may be represented by a phasor (recall that a phasor is a complex number which represents the amplitude and phase of a sinusoid; see *Phasors* above):

$$\mathbf{E}(t) = m\sqrt{P_{total}}\left[\text{Re}\left\{\varepsilon_x e^{-i\omega t}\right\}\hat{\mathbf{x}} + \text{Re}\left\{\varepsilon_y e^{-i\omega t}\right\}\hat{\mathbf{y}} + \text{Re}\left\{\varepsilon_z e^{-i\omega t}\right\}\hat{\mathbf{z}} \right]$$

$$= m\sqrt{P_{total}} \ \text{Re}\left\{\left(\varepsilon_x\hat{\mathbf{x}} + \varepsilon_y\hat{\mathbf{y}} + \varepsilon_z\hat{\mathbf{z}}\right)e^{-i\omega t}\right\}$$

where P_{total} is the total intensity of the wave in W/m^2

For elliptical polarization, both the magnitude and direction of the E-field are functions of time.

The propagation vector is $\mathbf{k}$ (*not to be confused with $\mathbf{k} \equiv \mathbf{z}$-hat*), whose magnitude is the spatial frequency in rad/m, and whose direction is the propagation direction. For a given E-field component $j \in \{x, y, z\}$, the intensity (power/area) carried in the $\mathbf{k}$ direction is proportional to E_j^2, where E_j is the (real) amplitude:

$$P_x = m^2 E_x{}^2 = |\varepsilon_x|^2 P_{total}, \quad P_y = m^2 E_y{}^2 = |\varepsilon_y|^2 P_{total}, \quad P_z = m^2 E_z{}^2 = |\varepsilon_z|^2 P_{total}$$

$$P_{total} = P_x + P_y + P_z \;\; \text{W/m}^2 \qquad \Rightarrow \qquad |\varepsilon_x|^2 + |\varepsilon_y|^2 + |\varepsilon_z|^2 = 1 \qquad \text{[Jac 7.13 p298]}$$

So $|\varepsilon_x|^2$ is the fraction of the total wave intensity carried by the E_x component, etc.

It is instructive to consider the real and imaginary parts of $\boldsymbol{\varepsilon}$ separately:

$$\boldsymbol{\varepsilon} = \boldsymbol{\varepsilon}_R + i\boldsymbol{\varepsilon}_I \qquad where \quad \boldsymbol{\varepsilon}_R \text{ and } \boldsymbol{\varepsilon}_I \text{ are real vectors}.$$

$\boldsymbol{\varepsilon}_R$ is the direction of the E-field at the reference time, $t = 0$, and all integer oscillation periods after, $t = nT = 2\pi n/\omega$ (Figure 4.3b, blue arrow). $\boldsymbol{\varepsilon}_I$ is the direction of the E-field at quadrature times, $t = (n + \frac{1}{4})T$ (Figure 4.3b, red arrow). The E-field is always perpendicular to $\mathbf{k}$, so both $\boldsymbol{\varepsilon}_R$ and $\boldsymbol{\varepsilon}_I$ lie in the plane perpendicular to $\mathbf{k}$.

$$\boldsymbol{\varepsilon}_R \cdot \mathbf{k} = \boldsymbol{\varepsilon}_I \cdot \mathbf{k} = 0 \qquad \Rightarrow \qquad \boldsymbol{\varepsilon} \cdot \mathbf{k} = 0 + 0i = 0.$$

Basis polarization vectors: Since EM waves are transverse, the E-field (and therefore polarization vector) lies in a plane perpendicular to $\mathbf{k}$. This plane is a 2-D space, and all vectors in that plane (for a given direction of $\mathbf{k}$) can be written as the linear combination of 2 basis vectors. For complex vectors like polarization vectors, the basis vectors can be either real or complex. We always choose the basis vectors to be orthonormal, i.e.

$$\mathbf{e}_1 \cdot \mathbf{e}_1 = \mathbf{e}_2 \cdot \mathbf{e}_2 = 1 \qquad \mathbf{e}_1 \cdot \mathbf{e}_2 = 0, \qquad where \quad \mathbf{a} \cdot \mathbf{b} \equiv a_x^* b_x + a_y^* b_y + a_z^* b_z.$$

[The dot product of complex vectors is just like a quantum inner product.]

The only two bases you ever see are linear polarization bases, and circular polarization bases. To simplify the discussion, we now focus on propagation in the z-direction ($\mathbf{k} = k\mathbf{e}_z$). There are 2 orthogonal directions of linear polarization, x and y. Therefore, our linear polarization basis vectors are simply:

$$\mathbf{e}_x = (1,0,0) \qquad \mathbf{e}_y = (0,1,0) \qquad \text{linear polarization basis vectors}.$$

For z propagation, the plane of the E-field is the x-y plane; thus every polarization vector for z propagation can be written as a (possibly complex) combination of $\mathbf{e}_x$ and $\mathbf{e}_y$. The polarization vector for 100% x polarization is just $\boldsymbol{\varepsilon} = \mathbf{e}_1 = (1, 0, 0)$. For 100% y polarization, $\boldsymbol{\varepsilon} = \mathbf{e}_2 = (0, 1, 0)$. For linear polarization at 45°, $\boldsymbol{\varepsilon} = (1/\sqrt{2}, 1/\sqrt{2}, 0)$, which means ½ the power is carried in E_x and half in E_y.

For right-hand-circular polarization (RHC), $\mathbf{E}(t)$ rotates counter clockwise (right hand rule applied to $\mathbf{k}$). This means $\mathbf{E}(t = 0)$ points in the x direction, so $\boldsymbol{\varepsilon}_R \sim (1, 0, 0)$. $\mathbf{E}(t = T/4)$ points in the y direction, so $\boldsymbol{\varepsilon}_I \sim (0, 1, 0)$. Then:

$$\boldsymbol{\varepsilon} \sim \boldsymbol{\varepsilon}_R + i\boldsymbol{\varepsilon}_I = (1,0,0) + i(0,1,0) = (1,i,0). \qquad Normalizing, \qquad \boldsymbol{\varepsilon} = \frac{1}{\sqrt{2}}(1,i,0).$$

For LHC, $\mathbf{E}(t = T/4)$ points in the $-y$ direction, so $\boldsymbol{\varepsilon} = (1/\sqrt{2}, -i/\sqrt{2}, 0)$. We have just derived the circular polarization basis vectors:

$$\mathbf{e}_R = \frac{1}{\sqrt{2}}(1,i,0) \qquad\qquad \mathbf{e}_L = \frac{1}{\sqrt{2}}(1,-i,0).$$

[Beware that in optics, RHC is called "left handed" and LHC is called "right handed." Beats me.]

Examples: Let's use our polarization basis vectors to find out how much power passes through some polarization filters.

(1) Given an elliptical wave with $\boldsymbol{\varepsilon} = (3/5, i4/5, 0)$, what fraction f of the power will survive a y-polarizing filter? The filter passes only the part of the wave with $\boldsymbol{\varepsilon} = (0, 1, 0)$. Therefore, we "project" our given electric field polarization onto this, i.e., we see what fraction of our wave is E_y-polarized. This is just like finding how much of a basis vector is in a quantum state. Then we square that, to get intensity:

$$f = \left|\mathbf{e}_y \cdot \boldsymbol{\varepsilon}\right|^2 = \left|(0,1,0)\cdot(3/5, i4/5, 0)\right|^2 = \left|i4/5\right|^2 = 16/25 .$$

A circular wave, $\boldsymbol{\varepsilon} = (1/\sqrt{2}, i/\sqrt{2}, 0)$, and a 45° linear wave, $\boldsymbol{\varepsilon} = (1/\sqrt{2}, 1/\sqrt{2}, 0)$, would both pass ½ the power.

(2) How much of an x polarized wave survives a chiral filter that passes only RHC?

$$f = \left|\mathbf{e}_R \cdot \boldsymbol{\varepsilon}\right|^2 = \left|\left(1/\sqrt{2}, i/\sqrt{2}, 0\right)\cdot(1,0,0)\right|^2 = \left|1/\sqrt{2}\right|^2 = 1/2 .$$

The answer is the same for any direction of linear polarization (axial symmetry):

$$f = \left|\mathbf{e}_R \cdot \boldsymbol{\varepsilon}\right|^2 = \left|\left(1/\sqrt{2}, i/\sqrt{2}, 0\right)\cdot(a,b,0)\right|^2 = \left|\frac{a+ib}{\sqrt{2}}\right|^2 = 1/2, \qquad \text{since} \qquad |\boldsymbol{\varepsilon}|^2 = 1 = |a+ib|^2 .$$

Notice that in these cases, linear polarized light went in, but circular polarized light came out.

(3) How much of an LHC wave would pass the RHC filter?

$$f = \left|\mathbf{e}_R \cdot \mathbf{e}_L\right|^2 = \left|\left(1/\sqrt{2}, i/\sqrt{2}, 0\right)\cdot\left(1/\sqrt{2}, -i/\sqrt{2}, 0\right)\right|^2 = \left|\frac{1}{2}-\frac{1}{2}\right|^2 = 0 .$$

Irrelevant point: Sometimes, you can find the direction of $\mathbf{k}$ from the polarization vector, sometimes not. If $\boldsymbol{\varepsilon}_R$ and $\boldsymbol{\varepsilon}_I$ point in different directions, then because they are both perpendicular to $\mathbf{k}$, $\boldsymbol{\varepsilon}_R \times \boldsymbol{\varepsilon}_I$ points along $\mathbf{k}$. But if $\boldsymbol{\varepsilon}_R$ and $\boldsymbol{\varepsilon}_I$ are parallel, or if $\boldsymbol{\varepsilon}_I = \mathbf{0}$, then $\boldsymbol{\varepsilon}_R \times \boldsymbol{\varepsilon}_I = 0$, and you can't tell anything. So far as I know, this fact is of no use at all.

TBS. Angular momentum of elliptically polarized waves.

Extension to other waves: For things like gravity waves, the field which varies sinusoidally is a rank-2 tensor, which can be written as a 2D matrix (below, left):

$$\varepsilon_{\mu\nu}(t,\mathbf{x}) = \mathrm{Re}\left\{ \begin{pmatrix} \varepsilon_{tt} & \varepsilon_{tx} & \varepsilon_{ty} & \varepsilon_{tz} \\ \varepsilon_{xt} & \varepsilon_{xx} & \varepsilon_{xy} & \varepsilon_{xz} \\ \varepsilon_{yt} & \varepsilon_{yx} & \varepsilon_{yy} & \varepsilon_{yz} \\ \varepsilon_{zt} & \varepsilon_{zx} & \varepsilon_{zy} & \varepsilon_{zz} \end{pmatrix} e^{-i\omega t} \right\} \qquad \varepsilon_{+}(t,\mathbf{x}) = \frac{1}{\sqrt{2}}\mathrm{Re}\left\{ \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} e^{-i\omega t} \right\} .$$

Above right is ε_+, the polarization tensor for "+" polarization propagating in the z-direction.

Sleepy Hollow: The Legend of the Headless Vector

½ and ¼ waves plates prove headfulness.??

Linear polarization direction is often called a "headless vector," because often only the plane of polarization is important (or known). For example, if we rotate a linear polarizer by 180°, it's behavior is not changed. Similarly, if we rotate the light incident on a polarizer by 180°, there is no change in the output light. However, these experiments do not explore the full "headedness" of polarization. We now show that polarization is, in fact, a "headful" vector, whose sense can be measured with interference. This is reasonable, because the polarization is fully described by the direction of the E-field, at a given reference time, and the E-field is a headful vector.

Recall every vector lies along an axis, and that the **sense** of a vector is simply which way the vector points along the axis. In some cases, including most cases of polarization, we know the axis of the E-field, but not its sense. Also, recall that the polarization vector specifies the complete ("headful") direction and phase of the E-field for a propagating wave.

Interference occurs between headful E-field vectors, including their sense. To illustrate, suppose we superpose two equal-amplitude linearly polarized EM waves, with polarization vectors $\boldsymbol{\varepsilon}_1$ and $\boldsymbol{\varepsilon}_2$. Then at the point of superposition, we get:

$$\mathbf{E}_1(t) = \text{Re}\left\{E_0 \mathbf{\epsilon}_1 e^{-i\omega t}\right\}, \qquad \mathbf{E}_2(t) = \text{Re}\left\{E_0 \mathbf{\epsilon}_2 e^{-i\omega t}\right\}, \qquad \mathbf{E}_1 + \mathbf{E}_2 = \text{Re}\left\{E_0 \left(\mathbf{\epsilon}_1 + \mathbf{\epsilon}_2\right)e^{-i\omega t}\right\}.$$

Suppose that $\mathbf{\epsilon}_1$ and $\mathbf{\epsilon}_2$ are real and parallel: $\mathbf{\epsilon}_1 = \mathbf{\epsilon}_2$. Then we get constructive interference. Now suppose we rotate *in space* the 2nd wave about its propagation direction by 180º. This is *not* a phase-shift; there is no time delay involved. After rotation, the polarization vector gets negated: $\mathbf{\epsilon}_2 \to -\mathbf{\epsilon}_1$. The interference is now destructive:

$$\mathbf{E}_1 + \mathbf{E}_2 = \text{Re}\left\{E_0 \left(\mathbf{\epsilon}_1 + \mathbf{\epsilon}_2\right)e^{-i\omega t}\right\} = 0 .$$

Therefore, we see that $\mathbf{\epsilon}_1$ and $-\mathbf{\epsilon}_1$ (or $\mathbf{\epsilon}_2$ and $-\mathbf{\epsilon}_2$) are *different* polarizations. One is the negative of the other. If $\mathbf{\epsilon}_2 = (\mathbf{e}_x, 0, 0)$ then it is horizontal polarization. Then $-\mathbf{\epsilon}_2$ is *also* horizontal polarization, but it's a *different* "horizontal:" it is the negative of $\mathbf{\epsilon}_2$'s "horizontal."

Thus the polarization vector describes not only the plane of polarization, but the sense of the E-field in that plane. For linear polarization, you can write the polarization as a unit vector parallel to the E-field. This is a simple case of a polarization vector.

Of course, the sense reverses every half-period, so the sense of the polarization vector is defined at some reference time. In fact, the polarization vector describes the full 360º phase (in time) of the E-field, and phase is always relative to some reference time (or phase), taken as zero. The phenomenon of constructive or destructive interference is, as always, independent of our reference time (or phase).

Summary: Some experiments are not sensitive to the sense of the polarization, and therefore measure only the plane of polarization. They often treat such a polarization measurement as a headless vector, which is good enough for some applications. However, we have shown that interference reveals that polarization is a headful vector, even if the sense is unknown.

Wave Packets

Often, one sends finite-time "messages" or signals with electromagnetic waves (or some other linear medium: wires, water, air, etc.). Such a signal can be decomposed into a sum of sinusoidal frequency components (i.e., the Fourier Transform), each frequency component with its own amplitude and phase. Since many propagation phenomena are frequency dependent, it is important to ask: What range of frequencies are needed to add together to form a finite time signal? We will argue that a continuous "band" of frequencies, of finite upper and lower bound, is   sufficient to construct a finite time signal. The difference between the upper and lower frequency in a signal is called the **bandwidth** (BW). (This term is much abused; in particular, it is frequently used (incorrectly) to mean "data capacity.") We give a hand-waving argument for a crude estimate of the bandwidth of a finite-time signal. When viewed as a sum of sinusoidal waves, such a signal is called a **wave-packet**.
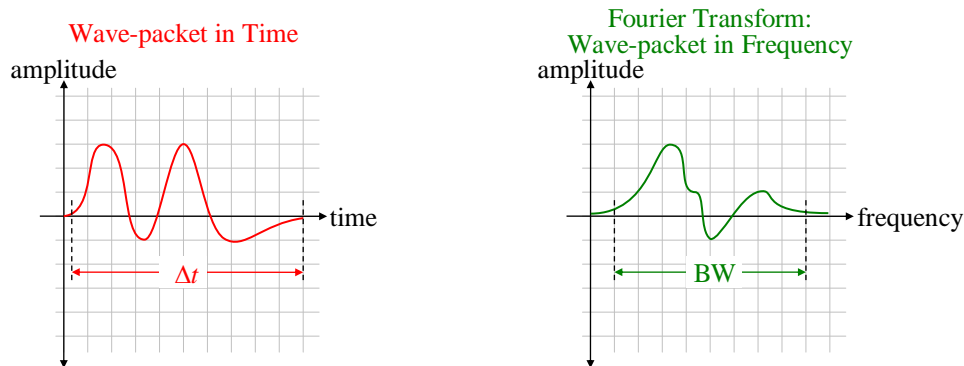

Figure 4.5 (Left) A wave-packet as a function of time. (Right) Frequency components of a wave-packet .

Note that when you transmit time-signals into a propagating medium, their extension in time becomes also an extension in space (e.g., wavelength). The following argument applies equally well to a signal at a

point in space that varies over time, and to a signal frozen in time that varies in space. Any reference to frequency can be equivalently thought of as time-frequency (ω), or spatial-frequency (wave-number, k).

When adding up lots of sinusoids, we consider the instantaneous phase of each component at some point (in either time or space). If there is a wide range of phases, spread over most of $[0, 2\pi)$, then the sinusoids will have lots of positive and negative components, and will add up to approximately zero. To have a significant contribution to our signal, most of the instantaneous phases need to be within a radian or two of each other. And this has to be true throughout the width (in time or space) of the wave packet.

Since the signal comprises different frequencies, the instantaneous phase difference between frequency components varies over time (or space). Components "in-phase" at one point will be "out-of-phase" at a distant point. (Below) the low- and high- frequency components (red and blue) start out in-phase, but drift to opposite phase over some time.
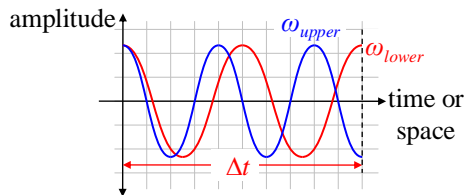

Figure 4.6 Wave components of different frequencies drift out of phase over time (or space).

The rate at which the phases drift apart is exactly the frequency difference between the two components:

$$\theta_1 \equiv \omega_1 t + \phi_1, \qquad \theta_2 \equiv \omega_2 t + \phi_2$$

$$\Delta\theta = \theta_1 - \theta_2 = \omega_1 t + \phi_1 - \omega_2 t - \phi_2 = (\omega_1 - \omega_2)t + const = (\Delta\omega)t + const, \quad \text{or} \quad \Delta\theta = (\Delta k)z + const$$

The constant term is not important, and simply depends on the shape of the signal and the origin of our t (or z) coordinate. What matters is that as t (or z) increases, so does the phase difference. Choose the coordinate origin to be the start of the wave-packet; then the width of the packet is Δt (or Δz). Clearly, the largest $\Delta\theta$ occurs when the frequency difference is maximum, i.e., between ω_{lower} and ω_{upper}:

$$\Delta\theta_{max} = (\omega_{upper} - \omega_{lower})\Delta t = (BW)\Delta t, \qquad \text{or} \qquad \Delta\theta_{max} = (k_{upper} - k_{lower})\Delta z = (BW)\Delta z .$$

Since $\Delta\theta$ must remain within 1 or 2 radians over the width of the signal, we have:

$$\Delta\theta \sim 1 \text{ or } 2 \text{ rad} = (BW)\Delta t \implies BW \sim \frac{1 \text{ or } 2 \text{ rad}}{\Delta t}, \quad \text{or} \quad BW \sim \frac{1 \text{ or } 2 \text{ rad}}{\Delta z} .$$

A more detailed analysis yields a theorem which states: a signal of finite time must (strictly speaking) have *infinite* bandwidth, and conversely, a signal of finite bandwidth must extend over *infinite* time. However, one can have an *approximately* finite-time and finite-bandwidth signal (for all practical purposes), because the amplitudes outside a finite interval are negligibly small.

Phase Velocity and Group Velocity

A hugely important concept now arises: what is the propagation speed of the wave packet? You might be tempted to say, "the same as the propagation speed of a single frequency." That's true if all frequencies propagate at the same speed. The speed of a single sinusoid is called the **phase velocity**, and can be seen from Figure 4.7.
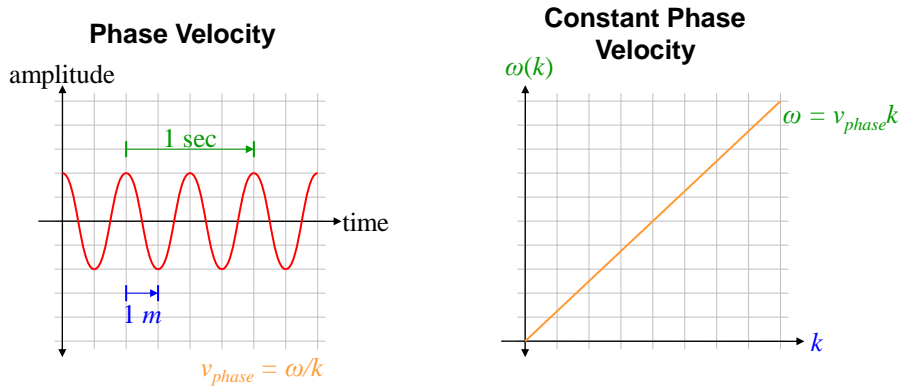
Phase Velocity

Constant Phase Velocity

$\omega(k)$

amplitude

1 sec

time

1 m

$v_{phase} = \omega/k$

$\omega = v_{phase}k$

k

Figure 4.7 Phase velocity.

In the figure, we have 4π rad/s, and π rad/m. The wave must then travel with:

$$v_{phase} = \frac{\omega}{k} = \frac{4\pi \text{ rad/s}}{\pi \text{ rad/m}} = 4 \text{ m/s} .$$

If all frequencies propagate at the same speed, then $\omega = v_{phase}k$ (Figure 4.7, right).

However, in most linear media, the propagation speed of a sinusoid *depends* on its frequency. Since a wave-packet is made up of many frequencies, which frequency should we use to determine its speed? You might guess the center frequency. Turns out, though, that what really matters is just how the speed *varies* with frequency. The wave-packet, or "group" of waves travels with a speed (**group-velocity**):

$$v_{group} = \frac{d\omega}{dk}. \qquad\qquad \text{Recall the phase velocity} \quad v_{phase} = \frac{\omega}{k} .$$

Actually, since each component travels at a different speed, the wave packet does not exactly travel "as a group." It gets "smeared" over time, because the relative phases of its components get misaligned. This smearing is called **dispersion**, and in some cases, is a serious limitation on a communication system. More on phase and group velocity in "Waveguides," later.

Poynting Vector For Linear Polarization

The Poynting vector describes the power density (= energy flux) of the propagating radiation: watts/m^2. It points in the direction of power flow, and is proportional to E and B:

$$\mathbf{S}(t,\mathbf{r}) = \mathbf{E}(t,\mathbf{r}) \times \mathbf{H}(t,\mathbf{r}) \qquad \text{SI units:} (V/m)(A/m) = W/m^2 .$$

Since $H \propto E$, and perpendicular to it, if either E or H is sinusoidal, then both are sinusoidal:

$$|\mathbf{S}(t)| = EH \sin^2 \omega t \qquad\qquad where \quad E, H \equiv \text{amplitudes} .$$

Quite often, we are more interested in the time-averaged power density, rather than the instantaneous power density. Then we use the fact that $\langle\sin^2\rangle = \frac{1}{2}$, over the long term (or an integral number of quarter-cycles):

$$\langle\mathbf{S}\rangle = \frac{1}{2} EH \underbrace{\left(\hat{\mathbf{E}} \times \hat{\mathbf{H}}\right)}_{\hat{\mathbf{k}}}, \qquad where \quad \hat{\mathbf{E}} \equiv \frac{\mathbf{E}}{E}, \ \hat{\mathbf{H}} \equiv \frac{\mathbf{H}}{H} \text{ are unit vectors in direction of } \mathbf{E} \text{ and } \mathbf{H} .$$

Beware of Solenoidal Poynting Vectors

In the radiation formula for a point particle [Jac ?? & Gri ??], there is a term called the "velocity term" or "velocity field," because it is proportional to the particle's velocity:

radiation formula here ??

This term seems to contribute to a Poynting vector, yet is said not to contribute to radiation. Some references' explanations are misleading: they say that the electromagnetic intensity due to the velocity term falls off as $1/r^4$, and so becomes insignificant compared to the $1/r^2$ "radiation" term. While the velocity term *does* drop as $1/r^4$, this is not the crux of the matter. The Poynting vector describes the power density (= energy flux density) of the propagating radiation: watts/m^2. Regardless of how the magnitude of the Poynting vector falls off, a radial component to it represents real energy flow. In the far field, we can imagine thick spherical shells around the radiation source. Since the shells are in vacuum (or non-absorbing material), the power (energy/time) flowing into any shell (from the source inside) must equal the power flowing out of that shell. Since the shell surface area increases as r^2, the power density must decrease as $1/r^2$, so that the product of the two, the total power, remains constant. This is just conservation of energy. Therefore:

> Conservation of energy requires that *all* radially directed power flows fall off *exactly* as $1/r^2$.

There *cannot be* any power flow that falls off as $1/r^4$, because it would violate conservation of energy.

So what of the velocity term? The real explanation is that the Poynting vector resulting from the velocity term is solenoidal, i.e. the lines of power flow close on themselves. No power flows out; it's almost as if the power just flows around in circles. The radial component of the velocity term Poynting vector is 0:

$$\mathbf{S}_{velocity} \cdot \hat{\mathbf{r}} = 0 \qquad \text{(velocity term Poynting vector)}.$$

Velocity fields have no outward power flow. The rate of fall off ($1/r^4$ or anything else) is irrelevant. We can see this by considering the well-known E-field of a particle at constant velocity. This field emanates from a single point in space called the "present position." That means the E-field everywhere points radially outward. It is also well-known that only the tangential component of E contributes to radiation, which you can readily see because only the perpendicular component of E contributes to the radial component of the Poynting vector (diagram??). The velocity field has no tangential component of E, and therefore no radiation. The Poynting vector field-lines make circles around the present position, and are completely solenoidal.

We can compute the radiated power out of an arbitrary closed surface (not necessarily a sphere, nor centered anywhere in particular) from:

$$P_{rad} = \oiint \mathbf{S} \cdot d\mathbf{a} .$$

While there may be patches (of a non-spherical surface) where the velocity terms have an outward component, their solenoidal nature insures that any such patch is exactly canceled by an inward component somewhere else.

In short, velocity terms don't radiate because their Poynting vectors go in closed loops.

TBS: Poynting Vector For Arbitrary Polarization

Realistic Waves: Unguided, Slow, and Twisting

The idealization of an infinite plane wave is useful for developing basic concepts in EM propagation. However, realistic waves have a finite cross-section in space. This has extremely important consequences: such waves are *not* transverse, travel slower than c, disperse over distance, and can carrying angular momentum. Furthermore, this *classical* angular momentum has a surprising tie to quantized radiation. [We hope it also has a tie to spin-1/2 particles, but that remains to be seen.]

Besides the practical applications, the analysis of realistic waves illustrates a number of important principles, e.g. sometimes idealized models unexpectedly obscure important realistic effects; advanced use of phasors, including time averages; and how nonlinear dependence produces new and nonobvious consequences.

We now consider some of these consequences for EM beams. This section requires that you understand linear and circular polarization, and their phasor (complex amplitude) representations, Poynting

vectors, and the principle of EM wave momentum. Some of the analysis is similar to that of waveguides, so familiarity with waveguides may be helpful, but is not necessary.

We follow these steps:

- Beams cannot be transverse.

- Therefore, beam group velocities are $< c$.

- Linearly polarized beams cannot have angular momentum.

- Circularly polarized beams *can* have angular momentum (unlike infinite plane waves).

- The seeds of quantized EM fields are hinted at even in classical electromagnetics.
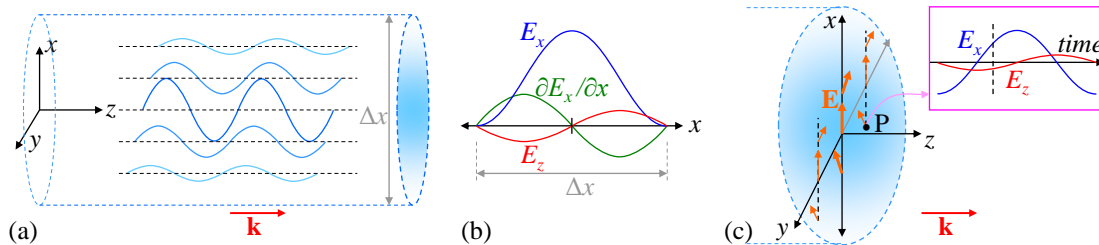
What's a Beam?

Figure 4.8 (a) A roughly circular beam of finite cross-section. (b) Qualitative example of E_x, $\partial E_x/\partial x$, and E_z across the beam face, at fixed time. (c) Example $\mathbf{E}(x, y)$ at a few points across the beam, at the time indicated on the slopes of both E_x and E_z (inset).

We consider a z-propagating wave with finite cross-section, i.e. a **beam** with finite spatial extent in x and y. A common example is a laser beam, say with linear polarization, and an axially symmetric gaussian intensity profile. Any beam must have the approximate form (similar to a waveguide):

$$\tilde{\mathbf{E}}(t,x,y,z) \approx E_0(x,y)\hat{\mathbf{n}}e^{i(kz-\omega t)} \qquad where \quad \tilde{\mathbf{E}} \equiv \text{complex}, \hat{\mathbf{n}} \equiv \text{complex polarization vector} . \quad (4.1)$$

$E_0(x, y)$ is arbitrary, not necessarily sinusoidal, or gaussian, or symmetric in any way. However, it must go to zero outside some widths Δx and Δy. Note 3 properties of such a wave:

1. At every point in space, $E_x(t)$, $E_y(t)$, and $E_z(t)$ are all sinusoidal in time, though we will show that E_z is of different phase.

2. At every point in in the beam face, $E_z(z)$ is sinusoidal in the z direction.

3. Fields across the beam face, $E_x(x, y)$ and $E_y(x, y)$, are *arbtrary*, and usually not sinusoidal functions of x or y. This implies that $E_z(x, y)$ is not sinusoidal in x or y.

We see immediately that the beam form (4.1) cannot be exact, because there is no diffraction. Maxwell's equations do not allow light to propagate in straight lines; it must diffract. However, by taking our cross-section to be large compared to the wavelength λ, we can make diffraction negligible. We confine our analysis to this approximation.

No Transverse Beams

We might expect a beam to be transverse, but even in the simple case of linear polarization, Maxwell's Equations do not permit this. Everywhere in charge-free space:

$$\nabla \cdot \mathbf{E} = 0 = \frac{\partial E_x}{\partial x} + \frac{\partial E_y}{\partial y} + \frac{\partial E_z}{\partial z} .$$

For linear polarization, we take $\mathbf{E}$ mostly in the x-direction, i.e. $\hat{\mathbf{n}} \approx (1,0,0)$. For a finite cross-section, $\partial E_x/\partial x \neq 0$ at least somewhere, and realistically:

$\dfrac{\partial E_x}{\partial x} \neq 0$, almost everywhere in the cross-section.

$E_y = 0$ everywhere for x polarization. Therefore:

$$\nabla \cdot \mathbf{E} = 0 \implies \qquad \frac{\partial E_z}{\partial z} = -\frac{\partial E_x}{\partial x} \neq 0 \ .$$

For linearly polarized beams of the nondiffracting form (4.1):

$$\frac{\partial E_z}{\partial z} = ikE_z, \qquad E_z(x,y) = -\frac{1}{ik}\frac{\partial E_0(x,y)}{\partial x} = \frac{i}{k}\frac{\partial E_0(x,y)}{\partial x} \ . \tag{4.2}$$

Thus the amplitude of E_z (including sign), at a point on the face (x, y), is $\partial E_0/\partial x/k$. Also, E_z is imaginary, so when $\partial E_0/\partial x > 0$, E_z *lags* E_x in time by a quarter cycle (Figure 4.8c, inset). For example, in *real* terms (i.e., sines and cosines, with no i):

$$E_x(t,x,y,z) = E_0(x,y)\cos(kz - \omega t) \qquad \implies \qquad E_z(t,x,y,z) = \underbrace{\frac{1}{k}\frac{\partial E_0(x,y)}{\partial x}}_{\text{amplitude}=\text{Im}\{E_z\}}\ \sin(kz - \omega t) \ .$$

Physically, Figure 4.8c shows a time when $E_x > 0$ (**E** points up), such as the time shown in the inset. Then in the lower half of the beam, $\partial E_x/\partial x > 0$, and $E_z < 0$. Furthermore, considered over time, when E_x is maximum, $\partial E_x/\partial x$ in time, and therefore $\partial E_z/\partial z$ has its maximum magnitude in time. $\partial E_z/\partial z$ is maximum at the zero crossings of E_z, hence the quarter cycle phase shift. (This phase shift will be crucial to angular momentum in circular polarization.)

In the upper half of the beam, $\partial E_x/\partial x < 0$, and $E_z > 0$. Along the y axis, $\partial E_x/\partial x = 0$, and $E_z = 0$.

The same argument, with $\nabla \cdot \mathbf{B} = 0$, shows that $B_z \neq 0$, as well. Thus, a linearly polarized beam cannot be transverse in either **E** or **B**, i.e. cannot be TEM, TE, nor TM. In fact, an identical derivation shows that the B-field looks the same as the E-field, except it is oriented mostly in the y direction, instead of x (Figure 4.9a). **E** lies in a plane parallel to x-z; **B** lies in a plane parallel to y-z.

The E and B fields are similar to those inside a waveguide, though with different boundary conditions. Nonetheless, some beam results are the same as waveguides: the group velocity $v_g < c$, and the phase velocity $v_p > c$.

Could a *circularly* polarized beam be transverse? No. A superposition of transverse waves is still transverse, and a linearly polarized wave can be constructed from two superposed circularly polarized waves. Transverse circularly polarized beams would contradict our proof that linearly polarized beams cannot exist.

In short, anywhere $\partial E_x/\partial x$ or $\partial E_y/\partial y$ are nonzero, E_z is nonzero.

A finite cross-section wave (a beam) has both E_z and B_z components, and therefore is not transverse in either **E** or **B**. Furthermore, its group velocity $v_g < c$, and its phase velocity $v_p > c$.
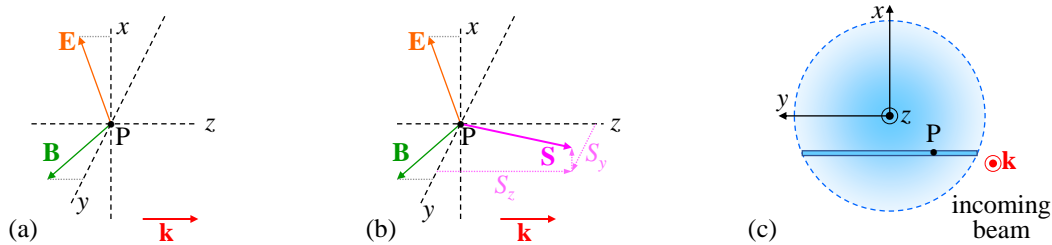

Figure 4.9 (a) Perspective view of **E** and **B** at a point P in the beam lower-right quadrant. Dashed lines are coordinate curves, not axes. (b) The Poynting vector poynts mostly along z, with small positive x and y components. (c) An integration stripe across the beam face.

How non-transverse is it? How big is a beam's z component? We'll use perturbation theory for this. As usual with perturbations, it is helpful to find a dimensionless, small expansion parameter. For our "wide" (non-diffracting) beam, the beam width Δx must be many wavelengths:

$$\varepsilon \equiv \frac{\lambda}{\Delta x} = \frac{2\pi}{k\,\Delta x} << 1 \qquad where \quad k \equiv k_z. \quad \text{Then:} \qquad \frac{1}{\Delta x} = \varepsilon \frac{k}{2\pi}\,.$$

For well-behaved (realistic) beams:

$$\frac{\partial E_x}{\partial x} \sim \frac{E_{x,\max}}{\Delta x} = \varepsilon \frac{k}{2\pi} E_{x,\max} << E_{x,\max}\,.$$

Then taking magnitudes from (4.2):

$$E_z \sim \frac{1}{k}\frac{\partial E_x}{\partial x} \sim \varepsilon \frac{1}{2\pi} E_{x,\max}, \qquad where \quad \varepsilon \equiv \lambda\,/\,\Delta x\,.$$

Thus a non-diffracting beam has a small z-component of $O(\varepsilon)$, and v_g is only slightly slower than c. [Giovannini, Romero, et. al., *Science* 2015/1/22] measured $v_g < c$ by a few ppm. (They also confirmed this effect is fully sustained even at the single-photon level, consistent with the quantum field equations of motion satisfying Maxwell's equations.)

Angular Momentum

It is well known that circularly polarized light carries angular momentum about the propagation axis **k**. There is often misunderstanding about how this angular momentum arises. Still less well-known is that the energy-to-angular-momentum ratio for circularly polarized *classical* EM waves mirrors the *quantum* relationship for photons: $U/L = \hbar\omega/\hbar = \omega$. Such a classical relationship provides a strong hint about how electromagnetism is quantized.

Linear polarization:

Recall L = r×p TBS??.

We first show that linear polarized beams have zero angular momentum at all times. At the point P in Figure 4.9c, both E_z and B_z are negative. Figure 4.9b shows that a negative E_z "pulls" the Poynting vector up above the y-z plane, i.e. S_x is positive. Similarly, a negative B_z pulls **S** out of the x-z plane toward the +y axis, i.e. S_y is positive. (The planes are not the actual x-z and y-z planes, because P is not the origin, but the planes are parallel to the x-z and y-z planes.) Since $L_z \propto xS_y - yS_x$, this shows that there are nonzero terms in at some points (x, y) on the face. However, we now show that when summed over the face, there is as much positive L_z as negative, so $L_z = 0$ at all times.

Consider S_x due to E_z in a beam; for illustration, assume an axially symmetric intensity. E_z is negative over the entire lower half of the beam (Figure 4.10a), because $\partial E_0/\partial x > 0$. E_z contribute to **S** only through B_y, and thus $E_z \times B_y$ points in the +x direction throughout the lower beam half. There is just as much clockwise momentum as counter-clockwise, and thus the contribution of the lower beam half to L_z is zero. In the upper beam half, $E_z > 0$, $S_x < 0$, but all the L_z cancels, just as in the lower half. The same idea insures L_z is zero even for non-axially symmetric beams, as will be rigorously shown later.
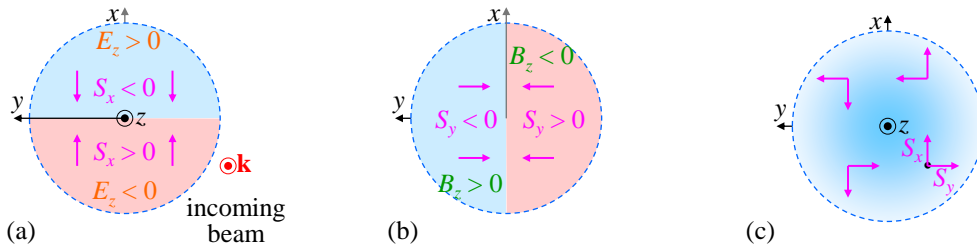
Figure 4.10 (a) E_z is positive in the top half, and negative in the bottom half. (b) B_z is positive in the left half, and negative in the right half. (c) RHC surviving contributions to **S**.

Now consider S_y due to B_z (Figure 4.10b). The same argument (rotated in space) shows that all the L_z from the left half cancel, as do those from the right half.

> For linear polarization, the angular momentum $L_z(t) = 0$, at all times.

Circular polarization: The circular polarization analysis is mathematically straightforward, if a little long-winded, because we also describe the physical significance of the math. The key result is that the angular momentum of the circularly polarized wave derives from the phase-shift of E_z and B_z compared to the x and y components, and from the nonlinear (quadratic) dependence of **S** on the fields **E** and **B**.

First, a circularly polarized *infinite* plane wave is transverse. Its Poynting vector **S** is in the z direction, so its angular momentum volume density:

$$\ell \equiv \frac{\mathbf{L}}{vol} = \mathbf{r} \times \mathbf{g} = \mathbf{r} \times \frac{\mathbf{S}}{c^2}, \qquad\qquad \mathbf{g} = \frac{\mathbf{S}}{c^2} \equiv \text{wave linear momentum density},$$

is perpendicular to **S**, and so has no z-component. In fact, all components are zero at all times: $\mathbf{L}(t) = \mathbf{0_v}$, because the x and y contributions on opposite sides of the z-axis all cancel. In general, axial symmetry does not allow a preferred x or y direction for *any* vector property.

> Circularly polarized *infinite* plane waves have *no* angular momentum.
> However, realistic beams *do*.

We can describe a circularly polarized (finite) beam as a superposition of two linearly polarized beams, which we here call the x-beam (**E** in x direction), and the y-beam. For RHC, the y-beam *lags* the x-beam by one-quarter cycle.

Recall that while **L** is a linear operation on **S**, **S** is *not* a linear operation on **E**; **S** is quadratic in **E**, so Poynting vectors from component waves *cannot*, in general, be simply added. If angular momentum were a linear operation on **E**, then the fact that linear polarization has $\mathbf{L} = \mathbf{0_v}$ would imply that all superpositions of linear polarization, including circular polarization, would have $\mathbf{L} = \mathbf{0_v}$. The nonlinear dependence of **S** on **E**, and therefore of **L** on **E**, makes angular momentum possible for circular polarization.

We now compute the angular momentum of a circularly polarized beam (this is [Jac Q7.29 p350]). We start with (4.1) for a z-traveling wave, but drop the tildes from the phasors. As with linear polarization, **E** is dominated by E_x and E_y, with a small E_z. For circular polarization, $E_0(x, y)$ is real, and the time evolution is simply that the entire physical state (**E**, **B**, **S**, and any other vectors) at each point (x, y) rotates around the propagation axis (an axis through (x, y) parallel to z). This is a kind of axial symmetry. Since all properties derive from the E and B fields, the time average of every vector (or tensor) property of the wave can have no x or y components, because they would cancel over one cycle.

The axial symmetry (over time) does not suggest converting to cylindrical coordinates, because $E_0(x, y)$ need not have circular (or any) symmetry.

We find the E-field components from:

$$E_x = E_0 e^{i(kz-\omega t)}, \qquad E_y = iE_0 e^{i(kz-\omega t)} \qquad \Rightarrow$$

$$\mathbf{E}(t,x,y,z) = \left[E_0(x,y)\left(\mathbf{e}_x + i\mathbf{e}_y\right) + E_{0,z}(x,y)\mathbf{e}_z \right] e^{i(kz-\omega t)}, \qquad \text{and} \qquad \frac{\partial}{\partial z} \to ik$$

Then:

$$\nabla \cdot \mathbf{E} = 0 = \frac{\partial E_0}{\partial x} + i\frac{\partial E_0}{\partial y} + \frac{\partial E_{0,z}}{\partial z} \qquad \Rightarrow \qquad E_{0,z}(x,y) = \frac{i}{k}\left(\frac{\partial E_0}{\partial x} + i\frac{\partial E_0}{\partial y} \right). \tag{4.3}$$

Note that, compared to linear polarization (eq. (4.2)), E_z now has an additional *real* term (the second term). This term comes from the y-beam. In the right half of the beam $\partial E_0/\partial y > 0$, so this term is negative and real, and *this term* (not all of E_z) is always of opposite sign to E_x. In the left half of the beam, this term is positive and real, and *this term* is the same sign as E_x.

Now having $\mathbf{E}$, we find $\mathbf{B}$ from Maxwell's Equation, $-\dfrac{\partial \mathbf{B}}{\partial t} = \nabla \times \mathbf{E}$:

$$+i\omega\mathbf{B} = \left(\frac{\partial E_z}{\partial y} - \frac{\partial E_y}{\partial z} \right)\mathbf{e}_x + \left(\frac{\partial E_x}{\partial z} - \frac{\partial E_z}{\partial x} \right)\mathbf{e}_y + \left(\frac{\partial E_y}{\partial x} - \frac{\partial E_x}{\partial y} \right)\mathbf{e}_z$$

$$\mathbf{B} = \frac{-i}{\omega}\left[\left(\frac{i}{k}\left(\frac{\partial^2 E_0}{\partial y\,\partial x} + i\frac{\partial^2 E_0}{\partial y^2} \right) - iki E_0 \right)\mathbf{e}_x + \left(ikE_0 - \frac{i}{k}\left(\frac{\partial^2 E_0}{\partial x^2} + i\frac{\partial^2 E_0}{\partial x\,\partial y} \right) \right)\mathbf{e}_y + \left(i\frac{\partial E_0}{\partial x} - \frac{\partial E_0}{\partial y} \right)\mathbf{e}_z \right]$$

We show that terms of (second derivative)/k are $O(\varepsilon^2)$ using essentially the same reasoning that above showed the first derivative terms are $O(\varepsilon)$:

$$\frac{1}{k}\frac{\partial^2 E_x}{\partial x^2} \sim \frac{1}{k}\frac{\left(\partial E_x / \partial x \right)_{\max}}{\Delta x} = \varepsilon\frac{1}{2\pi}\left(\frac{\partial E}{\partial x} \right)_{\max} << \left(\frac{\partial E}{\partial x} \right)_{\max}.$$

The same is true for $\partial/\partial y$. Then to first order in ε:

$$\mathbf{B}(t,x,y,z) = \frac{-i}{\omega}\left[kE_0\mathbf{e}_x + ikE_0\mathbf{e}_y + \left(i\frac{\partial E_0}{\partial x} - \frac{\partial E_0}{\partial y} \right)\mathbf{e}_z \right]$$

$$= \frac{-i}{c}\left[E_0\mathbf{e}_x + iE_0\mathbf{e}_y + \frac{i}{k}\left(\frac{\partial E_0}{\partial x} + i\frac{\partial E_0}{\partial y} \right)\mathbf{e}_z \right] = -\frac{i}{c}\mathbf{E}, \qquad \text{or} \qquad \mathbf{H} = -\frac{i}{Z}\mathbf{E}$$

where $\quad Z \equiv \dfrac{E}{H} = \sqrt{\dfrac{\mu}{\varepsilon}}$ or μc.

The factor $(-i)\mathbf{E}$ means that at every point in space, $\mathbf{B}/\mathbf{H}$ *leads* $\mathbf{E}$ by one quarter cycle, as expected for RHC polarization.

We now find the Poynting vector mathematically, and then explain what the math means physically. The Poynting vector $\mathbf{S} = \mathbf{E} \times \mathbf{H}$., where $\mathbf{E}$ and $\mathbf{H}$ are phasors. The time-averaged product of two sinusoids given by phasors A and B is Re$\{AB^*\}$ (see *Funky Electromagnetics Concepts*, Phasors), so the time-average Poynting vector is:

$$\langle \mathbf{S}(x,y) \rangle = \frac{1}{2}\text{Re}\left\{ \mathbf{E} \times \mathbf{H}^* \right\} = \frac{1}{2Z}\text{Re}\left\{ \mathbf{E} \times i\mathbf{E}^* \right\}.$$

Note that the *vector* $\mathbf{E} \times \mathbf{E}^* \neq |\mathbf{E}|^2$ (a scalar), and is not necessarily real. From $\mathbf{E}$ found in (4.3), we have straightforward (though slightly tedious) algebra, to find the time-average $<S_x(x,y)>$ and $<S_y(x,y)>$:

$$\langle S_x(x,y)\rangle = \frac{1}{2Z}\operatorname{Re}\left\{E_y iE_z^* - E_z iE_y^*\right\}$$

$$= \frac{1}{2Z}\operatorname{Re}\left\{\underbrace{(iE_0)}_{E_y}\underbrace{\frac{1}{k}\left(\cancel{\frac{\partial E_0}{\partial x}} - i\frac{\partial E_0}{\partial y}\right)}_{iE_z^*} - \underbrace{\frac{i}{k}\left(\cancel{\frac{\partial E_0}{\partial x}} + i\frac{\partial E_0}{\partial y}\right)}_{E_z}\underbrace{(E_0)}_{iE_y^*}\right\}$$

$$= \frac{1}{Zk}\operatorname{Re}\left\{E_0\frac{\partial E_0}{\partial y}\right\} = \frac{1}{Zk}E_0\frac{\partial E_0}{\partial y}.$$

$$\langle S_y(x,y)\rangle = \frac{1}{2Z}\operatorname{Re}\left\{E_z iE_x^* - E_x iE_z^*\right\} = \frac{1}{2Z}\operatorname{Re}\left\{\frac{i}{k}\left(\frac{\partial E_0}{\partial x} + i\cancel{\frac{\partial E_0}{\partial y}}\right)iE_0 - E_0\frac{1}{k}\left(\frac{\partial E_0}{\partial x} - i\cancel{\frac{\partial E_0}{\partial y}}\right)\right\}$$

$$= \frac{1}{Zk}\operatorname{Re}\left\{-E_0\frac{\partial E_0}{\partial x}\right\} = \frac{-1}{Zk}E_0\frac{\partial E_0}{\partial x}.$$

$$\langle S_z(x,y)\rangle = \frac{1}{2Z}\operatorname{Re}\left\{i\left(E_xE_y^* - E_yE_x^*\right)\right\} = \frac{1}{2Z}\operatorname{Re}\left\{i\left(E_0(-iE_0) - iE_0E_0\right)\right\} = \frac{E_0^{\ 2}}{Z}.$$

$S_z(x, y)$ is the usual power density ($dP/da \equiv I$, intensity) carried along with the wave.

What does this math mean? Consider the point P, at a point in time (say $t = 0$) when **E** points dominantly in the x-direction. The small E_z from the y-beam crosses into the large H_y from the x-beam giving a positive S_x contribution that is in *neither* the x-beam nor the y-beam by itself. The cross-product $\mathbf{S} \propto \mathbf{E}\times\mathbf{H} \propto \mathbf{E}\times\mathbf{E}^*$ has cross-terms that *mix* fields from the x-beam and y-beam, yielding results with no analog in either beam alone. This nonlinearity in **E** begets the surviving angular momentum L_z.

Figure 4.10c shows the contributions to L_z from S_x in all four quadrants. They all reinforce. On the right-half of the beam, the small E_z from the y-beam is negative, so S_x points up. On the left beam half, the small E_z is positive, so S_x points down. The "circulation" is always counter-clockwise, i.e. RHC. A detailed list of every part of the equation for S_x is in the "Loose Ends" below.

There are four more contribution to L_z, which arise from S_y: the large E_x of the x-beam crossed into the small H_z of the y-beam. In the lower half of the beam, the small $H_z > 0$, so S_y points right. In the upper half, the small $H_z < 0$, so S_y points left. Again, all quadrants circulate counter-clockwise, i.e. RHC.

The time-averaged angular momentum of the beam (per unit length Δz) is then the integral of the volume density ℓ over the beam face area:

$$\frac{\langle \mathbf{L}(x,y)\rangle}{\Delta z} = \frac{1}{c^2}\iint_{beam}dx\,dy\,\mathbf{r}\times\langle \mathbf{S}(x,y)\rangle.$$

For brevity, we now omit the "time average" brackets. For circular polarization, L_z is constant in time, anyway, which follows from the rotating time evolution mentioned earlier. But for arbitrary polarization, the time-averaged L_z is (using $E_0\,\partial E_0/\partial x = (1/2)\,\partial(E_0^2)/\partial x$):

$$\frac{L_z}{\Delta z} = \frac{1}{c^2}\iint_{beam}dx\,dy\left(xS_y - yS_x\right) = \frac{1}{c^2 Zk}\int_{-\Delta x/2}^{\Delta x/2}dx\int_{-\Delta y/2}^{\Delta y/2}dy\,\frac{1}{2}\left(-x\frac{\partial\left(E_0^{\ 2}\right)}{\partial x} - y\frac{\partial\left(E_0^{\ 2}\right)}{\partial y}\right).$$

Both terms in the integrand are even, so they contribute nonzero reinforcing values to L_z. In the example of Figure 4.8, the x-derivative is positive when x is negative, and vice-versa. Similarly for y in the second term. We can integrate each term by parts, and the boundary terms drop out, because the fields are zero at the beam edge (outside the regions of Δx and Δy):

$$\int_{-\Delta x/2}^{\Delta x/2} -x \frac{\partial \left(E_0{}^2\right)}{\partial x} dx = \underbrace{\left[-xE_0{}^2\right]_{-\Delta x/2}^{\Delta x/2} + \int_{-\Delta x/2}^{\Delta x/2} E_0{}^2 dx}_{\text{boundary term}},$$

and similar for the y term. Then total angular momentum (per unit length) is:

$$\frac{L_z}{\Delta z} = \frac{1}{c^2 k} \left[\int_{-\Delta x/2}^{\Delta x/2} dx \int_{-\Delta y/2}^{\Delta y/2} dy \frac{1}{2}\right] \left(2 \underbrace{\frac{E_0{}^2}{Z}}_{\text{intensity}}\right) = \frac{P}{c^2 k} = \frac{P}{c\omega}, \qquad P \equiv \text{power} .$$

This is positive, which means the angular momentum corresponds to right-handed rotation around **k**, the same sense that the E-field rotates. Also, the angular momentum depends only the total beam power, and has *no* dependence on the form of $E_0(x, y)$.

For any EM beam, the energy in a length Δz is:

$$U = P \, \Delta t = P \, \Delta z / c \qquad \Rightarrow \qquad \frac{U}{\Delta z} = \frac{P}{c} .$$

Then the ratio of energy to angular momentum is:

$$\frac{U / \Delta z}{L_z / \Delta z} = \frac{P}{c} \frac{c\omega}{P} = \omega ,$$

exactly as in quantum electrodynamics (photons), where $U/L = \hbar\omega/\hbar = \omega$.

Summary: Only *finite* beams with at least some circular polarization can carry angular momentum; idealized plane waves cannot. While EM beam angular momentum is linear in **S**, it is quadratic in the fields. Superposition does not apply to energy/momentum flow, nor angular momentum. When considering circular polarization as a superposition of two linearly polarized beams, the angular momentum arises from nonlinear cross-terms in the fields, producing a phenomenon not present at all in either component beam. (Quantum wave interference is another example of cross-terms leading to effects that are not present in either component alone; Pr(observation) $\propto |\psi|^2$.)

Furthermore, the angular momentum of a circularly polarized beam depends entirely on the beam energy, without regard to the detailed form of the intensity or field profile. Classical electromagnetics determines that the ratio U/L_z of energy to angular momentum is the wave frequency ω, exactly as it is for quantized photons. This classical result is a hint to the nature of the quantized EM field.

Loose ends: We easily show that the total (over the beam face) of S_x, and S_y, are zero at all times. This is expected, since energy does not flow out the sides of the beam (in the no-diffraction approximation of (4.1)). For S_x, consider any narrow stripe across the face in the y-direction (Figure 4.9c). Then:

$$S_{x,stripe}(t) \propto \int_{-\Delta y}^{\Delta y} E_0 \frac{\partial E_0}{\partial y} dy \propto \int_{-\Delta y}^{\Delta y} \frac{\partial \left(E_0{}^2\right)}{\partial y} dy \propto E_0{}^2 (\Delta y) - E_0{}^2 (-\Delta y) = 0 ,$$

because the beam fields are zero at its edges, by definition. Since all stripes have zero S_x, the total $S_x = 0$. A similar argument shows that $S_y = 0$ at all times.

Now we identify all the pieces of the equation for $\langle S_x(x, y)\rangle$, repeated here from above. Recall that the inner product of two complex numbers A and B is $A(B^*)$; the real part of the inner product is the product of the parallel (i.e., in-phase) components of A and B.

$$\left\langle S_x(x, y)\right\rangle = \frac{1}{2Z} \text{Re}\left\{E_y i E_z^* - E_z i E_y^*\right\} .$$

The first term is $E_y i E_z^*$, where:

E_y is the complex phasor for the *real* $E_y(t)$.

$-iE_z$ is proportional to the phasor for the *real* $H_z(t)$.

$iE_z^* = (-iE_z)^* \propto H_z^*$ is proportional to the conjugate phasor of H_z, needed to find the part parallel E_y, i.e. the part that contributes to the time average.

$(1/2)$ is the time average of $\cos^2$.

$\mathrm{Re}\{E_y iE_z^*\}$ is the product of in-phase components of the phasors E_y and H_z.

$Z = E/H$, so $H_z = E_z/Z$. Finally,

$(1/2Z)\mathrm{Re}\{E_y iE_z^*\}$ is the time average of the *real* sinusoids $E_y(t)$ and $H_z(t)$, i.e.:

$$\frac{1}{T}\int_{cycle}|E_y|\cos\left[\omega t - \arg(E_y)\right]|H_z|\cos\left[\omega t - \arg(H_z)\right]\ dt$$

The second term in $<S_x>$ is $E_z iE_y^*$, where:

E_z is the complex phasor for the *real* $E_z(t)$

$-iE_y$ is proportional to the phasor for the *real* $H_y(t)$

and the rest is the same reasoning as in the first term. The minus sign before the second term is simply the right-hand rule for $\mathbf{E} \times \mathbf{H}$ applied to E_z and H_y.

Elliptical polarization: Elliptical polarization can be written as a superposition of a linearly polarized wave and a circularly polarized wave. In a sense, only the circularly polarized component contributes to the angular momentum, but recall the nonlinear dependence of $\mathbf{L}$ on the fields. The linear and circular field components are *not* orthogonal, and so the total Poynting vector, and angular momentum, is *not* necessarily the sum of the component Poynting vectors and angular momenta. We do not pursue this analysis here.

Waveguides

We should first note that in the context of waveguides, the word "mode" is somewhat abused: in classical mechanics, and in E&M cavities, a "mode" implies a specific frequency of oscillation. In waveguides, the "mode" really refers to the eigenvalue of the waveguide transverse eigenfunction/eigenvalue equation (described below), and covers a *continuum* of temporal frequencies, ω, and spatial frequencies (wave numbers), k. One could say the waveguide "mode" refers only to the *transverse* mode, but not the whole of the EM wave(s) in the waveguide.

This section relies heavily on the methods developed in "Solving Laplace's Equation" and "Propagation in a Vacuum," above. You should read those sections first. Those sections concern electrostatics and magnetostatics. In the topic of waveguides, we extend those principles to electro-*dynamics*, where the fields are time-dependent. We therefore use the full, time-dependent Maxwell equations, including time derivatives.

Waveguides are concerned with the propagation of waves down a fixed (but arbitrary) cross sectioned cylinder. The most common cross sections are rectangles, circles, and fractions of a circle. Since we are concerned with wave propagation, prior experience strongly suggests decomposition of an arbitrary wave into sinusoidal components. [More rigorously, sinusoids compose a basis set of functions for the construction of any arbitrary function, and even better, are the eigenfunctions of linear integro-differential equations (such as the wave equations resulting from Maxwell's equations). This means we can study a single sinusoid of arbitrary frequency, and know that an arbitrary function can be made from a linear superposition of these sinusoids, and the sinusoids will not interact with each other. This is the standard method of eigenfunction decomposition and analysis, familiar from Fourier analysis and quantum mechanics.]

Under what conditions can a sinusoidal wave propagate down the waveguide, but remain otherwise unchanged? We assume an ideal conductor for the boundary of the waveguide, so there is no loss as the wave propagates. (Loss is a small correction to the lossless case [Jac p353m]). The choice of propagating sinusoids leads immediately to the use of phasors (complex numbers) for all components of E and B fields;

WLOG, we assume propagation in the $+z$ direction. Then all 3 components of both the E and B fields vary sinusoidally in time and z-direction:

$$\mathbf{E}(x, y, z, t) = \mathrm{Re}\left\{\mathbf{E}(x, y)e^{ikz - i\omega t}\right\}, \quad \mathbf{B}(x, y, z, t) = \mathrm{Re}\left\{\mathbf{B}(x, y)e^{ikz - i\omega t}\right\} \quad \text{[Jac 8.18 p357]}.$$

Note that ω is always real, and for lossless propagation, k must be real. So far, these are still fully 3-dimensional vector fields with components in all 3 directions. All the components propagate uniformly in the z direction. Because we have now defined the variation of the fields in z and t, we can find the time and z derivatives, and partially evaluate the Laplacian operator in the wave equations above:

$$\frac{\partial^2 \mathbf{E}}{\partial z^2} = -k^2 \mathbf{E}, \qquad \frac{\partial^2 \mathbf{E}}{\partial t^2} = -\omega^2 \mathbf{E}, \qquad \frac{\partial^2 \mathbf{B}}{\partial z^2} = -k^2 \mathbf{B}, \qquad \frac{\partial^2 \mathbf{B}}{\partial t^2} = -\omega^2 \mathbf{B}.$$

Since we know the variation in z, it becomes useful to separate the Laplacian operator into its "z-part" and its "transverse part" (i.e., its x-y part), so we define:

$$\nabla^2 = \nabla_t^2 + \frac{\partial^2}{\partial z^2} \quad \Rightarrow \quad \nabla_t^2 \equiv \nabla^2 - \frac{\partial^2}{\partial z^2} = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} = \frac{1}{r}\frac{\partial}{\partial r}\left(r\frac{\partial}{\partial r}\right) + \frac{1}{r^2}\frac{\partial^2}{\partial \phi^2} \quad \text{[Jac 8.20 p357]}.$$

Note the subscript "t" means "transverse" here, not "tangential." Also, we may use polar coordinates (r, ϕ) or (x, y) coordinates for the transverse part (or any other 2-D coordinates for that matter), whichever is easier. With these definitions, keeping in mind that all vector components are complex-valued phasors, the wave equations above become:

$$\left(\nabla_t^2 + \frac{\partial^2}{\partial z^2}\right)\mathbf{E} = \frac{1}{c^2}\frac{\partial^2}{\partial t^2}\mathbf{E} \quad \Rightarrow \quad \left(\nabla_t^2 - k^2\right)\mathbf{E} = -\frac{\omega^2}{c^2}\mathbf{E}, \qquad \nabla_t^2 \mathbf{E} = \left(k^2 - \frac{\omega^2}{c^2}\right)\mathbf{E},$$

or in components:

$$\nabla_t^2 E_x = \left(k^2 - \frac{\omega^2}{c^2}\right)E_x \qquad \nabla_t^2 E_y = \left(k^2 - \frac{\omega^2}{c^2}\right)E_y, \qquad \text{and} \qquad \nabla_t^2 E_z = \left(k^2 - \frac{\omega^2}{c^2}\right)E_z.$$

Also for **B**:

$$\left(\nabla_t^2 + \frac{\partial^2}{\partial z^2}\right)\mathbf{B} = \frac{1}{c^2}\frac{\partial^2}{\partial t^2}\mathbf{B} \quad \Rightarrow \quad \nabla_t^2 \mathbf{B} = \left(k^2 - \frac{\omega^2}{c^2}\right)\mathbf{B}$$

i.e., $\qquad \nabla_t^2 B_x = \left(k^2 - \frac{\omega^2}{c^2}\right)B_x \qquad \nabla_t^2 B_y = \left(k^2 - \frac{\omega^2}{c^2}\right)B_y, \qquad \text{and} \qquad \nabla_t^2 B_z = \left(k^2 - \frac{\omega^2}{c^2}\right)B_z$

Thus we see that the conditions for propagation down a waveguide are that all 6 E & M components, E_x, E_y, E_z, B_x, B_y, and B_z, must satisfy the same 2D eigenfunction/eigenvalue equation in the transverse plane. Note, however, that some of the 6 components may be everywhere zero, in which case those components identically satisfy *any* (linear) eigenfunction/eigenvalue equation. Each scalar field eigenfunction has an eigenvalue, which determines the dispersion relation relating k and ω for that eigenfunction. Further, to achieve propagation of such a mode, k and ω must be the same for all 6 E & M components. Therefore the eigenvalues (and thus dispersion relation) must be the same for all non-zero E- and B-field components of a single frequency. It is common to define the eigenvalue for a given waveguide frequency, and therefore the dispersion relation, as:

$$eigenvalue \equiv -\gamma^2 = k^2 - \frac{\omega^2}{c^2}, \qquad \gamma^2 = \frac{\omega^2}{c^2} - k^2 \qquad \Rightarrow \qquad \omega(k) = c\sqrt{k^2 + \gamma^2}.$$
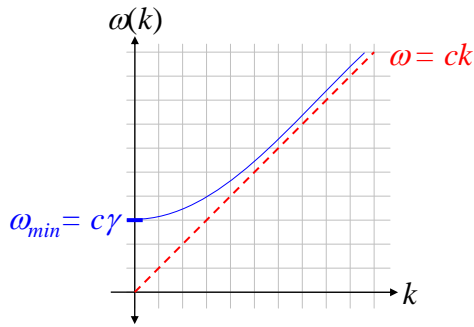
Figure 4.11 Dispersion relation for EM propagation in a waveguide.

Note that for hollow cross-sectional waveguides, the eigenvalues are always negative (why??), so $\gamma^2 > 0$. For propagation, k must be real > 0. Thus, a given propagation mode has a minimum frequency for propagation of

$$\omega_{\min} = c\gamma \ .$$

In free space, and with plane or spherical waves, $\gamma = 0$, the dispersion relation is linear: $\omega = ck$, and there is no frequency dispersion. Note that as $\omega \to \infty$, the dispersion relation for all waveguide modes is asymptotic to $\omega = ck$, which is the free space dispersion relation (of no actual dispersion).

Now that we have the basic equations of operation of a waveguide, we can consider the boundary conditions that exist in a waveguide.

Boundary Conditions and Propagation

The walls of a waveguide are good conductors, which imposes boundary conditions that further constrain the propagation of EM waves down the guide, compared to free-space propagation. In particular, a good conductor has surface charges and currents that eliminate all E (& D) and B (& H) fields inside it (except for a small skin depth) [Jac p358]. Also, as always, the $E_{parallel}$ (= E_z) and B_n (= $B_{transverse}$) are continuous at the boundaries. Therefore, at the boundaries,

$$E_z = 0, \quad \mathbf{B} \cdot \hat{\mathbf{n}} = 0. \qquad (NB : \mathbf{B}_t \equiv B_x\hat{\mathbf{x}} + B_y\hat{\mathbf{y}}) \ .$$

Note the subscript "t" still means "transverse" here, not "tangential."

TBS.

Phase and Group Velocity In a Waveguide

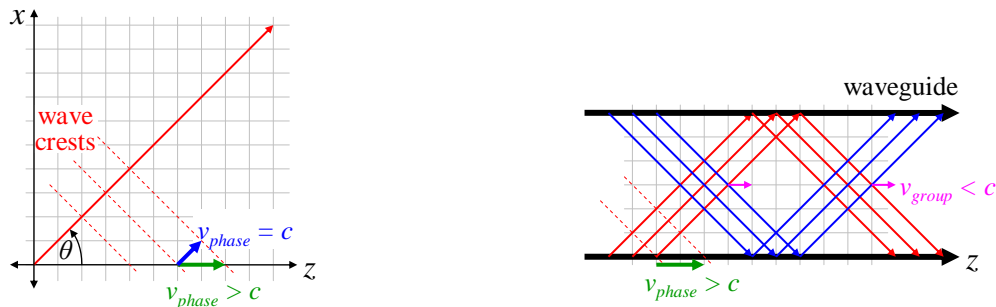First, consider a plane wave propagating in free space at an angle to the z-axis (Figure 4.12, left).

Figure 4.12 (Left) Phase velocity in free space. (Right) Phase and group velocity in a waveguide.

In the direction of propagation (along the red arrow), the phase velocity is just c, the speed of light. But suppose we set up electrometers to measure the electric field as it passes down the z-axis. The *time* between wave crests is the same, but the *distance* between wave crests along the z-axis is greater. So the wave crests are moving *faster than light* along the z-axis; this is required by the geometry. Hence, in the z-direction, $v_{phase} > c$.

The propagation down a waveguide can be written as a superposition of two constituent waves traveling at an angle to the waveguide axis (Figure 4.12, right, blue and red). The constituent waves reflect off the walls of the waveguide, and superpose to produce a net propagation in the z direction. The phase velocity of both the blue and red constituents is the same as in the left diagram:

$$v_{phase} = c/\cos\theta > c .$$

However, the net propagation speed in the z direction is just the z component of the red (or blue) constituent:

$$v_{group} = c(\cos\theta) < c .$$

Therefore, in a waveguide, the phase velocity is always $> c$, and the group velocity is always $< c$.

Note that the constituents are *not* simply two uniform plane waves. The boundary conditions must be met on each constituent wave: namely, that $E_z = 0$ at the walls, and that $B_n = 0$ (the normal component of the B-field). So the constituent waves also taper to zero at the edges, just like the total wave does.??

The angle θ can be determined from the z-component of the spatial frequency (the wave-vector), k_z. We find k_z from solving Maxwell's equations, and θ from:

$$k_z = \frac{\omega}{c\cos\theta} .$$

Multiple modes of one frequency: It is generally the case that a single frequency can propagate in multiple modes. The modes typically have different group velocities, though, so they are dispersed in time and space as they propagate. This is usually undesirable. To avoid this, the transmitter must excite only a single desired mode. Methods of doing this are beyond our scope.
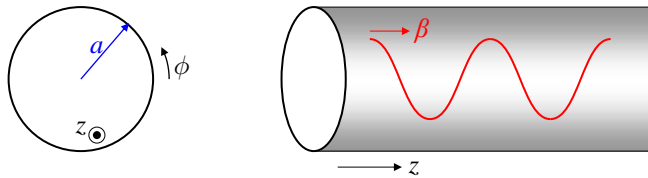
Cylindrical Hollow Waveguides

Figure 4.13 Cylindrical waveguide.

Cylindrical waveguides are common: they are just metal pipes, possibly gently bent to guide the waves along a path. The propagation modes are different than those of rectangular waveguides, but are derived from the same principles. In cylindrical coordinates, just as in rectangular, we generalize Laplace's equation (3.5) to time-dependent fields by including the time-dependent factor $e^{-i\omega t}$. This gives the wave equation:

$$\nabla^2 B_z = \frac{-\omega^2}{c^2} E_z . \qquad \nabla^2 = \frac{1}{r}\frac{\partial}{\partial r}\left(r\frac{\partial}{\partial r}\right) + \frac{1}{r^2}\frac{\partial^2}{\partial\phi^2} + \frac{\partial^2}{\partial z^2} \quad \Rightarrow \tag{4.4}$$

$$B_z(r,\phi,z) = R(r)Q(\phi)Z(z)e^{-i\omega t} \qquad \text{(and similar for } E_z\text{).}$$

As before, we seek propagation modes down the pipe, so:

$$Z(z) = e^{ik_z z}, \qquad and \qquad \frac{\partial^2}{\partial z^2} Z(z) = -k_z{}^2 Z(z) \qquad where \quad k_z \equiv \text{wave number} .$$

Solutions: As when solving Laplace's equation, the azimuthal function must satisfy:

$$\frac{1}{r^2} \frac{\partial^2}{\partial \phi^2} Q(\phi) = -\frac{k_\phi{}^2}{r^2} Q(\phi) \;\Rightarrow\; \frac{\partial^2}{\partial \phi^2} Q(\phi) = -k_\phi{}^2 Q(\phi), \quad Q(\phi) = e^{ik_\phi \phi} .$$

It is conventional to call k_ϕ 'm': $m \equiv k_\phi, \;\; Q(\phi) = e^{im\phi}$.

Then $R(r)$ must satisfy:

$$\frac{1}{r} \frac{\partial}{\partial r} \left(r \frac{\partial}{\partial r} \right) R(r) = \left(-k_r{}^2 + \frac{m^2}{r^2} \right) R(r), \quad or \quad \left[\frac{1}{r} \frac{\partial}{\partial r} \left(r \frac{\partial}{\partial r} \right) + k_r{}^2 - \frac{m^2}{r^2} \right] R(r) = 0 .$$

Notice that the second term on the right cancels the azimuthal term from the ϕ contribution. As with Laplace's equation, this equation is singular at $r = 0$, which is often a physically valid radius. The singularity is easily removed by defining $R(r = 0)$ as the limit:

$$R(0) \equiv \lim_{r \to 0} R(r) .$$

The differential equation is the famous Bessel equation, solved by Bessel functions of the first kind, $J_m(x)$:

$$R(r) = J_m(k_r r) .$$

Quantization: In the rectangular case, k_x and k_y were quantized by the boundary conditions of the guide. In the cylindrical case, k_r and m are quantized. With the cancellation of the azimuthal terms, $k_r{}^2$ acts like $(k_x{}^2 + k_y{}^2)$, and puts a lower limit ω_{min} on possible solutions to Maxwell's equations in the guide. The constraint is:

$$-k_r{}^2 - \beta^2 = -\frac{\omega^2}{c^2} \quad or \quad \beta^2 = \frac{\omega^2}{c^2} - k_r{}^2 > 0 .$$

The azimuthal parameter m must make the wave-function single-valued as ϕ goes across 2π. Therefore, m must be an integer. (This is exactly the same quantization and azimuthal function as cylindrical and spherical systems in quantum mechanics.) In our case, we can take $m \geq 0$, so:

$$m = 0, 1, 2, \ldots$$

In preparation for considering the radial modes, we recall the form of Bessel functions of first kind (Figure 4.14). They oscillate, with an infinite number of zeros, and extrema.
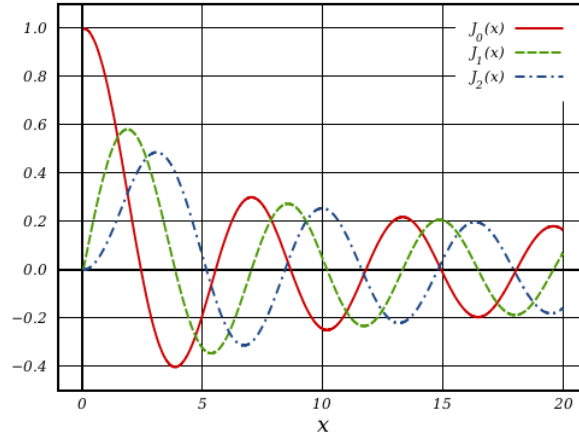
Figure 4.14 Some Bessel functions of the first kind.

The radial function $R(r)$ is quantized by the boundary condition at the walls of the pipe. For TE modes, $\mathbf{B}_\parallel$ must be stationary at the edge $r = a$, so:

$$\frac{\partial B_z}{\partial r} \propto \frac{dR(a)}{dr} = 0 \Rightarrow \qquad J'_m(k_r a) = 0, \quad k_r = \frac{x'_{mj}}{a}, \quad x'_{mj} \equiv j^{\text{th}} \text{ zero of } J'_m, \quad j = 1, 2, \ldots$$

$$B_z(t, r, \phi, z) = J_m(k_r r) e^{im\phi} e^{i(k_z z - \omega t)}, \qquad E_z(t, r, \phi, z) = 0.$$

These solutions are the TE_{mj} modes.

For **TM modes**, $\mathbf{E}_\parallel = 0$, so $E_z = 0$, and:

$$E_z(z) \propto R(a) = 0 \Rightarrow \qquad J_m(k_r a) = 0, \quad k_r = \frac{x_{mj}}{a}, \quad x_{mj} \equiv j^{\text{th}} \text{ zero of } J_m, \quad j = 1, 2, \ldots$$

$$E_z(t, r, \phi, z) = J_m(k_r r) e^{im\phi} e^{i(k_z z - \omega t)}, \qquad B_z(t, r, \phi, z) = 0.$$

These solutions are the TM_{mj} modes.

Order of the modes: Within the TE cases, and within the TM cases, except for the $m = 0$ case, k_r increases as m increases, i.e. k_r increases with the azimuthal mode, so the cutoff frequency increases with both m and j, similar to the rectangular waveguide. You can see from Figure 4.14 that J_0 is special, because mathematically, its first stationary point is $x = 0$, but that is not a physical mode. Therefore, the fastest mode (aka "dominant mode", with smallest k_r) is TE_{11}. However, the TE modes interleave with TM modes, so the sequence of modes with increasing cutoff frequency $\omega_{\min}$ is somewhat awkward:

m	TE (x'_{m1})	TM (x_{m1})
0	3.8317	2.4048
1	1.8412	3.8317
2	3.0542	5.1356
3	4.2012	6.3802

Guiding "Light"

There is often confusion regarding the differences between transmission lines, waveguides, and fiber optics. All three devices are ways of transmitting electromagnetic power along a guided path, but all three are *different*, and operate on different principles. (Many internet arguments are rendered moot by this fact.) All three transmit information in essentially analog form (continuously variable). Nowadays, digital

information is often modulated onto analog "carriers" (sinusoids), for digital communication. We now consider each of these devices, in turn.
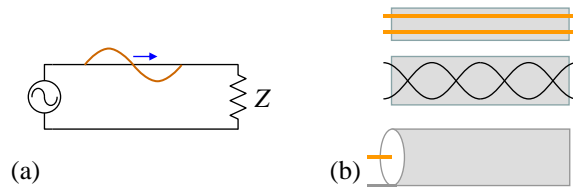
Transmission Lines

Figure 4.15 (a) A source feeds a transmission line, and then a load. (b) Parallel wire, twisted pair, and coaxial transmission lines.

Transmission lines are probably the simplest of these three devices. They transmit power by conduction of voltages and currents through two long conductors, shown schematically in Figure 4.15a. At the end of the line, a voltage drives a current through the load, delivering power. In a transmission line, the conductors are specifically arranged in a geometry that favors clean propagation of energy and/or information. Three common geometries are parallel wire, twisted pair, and coaxial (Figure 4.15b). Other transmission lines are frequently fabricated on circuit boards by controlling the geometry of the conductors and properties of the insulators.

Transmission lines are commonly used for transmission of either large power, or information, or both. Over-the-air television reception (54 - ~150 MHz) often uses parallel wires from the antenna to the receiver, in a flat ribbon of plastic about 1 cm wide with wires embedded along the edges, called "twin-lead". Category 3/5/6 communication wiring is twisted pair of 24-26 gauge (0.4 - 0.5 mm dia.). Residential cable TV companies provide information over coaxial cable, but do not deliver significant power for doing work. Microphone cords are usually coaxial transmission lines delivering tiny signals to an amplifier. Headphone cords are often coaxial transmission lines delivering less than a watt of power. Transmission lines to radio-frequency transmitter antennas deliver a signal of up to many kilowatts for radiation. Coaxial lines are commonly between about 2 mm to several cm in diameter. An old Ethernet coaxial line was about 2 cm diameter semi-rigid (bendable) aluminum tubing, with an air dielectric, and a central copper wire suspended by thin plastic discs every 10-20 cm.

Additional information: In the parallel wire and twisted pair geometries, the two conductors are identical, and interchangeable. We can somewhat arbitrarily define one conductor as "signal" and the other as "return." When the connecting circuitry also treats the wires identically, the transmission line is called "balanced." An unbalanced line can be connected to a balanced line by a simple isolation transformer, often with a grounded center tap.

In the coaxial configuration, the two conductors are physically different, and (by design) couple differently to the outside world. The inner conductor is often called the "signal." The outer conductor is often called a "shield," though purists call it a "return," because current flowing into the signal wire returns through the outer conductor. Coaxial transmission lines cannot be balanced, because the conductors are physically different. Some coaxial lines have a third conductor, also coaxial, surrounding the signal and return conductors. This true "shield" carries no signal current, and provides isolation from outside interference.

Waveguides

Waveguides are a *single* hollow conductor, usually with a rectangular or circular cross-section. They are mostly used for transmission to or from high-frequency antennas (> ~ 1 GHz). At such frequencies, waveguides have lower loss than transmission lines. Waveguides are typically a few cm across. For transmitters, the waveguide transfers signal and power to the antenna for radiation. For receivers, the power is tiny, but waveguides transfer as much signal as possible to maximize signal-to-noise ratio (SNR).

In a waveguide, all of the energy is transferred in the E and B fields of the guide. There are *no* long-distance voltages or currents, and no significant forces on charges.

In practice, there is no such thing as a "coaxial waveguide," because realistic waveguides are hollow, with no central conductor. Some references may be confusing because they include a subsection on coaxial transmission lines within a section on waveguides; the two are very different.

Within a waveguide, as in all electromagnetics, one can compute the power transferred from the Poynting vector. Of course, this is also true in a transmission line, even though the physics of the power transfer is very different. The Poynting vector works both in the presence and in the absence of driven charges, but it is not the whole story.

Fiber Optics

Communications fiber optic cables are flexible thin fibers that transmit small amounts of light. High quality communication fibers provide information only, as it is not practical to deliver significant power through communication cables. (However, some communication light sources are lasers with many milliwatts, and can be an eye-hazard. Never look into a fiber optic cable.)

Some room lighting systems and toys use fibers to deliver light, a form of power, but not information. Research is also being done to transmit usable power by light in cables. We do not consider such systems here.

Communication fibers are typically 125 μm in diameter. They have inner cores that carry the light; the outer material is "cladding." Inner cores are either around 60 μm in diameter (called "multi-mode fiber"), or around 9 μm ("single-mode fiber"). Common wavelengths are 1.31 μm and 1.55 μm (both infrared), so the single-mode fiber core is only a few wavelengths wide.

Fibers do not conduct electricity at all, so all the power is carried in the EM fields of the light beam. This is similar to a waveguide, but with a big difference: an optical fiber "guides" the light beam with controlled indexes of refraction in the fiber. There is no conducting edge to the wave, as there is in a waveguide where the conducting tube abruptly cuts off the EM fields.

Summary of Guided Light

Transmission lines, waveguides, and fiber optics all transfer information via electromagnetic energy. Transmission lines and waveguides may do so at large power, for applications such as radio transmission. All three devices operate on distinct physical principles, though all within the umbrella of classical electromagnetism.

The Speed of Electricity

Questions, and Short Answers

In a Veritasium video, Derek Muller asks: (1) Consider a circuit layout of two long pairs of wires (Figure 4.16a), with a signal feed in the middle of one wire, and a load in the middle of the other. The "wings" are long, but the physical distance from source to load is quite small. How long does it take for the signal to reach the load? We performed the experiment to find out.

The video also asks another question: (2) What is the path through space of the energy from source to load? It then describes the Poynting vector.

As we show below, the answer to question (1) is "it depends," but in the simplest case, the load receives nearly the full voltage and current almost immediately, without any delay from the long wings. In this simplest case, the answer to question (2) is "At first, through the capacitance between the wires, transitioning gradually to current flow in the wires." (Radio waves are not a part of the power delivered to the load.)

On the Poynting vector **S**, we show the well-known fact that Poynting's theorem does *not* necessarily give the physical flow of energy through space [Jac p259][Gri p358]. We show that the Poynting vector is always *related to* the energy flow, and sometimes it *is* the energy flow, but it many common cases, it is implausible for the Poynting vector to be actual energy flow. In a steady circuit, the energy flows down the wires, as confirmed definitively by these two facts:

- If the wires are left intact, then no stationary intervention of any kind into the space between and around the wires will perceptibly change the power delivered to the load, even if it radically alters the Poynting field.

- If the wires are severed, no stationary intervention of any kind will enable power to be delivered to the load.

In other words, in a steady state circuit, the wires alone deliver power to the load.

We think the weakest points in claiming that **S** is *always* the energy flow are:

1. In a steady state circuit, the claim is that the current is driven by the **S**-field created by the B-field *created by the current*! In other words, the energy to drive the current comes indirectly from the current itself.

2. Poynting's theorem *starts* from the fact that the electrons are driven by the E-field around them, and nothing else.

3. The Poynting vector in steady state cannot be measured, even in principle, by any electromagnetic experiment. In science, a thing that can never be detected does not exist.

Any one of these is enough to refute the claim that in steady state, the Poynting vector is carrying the energy delivered to the load. We examine these ideas in more detail below.

Some physicists still debate the meaning of the Poynting vector, so there is not a generally accepted interpretation, but we think the reasoning in this section is definitive.

Burden of Proof

Traditionally, many physicists have simply accepted that the Poynting vector **S** is the actual energy flux density (though many have not). But science is not decided by tradition; it is decided by experiment and logic. As we show, Poynting's theorem was never a reason to accept the traditional view in the first place, and we give many arguments against the claim that **S** *is* the energy flow. The traditional view holds that **S** is the true energy flow, but it is unmeasurable, even in principle. The burden of proof lies with the traditionalists: what is the evidence? How do they answer the 3 refutations above?

Outline

The concepts underlying the answers are nontrivial, and sometimes subtle. They are commonly misunderstood, even among experts. This discussion is somewhat long, but starts at a simple level, and increases in technical detail as it progresses. We follow these steps:

- The experiment: nearly full voltage and current, almost immediately.
- Variation of the experiment: alternating voltage steps approaching steady state.
- There can be multiple correct ways of understanding physics.
- Introduction to the Poynting vector field.
- The Poynting vector field and steady state currents: less than meets the eye.
- Tiny fields? Not the answer.
- Relationship to General Relativity (not testable with foreseeable technology).
- Appendix: Detailed discussion and derivation of Poynting's theorem.

Momentum? There is another whole story, even longer, about momentum in electromagnetism. Some aspects of EM momentum are also still debated by physicists. There are some relations between energy and momentum, but we do not discuss momentum here. However, everything here is fully consistent with all EM momentum experiments.
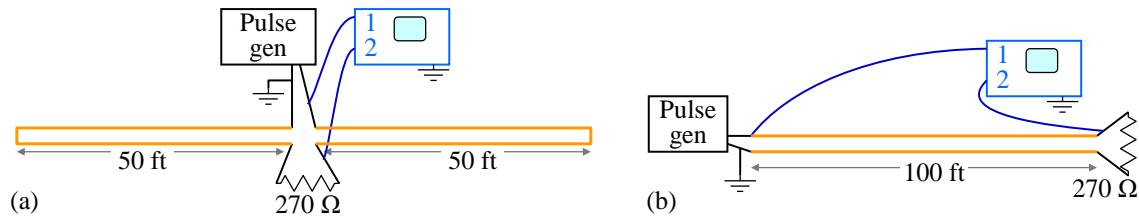
Figure 4.16 Test setup schematic. (a) The experiment. (b) Propagation speed test.

The Experiment

There are some variables that are unspecified in the question, and they matter, but we focus on the *simplest* and most fun case, where the load receives nearly the full voltage almost immediately. We'll discuss a more-complicated case afterward.

First, a definition: the term "transmission line" has a specific meaning in this context: a **transmission line** is a pair of conductors with significant length, and uniform cross section along that length. The two wings in question are transmission lines. Ideal transmission lines propagate electricity at a finite speed which is measurably slower than the speed of light, with no frequency dependence. When a signal reaches the end of a transmission line, a fraction of it is usually reflected back toward the source; then at the source, another fraction is reflected back, etc. Transmission lines are electrically identical to a resistor for times less than a round-trip time. Transmission lines necessarily have inductance per unit length (L_0 in H/m) and capacitance per unit length (C_0 in F/m). The small resistance of the conductors is often not relevant, and we can safely take it to be zero here. An ideal transmission line transmits a signal undistorted.

Figure 4.17 shows the equipment setup. Our wings are 12-gauge hardware-store extension cords, and had to be coiled to fit in the room. Our propagation speed measurement shows that this does not affect the result, as expected.

Figure 4.17 Test setup.

First, to verify the equipment, we made a simple propagation delay measurement through the two extension cords plugged together (Figure 4.16b). Combined, they are 100 ft long. Figure 4.18a shows the

result: the top trace is the input voltage; the bottom is the load voltage. Look at the falling edges. The time base of the oscilloscope is 0.2 μs/div. The propagation delay is 0.19 μs, for a propagation speed of:

$$\frac{100\,\text{ft}}{0.20\,\mu s} = \frac{30.5\,\text{m}}{0.20\,\mu s} = 161\,\text{m}/\mu s = 0.54c\,.$$

This is just over half the speed of light, and in the expected range of about 50% - 95% that of light. This measurement confirms the equipment's ability to measure the desired delay, with a resolution of about 0.01 μs (10 ns). The clean signal at the load (lower oscilloscope trace) confirms the transmission has finite speed and little distortion, as expected.
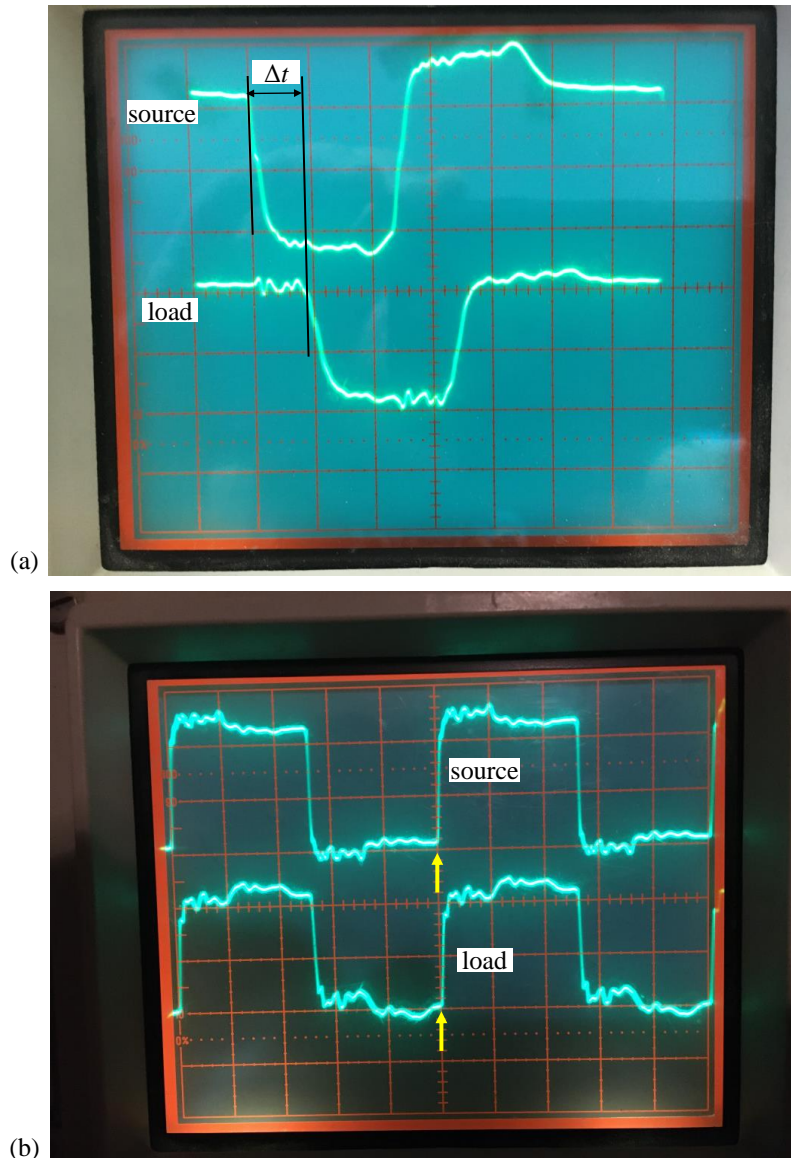

(a)

(b)

Figure 4.18 (a) Propagation delay verifies equipment. (b) Experiment in question shows no measurable delay.

The Fun Case: Nearly Full Voltage, Almost Immediately

Second, the experiment in question: Figure 4.18b shows the measurement (rising edge). Again, the top trace is the input voltage; the bottom is the load voltage + left-wing voltage. (Grounding requirements prevent us from measuring the load voltage by itself with our equipment.) The load receives the signal essentially instantaneously, with no more than ~0.01 µs delay, far faster than the propagation time down the wings and back, which is about 0.2 µs.

How is this possible? The key is that an (ideal) transmission line, measured from the input terminals, appears to be a simple resistance for a finite time. This resistance is called the **characteristic impedance**, Z_0. Therefore, until one round-trip time (0.2 µs here), our setup has an equivalent circuit of Figure 4.19a. Conveniently, the characteristic impedance of heavy-gauge extension cords is low, measured at ~53 Ω. Our load is 270 Ω, roughly the on-resistance of a 50 W incandescent bulb, so Z_0 is significantly less than R_{load}. Thus, nearly the full voltage appears on the load almost immediately, as experimentally verified.
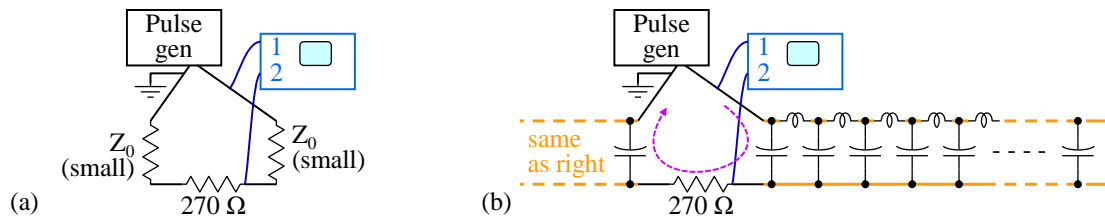

Figure 4.19 (a) Test setup short-time equivalent schematic. (b) Discrete component approximation to the continuous transmission line.

More Details

Figure 4.19b shows the discrete-component approximation to the continuous transmission line. For times < round-trip time and for ($R_{load} \gg Z_0$), the inductive effect is small compared to the capacitive effect. When power is applied, there is a physically short current path through the two nearest capacitors (dotted arrow), which almost immediately powers the load. As the voltage/current wave propagates out the wings, the successive capacitors charge one after another. This provides continuing current through the load.

When the voltage step hits the ends, it is inverted and reflected, and propagates back toward the source. This slightly decreases the voltage across the wing, and therefore *increases* the voltage across the load. You can see this in the lower trace of Figure 4.18b, and it occurs at exactly the round-trip time (corresponding to 100 ft of cable delay).

Question: What determines the cable characteristic impedance Z_0, and how do we make it small? Standard analysis shows:

$$Z_0 = \sqrt{L_0 / C_0} \ .$$

Low Z_0 comes from low L_0 and high C_0, consistent with the capacitance dominating the effect. Low L_0 comes from fat wires close together, as does high C_0. Just like a 12-gauge hardware-store extension cord.

Mind the Gap

Note that none of this explanation involves the Poynting vector field, $\mathbf{S(r)}$. However, we can compute the Poynting field in the capacitor gaps (between the wires) during the startup transient. It will show anomalies similar to a steady state current in a resistor, which we analyze in detail later. However, the current during startup takes a different physical path than that in steady state, and so the Poynting field during startup is different than that in steady state. Hence we cannot use the steady state Poynting field to explain how the startup happens.

Approach to Steady State

As the reflections occur and weaken, the system approaches steady state, the E-field in the wire gap stabilizes, and current in the wires stabilizes. At this point, all the current is in the wires, and the capacitance between them is fully charged; the physical length is then irrelevant. Compared to startup, the

Poynting vector everywhere (in the gap and around the wires) has a different form in steady state, which we discuss in detail later.

The Opposite Case: Slow, Stepped Turn On

The opposite case is for high Z_0, which comes from small wires far apart. Recall, though, that "high" and "low" are relative to the load resistance. We verified the following analysis experimentally (scope trace not shown). This is more difficult to compute analytically, so we provide just a qualitative description.

For high Z_0, Figure 4.19a shows that at first, most of the voltage is dropped across the wings, and the load sees only a fraction. When the reflected wave returns to the source, it has lowered the wing voltage, so the load voltage steps up. At the source, the traveling pulse is weakened (by the impedance mismatch) and reflected again out the wings. This reflection is not inverted, so the reflected pulse remains negative. At the far end, the pulse is reflected and inverted back to positive. When this pulse hits the source, the transmission line voltage increases, causing the load voltage to decrease. Then repeated reflections will alternately decrease and increase the wing voltage and therefore the load voltage, in steps. As the reflections die out, the system approaches the steady-state DC condition, where again, the load sees full voltage, and the wires carry full current. Note that the approach to steady state is not smooth, but a series of discrete steps in the load voltage, spaced by the round-trip travel time of electricity on the transmission line.

The Difference Between Right and Right

Here are some examples of subtleties in electromagnetic theory: Do two electrons repel each other? (a) Most people would say yes, and I agree. But one could argue: (b) No, one electron creates an electric field, and that field pushes on the other electron. This is also correct; it's slightly more detailed, and from a somewhat different viewpoint, but (a) is still correct, as well. But: (c) In calculating the force of (b), we use only the E-field from one electron, even though we know both produce E-fields. To use the full E-field, we have to compute force with the Maxwell stress tensor; this is also correct.

<div style="border:1px solid blue; text-align:center; color:blue;">There are multiple correct views one can take.</div>

In a steady state electric circuit, it is valid and correct to say that electrons enter the load with high potential energy, and (different) electrons leave with low potential energy. The electric signal travels through the wires, as electrons push on other electrons along the way. As each electrons push others through the load, their potential energy is given to the load. Another viewpoint is to say electrons (and protons) make E-fields in the wire which in turn push other electrons; this is just as valid.

Similarly, we can use the Poynting vector field to compute where in space the energy is increasing or decreasing or being dissipated, or we can just as well use $P = VI$, or $P/vol = \mathbf{E} \cdot \mathbf{J}$. The results are the same. However, this does *not* justify the claim that the energy flow in space is actually $\mathbf{S}(\mathbf{r})$.

The Poynting Vector Field

For propagating waves, it is empirically true that the Poynting vector $\mathbf{S}$ gives the magnitude and direction of energy flow. This is also easy to show theoretically, by computing the volumetric energy density of the wave, and multiplying by the speed of light. For other cases of EM fields, it is a common misperception that Poynting's theorem demands that $\mathbf{S}$ still gives the energy flow, but as we show below, it does not. This fact is well-established [Jac p259].

Poynting's theorem merely states that given some closed surface in space, the energy delivered to any charges in the volume equals the energy decrease of the fields in the volume, minus the field energy leaving the volume. As a slightly imperfect analogy, consider a coffee cup with a travel lid (a closed surface in space). You sip from it, but are clumsy, so some of the coffee spills down your chin. The amount of coffee delivered to your mouth (energy delivered to charges) equals the amount of coffee leaving the cup (energy decrease of fields in the volume), minus the spilled coffee (energy leaving the volume). Detail: Poynting's theorem is usually stated as a rate, which is energy per unit time or power.

Now Poynting's theorem follows quickly from Maxwell's equations, with currents or without (see below). But the theorem does *not* say that the Poynting vector *is* the energy flow. Technically, Poynting's theorem says the divergence of the Poynting vector equals the divergence of the energy flow, as we explain next.

> What we know for sure: Poynting's Theorem is always satisfied, currents or not.
> What is less clear: what is the physical meaning of the Poynting vector
> in the presence of steady currents, or steady magnetic fields?
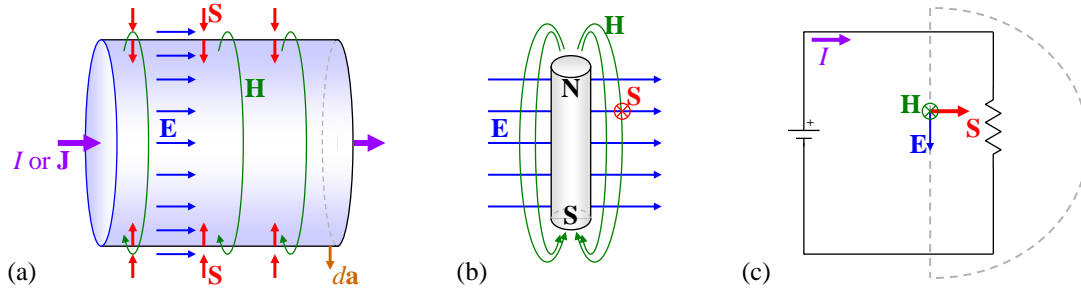
Case In Poynt: Steady State Systems

Figure 4.20 (a) Poynting vector for a volume which could be a wire, resistor, or charge density in free space. (b) Permanent magnet in an E-field. (c) Steady state source delivering power to a load.

Consider an (idealized) steady state current in a resistor (Figure 4.20a). The resistor continuously dissipates power, and there are no changing fields to provide energy (steady state). The Poynting vector is formed from the B-field *created by the current*, and points inward. Does the energy driving the current come from the B-field created by the current? It seems hard to imagine that this is how the energy flows. (Richard Feynmann called this idea "crazy" [Feynmann Lectures, Vol. II p27-8], though he did not dispute it.)

Let us consider what Poynting's theorem actually says about this. Assume that the true energy flux density (watts/m^2) is a well-defined function of space, call it $\mathbf{Q(r)}$. Separately, we define the Poynting vector field $\mathbf{S(r)} \equiv \mathbf{E(r)} \times \mathbf{H(r)}$. Poynting's theorem for an infinitesimal volume is:

$$\underbrace{\frac{d}{dt}\frac{W_{\text{charges}}}{vol}}_{\substack{\text{rate of work}\\\text{done on charges}}} = \underbrace{-\frac{d}{dt}u_{em}}_{\substack{\text{rate of field}\\\text{energy decrease}}} - \underbrace{\nabla \cdot \mathbf{S}}_{\substack{\text{rate of field}\\\text{energy outflow}}} \qquad \left[\text{cf. Gri 8.9 p358}\right]. \qquad (4.5)$$

The derivation of Poynting's theorem starts with the work-energy theorem, and the rate of work done *by the E-field* on a charge density (if any) at a point in space:

$$\frac{Power}{vol} = \mathbf{E} \cdot \underbrace{\rho\mathbf{v}}_{\mathbf{J}} = \mathbf{E(r)} \cdot \mathbf{J(r)} \ .$$

The E-field above is some *external applied* E-field, and *any applied H-field does no work* on the charges. The work going into the charges comes from the E-field, and the E-field alone. In the simple case of $\mathbf{H}_{applied} = \mathbf{0_v}$, we have $\mathbf{S} = \mathbf{E} \times \mathbf{H}_{applied} = \mathbf{0_v}$. The work going into the charges is not represented by any *applied* Poynting vector.

In vacuum, there are no charges, no work done, and the left-hand-side (LHS) is zero. Note that the above equation does *not* say that $\mathbf{S}$ *is* the energy flux density $\mathbf{Q}$; it says that the divergence of $\mathbf{S} = $ divergence of $\mathbf{Q}$: $\nabla \cdot \mathbf{S} = \nabla \cdot \mathbf{Q}$. This allows $\mathbf{S}$ to vary dramatically in both magnitude and direction from $\mathbf{Q}$, and in fact, such differences appear in many simple situations shown below. Poynting's theorem contains only $\nabla \cdot \mathbf{S}$, but not $\mathbf{S}$ itself, so Poynting's theorem cannot say that $\mathbf{S} = \mathbf{Q}$ [Jac p259].

Now consider five examples. Case (1): A resistor like Figure 4.20a. The current flows solely because of the applied E-field. The applied $\mathbf{H}_{applied} = \mathbf{0_v}$. The current *itself* produces the resulting H-field; this H-field is not a *source* of energy; it results from the charges *getting* energy from the applied E-field. Poynting's theorem tells us that the H-field is generated in just such a way that (4.5) is satisfied. In this case, the Poynting vector $\mathbf{E} \times \mathbf{H}_{resulting}$ represents just a mathematical identity such that:

$$\oiint_{resistor} \left(\mathbf{E} \times \mathbf{H}_{resulting} \right) \cdot d\mathbf{a} = IV = Power \ .$$

In other words, $\mathbf{S}$ *inside* a charge density does *not* represent the flow of energy from the external source into the charges. But its surface integral happens to equal the power delivered from the external source.

Case (2): A vacuum region of space. In the absence of charges, the LHS of (4.5) is zero, and we have simple continuity of electromagnetic field energy:

$$\underbrace{\frac{d}{dt} u_{em}}_{\substack{\text{rate of field} \\ \text{energy change}}} = - \underbrace{\nabla \cdot \mathbf{S}}_{\substack{\text{rate of field} \\ \text{energy outflow}}} \ .$$

This *still* does not mean $\mathbf{S} = \mathbf{Q}$; it means only that $\nabla \cdot \mathbf{S} = \nabla \cdot \mathbf{Q}$.

Case (3): Figure 4.20b shows a permanent magnet (or a steady-state electromagnet) in a stationary E-field. $\mathbf{H}$ points down the page, and $\mathbf{S}$ points into the page. Clearly, there is no energy flowing into the page, even though $\mathbf{S} \neq \mathbf{0_v}$. However, $\nabla \cdot \mathbf{S} = 0$, and that is consistent with the true energy flow $\mathbf{Q} = \mathbf{0_v}$, and $\nabla \cdot \mathbf{Q} = 0$.

Case (4): What if we apply both an external E-field and an external H-field to a charge density? Then the total $\mathbf{H}$ is just the sum $\mathbf{H}_{total} = \mathbf{H}_{applied} + \mathbf{H}_{resulting}$. Then $\mathbf{S}$ gets a contribution from $\mathbf{H}_{applied}$ (which can do no work on charges), plus another contribution from $\mathbf{H}_{resulting}$. As in Case (1), the latter contribution is just a mathematical identity, and not the flow of anything. Therefore, this $\mathbf{S}$ is *not* the flow of EM field energy.

Case (5): Figure 4.20c depicts a simple laboratory-scale circuit, in steady state. The voltage between the top and bottom wires means there is an E-field as shown. The current in the top and bottom wires creates an H-field, as shown. This yields an $\mathbf{S}$ vector field. We take our volume as the large ($\rightarrow \infty$) half-ball enclosing the load. The fields at infinity decrease with distance such that the integral of $\mathbf{S}$ over the dome is zero. Therefore, the surface integral is just that over the infinite plane. Analysis shows that Poynting's theorem is satisfied: the $\mathbf{S}$ vector integrated over the plane does equal *VI*, the power delivered to the load. But again, the resistor current flows as a result of the energy delivered by the E-field in the resistor to the mobile charges there. The energy flow through the vacuum around the wires is zero. In the space outside the wires, $\mathbf{S} \neq \mathbf{0_v}$, but $\nabla \cdot \mathbf{S} = 0$. This is consistent with the true energy flow outside the wires of $\mathbf{Q} = \mathbf{0_v}$, and $\nabla \cdot \mathbf{Q} = 0$.

We could test for energy flow in the space around the wires by inserting small dipole or loop antennas. It is well established experimentally and theoretically that such measurements yield zero signal. It's hard to justify the claim that energy flows in that region, when it cannot be detected by any measurement, even in principle. In contrast, radiation fields are easily measured with antennas, or thermal absorbers. Furthermore, no material can be placed in the space between or around the battery and load to block or alter the power delivered in any perceptible way: no shield, dielectric, permeable metal, conductor, or superconductor will alter the power at all. We can immerse the entire circuit in static electric fields, magnetic fields, or both, with no change in power, even though such fields change the Poynting vector field. In fact, no stationary intervention of any kind that leaves the wires intact will change the power. The wires are the *only element* needed to deliver the power.

There is no question that the charges in the wires and resistor are driven by electric fields in the wires and resistor. In steady state, these fields must arise from stationary charges, driven by the battery to points around the circuit until steady state is reached.

Case Summary: In all the above cases, **S** does *not* represent the true flow of EM field energy into the charges. In fact, of the seven cases we've mentioned (including the propagating wave and the gap between capacitor plates), **S** = **Q** (the true flow of energy) *only* in the case of a propagating EM wave.
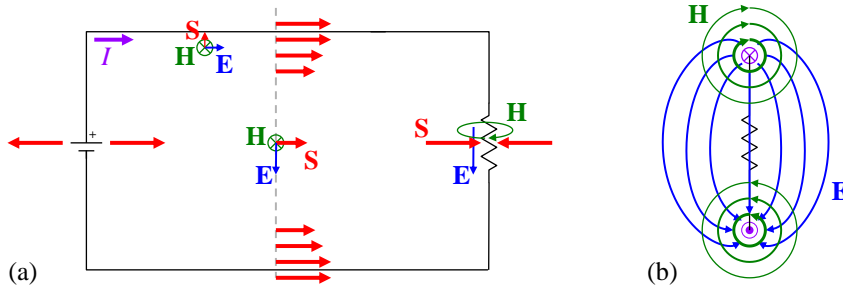
Simple Circuit: A Closer Look

Figure 4.21 (a) Poynting vector components for a steady state circuit. (b) Looking into the wires toward the resistor. The **E** and **H** fields are both stronger closer to the wire; thus, so is **S**.

Let's understand the simple resistor circuit a little better. We must be careful when computing components of **S**, because superposition does not apply to **S** (**S** is quadratic in the fields, not linear). First, (Figure 4.21a) shows the tiny E-field just outside the connecting wire that is related to the small voltage drop across the wire driving the current there. The wires are very low resistance, so this E-field is small. Standard boundary conditions at an interface show that the E-field parallel to the wire is the same just outside the wire as just inside it. This produces a component of **S** directed in toward the wire, which is small because the work done inside the wire itself is small (small resistance, small heat loss).

In contrast, the E-field between the wires (see figure) is much larger, because it has to span the full voltage of the load. Furthermore, **E** and **H** are both stronger near the wire (Figure 4.21b). This is where the bulk of the Poynting vector that pierces the plane is located.

The E-field along the resistor is also the same just inside it as just outside it (Figure 4.20a), and is large. The work is done by the E-field inside driving the current. **S** points inward toward the resistor. We haven't drawn it, but all of the **S** field lines that pierce the plane ultimately trace a path to the resistor, and terminate inside it. The details depend on how the circuit is physically laid out in space.

Three Categories of Poynting Field Lines
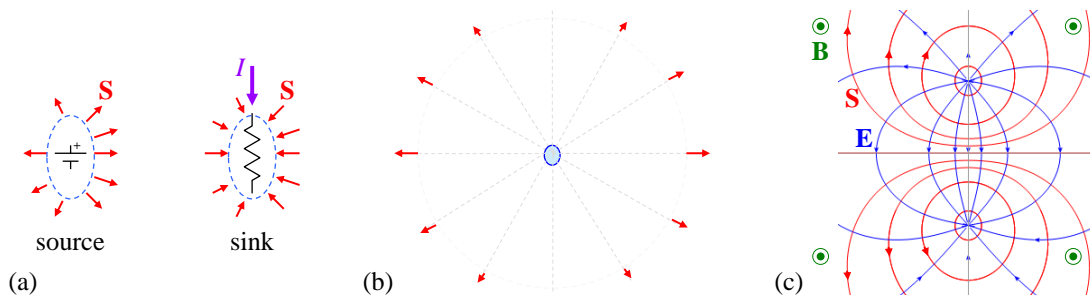
Figure 4.22 Poynting field line categories (a) sources and sinks; (b) far-field radiation; and (c) solenoidal; the B-field is uniform out of the page.

Poynting field lines come in 3 categories (Figure 4.22): (1) sources and sinks; (2) radiation; and (3) solenoidal (closed loop). Source Poynting field lines begin at sources of energy, such as batteries or transmitters. Sink Poynting field lines end at absorbers of energy, such as resistors, motors, etc. Radiation Poynting field lines start at the source and go radially outward to far field propagating waves, which carry

energy (and momentum) out to infinity. Solenoidal (close loop) Poynting field lines close on themselves, have no beginning or end, and therefore ultimately go nowhere.

Note that radiation fields are easily measured with antennas, or thermal absorbers. Sources and sinks, and solenoidal Poynting fields are unmeasurable by any electromagnetic process. There is no evidence they correspond to anything real.

Tiny Fields? No

Some physicists have speculated that perhaps there are tiny electromagnetic fields near the wires, that are "idealized away" in our idealized analysis, and that perhaps such fields make a more sensible Poynting vector. However, this idea doesn't work quantitatively. In case 5 above (Figure 4.20c), the usual idealizations for voltage, current, and fields yield quantitative agreement with Poynting's theorem for energy delivered to the load. This means the effects of any tiny fields cannot be significant.

General Relativity to the Rescue?

The only known noninvasive way, in *principle*, to experimentally locate energy-flow in space is through gravity, i.e. general relativity (GR). The source of gravity is the stress-energy tensor field $T_{\mu\nu}(\mathbf{r})$, which gives the energy and energy-flow (as well as momentum) as a function of position. Therefore, precise measures of gravity throughout a volume gives information about the stress-energy tensor. Unfortunately, the gravitational effect of realizable electromagnetic fields is far too small to be measured with any foreseeable technology.

Furthermore, the canonical process for deriving the stress energy tensor from a lagrangian fails for electromagnetism: the resulting tensor is not symmetric [Jac p605+]. GR demands that $T_{\mu\nu}$ be symmetric. (Conservation of angular momentum also demands it [Jac p608].) So (without any experimental evidence) physicists add a term to the canonical $T_{\mu\nu}{}^{(can)}$, using other relativistic criteria. The result is the (somewhat ad hoc) "symmetrized electromagnetic stress-energy tensor". Why should we believe that this tensor (and therefore the related Poynting vector $\mathbf{S} = T^{i0}$) physically describes the flow of electromagnetic energy in space? What would happen if we used a different symmetrization method, such as just taking the symmetric part of the canonical stress-energy tensor $T_{\mu\nu}{}^{(can)}$?

And finally, it is not experimentally established that electromagnetic energy and energy-flow act as a source of gravity at all. Most physicists believe it does, but there is little evidence. Some suggestion comes from cosmology, but the early stages of the universe's evolution are still uncertain. [Interestingly, there is little evidence that electrons act as a source of gravity, either. Lunar Laser Ranging (LLR) indicates that they do, but some physicists do not find this definitive [Unnik].]

Observations and Speculations on the True Energy Flow, Q(r)

One can legitimately ask, "Is the energy flux density (W/m^2) a well-defined concept?" Recall that there is no such thing as a moving EM field: E-fields, B-fields, A-fields have no velocity; it's not that their velocity is zero, it's that the concept of velocity does not apply to them. This is evident because Maxwell's equations and the Lorentz force law have no field velocities in them. **E** and **B** are simply functions of time and space, $\mathbf{E}(t, \mathbf{r})$, etc. If the vector fields **E** and **B** have no velocity, then do the scalar fields $E^2(t, \mathbf{r})$ or $B^2(t, \mathbf{r})$ have a velocity? If not, then does the energy given by $u(t, \mathbf{r}) \propto (E^2 + B^2)$ have a velocity?

For simplicity, consider a region of space with no charges. We think it's clear that if **E** has no velocity, then neither does E^2. What then is the velocity (really, flux density $\mathbf{Q}(t, \mathbf{r})$) of $u(t, \mathbf{r})$? Like Poynting's theorem, all we can say for sure about the true energy flux density is that it satisfies the continuity equation, i.e. conservation of energy:

$$\frac{\partial u}{\partial t} = -\nabla \cdot \mathbf{Q} \ .$$

This is a well-known differential equation, and given $u(t, \mathbf{r})$, it defines **Q** only up to a curl plus a constant vector. This equation alone is insufficient to uniquely define **Q** from u. There are several cases where standard electromagnetics produces **S** fields that have non-zero curl. For example, the so-called "velocity

fields" of a moving point charge, a permanent magnet in an E-field (Figure 4.20b), or a charge in a B-field. These solenoidal (closed loop) **S** components supposedly represent eddies of energy, circulating around forever, with no observable consequences (except *maybe* in GR). One can define a unique **Q** by specifying an auxiliary condition on its curl. Perhaps we might specify curl **Q** = $\mathbf{0_v}$. Then **Q**(**r**) ≡ curlfree(**S**) (where 'curlfree' is a linear operator taking only the "noncurly" component, and fixing **S** = $\mathbf{0_v}$ at infinity) yields a unique field satisfying the continuity equation for the given *u*.

This would not, however, resolve the funny issue of the Poynting vector pointing inward toward a power sink, such as a resistor.

So which **Q** (which auxiliary condition) should we use? Perhaps GR requires the original **Q** = **S**, given by naive equations of electromagnetics formed before GR was known. But one can easily imagine that GR might actually require a curl-free **Q**. This leads to a simple set of equations, and avoids the weird "ghost energy flow" that is physically implausible. Note that GR routinely performs operations on tensors such as symmetrizing, or anti-symmetrizing, so removing a curl is perfectly reasonable. Especially compared to the shenanigans that go on in QFT every day, this seems like a very simple and plausible physical theory.

We think that only GR experiments can resolve the question "Which **Q**?", and such experiments are out of technological reach for the foreseeable future.

Appendix: Detailed Discussion and Derivation of Poynting's Theorem

Poynting's theorem follows directly from the work-energy theorem and Maxwell's equations. The force (per unit volume) on a charge density by fields is:

$$\frac{\mathbf{F(r)}}{vol} \equiv \mathbf{f(r)} = \rho(\mathbf{r})\mathbf{E(r)} + \mathbf{J(r)} \times \mathbf{B(r)} \quad \left[\text{Gri 8.14 p362}\right].$$

The applied magnetic field does no work on moving charges, so the power delivered to charges comes from the E-field *alone*. The power density (work per unit time per unit volume) on the charges by the E-field is:

$$\frac{Power}{vol} = \mathbf{E} \bullet \underbrace{\rho\mathbf{v}}_{\mathbf{J}} = \mathbf{E(r)} \bullet \mathbf{J(r)} .$$

The goal of the theorem is to describe this power density in terms of just fields, without reference to the charge or current density. We can use Maxwell's equations to eliminate **J** in favor of fields:

$$\frac{Power}{vol} = \mathbf{E} \bullet \underbrace{\left(\nabla \times \mathbf{H}_{resulting} - \frac{\partial \mathbf{D}}{\partial t} \right)}_{\mathbf{J(r)}} = \mathbf{E} \bullet \left(\nabla \times \mathbf{H} - \varepsilon \frac{\partial \mathbf{E}}{\partial t} \right),$$

where **H**$_{resulting}$ is the field resulting from the current density **J** only. Any *applied* H-field has $\nabla \times \mathbf{H}_{applied} = \mathbf{0_v}$, so we can define **H** = **H**$_{resulting}$ + **H**$_{applied}$, and replace **H**$_{resulting}$ with just **H**. Neither of the components of **H** delivers any power to anything.

Now the second term can be written in terms of $\partial/\partial t$ of the electric field energy density, $\varepsilon E^2/2$, so we naturally ask, "Can we make the first term look like a *magnetic* energy density?" We need to replace **E** with **B** or **H**. This first term is a dot product, so we examine the product-rule vector identities. We find:

$$\mathbf{E} \bullet \left(\nabla \times \mathbf{H} \right) = \mathbf{H} \bullet \left(\nabla \times \mathbf{E} \right) + \nabla \bullet \left(\mathbf{H} \times \mathbf{E} \right) = \mathbf{H} \bullet \left(\nabla \times \mathbf{E} \right) - \nabla \bullet \left(\mathbf{E} \times \mathbf{H} \right).$$

Then Faraday's Law $\nabla \times \mathbf{E} = -\partial \mathbf{B}/\partial t$ makes the first term entirely magnetic:

$$\frac{Power}{vol} = \mathbf{H} \cdot (\nabla \times \mathbf{E}) - \nabla \cdot (\mathbf{E} \times \mathbf{H}) - \varepsilon \mathbf{E} \cdot \frac{\partial \mathbf{E}}{\partial t} \qquad\qquad \text{Use:} \ \mathbf{B} = \mu \mathbf{H}$$

$$= -\frac{\mathbf{B}}{\mu} \cdot \left(\frac{\partial \mathbf{B}}{\partial t} \right) - \varepsilon \mathbf{E} \cdot \frac{\partial \mathbf{E}}{\partial t} - \nabla \cdot (\mathbf{E} \times \mathbf{H})$$

$$= -\frac{\partial}{\partial t} \frac{1}{2} \underbrace{\left(\varepsilon E^2 + \frac{B^2}{\mu} \right)}_{\equiv u_{em}} - \nabla \cdot \mathbf{S} \qquad\qquad (4.6)$$

$$\underbrace{\frac{Power}{vol}}_{\substack{\text{rate of work} \\ \text{done on charges}}} = \underbrace{-\frac{\partial u_{em}}{\partial t}}_{\substack{\text{rate of field} \\ \text{energy decrease}}} - \underbrace{\nabla \cdot \mathbf{S}}_{\substack{\text{rate of field} \\ \text{energy outflow}}}$$

Thus, in a volume:

> The rate of work done on the charges equals the rate of energy loss of the fields in the volume, minus the *outflow* rate of energy from the volume.

Conservation of energy then implies that $\nabla \cdot \mathbf{S} = \nabla \cdot \mathbf{Q}$, where $\mathbf{Q}$ is the true flow of energy in the fields. But *nothing* in this derivation says that $\mathbf{S} = \mathbf{Q}$, even though this Poynting vector $\mathbf{S}$ is exactly the same as what we needed to compute the energy flow in a propagating wave.

A similar analysis for momentum defines the Maxwell stress tensor field $\mathbf{T}(\mathbf{r})$.

Acknowledgements

5   Relativistic Electromagnetics

Relativity is second only to statistics in conceptual difficulty, because it is so contrary to common experience. However, when approached systematically, it can be understood, eventually. Here we not only present valid reasoning, but also describe the failure of common *invalid* reasoning. Topics:

1. Valid frame of reference

2. Time dilation and length contraction

3. Transformation of E and B fields

4. Doppler effect

5. Transformation of electromagnetic intensity

Construction of a Valid Frame of Reference

How to construct a valid reference frame:

(1) Fill space with stationary clocks. Two clocks can be synchronized by putting a flashing light source halfway between them. Synchronize all your clocks; they remain synchronized.

(2) Fill space with a coordinate system. Measure the distances to vertices in the coordinate system with *stationary* measuring rods.

(3) Measure any event as its *position* on the coordinate system, and its *time* measured by a clock *right at* the event. Therefore, there is no propagation delay between the event and the clock, and no other funny relativistic effect, either.

Example: Measuring speed: The wrong way: we're used to marking off miles on the road, driving over the road in our car, and timing the distance between mile markers. But this overstates our speed. Why? Because once we start moving in the car, the miles on the road shrink from length contraction, and we think we're traveling farther than we really are. Remember, "really" means "distances as measured in our frame of reference," i.e. measured by rods at rest with respect to us. No other method of measuring distance is valid. **The right way**: we measure our speed relative to the road by measuring how long it takes a *single* point on the road to pass between two known points in *our* reference frame. The two known points are stationary relative to us, and we use two clocks, one at each point, to measure the time interval. Since all our clocks are synchronized, we can always use different clocks to measure time at different events.

By symmetry, if we measure our speed relative to the road, and someone on the road measures our speed relative to the road, we must get the same answer. Since either frame is valid, there is no basis for one frame to measure faster than the other, so the two observers must measure the same speed.

TBS: distance perpendicular to the motion must be the same for both observers. [Helliwell p??].

Time Dilation and Length Contraction

Alice and Bob are moving relative to each other, but of course, each has a measuring system (reference frame) that is stationary relative to themselves.
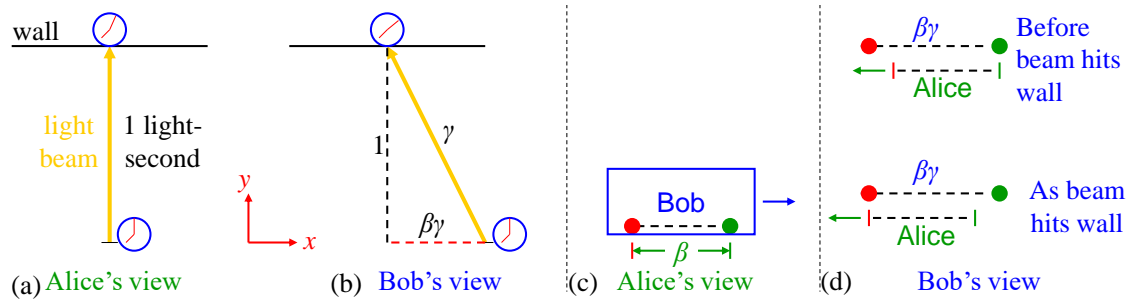
Figure 5.1 (a, b) Alice shines a flashlight in the *y* direction. It travels 1 light-second (in, of course, 1 second), and hits a wall. The two defining events are (1) Alice turns on the flashlight, and (2) the beam hits the far wall. (c, d) Length contraction: see text.

Bob travels in the $+x$ direction (parallel to the wall) at speed $\beta = v / c$. What does Bob measure? First, Bob sees himself as stationary, and Alice as moving to the left. He sees that the light beam is not perpendicular to the wall, so we conclude that the angle at which a light beam travels is relative; i.e., different observers measure different angles. (This is true of baseballs or anything else that moves.) Also, Bob sees the light travel a larger distance than Alice does; since the speed of light is the same for Alice and Bob, Bob clocks a longer time between the two events. Call Bob's time γ (in seconds). During that time, Alice moves left a distance (by his measure) of $\beta\gamma$ light seconds. Since distances perpendicular to the motion are the same for Alice and Bob, Bob measures a *y* displacement of 1 (light-second). Therefore, Bob's time is:

$$\gamma^2 = \left(\beta\gamma\right)^2 + 1^2, \qquad \gamma^2 - \beta^2\gamma^2 = 1, \qquad \gamma = \frac{1}{\sqrt{1-\beta^2}}.$$

During this time, a clock stationary in Alice's frame measures only 1 second. Therefore, Bob sees a moving clock as ticking more slowly than his own (stationary) clocks. This is **time dilation**. By symmetry, Alice (and everyone else in the universe) must see the same phenomenon.

> Moving clocks tick more slowly than stationary ones.

This is *not* an illusion caused by the finite speed of light (see discussion of valid reference frames above). This is real.

> Do not be misled by this erroneous argument: Alice sees one second on her clock, and sees Bob time $\gamma > 1$ seconds on his clocks, therefore Alice thinks Bob's moving clocks run *faster*. The error is that Bob used two clocks, at different *x* positions, to make his measurement. In Bob's frame, those clocks are synchronized, but in Alice's frame they are not.

If Alice looks at any *one* of Bob's clocks (not two clocks in two different *x* positions), then she will indeed see Bob's clock run more slowly.

Length contraction: By symmetry, Alice and Bob agree that they are moving at a speed β relative to each other. When Alice turns on her flashlight, she simultaneously marks a green dot on Bob's moving frame. When she detects the light on the wall, she marks a red dot on Bob's frame. During the interval, Alice measures Bob move $\beta\Delta t = \beta$ light-seconds across her frame; i.e. Alice measures the *x* distance between two moving dots as β light-seconds. The dots are stationary in Bob's frame, and as already noted, Bob measures their separation as $\beta\gamma$ light-seconds. Therefore, Alice measures the length of moving things as shorter than someone at rest with respect to those things, by the factor $1/\gamma$.

> Do not be misled by this erroneous argument: Look at the diagram of Alice's measurement of the distance between the dots she painted on Bob's reference frame. As noted, Alice measures that distance as β. Suppose Alice has a ruler on her ground, with which she measures that distance. One might (incorrectly) conclude that at that moment, Bob sees her moving ruler aligned with his dots, and conclude that in his frame, the ruler measures $\gamma\beta$, and thus moving rulers *grow* instead of

> *contract*. This is wrong because in Bob's frame, the dots do *not* align with Alice's ruler at a single instant in time.

For the correct reasoning, see Figure 5.1d: In Bob's frame, Alice is moving to the left, and her ruler is short. Bob measures the red dot aligns with the red end of Alice's ruler *after* the green dot aligns with the green end of Alice's ruler.

> Moving objects are measured *shorter* than when at rest.

The Electric Field of a Relativistic Charge

Since at least 1959, many references reproduce a misleading diagram of the E-field of a relativistically moving charge (Figure 5.2a, crossed out). The radial lines look like electric field vectors, but they can't be: a single point in space has only one **E** vector associated with it, not multiple, and the E-field is undefined *at* a point charge.

Instead, the radial lines are field lines, more clearly shown in Figure 5.2a, right. Recall that the density of field lines gives the strength of electric field, and the field is stronger at the top and bottom, and weaker along the line of motion.

We may consider **E** at points on a sphere some distance from the charge (Figure 5.2b). This shows the field using standard vector depictions: longer vectors are higher magnitude. In the static case, all the vectors are equal magnitude at equal distance. For a relativistic charge (Figure 5.2c), the field is stronger at the top and bottom, and weaker along the line of motion. The field strength is symmetric about the vertical line through the charge.
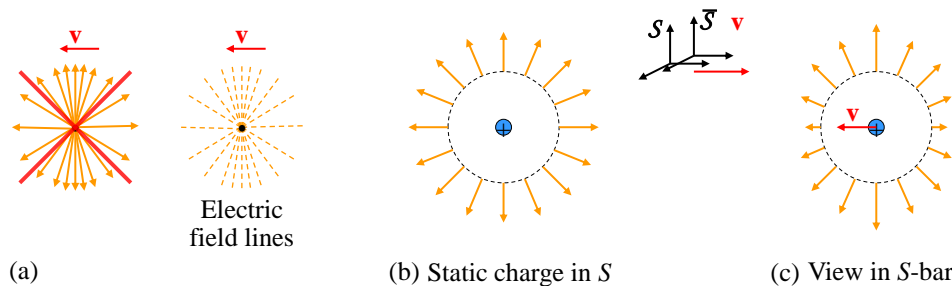

Electric field lines

(a)   (b) Static charge in S (c) View in S-bar

Figure 5.2 (a) Misleading, and better, diagram of field lines of a relativistic charge. (b) Electric field vectors for a static charge in S; and (c) for the moving charge observed in S-bar.

Some references have an even worse diagram that shows a circle of electric field vectors compressed by length contraction! The variation of electric field with angle θ is *not* due to length contraction. For one thing, length contraction doesn't contract vectors in general; it contracts lengths. Forces, fields, and other mathematical objects transform differently (don't even get me started on angular momentum). In this case, **E** is proportional to force, which depends on rate-of-change of momentum, which depends on time dilation, so relativistic E-fields are more complicated than just length contraction.

Relativistic electrons are not hard to come by. Cathode ray tubes (CRTs) are old fashioned video displays that used to be in every television. Many TVs accelerated electrons to ~25 keV. This corresponds to $\gamma = 1.05$, and $v/c = 0.3$.

## Transformation of E & B Fields

There's a simple picture of the relativistic transformation of E & B fields. We seek **E'** and **B'**, the fields in the moving frame, in terms of E and B, the fields in the stationary (lab) frame. First, we'll find **E'**, the E-field in the moving frame. We start with transformation of the E-field alone. Imagine an E-field created by a sheet charge:
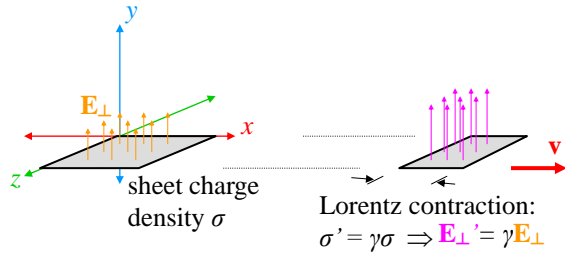
In the moving frame, the sheet charge is compressed to a smaller area, by the length contraction factor γ. Therefore, $\mathbf{E'}$ is increased by the same factor.

Now we add a B-field to the stationary frame:
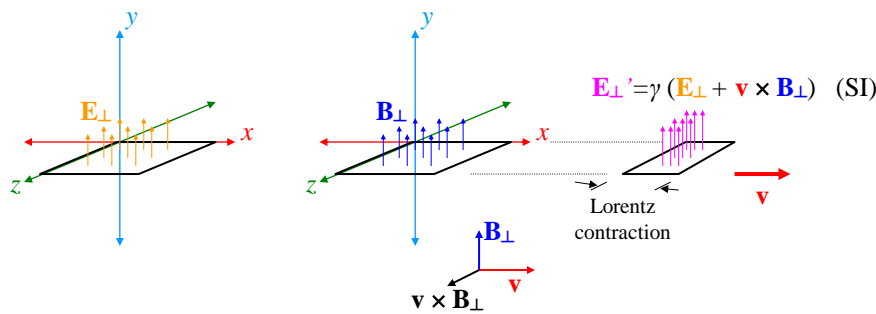
A stationary charge in the moving frame, is (in the lab frame) moving relative to $\mathbf{B}$. Nonrelativistically, the charge feels a force of $\mathbf{F} = q\mathbf{v} \times \mathbf{B}$, just like an E-field of $\mathbf{v} \times \mathbf{B}$ (SI units), simply from the Lorentz force law. This appears in the moving frame to be an E-field: $\mathbf{E'} = \mathbf{v} \times \mathbf{B}$. Boosting to relativistic speeds, picture that the E-field can be thought of as "lines of force" whose density per unit area is proportional to the field strength. Because of Lorentz contraction, the same lines of force occupy a smaller area in the moving frame, and so their density is increased by γ:

$$\mathbf{E}_\perp{}' = \gamma\left(\mathbf{E}_\perp + \mathbf{v} \times \mathbf{B}_\perp\right), \qquad E_\parallel{}' = E_\parallel \qquad \text{[Jac 11.148 p558]}.$$

There is no contraction of the plane perpendicular to the direction of motion, and no Lorentz force from $B_\parallel$ (parallel to the velocity). Hence, $E_\parallel{}' = E_\parallel$.

$\mathbf{B'}$ is similar. TBS??

$$\mathbf{B}_\perp{}' = \gamma\left(\mathbf{B}_\perp - \mathbf{v} \times \mathbf{E}_\perp\right), \qquad B_\parallel{}' = B_\parallel \qquad \text{[Jac 11.148 p558]}.$$

We can combine the E (and B) transformations above into a single vector equation for each. If we start with the perpendicular formulas, and note that the cross-product term does not contribute anything to the parallel component, then replacing $\mathbf{B}_\perp \rightarrow \mathbf{B}$ would incorrectly inflate the parallel component by γ. We can fix this by just subtracting off $(\gamma - 1)E_\parallel = (\gamma - 1)\hat{\boldsymbol{\beta}} \cdot \mathbf{E}$:

$$\mathbf{E'} = \gamma\left(\mathbf{E} + \boldsymbol{\beta} \times \mathbf{B}\right) - (\gamma - 1)\hat{\boldsymbol{\beta}} \cdot \mathbf{E} \qquad\qquad \mathbf{B'} = \gamma\left(\mathbf{B} - \boldsymbol{\beta} \times \mathbf{E}\right) - (\gamma - 1)\hat{\boldsymbol{\beta}} \cdot \mathbf{B}.$$

Jackson complicates this formula with the identity $(\gamma - 1)\hat{\boldsymbol{\beta}} \cdot \mathbf{E} = \dfrac{\gamma^2}{\gamma + 1}\boldsymbol{\beta}(\boldsymbol{\beta} \cdot \mathbf{E}) \Rightarrow$

$$\mathbf{E'} = \gamma\left(\mathbf{E} + \boldsymbol{\beta} \times \mathbf{B}\right) - \frac{\gamma^2}{\gamma + 1}\boldsymbol{\beta}(\boldsymbol{\beta} \cdot \mathbf{E}), \quad \mathbf{B'} = \gamma\left(\mathbf{B} - \boldsymbol{\beta} \times \mathbf{E}\right) - \frac{\gamma^2}{\gamma + 1}\boldsymbol{\beta}(\boldsymbol{\beta} \cdot \mathbf{B}) \qquad \text{[Jac 11.149 p558]}.$$

Acceleration Without Force

Consider a particle moving in the *x*-direction, and a force pushing it in the +*y*-direction. The particle accelerates up, but it also *decelerates* in the *x*-direction, even though there is no force in the *x*-direction. That's because the acceleration in *y* increases the particle's magnitude of velocity (speed), and therefore the particle's $\gamma = 1/\sqrt{(1 - v^2/c^2)}$. The *x-momentum* doesn't change, but when γ increases, v_x must *decrease* to keep the same momentum:

$$F_y = \frac{dp_y}{dt} \qquad \Rightarrow \qquad v_y \text{ increases, and } |\mathbf{v}| \text{ and } v^2 \text{ increase.}$$

$$F_x = 0 = \frac{dp_x}{dt} \qquad \Rightarrow \qquad p_x = \gamma m v_x = const. \qquad \gamma \text{ increasing} \qquad \Rightarrow \qquad v_x \text{ decreases!}$$

On-Axis Doppler
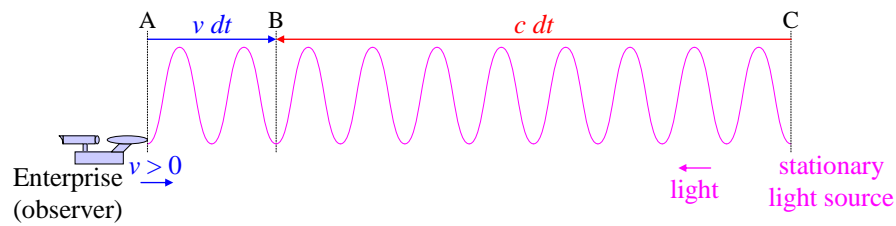
We first consider the simpler case of motion along the line of propagation. We are a stationary light source, and there is a moving observer traveling directly toward us (above). We can deduce what the observer sees from our own observations. Note that wave crests are countable things, and therefore the same in all frames. In an interval of our time *dt*, the observer starts at A and ends at B, and simultaneously, the wave train starts at C and also ends at B. During this interval, the observer sees all the crests between A and C, i.e.

$$n = (f\ dt)\left(\frac{c+v}{c}\right) = (f\ dt)\left(1 + \frac{v}{c}\right) \qquad where \quad v \text{ is the speed of } approach .$$

However, his clock advances only (*dt* / γ) seconds, due to time dilation. Without loss of generality, let *dt* = 1 second. Then in one of the observer's seconds, he sees

$$n' = f' = \gamma f\left(1 + \frac{v}{c}\right).$$

This formula is valid for positive and negative *v*. And of course, only the *relative* motion of the source and observer matter (after all, this *is relativity*, isn't it?)

Off-Axis Doppler

For variety, we now take the reverse (but equivalent) view that you are now the observer, you are stationary, and the light source is moving. (I think this is a little harder to follow). You are in the enterprise, and a star has an off-axis relative motion of **v** (see diagram).

Let   *r* = distance from enterprise to star
        x = "horizontal" distance to star
        v = speed of star in x direction (*v* < 0)
        f_s = frequency at which star emits light (in its own frame)
        f_o = frequency at which enterprise sees light
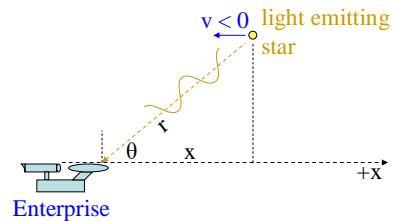
From the diagram, and recalling that $v < 0$,

$$r^2 = x^2 + d^2$$

$$2r \frac{dr}{dt} = 2x \frac{dx}{dt} = 2r(\cos \theta) v$$

$$dr = v(\cos \theta) dt$$

Because of time dilation, the emitted frequency in the Enterprise frame is multiplied by a factor of $(1/\gamma)$. In addition, the waves are "compressed" by the changing separation between the star and Enterprise. In a time interval dt, the Enterprise sees the star emit $(f_s/\gamma)\, dt$ cycles of EM wave. But those cycles get squeezed into a space of $(c\, dt + dr)$ distance $(dr < 0)$. Thus, the number of cycles the enterprise sees in a full distance of $(c\, dt)$ is (recalling again that $v < 0$):

$$f_o\, dt = \frac{f_s}{\gamma} dt \left(\frac{c\, dt}{c\, dt + dr} \right) = \frac{f_s}{\gamma} dt \left(\frac{c\, dt}{c\, dt + v \cos \theta\, dt} \right)$$

$$f_o = \frac{f_s}{\gamma} \left(\frac{c}{c + v \cos \theta} \right) = \frac{f_s}{\gamma(1 + \beta \cos \theta)}$$

This reduces to the well known simpler formula for on-axis motion when $\theta = 0$:

$$f_o = \frac{f_s}{\gamma(1 + \beta \cos \theta)} = \frac{f_s}{\gamma(1 + \beta)} .$$

Transformation of Intensity (Power Density)

If a light wave has an intensity I in the stationary frame, what is its intensity I' as seen in an on-axis moving frame? We first find this from the transformation of the E field, and noting that $I \propto E^2$. Then we show another, more intuitive, way to find the intensity transformation.

First, the transforming E-field approach: In gaussian units, a free space wave as $B = E$. For on-axis motion:

$$\mathbf{E}' = \gamma \left(\mathbf{E} + \frac{\mathbf{v}}{c} \times \mathbf{B} \right) \qquad \Rightarrow \qquad E' = \gamma \left(E + \frac{v}{c} B \right) = \gamma \left(E + \frac{v}{c} E \right) = E\, \gamma \underbrace{\left(1 + \frac{v}{c} \right)}_{\text{Doppler shift}} .$$

> Electric fields in propagating waves transform exactly like the Doppler frequency shift.

Then, for source and observer *approaching*:

$$I \propto E^2 \quad \Rightarrow \quad I' = \gamma^2 \underbrace{\left(1 + \frac{v}{c} \right)^2}_{\substack{\text{square of} \\ \text{Doppler shift}}} I = \frac{(1 + v/c)^2}{(1 - v/c)(1 + v/c)} I = \underbrace{\frac{1 + v/c}{1 - v/c}}_{\substack{\text{also square of} \\ \text{Doppler shift}}} I .$$

> Intensity transforms as the square of the Doppler shift.

It is perhaps more intuitive (though less theoretically sound) to consider the light wave as a "flux of photons," i.e. some number of photons per unit area per unit time, impinging on the observer. The flux times the photon energy is proportional to intensity. If the flux seen by a stationary observer is f, what is the flux seen by a moving observer, f'? The picture is essentially identical to that we used for deriving on-axis Doppler shift:
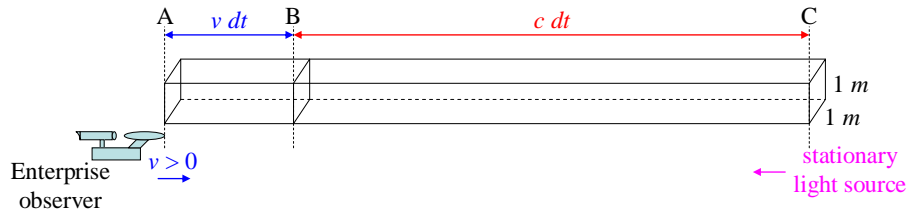
Without loss of generality (wlog), the tube has cross-sectional area of 1m^2. As seen by you (stationary), the moving observer intercepts $n_1 = (f\ dt)\left(1 + \dfrac{v}{c}\right)$ photons in the time interval dt. But again, his clock only advanced $(dt\ /\ \gamma)$ seconds. Therefore, the observer's photons per unit time (the cross sectional area for him is still 1m^2) is $n' = f' = \gamma f\left(1 + \dfrac{v}{c}\right)$. The flux transforms like Doppler shift. (Note also that as the moving observer looks ahead at the photon "density" in space, his density is greater than yours by a factor of γ, because the tube is shorter by a factor of γ due to length contraction.)

But each photon's energy is proportional to its frequency, $E = \hbar\omega$, and each photon's frequency is Doppler shifted by $\gamma(1 + v/c)$, as shown earlier. Ergo,

$$I' \propto n'E' \qquad\Rightarrow\qquad I' = I\gamma\left(1 + \frac{v}{c}\right)\gamma\left(1 + \frac{v}{c}\right) = I\gamma^2\left(1 + \frac{v}{c}\right)^2 \qquad v \equiv \text{approach speed},$$

as before.

How Big Is a Photon?

So what's the theoretical objection to the above photon-flux model of intensity transformation? The problem is that photons are *not* point-like particles of small dimensions. The electromagnetic wave is a widely distributed quantum-mechanical wave-function (of sorts) of the photon. However, in quantum field theory (the only valid quantum theory for photons), the photon density is proportional to the square of the electromagnetic intensity, just as with massive particles. The photon itself, though, cannot be localized, or thought of as a point, but the density approach is rigorously valid.

There are many real-world situations in which a single photon is spread over many meters or kilometers, e.g., lunar laser ranging which detects an individual photon spread laterally over many square km of the earth, and a fraction of that photon is spread over many square meters of the receiving telescope.

The Pressure of Light

We can derive the pressure of light from the fundamental relativistic energy momentum relationship:

$$E^2 = \left(mc^2\right)^2 + \left(pc\right)^2.$$

Since light is (i.e., photons are) massless, we have:

$$E = pc \qquad \text{(light or other massless particles)}.$$

Now force $F = dp/dt$, so for normal incidence light absorbed onto a surface:

$$F = \frac{dp}{dt} = \frac{1}{c}\frac{dE}{dt} = \frac{P}{c} \qquad where \quad P \equiv \text{power of radiation}.$$

Taken per unit area:

$$Pressure = \frac{I}{c} \qquad where \quad I \equiv \text{intensity} = \text{power per unit area} .$$

Example: Reflection Off a Moving Mirror

We show here a novel solution to the classic test question of light reflecting off a moving mirror, using conservation of energy entirely in the lab frame.

The question: An ideal mirror moves to the right, perpendicularly to its face. A physicist shines a light on it with normal incidence from the left (Figure 5.3). Find the power reflection coefficient of the moving mirror $\Gamma \equiv (P_r / P_i)$.
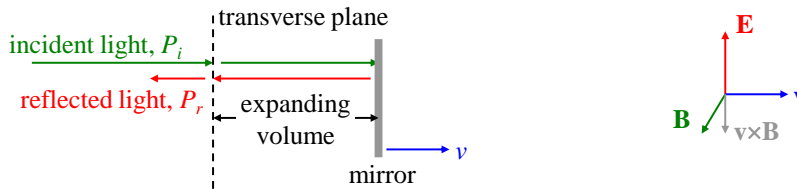

Figure 5.3 Light reflecting off a moving mirror.

The intensity transformation way: In previous sections, we derived (in two different ways) that E and B fields transform like the Doppler shift, and therefore intensity transforms like the square of Doppler (the minus sign below is in the numerator because source and observer are receding):

$$I' = I_i \left(\frac{1 - v/c}{1 + v/c} \right) \qquad v \equiv \text{recession speed} .$$

In this problem, we have two intensity transformations of the same factor: once into the mirror frame, and once again from the mirror frame to the lab frame. The result is then simply

$$I_r = I' \left(\frac{1 - v/c}{1 + v/c} \right) = I_i \left(\frac{1 - v/c}{1 + v/c} \right)^2 \qquad \Rightarrow \qquad \Gamma = \left(\frac{1 - v/c}{1 + v/c} \right)^2 .$$

The energy way: It is perhaps instructive to consider conservation of energy in the lab frame: the energy crossing a transverse plane (parallel to the mirror) goes into (1) filling up the increasing distance to the mirror with electromagnetic energy, (2) doing work on the mirror from the pressure of light, and (3) transmitting light across the boundary back into the laboratory. Without loss of generality, we can use an incident power of 1 (in arbitrary units). In some time interval, the fraction of the incident energy that gets stored in the expanding space up to the mirror is then v/c. The rest of the light hits the mirror, and reflects. Similarly, the fraction of reflected energy that gets stored in the expanding space is $\Gamma v/c$.

The work done on the mirror by the incident light is $F \Delta x$, where $F = P/c$, and Δx in a unit time is v. Then $W = Pv/c$. However, P is not $P_i = 1$, but just the fraction of P_i that is not stored in the expanding volume. We found above that the stored energy is v/c, so that which hits the mirror is $(1 - v/c)$, and the work is $W = (1 - v/c)v/c$.

Similarly for the work of the reflected light, but the total work is from the light stored in the expanding volume *plus* the light sent back through the transverse plane into the lab. The stored energy is $\Gamma v/c$, and the reflected light power is Γ. Hence the work is $W = \Gamma(1 + v/c)v/c$.

Then by conservation of energy, the incident energy equals the sum of stored incident plus reflected energy, plus the incident plus reflected work:

$$1 = \frac{v}{c} + \left(1 - \frac{v}{c}\right)\frac{v}{c} + \Gamma\left(1 + \frac{v}{c}\right)\frac{v}{c} + \Gamma\frac{v}{c} + \Gamma \qquad \Rightarrow \qquad 1 - 2\frac{v}{c} + \left(\frac{v}{c}\right)^2 = \Gamma\left(1 + 2\frac{v}{c} + \left(\frac{v}{c}\right)^2\right)$$

stored in space   incident work on mirror   reflected work on mirror   stored in space   reflected into lab

$$\Gamma = \left(\frac{1 - v/c}{1 + v/c}\right)^2$$

Beaming

Both Jackson and Griffiths have misleading diagrams for relativistic beaming at high γ. Both their diagrams show no backward emission components at all. Inspection of the equation shows immediately that this is not the case; the power is finite for all angles up to π. Here are more suggestive diagrams:
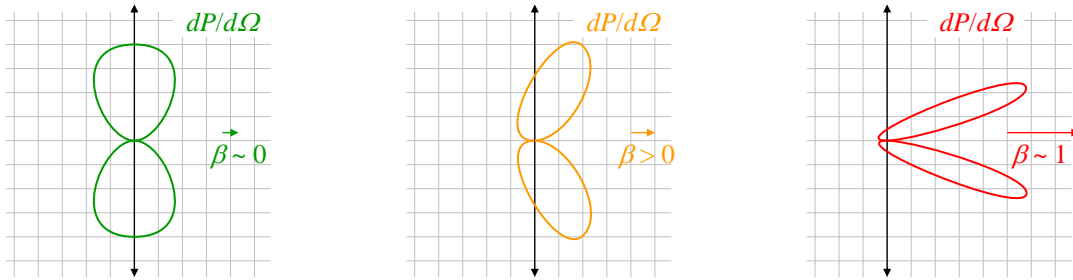

Figure 5.4 Radiation pattern (cross-section) for transverse acceleration at various velocities.

Remember that the radiation pattern is axially symmetric about the velocity vector, so the "beaming" is more like a big hollow *cone* of light than a pair of headlights.

Covariant Form of Maxwell's Equations

TBS. Given that $F^{\mu\nu}$ is a rank-2 tensor, we can show that J^{μ} is a 4-vector:

Define $\Lambda^{\mu}_{\delta} = \dfrac{\partial \overline{x}^{\mu}}{\partial x^{\delta}} \equiv$ Lorentz transformation matrix from s to $\overline{s}$

$\lambda^{\alpha}_{\nu} = \dfrac{\partial x^{\alpha}}{\partial \overline{x}^{\nu}} = \left(\Lambda^{\mu}_{\delta}\right)^{-1} \equiv$ Lorentz transformation matrix from $\overline{s}$ to s

Then $\lambda^{\alpha}_{\nu} \Lambda^{\nu}_{\lambda} = \delta^{\alpha}_{\lambda}$ the identity transformation

Note $\overline{\partial}_{\mu} = \lambda^{\nu}_{\mu} \partial_{\nu} = \partial_{\nu} \lambda^{\nu}_{\mu}$ *from* $\dfrac{\partial}{\partial \overline{x}^{\mu}} = \dfrac{\partial x^{\nu}}{\partial \overline{x}^{\mu}} \dfrac{\partial}{\partial x^{\nu}}$

In s: $\partial_{\nu} F^{\mu\nu} = \mu_0 J^{\mu}$

In $\overline{s}$: $\overline{\partial}_{\nu} \overline{F}^{\mu\nu} \equiv \dfrac{\partial \overline{F}^{\mu\nu}}{\partial \overline{x}^{\nu}} = \partial_{\alpha} \underbrace{\lambda^{\alpha}_{\nu} \Lambda^{\nu}_{\lambda}}_{=\delta^{\alpha}_{\lambda}} \Lambda^{\mu}_{\delta} F^{\delta\lambda} = \Lambda^{\mu}_{\delta} \underbrace{\partial_{\alpha} F^{\delta\alpha}}_{=\mu_0 J^{\delta}} = \mu_0 \Lambda^{\mu}_{\delta} J^{\delta} = \mu_0 \overline{J}^{\mu}$

6 Shorts

Bound and Gagged

Are bound charges real? To understand this, we must be more specific: there are *real* bound charges that are responsible for permittivity (and equivalently, for dielectric constant). However:

> The so-called "bound charge distributions" that one can calculate *are not real*
> (despite what some references say).

[e.g., Culwick 1959 p??]. A simple example proves the point. Consider a broad (i.e., infinite) flat sheet of dielectric material:
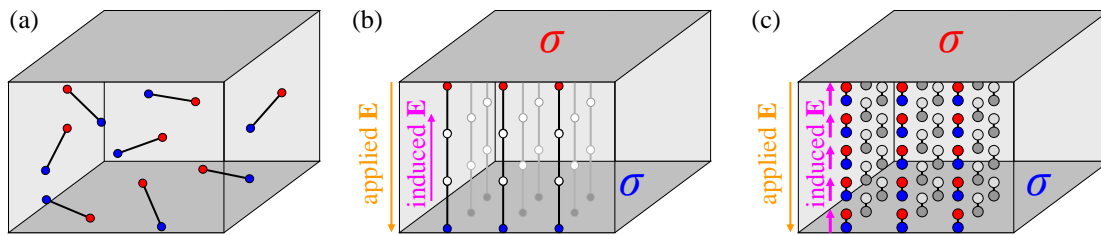

Figure 6.1 (a) Dielectric with no E-field. Blue is positive, red is negative. (b) Dielectric with long dipoles in an E-field. (c) Dielectric with higher volume density of shorter dipoles in E-field.

Figure 6.1 shows only a representative volume, but the top and bottom sheets are given as broad. Inside the dielectric are molecules with electric dipole moments. In Figure 6.1a, absent an E-field, the dipoles are randomly oriented. They therefore sum to zero total dipole moment. Figure 6.1b, in an E-field, the dipoles align. In this example, the dipoles are unrealistically long, and the positive (blue) ends of the top two rows land on top of, and cancel, the negative (red) ends of the bottom two rows. The net result is that all inside charges cancel, and we are left with two surface charges, negative on top, and positive on the bottom. These charge densities are what one computes with the standard formula:

$$\sigma_{bound} = \mathbf{P} \cdot \hat{\mathbf{n}} \qquad \text{[Gri ??]} .$$

The induced electric field from the dipoles is uniform throughout the slab.

However, this is clearly an unrealistic depiction. Real molecules aren't that close together; their dipole moments are not ideal point charges separated by vacuum; and adjacent charges don't exactly cancel.

More realistically (but still idealized), above right, the dipoles don't overlap (they're shorter); each one has a larger dipole moment, so the endpoint charges are larger; and there are more of them. However, the dipole *moment* per unit volume is the same as the middle diagram. Therefore, these dipoles add up to the same total dipole moment in the slab as the middle diagram, but they result in *larger real* surface charge densities. In addition, there are sheet charges at other layers below the surface, which cause the induced E-field to be large between the charges of each dipole, and zero in the gaps between layers of dipole. That's because a *real* dipole moment is spread throughout the volume of the dielectric. In fact, the standard definition of polarization for a dielectric (in an E-field) is dipole moment *density* (dipole moment *per unit volume*). The total dipole moment in the slab is the product of moment-density times volume:

$$\mathbf{p} = (\mathbf{P})(Volume) \qquad where \quad \mathbf{p} \equiv dipole\ moment;\ \mathbf{P} \equiv dipole\ moment\ density .$$

Therefore:

> The bound surface charge densities we compute for dielectrics are *model* densities which give
> *equivalent dipole moment* to the real dielectric, but *are not real*.

Finally, Figure 6.1 shows cartoon dipoles for illustration, but actual atomic/molecular dipoles are more like Figure 6.2. Influenced by an external E-field, the wave-function of the electron(s) shifts slightly to create a small dipole moment of the atom or molecule. In quantum perturbation theory, wave-function

shifts occur at 2nd-order in energy, so calculating the dipole moment requires finding 2nd-order perturbations.
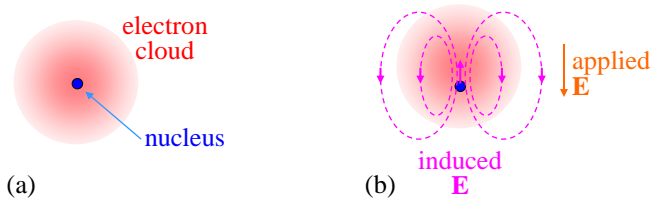

Figure 6.2 (a) Unpolarized atom, with symmetric electron "cloud." (b) Polarized atom with displaced nucleus/cloud.

Pole to Pole: Current Loops Are Approximate Dipole Moments

A current loop has mostly a dipole moment, but any finite-sized loop also has other multipole moments in it. A pure dipole results only in the limit that the loop size goes to zero, while the current increases to infinity, in such a way that the product *IA* remains constant. (This is similar to an electric dipole, which is a pure dipole only in the limit of the charge $q \to \infty$, while separation $d \to 0$, such that the product $qd =$ constant.) For example, in a pure magnetic dipole pointing up, all the magnetic field in the equatorial plane points down (except exactly at the origin):

$$\text{dipole:} \quad \mathbf{B}(\mathbf{r}) = \frac{3(\mathbf{M}\cdot\hat{\mathbf{r}})\hat{\mathbf{r}} - \mathbf{M}}{r^3} \qquad \text{[Jac 5.56 p186, Gri 3.103 p153]}$$

But in a finite loop, the field inside the loop points up. Hence, a finite loop cannot be a pure dipole.

(a) Pure dipole   (b) Finite size current loop

Figure 6.3 (a) Pure dipole B-field points down everywhere in equatorial plane. (b) Finite-size current loop B-field points up inside loop.

The Mythical Moving B-field

Can B-fields move? No. B-fields must be computed either (1) in the frame in which they are used (the lab frame), or (2) in some other frame, and then transformed into the lab frame. In either case, there is no such thing as a "moving" B-field, although the latter case can sometimes be profitably thought of as such. Note that all of electromagnetics is contained in the Lorentz force law and Maxwell's Equations. Neither of these has any place for a B-field velocity.

In the first case (computing a B-field in the lab frame), B-fields may vary in space and time, because they are derived from currents which vary in space and time. However, they cannot have any "velocity" because there is no such concept in the Biot-Savart Law, or in Ampere's Law, which are the two primary ways of computing B-fields.

In the 2nd case, we may compute B-fields in a moving frame (in which frame they are not moving), and then transform them into the lab frame. Once in the lab frame, though, they again have no velocity. In the nonrelativistic version of this case, one can sometimes reasonably visualize such B-fields as "moving," which is a shortcut for transforming the fields into E-fields in the lab frame. However, there is no support for moving B-fields in either Maxwell's equations or the Lorentz force law.

Similar considerations apply the vector potential, either 3-vector $\mathbf{A}(t, \mathbf{r})$ or 4-vector A^μ. There is no model of a moving vector potential in Maxwell's equations, or in Lagrangian or Hamiltonian formulations of EM fields and charges. The B-field (and the 3-vector A-field) is relative: it is frame dependent. At relativistic speeds, this is unarguable. Nonrelativistic speeds are just the low-speed limit of the full relativistic theory.

There are no EM equations in which the velocity of the B-field (or A-field) is either defined or used.

Check Your Gauges

Gauge freedom, aka gauge invariance, is a very important theoretical tool for electromagnetics. It also has important implications for quantum electrodynamics (QED), and all quantum field theories. Potentials, both scalar and vector potentials, often simplify the calculations of real problems. But neither the scalar nor vector potentials are unique functions for a given system. Some choices of potentials are easier to work with than others, and different questions lead to different choices for the "simplest" potentials. This freedom of choice in the potentials is called **gauge invariance**. It turns out that with the concept of a vector potential, inescapably goes the concept of gauge invariance.

We present a brief review of gauge invariance, and then the two most common gauges, "Lorenz gauge" and "radiation gauge" (aka "Coulomb gauge" or "transverse gauge"). We clarify that both gauges are actually families of gauges. This section assumes a thorough understanding of vector calculus, and that you have seen the concept of a vector potential.

> Gauge invariance is based on the fact that all physical predictions
> can be made from knowledge of the **E** and **B** fields.

The $\mathbf{E}(t, \mathbf{x})$ and $\mathbf{B}(t, \mathbf{x})$ fields have simple physical meanings which are summarized by the Lorentz force on a test charge at some point in space:

$$\mathbf{F} \equiv \frac{d\mathbf{p}}{dt} = q\left(\mathbf{E} + \frac{\mathbf{v}}{c} \times \mathbf{B} \right) \qquad q \equiv \text{test charge. gaussian units, relativistically valid .}$$

This says that the **E** field acts on any charge, moving or not. The **B** field acts only on moving charges. The **E** field can do work on a charge, but the **B** field cannot directly do work on a charge. The **E** and **B** fields are directly measurable by their effects on a test charge. Physically, the sources of **E** fields are (1) charges (moving or not), and (2) changing **B** fields. The source of **B** fields is charge currents (i.e., moving charges).

Sometimes however, instead of working directly with **E** and **B** fields, it is more convenient to work with potentials, which generate the **E** and **B** fields by a simple mathematical operation. Two such potentials are needed: a scalar potential $\Phi(t, \mathbf{x})$, and a vector potential $\mathbf{A}(t, \mathbf{x})$. These potential functions are not unique: different choices can produce the exact same **E** and **B** fields, and thus the same physical predictions. N

> All known laws of physics, classical and quantum, are gauge invariant.

The statement of gauge invariance is nothing more than the mathematical fact that I can construct multiple pairs of functions, $\Phi(t, \mathbf{x})$ and $\mathbf{A}(t, \mathbf{x})$, that generate the *same measurable* **E** and **B** fields. The fact that many such pairs of functions exist is gauge invariance. There is no mathematical requirement that these two functions be partitioned in such a way as to have any immediate physical meaning.

Static Fields

> The *defining* property of the magnetic vector potential $\mathbf{A}(t, \mathbf{x})$ is that $\mathbf{B}(t, \mathbf{x}) = \nabla \times \mathbf{A}(t, \mathbf{x})$.

Static fields are not functions of time, so in the static case, the above definition can be reduced to:

$$\mathbf{B}(\mathbf{x}) = \nabla \times \mathbf{A}(\mathbf{x}) \qquad \text{(static fields) .}$$

Then we can define an electric potential $V(\mathbf{x})$ that directly relates to energy:

Define $V(\mathbf{x})$ such that $\mathbf{E}(\mathbf{x}) = -\nabla V(\mathbf{x})$ $\Rightarrow$ $V(\mathbf{x}) = -\int_{\mathbf{x}_0}^{\mathbf{x}} \mathbf{E}(\mathbf{x}) \cdot d\mathbf{x}$

where we choose $V(\mathbf{x}_0) \equiv 0$.

This electric potential has no gauge freedom (only an arbitrary additive constant). The **E**-field is given by the gradient of the electric potential, which in the static case also serves as a scalar potential:

$$\mathbf{E}(\mathbf{x}) = -\nabla V(\mathbf{x}) = -\nabla \Phi(\mathbf{x}) \qquad where \quad V(\mathbf{x}) \equiv \text{electric potential} \quad \text{(static fields)}.$$

$V(\mathbf{x})$ has units of joules/coulomb, or volts. Both of these equations involve only derivatives of the potentials, so adding a constant to Φ or **A** doesn't change the fields. This is a trivial form of gauge invariance, which we ignore.

More importantly, the **A**-field is not unique because there are many functions $\mathbf{A}(\mathbf{x})$ which have the same curl. In particular, any irrotational (i.e., curl-free) vector field can be added to **A**, and it doesn't change the curl.

> Any irrotational field can be written as the gradient of a scalar field,
> and the gradient of any scalar field is irrotational.

Our **gauge freedom** for static fields then means we can add the gradient of any scalar function to **A**, and get another, physically equivalent, vector potential:

$$\mathbf{A}'(\mathbf{x}) = \mathbf{A}(\mathbf{x}) + \nabla \Lambda(\mathbf{x}) \qquad \text{(gauge invariance for static fields)},$$

where **A'** is physically equivalent to **A**, and $\Lambda(\mathbf{x})$ is any scalar function.

Time Varying (Dynamic) Fields

Things get more complicated in the dynamic case. Following [Jac sec. 6.2-6.5 p239-48], electric and magnetic fields can still be specified in terms of a scalar potential and vector potential, but now a time-varying **B**-field, and thus a time-varying **A**-field, contributes to the **E**-field:

$$\nabla \times \mathbf{E}(t, \mathbf{x}) = -\frac{\partial \mathbf{B}(t, \mathbf{x})}{\partial t} = -\frac{\partial}{\partial t}\left(\nabla \times \mathbf{A}(t, \mathbf{x})\right) \qquad \Rightarrow \qquad \nabla \times \left(\mathbf{E}(t, \mathbf{x}) + \frac{\partial \mathbf{A}(t, \mathbf{x})}{\partial t}\right) = 0.$$

Note that the curl and time derivatives commute, so we can bring the $\partial/\partial t$ inside the curl operator. Since the curl of $\mathbf{E} + \partial \mathbf{A}/\partial t$ is zero, it can be written as the gradient of a scalar function which we call $-\Phi$:

$$\mathbf{E}(t, \mathbf{x}) + \frac{\partial \mathbf{A}(t, \mathbf{x})}{\partial t} = -\nabla \Phi(t, \mathbf{x}) \qquad\qquad \mathbf{B}(t, \mathbf{x}) = \nabla \times \mathbf{A}(t, \mathbf{x})$$

$$\mathbf{E}(t, \mathbf{x}) = -\nabla \Phi(t, \mathbf{x}) - \frac{\partial \mathbf{A}(t, \mathbf{x})}{\partial t}$$

> For dynamic fields, $\Phi(t, \mathbf{x})$ is *not* the electric potential. It is the **scalar potential**, chosen for mathematical convenience, which, along with a matching $\mathbf{A}(t, \mathbf{x})$, generates $\mathbf{E}(t, \mathbf{x})$.

In fact, the electric potential *does not exist* (in general) in the time varying case, since the electric field has a non-zero curl. This means that the energy ΔE required to move a charge from a point $\mathbf{x}_1$ to $\mathbf{x}_2$ *depends on the path*, and hence there cannot exist any function $V(\mathbf{x})$ such that $\Delta E = V(\mathbf{x}_2) - V(\mathbf{x}_1)$. Nonetheless, it is possible to define the *electrostatic* potential, which accounts for only the coulomb forces of source charges, but not the dynamics of varying **B**-fields.

Since both **E** and **B** are found from derivatives of **A** and Φ, **A** and Φ are not unique. **Gauge freedom** (or **gauge invariance**) is the ability to impose additional mathematical constraints on **A** and Φ, that don't change the physical system. [It is sometimes said that **E** and **B** are the "physical" fields, and **A** is not; this point is debatable, but the preceding definition of gauge invariance holds either way.]

Lorenz Family of Gauges

You may hear talk of "the" Lorenz gauge (often incorrectly referred to as "Lorentz" gauge [Jac p294]). **Lorenz gauges** are actually a family of gauges, which satisfy

$$\nabla \cdot \mathbf{A}(t, \mathbf{x}) + \frac{1}{c^2} \frac{\partial \Phi(t, \mathbf{x})}{\partial t} = 0 \qquad \text{Lorenz gauge} \qquad \text{[Jac 6.14 p240]}$$

$$or \qquad \partial_\mu A^\mu(t, \mathbf{x}) \equiv \sum_{\mu=0}^{3} \frac{\partial A^\mu}{\partial x^\mu} = 0 \qquad \text{Relativistic form} \qquad \text{[Jac 11.133 p555]}$$

[Aside: As a simple example of gauge invariance, look at the first equation above. We can generate another Lorenz gauge vector potential **A**'(t, **x**) by adding any static (i.e., time-independent) divergenceless vector field **G**(**x**) to a given Lorenz gauge field **A**. Since the curl of **G** must be 0 (to preserve the **B**-field), **G** can be written as the gradient of a scalar field:

$$\mathbf{G}(t, \mathbf{x}) = \nabla \Lambda(t, \mathbf{x}) .$$

Since we chose ∇·**G** = 0 (divergenceless), Λ is a harmonic function (it satisfies Laplace's equation):

$$\nabla \cdot \mathbf{G}(t, \mathbf{x}) = 0 \qquad \Rightarrow \qquad \nabla \cdot \nabla \Lambda(t, \mathbf{x}) = \nabla^2 \Lambda(t, \mathbf{x}) = 0 .$$

For boundary conditions at infinity, this implies that Λ(t, **x**) = constant. For restricted regions of space (and therefore local boundary conditions), Λ may be non-trivial.]

However, in general, a significant gauge transformation involves changing both **A** and Φ. We can generate another Lorenz-family gauge with any transformation of the following form:

$$Given \ that \qquad \left(\nabla^2 - \frac{1}{c^2} \frac{\partial^2}{\partial t^2} \right) \Lambda(t, \mathbf{x}) \equiv \Box^2 \Lambda(t, \mathbf{x}) = 0, \qquad then$$

$$\mathbf{A}'(t, \mathbf{x}) = A(t, \mathbf{x}) + \nabla \Lambda(t, \mathbf{x}) \qquad \Phi'(t, \mathbf{x}) = \Phi(t, \mathbf{x}) - \frac{\partial \Lambda(t, \mathbf{x})}{\partial t} \qquad \text{[Jac 6.19 p241]}$$

We can see that the new fields are also Lorenz gauge by simply plugging into the gauge definition above:

$$\nabla \cdot \mathbf{A}' + \frac{1}{c^2} \frac{\partial \Phi'}{\partial t} = \nabla \cdot (\mathbf{A} + \nabla \Lambda) + \frac{1}{c^2} \frac{\partial}{\partial t} \left(\Phi - \frac{\partial \Lambda}{\partial t} \right) = \nabla \cdot \mathbf{A} + \nabla^2 \Lambda + \frac{1}{c^2} \frac{\partial \Phi}{\partial t} - \frac{1}{c^2} \frac{\partial^2 \Lambda}{\partial t^2}$$

$$= \left(\nabla \cdot \mathbf{A} + \frac{1}{c^2} \frac{\partial \Phi}{\partial t} \right) + \left(\nabla^2 \Lambda - \frac{1}{c^2} \frac{\partial^2 \Lambda}{\partial t^2} \right) = 0$$

This is much easier to see relativistically, where the constraint on Λ is:

$$\partial_\mu \partial^\mu \Lambda(t, \mathbf{x}) = 0. \qquad Then$$

$$A'^\mu = A^\mu + \partial^\mu \Lambda \qquad and \qquad \partial_\mu A'^\mu = \partial_\mu A^\mu + \partial_\mu \partial^\mu \Lambda = 0$$

Lorenz gauges are Lorentz covariant: the 4-divergence $\partial_\mu A^\mu = 0$. The Lorenz gauge makes the wave equations nice and symmetric, and has a simple, covariant relativistic form [Jac p241].

Transverse, Radiation, or [gag] "Coulomb" Gauges

Another family of gauges is **radiation gauge**, aka **transverse gauge**, or (misleadingly) **Coulomb gauge**, defined by

$$\nabla \cdot \mathbf{A}(t, \mathbf{x}) = 0 \qquad \text{radiation (aka transverse) gauge} \qquad \text{[Jac 6.21 p241]},$$

which implies [Jac p241]:

$$\Rightarrow \qquad \nabla^2 \Phi(t, \mathbf{x}) = -4\pi \rho(t, \mathbf{x}) \qquad \text{(Gaussian units)} \qquad \text{[Jac 6.22 p241]} \,.$$

> The term "Coulomb" gauge can be confusing. It suggests that $\Phi(t, \mathbf{x})$ is the electric (so-called Coulomb) potential, which it is not. Further, the relation of $\Phi(t, \mathbf{x})$ to $\rho(t, \mathbf{x})$ looks the same as the *static* field, but that is not physically significant.

[Jac p242b] says that Coulomb gauge "indicates" that Coulomb forces propagate instantly. This is misleading, since Φ is a mathematical field, not a physical one. However, [Jac] is quite clear that all EM fields propagate at the *finite* speed of light, and all behavior is consistent with Special Relativity.

We can generate another radiation-family gauge by adding a static gradient of a harmonic function to **A**:

$$\mathbf{A}(t, \mathbf{x}) \rightarrow \mathbf{A}(t, \mathbf{x}) + \nabla \Lambda(\mathbf{x}) \qquad where \quad \nabla^2 \Lambda(\mathbf{x}) = 0 \text{ preserves radiation gauge}\,.$$

Since $\nabla \Lambda$ is irrotational, it preserves **B**. Since $\partial(\nabla \Lambda)/\partial t = 0$, it preserves **E**. [As before, for *all* of space, Λ being harmonic implies that $\Lambda(\mathbf{x}) = $ constant. For restricted regions of space, Λ may be non-trivial.]

The radiation gauge is often used when no significant sources are nearby, such as the radiation fields far from the source. Then $\Phi \rightarrow 0$ in the radiation gauge, and **A** satisfies the homogeneous (i.e. no-source) wave-equation:

$$\nabla^2 \mathbf{A}(t, \mathbf{x}) - \frac{1}{c^2}\frac{\partial^2 \mathbf{A}}{\partial t^2} = 0 \quad \text{[from Jac 6.30 p242]} \qquad \mathbf{E}(t, \mathbf{x}) = -\frac{\partial \mathbf{A}}{\partial t}, \quad \mathbf{B}(t, \mathbf{x}) = \nabla \times \mathbf{A} \quad \text{[Jac 6.31 p242]}\,.$$

With this additional condition that $\Phi = 0$, the **E**-field is given by $\mathbf{E} = -(\partial \mathbf{A}/\partial t)$. Note that:

> Far from sources, when $\Phi = 0$, the radiation gauge is also a Lorenz gauge.

This follows immediately from the definitions:

$$\underbrace{\nabla \bullet \mathbf{A}(t, \mathbf{x}) = 0}_{\textit{radiation gauge}}, \qquad \underbrace{\Phi(t, \mathbf{x}) = 0}_{\textit{far field}} \qquad \Rightarrow \qquad \underbrace{\nabla \bullet \mathbf{A}(t, \mathbf{x}) + \frac{1}{c^2}\frac{\partial \Phi(t, \mathbf{x})}{\partial t} = 0}_{\textit{Lorenz gauge}}\,.$$

Furthermore, the Lorenz gauge is Lorentz covariant, but neither of the radiation gauge conditions $\Phi = 0$, nor $\nabla \bullet \mathbf{A} = 0$ is covariant (i.e., they are frame dependent).

Even a static **E**-field can use a gauge in which $\Phi = 0$. This implies that **A** contains an irrotational part which grows unbounded (linearly) with time. (This part has no curl, and so contributes nothing to the **B** field.) Such an unbounded **A**-field, in this gauge, makes it hard to believe the **A**-field is a physical field [but see the Bohm-Aharonov effect in quantum mechanics].

In general, for dynamic fields, the scalar potential $\Phi(t, \mathbf{x})$ is *not* the electric potential. In fact, in general, no time-varying "electric potential" is possible (why not?? Note that such a field is usually not conservative). The radiation gauge is no exception. $\Phi(t, \mathbf{x})$ is a mathematical convenience, whose relation to $\rho(t, \mathbf{x})$ *looks* the same as a *static* field, but whose meaning is very different. An electric potential cannot be constructed, and is certainly *not* simply the integral of the static Coulomb potential over the charge distribution [ref ??].

Canonical Momentum Is Gauge Dependent

Gauge choices often help tremendously in revealing physics properties. Many calculations are greatly simplified by a good gauge choice, which makes it much easier to understand some phenomena. For example, in quantum field theory, there is one property that is very clear in one gauge, and very obscure in a 2nd gauge. But there is a 2nd property that is clear in the 2nd gauge, but obscure in the first. However, because we know the two results are related by a gauge transformation, we know that both properties must always hold. It is important to be clear, though, which quantities are gauge invariant, and which are gauge dependent. Canonical momentum is gauge dependent.

Recall that the EM Lagrangian for a charged particle, and its canonical momentum are (gaussian units):

$$L(\mathbf{q}, \dot{\mathbf{q}}) = T(\dot{\mathbf{q}}) - V(\mathbf{q}) + \frac{e}{c}\dot{\mathbf{q}} \cdot \mathbf{A}(\mathbf{q}) \quad or \quad L = T(\dot{q}_1, \dots \dot{q}_n) - V(q_1, \dots q_n) + \frac{e}{c}\sum_{j=1}^{n} \dot{q}_j A_j(q_1, \dots q_n)$$

where $\mathbf{q} \equiv (q_1, \dots q_n)$ are the set of generalized coordinates for the particle

$$\mathbf{p} \equiv \frac{\partial L}{\partial \dot{\mathbf{q}}} = m\dot{\mathbf{q}} + \frac{e}{c}\mathbf{A} \qquad\qquad or \qquad p_i = m\dot{q}_i + \frac{e}{c}A_i$$

where e is the charge of a partical.

Because of gauge freedom in choosing $\mathbf{A}(\mathbf{q})$, the canonical momentum is gauge dependent, and even which components of momentum are conserved is gauge dependent. Consider a positive charge moving in a uniform B-field directed out of the page. Here are 2 different $\mathbf{A}$ fields that describe this situation:
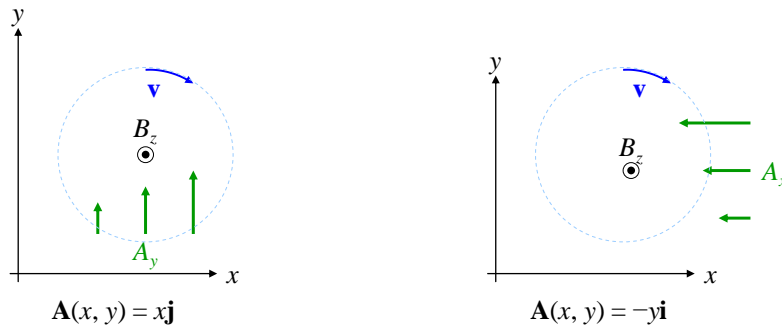

Figure 6.4   Two different gauges for $\mathbf{A}$ that describe the same physical situation. On the left, p_y is conserved and p_x is not. On the right, p_x is conserved and p_y is not.

Recall from the Euler-Lagrange equations of motion for any Lagrangian:

$$\frac{d}{dt}\frac{\partial L(q_i, \dot{q}_i, t)}{\partial \dot{q}_j} = \frac{\partial L}{\partial q_j} \qquad \Rightarrow \qquad If \;\; \frac{\partial L}{\partial q_j} = 0, \;\; then \qquad \frac{\partial L}{\partial q_j} = 0 = \frac{d}{dt}\frac{\partial L}{\partial \dot{q}_j} \equiv \frac{d}{dt}p_j.$$

In the situation above, the particle motion is a circle. In the left case,

$$\mathbf{A}(x, y) = x\mathbf{j}, \qquad L(x, y) = \frac{1}{2}mv^2 + \mathbf{v} \cdot \mathbf{A} = \frac{1}{2}m\left(\dot{x}^2 + \dot{y}^2\right) + \dot{y}x$$

$$\frac{\partial L(x, y)}{\partial y} = 0 \qquad \Rightarrow \qquad p_y = const = \frac{\partial L}{\partial \dot{y}} = m\dot{y} + \frac{e}{c}A_y = m\dot{y} + \frac{e}{c}x$$

and p_x is *not* conserved. You can see this on the left and right edges of the circle of motion. On the left side, $\mathbf{v}$ and $\mathbf{A}$ point up, and add. On the right, $\mathbf{v}$ opposes $\mathbf{A}$, and they subtract. The sum in both cases is the same, i.e. p_y is conserved. But in the right case,

$$\mathbf{A}(x, y) = -y\mathbf{i}, \qquad L(x, y) = \frac{1}{2}m\left(\dot{x}^2 + \dot{y}^2\right) - \dot{x}y$$

$$\frac{\partial L}{\partial x} = 0 \quad \Rightarrow \qquad p_x = const = \frac{\partial L}{\partial \dot{x}} = m\dot{x} + \frac{e}{c}A_x = m\dot{x} - \frac{e}{c}y$$

and p_y is not conserved. Here, the top and bottom edges of the circle are examples of $\mathbf{v}$ and $\mathbf{A}$ combining to the same p_x value at both points.

Note that any rotation of the $\mathbf{A}$ field would produce exactly the same physics, and would be just another gauge change. If we made the $\mathbf{A}$ field point at 45°, then the linear combination $p_x + p_y$ would be conserved, but neither one separately would be conserved.

Reflection Symmetry

When considering magnetic systems, one might encounter reflection symmetry, which is essentially a parity transformation. In such cases, we must distinguish between "polar" vectors and "axial" vectors. In short, **polar vectors** flip direction under a parity transformation, and **axial vectors** do not. So position, momentum, and E-fields, for example, are polar vectors. Angular momentum and B-fields are axial vectors (some axial vectors point along an axis of rotation, e.g. angular momentum, hence the name "axial"). This means that when you consider "reflection" symmetry (i.e., parity transformation) in E&M, you have to flip the E-field, but *not* the B-field. This is a little weird. You have to be careful with reflection symmetry and magnetic fields.

Bandwidth

[Taken from my original "Simple English Wikipedia: Bandwidth" page. As such, this section, and *only* this section, is public domain.] Bandwidth is a measure of how much frequency space of a spectrum is used. To clarify this, we must define some terms. Many systems work by means of vibrations, or **oscillations**. Vibrations are something that goes back and forth at a regular rate. Each complete cycle of "back and forth" is called, simply enough, a **cycle**. The number of cycles per second of a system is its **frequency**.

Frequency is measured in cycles per second, usually called "Hertz", and abbreviated "Hz".

However, most systems don't operate at just a single frequency. They operate at many different frequencies. For example, sound is vibrations. Therefore, it has at least one frequency, and usually many different frequencies. People can hear sound frequencies as low as about 20 Hz, and as high as about 20,000 Hz. A **band** of frequencies is a continuous range of frequencies; in this example, the band of frequencies people can hear is from 20 Hz to 20,000 Hz.

Finally, **bandwidth** is how wide the frequency band is, that is, the highest frequency minus the lowest frequency. In the hearing example, the bandwidth of a person's ears is about 20,000 Hz - 20 Hz = 19,980 Hz.

Bandwidth is often applied to the electromagnetic spectrum: radio waves, light waves, X-rays, and so on. Radio waves are oscillations of electric and magnetic fields. For example, the lowest United States AM radio channel covers the band of frequencies from 535,000 Hz to 545,000 Hz. It therefore has a bandwidth of 10,000 Hz (10 kHz). All United States AM radio stations have a bandwidth of 10,000 Hz. The lowest United States FM radio channel covers the band from 88,000,000 Hz (88 MHz) to 88,200,000 Hz (88.2 MHz). It therefore has a bandwidth of 200,000 Hz (200 kHz).

The term "bandwidth" has been misappropriated into the field of digital data communication. It is often incorrectly used to mean "data carrying capacity". However, there is no such thing as "digital bandwidth". The proper term for the data carrying capacity of a communication channel is **channel capacity**.

It is true that, in general, the channel capacity of a system increases with the bandwidth used for communication. However, many other factors come into play. As a result, in many (if not most) real systems, the channel capacity is not easily related to the channel bandwidth.

Is an Electron's Mass Due to Its E-field?

No.

There is much speculation about whether the mass of fundamental particles comes from the energy of their inherent fields, but it is fairly easy to see that that is not plausible.

Consider the hypothesis that an electron's mass is entirely due to the energy of its electrostatic field, through $m = $ (energy)$/c^2$. It is reasonable to also assume the positron must be the same. When we bring an electron and a positron close together, their fields cancel to some extent. The closer the particles, the more cancellation, and we can achieve any degree of cancellation we choose. Therefore, the mass of an electron-positron system should be less than 2x511 keV, and for some collisions, it would be *much* less.

Electron-positron systems have been widely studied in particle physics experiments for many decades. All such experiments show that *e-p* systems always have rest-mass of 2x511 keV. Therefore, we conclude that experiments rule out the hypothesis.

In fact, in any system of particles, the energy of any fields around charged particles must be *added* to the rest mass to get the total mass-energy of the system. The field energy is typically quite small. For an isolated electron, we can calculate the so-called "classical radius:" the radius a spherically symmetric charge of an electron would be to have an E-field energy equal to its mass; it is about 3×10^{-15} m.

The energy in the field of a spherical shell of uniform charge density $\propto 1/r$:

$$\mathcal{E} \propto \int_{r_{classical}}^{\infty} |\mathbf{E}|^2 \, d\tau \propto \int_{r_{classical}}^{\infty} \underbrace{\left(\frac{1}{r^2}\right)^2}_{|\mathbf{E}|} 4\pi r^2 \, dr \propto \int_{r_{classical}}^{\infty} r^{-2} \, dr = \frac{1}{r_{classical}} \ .$$

A typical atomic electron is spread over an orbit of about 10^{-10} m, more than 4 orders of magnitude larger than its hypothetical "classical radius." Its (uncancelled) field energy would then be more than 4 orders of magnitude *smaller than* its rest mass. By itself, this would be readily measurable with modern methods, but the field of a neutral atom is almost entirely cancelled by the nucleus; the resulting field energy is negligible. The electron remains fully massive, however, at 511 keV. Despite all these field configuation considerations, the quantum mechanical calculation of the size of a hydrogen atom gives the correct, measured result only when supplied with the well-known rest-mass.

Furthermore, in Quantum Electrodynamics (QED), one can compute the gyromagnetic ratio (essentially, the magnetic field) of an electron to 14 digits, and it agrees with experiment to the last digit. This calculation also uses the standard rest-mass of the electron. QED is the most accurately verified theory known to humans, and it uses the well-known rest-mass in all its calculations, regardless of field configurations.

Note that there is no evidence that an electron has any property related to the hypothetical "classical radius." About the most we can say is that to confine an electron's wave-function to (half) this size, you need an energy equivalent to (twice) its mass, and you may well generate an electron-positron pair in the process. The created pair's rest-mass steals the energy from the original electron, and so you can't really squeeze an electron much smaller than (half) its classical radius.

Still more: A proton is roughly 2000 times heavier than an electron. If mass were due to a charge's electric field, a proton would be about 2000 times *smaller* than an electron. Scattering experiments clearly show a proton size on the order of 1 fm, whereas electrons have no discernable fundamental size at all. (Their size depends entirely on the spatial extent of their wave-functions.)

Furthermore, what would be the origin of the mass of a neutron? Or a neutral pion? We might guess quarks, but neutrons and neutral pions are like "atoms" of quarks, and the energy of any supposed internal EM fields would be many orders of magnitude less than the particles' known masses.

All experimental evidence is consistent with the notion that a fundamental particle's mass is a constant of nature. In the standard model, all rest masses are due to interactions with the Higgs boson. The strength of the interaction (a constant of nature) determines the mass. The standard model is well supported by a huge base of varied experiments, which further rule out the "electric field as mass" hypothesis. In my opinion, the *origin* of mass is less well established than most of the standard model, but the data ruling out the hypothesis still stand.

Skin Depth

It may seem odd that the skin-depth formula:

$$\delta = \sqrt{\frac{2}{\omega \sigma \mu}}$$

contains μ, but not ε. The reason is that a conductor ($\sigma > 0$) has, at low frequency, infinite permittivity: $\varepsilon \to \infty$. When ω is high enough that the finite conductivity prevents the conductor from fully expelling the E-field, then the relevant electrical parameter is σ. So there *are* both an electrical and a magnetic parameter in the skin depth; it's just that the relevant electrical parameter for a conductor is σ, not ε.

Future Funky Electromagnetic Topics

Let me know what topics you'd like to see.

It seems that in the universe, there are fundamental electric monopoles (charges), and fundamental magnetic dipoles (intrinsic dipoles of charged fermions). This mean we should include a source to the sourced (but otherwise vacuum) Maxwell's Equations, $\boldsymbol{\rho}_M$, for magnetic dipole moment density (a vector field), or $\mathbf{m}_q$ for a point-charge dipole moment. In practice, though, for fluids (e.g., gasses), the magnetic dipole moments (being vectors) average to near zero. In contrast, the charges sum. To get the magnetic moments to be significant, they must align, which typically requires either a solid material, or an outside field to influence the fluid.

1. Do W^+/W^- have magnetic moments? Yes [ref??]

2. Force between two capacitor plates is ½ what you might think from σAE; must look at it on a mesoscopic scale (bigger than atoms, smaller than capacitor). Also, a charging capacitor *always* dissipates the same power in its resistor as is left on the capacitor. This means the idealized approximation of zero-loss wires is impossible.

3. Hidden momentum.

7 Appendices

References

Arfken, George B. and Hans J. Weber, *Essential Mathematical Methods for Physicists*, Elsevier Science, 2004. ISBN 0-12-059877-9.

[Feynmann]   Feynman, Leighton, Sands, *The Feynman Lectures on Physics*, Addison-Wesley, 1964, ISBN-13: 978-0465023820.

[Gri 2013]   David J. Griffiths, *Introduction to Electrodynamics*, 4th ed., Pearson, 2013. ISBN-13 978-0-321-85656-2.

Helliwell, *Special Relativity*, ??

[Jac 1999]   John David Jackson, *Classical electrodynamics*, 3rd ed., John Wiley & Sons, Inc., 1999. ISBN 0-471-30932-X.

[L&L]   Landau & Lifshitz, *The Classical Theory of Fields*, Butterworth-Heinemann, 1975 (reprinted with corrections 1987). ISBN 0 7506 2768 9.

Marion & Thornton, 4th ed.

[Unnik]   C.S. Unnikrishnan, G.T. Gillies, "Do leptons generate gravity? First laboratory constraints obtained from some G experiments and possibility of a new decisive constraint", Physics Letters A 288 (2001-09-24) p161–166.

Glossary

Definitions of common Electro-magnetic terms:

$<x>$ the average (sometimes called "expectation") value of 'x'

current   In a wire, current is flow rate of charge *passing a point*: $i \equiv \dfrac{dq}{dt}$. In contrast to voltage, which is *between two points*.

fact   A small piece of information backed by solid evidence (in hard science, usually repeatable evidence). If someone disputes a fact, it is still a fact. "If a thousand people say a foolish thing, it is still a foolish thing." See also "speculation," and "theory."

linear   scales with multiplication by a constant, and summing of two parts. In other words, it commutes with scalar multiplication, and distributes over addition. E.g., a linear function satifies $f(ka+b) = kf(a) + f(b)$.

sense   Every vector lies along an axis. The sense of a vector is simply which way the vector points along the axis. In some cases, we know the axis of a vector, but not its sense.

speculation   A guess, possibly hinted at by evidence, but not well supported. Every scientific fact and theory started as a speculation. See also "fact," and "theory."

theory   the highest level of scientific achievement: a quantitative, predictive, testable model which unifies and relates a body of facts. A theory becomes accepted science only after being supported by overwhelming evidence. A theory is not a speculation, e.g. Maxwell's electromagnetic theory. See also "fact," and "speculation."

voltage   Electrostatic potential energy difference per unit charge between two points. E.g., if V_{AB} = +2 V, it means point A has 2 joules per coulomb higher potential energy than point B. If I move 3 C of charge from B to A, I must do 6 J of work on them, which is stored as 6 J of potential energy. If I let the charges run free from A to B, the E-field does 6 J of work on them, so they acquire 6 J of energy (in this, kinetic energy).

WLOG or WOLOG   without loss of generality

Formulas

Basics

SI		Gaussian	
$\nabla \cdot \mathbf{D} = \rho$	$\mathbf{E} = \mathbf{D} - \mathbf{P} = \dfrac{\mathbf{D}}{\varepsilon}$	$\nabla \cdot \mathbf{D} = 4\pi\rho$	$\mathbf{E} = \mathbf{D} - 4\pi\mathbf{P} = \dfrac{\mathbf{D}}{\varepsilon}$
$\nabla \times \mathbf{E} = -\dfrac{\partial \mathbf{B}}{\partial t}$	$\mathbf{B} = \mathbf{H} + \mathbf{M} = \mu\mathbf{H}$	$\nabla \times \mathbf{E} = -\dfrac{1}{c}\dfrac{\partial \mathbf{B}}{\partial t}$	$\mathbf{B} = \mathbf{H} + 4\pi\mathbf{M} = \mu\mathbf{H}$
$\nabla \cdot \mathbf{B} = 0$	$\mathbf{F} = q\mathbf{v} \times \mathbf{B}$	$\nabla \cdot \mathbf{B} = 0$	$\mathbf{F} = \dfrac{q}{c}\mathbf{v} \times \mathbf{B}$
$\nabla \times \mathbf{H} = \mathbf{J} + \dfrac{\partial \mathbf{D}}{\partial t}$		$\nabla \times \mathbf{H} = \dfrac{4\pi}{c}\mathbf{J} + \dfrac{1}{c}\dfrac{\partial \mathbf{D}}{\partial t}$	

$\mathbf{A}(\mathbf{r}) = \dfrac{\mu_0}{4\pi}\displaystyle\int \dfrac{\mathbf{J}(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} d^3r'$ [Jac 5.32 p181] $A(\mathbf{r}) = \dfrac{1}{c}\displaystyle\int d\mathbf{r}' \dfrac{\mathbf{J}(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|}$ [LL 43.5 p111]

	SI	Gauss	SI $\to$ Gauss
(1)	$\phi(r) = q/4\pi\varepsilon_0 r \quad \to$	$\phi(r) = q/r \quad \Rightarrow$	$\varepsilon_0 \to 1/4\pi$.
(2)	$c^2\varepsilon_0\mu_0 = 1 \quad \to$	$\varepsilon_0\mu_0 = 1 \quad \Rightarrow$	$\mu_0 \to 4\pi/c^2$.
(3)	$\mathbf{F} = q\mathbf{v} \times \mathbf{B} \quad \to$	$\mathbf{F} = q\dfrac{\mathbf{v}}{c} \times \mathbf{B} \quad \Rightarrow$	$\mathbf{B} \to \mathbf{B}/c$.

Vector Stuff

$$\nabla\Phi = \mathbf{e}_r \frac{\partial\Phi}{\partial r} + \mathbf{e}_\phi \frac{1}{r\sin\theta}\frac{\partial\Phi}{\partial\phi} + \mathbf{e}_\theta \frac{1}{r}\frac{\partial\Phi}{\partial\theta} = \mathbf{e}_r \frac{\partial\Phi}{\partial r} + \mathbf{e}_\phi \frac{1}{r}\frac{\partial\Phi}{\partial\phi} + \mathbf{e}_z \frac{\partial\Phi}{\partial z}$$

$$\nabla\cdot\mathbf{A} = \frac{1}{\rho}\frac{\partial}{\partial\rho}\rho A_\rho + \frac{1}{\rho}\frac{\partial A_\phi}{\partial\phi} + \frac{\partial A_z}{\partial z}$$

$$= \frac{1}{r^2}\frac{\partial}{\partial r}r^2 A_r + \frac{1}{r\sin\theta}\frac{\partial}{\partial\theta}(\sin\theta A_\theta) + \frac{1}{r\sin\theta}\frac{\partial A_\phi}{\partial\phi}$$

$$\nabla\times\mathbf{A} = \mathbf{e}_x\left(\frac{\partial A_z}{\partial y} - \frac{\partial A_y}{\partial z}\right) + \mathbf{e}_y\left(\frac{\partial A_x}{\partial z} - \frac{\partial A_z}{\partial x}\right) + \mathbf{e}_z\left(\frac{\partial A_y}{\partial x} - \frac{\partial A_x}{\partial y}\right)$$

$$= \mathbf{e}_r\left(\frac{1}{r}\frac{\partial A_z}{\partial\phi} - \frac{\partial A_\phi}{\partial z}\right) + \mathbf{e}_\phi\left(\frac{\partial A_r}{\partial z} - \frac{\partial A_z}{\partial r}\right) + \mathbf{e}_z\frac{1}{r}\left(\frac{\partial}{\partial r}(rA_\phi) - \frac{\partial A_r}{\partial\phi}\right)$$

$$= \mathbf{e}_r\frac{1}{r\sin\theta}\left(\frac{\partial}{\partial\theta}(\sin\theta A_\phi) - \frac{\partial A_\theta}{\partial\phi}\right) + \mathbf{e}_\theta\left(\frac{1}{r\sin\theta}\frac{\partial A_r}{\partial\phi} - \frac{1}{r}\frac{\partial}{\partial r}rA_\phi\right) + \mathbf{e}_\phi\frac{1}{r}\left(\frac{\partial}{\partial r}(rA_\theta) - \frac{\partial A_r}{\partial\theta}\right)$$

$$\nabla^2 = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2} \qquad = \frac{1}{r}\frac{\partial}{\partial r}\left(r\frac{\partial}{\partial r}\right) + \frac{1}{r^2}\frac{\partial^2}{\partial\phi^2} + \frac{\partial^2}{\partial z^2}$$

$$= \frac{1}{r^2}\frac{\partial}{\partial r}\left(r^2\frac{\partial}{\partial r}\right) + \frac{1}{r^2\sin\theta}\frac{\partial}{\partial\theta}\left(\sin\theta\frac{\partial}{\partial\theta}\right) + \frac{1}{r^2\sin^2\theta}\frac{\partial^2}{\partial\phi^2}, \qquad NB: \frac{1}{r^2}\frac{\partial}{\partial r}\left(r^2\frac{\partial}{\partial r}\right) = \frac{1}{r}\frac{\partial^2}{\partial r^2}(r\bullet)$$

$$\nabla\times(\nabla\times\mathbf{a}) = \nabla(\nabla\cdot\mathbf{a}) - \nabla^2\mathbf{a}, \qquad \mathbf{a}\times(\mathbf{b}\times\mathbf{c}) = \mathbf{b}(\mathbf{a}\cdot\mathbf{c}) - \mathbf{c}(\mathbf{a}\cdot\mathbf{b})$$

Solutions to Laplace's Equations

2D rectangular : $\Phi(x, y) = Ax + By + D$

or $\Phi(x, y) = X(x)Y(y)$ $\nabla^2 \Phi = 0$ $\Rightarrow$ $\dfrac{1}{X(x)} X'' + \dfrac{1}{Y(y)} Y'' = 0$

$$\begin{cases} X(x) = A\sin\alpha x + B\cos\alpha x, & \text{far-field} \rightarrow \text{lowest } \alpha \\ Y(y) = C\sinh\alpha y + D\cosh(\alpha x) & \text{or} \qquad Y(y) = C\exp(\alpha y) + D\exp(-\alpha y) \end{cases}$$

or $$\begin{cases} X(x) = A\sinh\alpha y + B\cosh(\alpha x) & \text{or} \qquad X(x) = A\exp(\alpha y) + B\exp(-\alpha y) \\ Y(y) = C\sin\alpha x + D\cos\alpha x, & \text{far-field} \rightarrow \text{lowest } \alpha \end{cases}$$

2D polar : $\Phi(r, \phi) = B_0 \ln\dfrac{r}{A_0} + \displaystyle\sum_{\nu=1}^{\infty} \left[\left(A_\nu r^\nu + B_\nu r^{-\nu} \right)\left(C_\nu \sin\nu\phi + D_\nu \cos\nu\phi \right) \right]$

3D cylindrical : $\Phi(r, \phi, z) = R(r)Q(\phi)Z(z)$

$R(r) = CJ_m(kr) + DN_m(kr)$ $Q(\phi) = e^{im\phi} = A\sin m\phi + B\cos m\phi$

$Z(z) = C\sinh\alpha y + D\cosh(\alpha x)$ or $Z(z) = C\exp(\alpha y) + D\exp(-\alpha y)$

$r = 0 \Rightarrow D = 0, \quad \Phi(r = a, \phi) = 0 \quad \Rightarrow \quad k_{mn} = \dfrac{x_{mn}}{a}$

$f(r) = \displaystyle\sum_{n=1}^{\infty} A_{\nu_n} J_\nu\left(x_{\nu_n} \dfrac{r}{a} \right)$ $A_{\nu_n} = \dfrac{2}{a^2 J_{\nu+1}^2(x_{\nu_n})} \displaystyle\int_0^a dr\, r\, f(r) J_\nu\left(x_{\nu_n} \dfrac{r}{a} \right)$

3D rectangular : $\Phi(x, y, z) = Ax + By + Cz + D$

or $\Phi(x, y, z) = X(x)Y(y)Z(z)$ $\nabla^2 \Phi = 0$ $\Rightarrow$ $\dfrac{1}{X(x)} X'' + \dfrac{1}{Y(y)} Y'' + \dfrac{1}{Z(z)} Z'' = 0$

$X(x) = A\sin\alpha x + B\cos\alpha x$ $\Rightarrow$ $\dfrac{1}{X(x)} X'' = -\alpha^2$

$Y(y) = C\sin\beta x + D\cos\beta x$ $\Rightarrow$ $\dfrac{1}{Y(y)} Y'' = -\beta^2 \left.\right\}$ $\alpha^2 + \beta^2 = \gamma^2,$ far-field $\rightarrow$ lowest γ

$Z = \exp(\pm\gamma z)$ or $\sinh/\cosh(\gamma z) \Rightarrow$ $\dfrac{1}{Z(z)} Z'' = +\gamma^2$

$f(x) = \dfrac{A_0}{2} + A_m \cos(x) + B_m \sin(x)$ $\begin{Bmatrix} A_m \\ B_m \end{Bmatrix} = \dfrac{2}{a} \displaystyle\int_{-a/2}^{a/2} dx\, f(x) \begin{Bmatrix} \cos \\ \sin \end{Bmatrix}\left(\dfrac{2\pi mx}{a} \right)$

3D spherical : $\Phi(r,\theta,\phi) = R(r)Q(\phi)P(\theta) \Rightarrow$

$$\frac{U(r)}{r} = R(r) = Ar^l + Br^{-(l+1)}, \qquad Q(\phi) = e^{im\phi} \qquad U_l(r) = \sqrt{\frac{2l+1}{2}}P_l(r) \qquad P_0(x) = 1$$

$$P_1(x) = x \qquad P_2(x) = \frac{1}{2}\left(3x^2 - 1\right) \qquad P_3(x) = \frac{1}{2}\left(5x^3 - 3x\right) \qquad P_4(x) = \frac{1}{8}\left(35x^4 - 30x^2 + 3\right)$$

Axial symmetry: $\Phi(r,\theta) = \sum_{l=0}^{\infty}\left(A_l r^l + B_l r^{-(l+1)}\right)P_l(\cos\theta) \qquad A_l = \frac{2l+1}{2}\int_0^{\pi} V(\theta)\, P_l(\cos\theta)\sin\theta\, d\theta$

$$Y_{lm}(\theta,\phi) = \sqrt{\frac{2l+1}{4\pi}\frac{(l-m)!}{(l+m)!}}P_l^m(\cos\theta)e^{im\phi} \qquad\qquad Y_{l,-m} = (\theta,\phi) = (-1)^m Y_{lm}^*(\theta,\phi)$$

$$Y_{00} = \frac{1}{\sqrt{4\pi}} \qquad\qquad Y_{11} = -\sqrt{\frac{3}{8\pi}}\sin\theta\, e^{i\phi} \qquad\qquad Y_{10} = \sqrt{\frac{3}{4\pi}}\cos\theta$$

$$Y_{22} = \frac{1}{4}\sqrt{\frac{15}{2\pi}}\sin^2\theta\, e^{2i\phi} \qquad Y_{21} = -\sqrt{\frac{15}{8\pi}}\sin\theta\cos\theta\, e^{i\phi} \qquad Y_{20} = \sqrt{\frac{5}{4\pi}}\left(\frac{3}{2}\cos^2\theta - \frac{1}{2}\right)$$

$$g(\theta,\phi) = \sum_{l=0}^{\infty}\sum_{m=-l}^{l} A_{lm}Y_{lm}(\theta,\phi) \qquad\qquad A_{lm} = \int_{sphere} d\Omega\, Y_{lm}^*(\theta,\phi)g(\theta,\phi)$$

$$\Phi(r,\theta,\phi) = \sum_{l=0}^{\infty}\sum_{m=-l}^{l}\left(A_{lm}r^l + B_{lm}r^{-(l+1)}\right)Y_{lm}(\theta,\phi) \qquad\qquad \frac{1}{|\mathbf{x}-\mathbf{x}'|} = \sum_{l=0}^{\infty}\frac{r_<^l}{r_>^{l+1}}P_l(\cos\theta)$$

Relativity and Mechanics

$$d\tau = \frac{ds}{c} \qquad\qquad t' = \gamma\left(t - vx/c^2\right) \qquad\qquad x' = \gamma(x - vt) \qquad \text{[Bra 1.55-56 p 44]}$$

$$1/\gamma^2 + \beta^2 = 1 \qquad\qquad u^i = \frac{dx^i}{d\tau} = \gamma(c,\mathbf{v}) \qquad\qquad \frac{dt}{dt'} = \frac{1 + \dfrac{\mathbf{u'}\cdot\mathbf{v}}{c^2}}{\sqrt{1 - v^2/c^2}} = \gamma\left(1 + \frac{\mathbf{u'}\cdot\mathbf{v}}{c^2}\right)$$

$$\mathbf{x}' = L\mathbf{x} = \begin{bmatrix} \gamma & -\gamma\beta_1 & -\gamma\beta_2 & -\gamma\beta_3 \\[2mm] -\gamma\beta_1 & 1 + \dfrac{(\gamma-1)\beta_1^2}{\beta^2} & 1 + \dfrac{(\gamma-1)\beta_1\beta_2}{\beta^2} & 1 + \dfrac{(\gamma-1)\beta_1\beta_3}{\beta^2} \\[3mm] -\gamma\beta_2 & 1 + \dfrac{(\gamma-1)\beta_1\beta_2}{\beta^2} & 1 + \dfrac{(\gamma-1)\beta_2^2}{\beta^2} & 1 + \dfrac{(\gamma-1)\beta_2\beta_3}{\beta^2} \\[3mm] -\gamma\beta_3 & 1 + \dfrac{(\gamma-1)\beta_1\beta_3}{\beta^2} & 1 + \dfrac{(\gamma-1)\beta_2\beta_3}{\beta^2} & 1 + \dfrac{(\gamma-1)\beta_3^2}{\beta^2} \end{bmatrix}\mathbf{x}$$

$$v_{total} = \frac{v_1 + v_2}{1 + v_1 v_2/c^2} \qquad\qquad v_y = v'_y\frac{\sqrt{1 - V^2/c^2}}{1 + v'_x V/c^2} \qquad \text{[L\&L 5.1 p13]}$$

$$a_{\parallel} = \frac{\left(1 - \dfrac{v^2}{c^2}\right)^{3/2}}{\left(1 + \dfrac{\mathbf{u'}\cdot\mathbf{v}}{c^2}\right)^3}a'_{\parallel} \qquad\qquad a_{\perp} = \frac{1 - \dfrac{v^2}{c^2}}{\left(1 + \dfrac{\mathbf{u'}\cdot\mathbf{v}}{c^2}\right)^3}\left[a'_{\perp} + \frac{\mathbf{v}\times(\mathbf{a}\times\mathbf{u'})}{c^2}\right] \qquad \text{[Jac p569]}$$

$$E = \gamma mc^2 \qquad E^2 = \left(mc^2\right)^2 + p^2 c^2 \qquad p^i = \left(E/c, p^x, p^y, p^z\right) = \left(E/c, \mathbf{p}\right)$$

$$\mathbf{p} = \gamma m\mathbf{v} = E\mathbf{v}/c^2 \qquad\qquad \mathbf{P} = \frac{\partial L}{\partial \mathbf{v}} = \gamma m\mathbf{v} + \frac{q}{c}\mathbf{A} = \mathbf{p} + \frac{q}{c}\mathbf{A}$$

$$\omega' = \gamma\omega\left(1 - \beta\cos\theta\right) \qquad\qquad k^i = \left(\omega = ck^0, \mathbf{k}\right)$$

$$\frac{d\gamma}{d\mathbf{v}} = \gamma^3 \frac{\mathbf{v}}{c^2} \qquad\qquad \int \frac{dv}{\gamma^3} = \int \frac{dv}{\left(1 - v^2/c^2\right)^{3/2}} = \frac{v}{\sqrt{1 - v^2/c^2}} = \gamma v$$

$$\Omega_c = \frac{eB}{\gamma mc} = \frac{ecB}{E} \qquad\qquad \frac{1}{h\nu'} - \frac{1}{h\nu} = \frac{2}{mc^2}\sin^2\frac{\theta}{2}$$

$$L = -mc^2\sqrt{1 - v^2/c^2} - e\phi + \frac{e}{c}\mathbf{A}\bullet\mathbf{v} \quad \left[\text{Jac 12.12 p582}\right] \quad H = \frac{1}{2m}\left(\mathbf{P} - \frac{e}{c}\mathbf{A}\right)^2 + e\phi \quad \left[\text{LL 16.10 p49}\right]$$

$$E_\parallel' = E_\parallel \qquad \mathbf{E}_\perp' = \gamma\left(\mathbf{E}_\perp + \boldsymbol{\beta}\times\mathbf{B}_\perp\right) \qquad\qquad B_\parallel' = B_\parallel \qquad \mathbf{B}_\perp' = \gamma\left(\mathbf{B}_\perp - \boldsymbol{\beta}\times\mathbf{E}_\perp\right)$$

Combined: $\qquad \mathbf{E}' = \gamma\left(\mathbf{E} + \boldsymbol{\beta}\times\mathbf{B}\right) - (\gamma - 1)\hat{\boldsymbol{\beta}}\cdot\mathbf{E} \qquad \mathbf{B}' = \gamma\left(\mathbf{B} - \boldsymbol{\beta}\times\mathbf{E}\right) - (\gamma - 1)\hat{\boldsymbol{\beta}}\cdot\mathbf{B}$

$or:\quad \mathbf{E}' = \gamma\left(\mathbf{E} + \boldsymbol{\beta}\times\mathbf{B}\right) - \dfrac{\gamma^2}{\gamma + 1}\boldsymbol{\beta}\left(\boldsymbol{\beta}\cdot\mathbf{E}\right) \qquad \mathbf{B}' = \gamma\left(\mathbf{B} - \boldsymbol{\beta}\times\mathbf{E}\right) - \dfrac{\gamma^2}{\gamma + 1}\boldsymbol{\beta}\left(\boldsymbol{\beta}\cdot\mathbf{B}\right) \qquad \left[\text{Jac 11.149 p558}\right]$

$$\mathbf{v}_D = c\frac{\mathbf{E}\times\mathbf{B}}{B^2} \qquad \mathbf{v}_G = \frac{\Omega_c a^2}{2B^2}\left(\mathbf{B}\times\nabla_\perp B\right) \qquad \mathbf{v}_C \approx c\frac{\gamma m}{e}v_\parallel^2\frac{\mathbf{R}\times\mathbf{B}_0}{R^2 B_0^2} = \frac{v_\parallel^2}{\Omega_c R}\frac{\mathbf{R}\times\mathbf{B}_0}{RB_0} \quad \left[\text{Jac 12.57-58 p590}\right]$$

$$\mathbf{v}_G + \mathbf{v}_C = \frac{1}{\Omega_c R}\left(v_\parallel^2 + \frac{1}{2}v_\perp^2\right)\hat{\mathbf{R}}\times\hat{\mathbf{B}} = \frac{mc}{eBR}\left(v_\parallel^2 + \frac{1}{2}v_\perp^2\right)\hat{\mathbf{R}}\times\hat{\mathbf{B}}$$

$$\mu = \frac{mv_\perp}{2B} \qquad\qquad \mathbf{F} = \nabla\left(\boldsymbol{\mu}\cdot\mathbf{B}\right) \qquad \left[\text{Jac 5.69 p 189}\right]$$

Adiabatic invariant: $\qquad Ba^2, \; v_\perp^2/B, \; p_\perp^2/B, \; \gamma\mu, \; \mu = e\Omega_c a^2/2c$

$$\text{NonRel}: \frac{d}{dt}mv_\parallel = -\mu\nabla B - e\nabla\phi = -\mu\frac{dB}{ds} - e\frac{d\phi}{ds} = \frac{\partial}{\partial s}\left(\frac{m}{2}v_\parallel^2\right) \qquad\qquad \text{Rel}: \frac{d}{dt}\gamma mv_\parallel = v_\parallel\frac{\partial}{\partial s}\left(\gamma mv_\parallel\right)$$

$$\gamma v_\parallel\frac{\partial}{\partial s}\left(\gamma mv_\parallel\right) = \frac{\partial}{\partial s}\left(\frac{m}{2}\gamma^2 v_\parallel^2\right) \qquad \ddot{z} \approx -\frac{v_{\perp 0}}{2B_0}\frac{\partial B(z)}{\partial z} \quad \left[\text{Jac 12.75 p595}\right]$$

12/21/2021 17:12   Page 104 of 108

$$F_{ik} \equiv \frac{\partial A_k}{\partial x^i} - \frac{\partial A_i}{\partial x^k} = \begin{bmatrix} 0 & E_x & E_y & E_z \\ -E_x & 0 & -B_z & B_y \\ -E_y & B_z & 0 & -B_x \\ -E_z & -B_y & B_x & 0 \end{bmatrix} \qquad F^{ik} = \begin{bmatrix} 0 & -E_x & -E_y & -E_z \\ E_x & 0 & -B_z & B_y \\ E_y & B_z & 0 & -B_x \\ E_z & -B_y & B_x & 0 \end{bmatrix}$$

$$T^{ik} = \begin{bmatrix} w & S_x/c & S_y/c & S_z/c \\ S_x/c & -\sigma_{xx} & -\sigma_{xy} & -\sigma_{xz} \\ S_y/c & -\sigma_{yx} & -\sigma_{yy} & -\sigma_{yz} \\ S_z/c & -\sigma_{zx} & -\sigma_{zy} & -\sigma_{zz} \end{bmatrix} \qquad \begin{cases} w = \dfrac{E^2 + B^2}{8\pi} & \text{[LL 31.5 p81]} \\[2mm] -\sigma_{\alpha\beta} = \dfrac{-1}{4\pi}\left(E_\alpha E_\beta + B_\alpha B_\beta - \dfrac{1}{2}\delta_{\alpha\beta}\left(E^2 + B^2\right)\right) \end{cases}$$ **??**

$$\begin{aligned} d\mathbf{f} &= d\mathbf{A} \cdot \ddot{\mathbf{T}} \\ \mathbf{f} &\equiv force \\ \mathbf{A} &\equiv area \end{aligned} \qquad \begin{cases} -\sigma_{xx} = \dfrac{1}{8\pi}\left(E_y{}^2 + E_z{}^2 - E_x{}^2 + B_y{}^2 + B_z{}^2 - B_x{}^2\right) \\[2mm] -\sigma_{xy} = \dfrac{-1}{4\pi}\left(E_x E_y + B_z B_y\right) \qquad \text{[LL 33.3 p87]} \end{cases}$$

Faraday?

$$j^j = \left(\rho c, \mathbf{J}\right)^T \qquad \frac{\partial j^i}{\partial x^i} = \frac{\partial \rho}{\partial t} + \nabla \cdot \mathbf{J} = 0 \qquad \frac{4\pi}{c} j^j = \frac{\partial}{\partial x^i} F^{ij} \qquad p^i = \left(E/c, \mathbf{p}\right) \qquad A^i = \left(\Phi, \mathbf{A}\right)$$

4-force: $\quad f^i = \dfrac{1}{c} F_{ik} j^k = \dfrac{dp^i}{d\tau} = c\dfrac{d}{ds} mu^i = \gamma\dfrac{d}{dt}\left(\gamma mc, \gamma mv\right) = \dfrac{eF^{ik}u_k}{c}$

$$\frac{\partial F^{ij}}{\partial x^k} + \frac{\partial F^{ki}}{\partial x^j} + \frac{\partial F^{jk}}{\partial x^i} = 0 \qquad \text{Scalars}: \quad F_{ij}F^{ij} = 2\left(B^2 - E^2\right) \qquad \varepsilon^{iklm} F_{ik} F_{lm} = -4E \cdot B$$

Lagrangian density: $\qquad L_f = \dfrac{-1}{16\pi} F_{ij}F^{ij} = \dfrac{1}{8\pi}\left(B^2 - E^2\right)$

Fields: Sheet: $E_{up} = E_{down} = 2\pi\sigma$ $E_{conductor} = 4\pi\sigma$ $\mathbf{K}_b = -c\hat{\mathbf{n}} \times \mathbf{M} = c\mathbf{M} \times \hat{\mathbf{n}}$

Dipoles

$$\mathbf{E}(\mathbf{r}) = \frac{\mathbf{P}}{r^3}\left(2\cos\theta\,\hat{\mathbf{r}} + \sin\theta\,\hat{\boldsymbol{\theta}}\right) \qquad \text{[Gri 3.103 p153]} \qquad\qquad \mathbf{M} = \frac{I}{c} area \qquad \boldsymbol{\Gamma} = \mathbf{M} \times \mathbf{B}_e$$

$$\mathbf{B}(\mathbf{r}) = -\nabla\frac{\mathbf{M}\cdot\mathbf{r}}{r^3} = \frac{3\left(M\cdot\hat{r}\right)\hat{r} - M}{r^3} = \frac{\mathbf{M}}{cr^3}\left(2\cos\theta\,\hat{\mathbf{r}} + \sin\theta\,\hat{\boldsymbol{\theta}}\right) \qquad \text{[Jac 5.56 p186, Gri 3.103 p153]}$$

Energy and Work

$$\mathbf{P} = \chi_e\mathbf{E}, \qquad \mathbf{D} = \mathbf{E} + 4\pi\mathbf{P} = \left(1 + 4\pi\chi_e\right)\mathbf{E} = \varepsilon\mathbf{E} \qquad \mathbf{M} = \chi_m\mathbf{H}, \qquad \mathbf{B} = \mathbf{H} + 4\pi\mathbf{M} = \left(1 + 4\pi\chi_m\right)\mathbf{H} = \mu\mathbf{H}$$

$$W = \sum_n \phi_n \oiint_{s_n} \frac{d\mathbf{S}\cdot\nabla\phi}{8\pi} + \frac{1}{2}\int d^3r\,\rho\phi = \frac{1}{2}\sum_n \phi_n Q_n + \frac{1}{2}\int d^3r\,\rho\phi = \int_\infty d^3r\,\frac{E^2 + B^2}{8\pi} = \int_\infty d^3r\,\frac{|\nabla\phi|^2 + B^2}{8\pi}$$

$$W = \frac{1}{2}\sum_{i,j=1}^n C_{ij}\phi_i\phi_j = \frac{1}{2}CV^2 = \frac{1}{2}\frac{Q^2}{C} \qquad\qquad C = \frac{2W}{V^2} = \frac{Q^2}{2W} \qquad\qquad w = \frac{1}{2}\mathbf{J}\cdot\mathbf{A}\ \left(\text{units ??}\ \frac{1}{c}cf\ \frac{\mathbf{F}}{V} = \frac{\mathbf{J}\times\mathbf{B}}{c}\right)$$

$$\nabla^2 \frac{1}{|r|} = \delta(0) \qquad Q_{\alpha\beta} = \int_\infty d^3 r' \left(3x'_\alpha x'_\beta - \delta_{\alpha\beta} r'^2 \right) \rho(r') \qquad \left[\text{Jac 4.9 p146, LL 41.3 p105} \right]$$

$$\phi_\rho(r) = \frac{q}{r} + \frac{\hat{\mathbf{r}} \cdot \mathbf{d}}{r^2} + \frac{1}{2} \sum_{\alpha,\beta} Q_{\alpha\beta} \frac{x_\alpha x_\beta}{r^5} + ... \qquad \left[\text{Jac 4.10 p146} \right]$$

$$F = qE(0) + \nabla \left(\mathbf{P} \cdot \mathbf{E} \right)_{r=0} + \left[\nabla \left(\frac{1}{6} \sum_{\alpha,\beta} Q_{\alpha\beta} \frac{\partial E_\alpha}{\partial x_\beta} \right) \right]_{r=0} + ... \qquad \left[\text{Jac q4.5 p171} \right]$$

$$\phi_e(r) = \phi_e(0) + r \cdot \nabla \phi_e(0) + \frac{1}{2} \sum_{\alpha,\beta} x_\alpha x_\beta \frac{\partial^2}{\partial x_\alpha \partial x_\beta} \phi_e(0) + ... \qquad \left[\text{Jac 4.22} \right]$$

$$W_e = \int_\infty d^3 r \, \rho(r) \phi_e(r) = q\phi_e(0) + \mathbf{p} \cdot \nabla \phi_e(0) + \frac{1}{6} \sum_{\alpha,\beta} Q_{\alpha\beta} \frac{\partial^2}{\partial x_\alpha \partial x_\beta} \phi_e(0) + ...$$

Maxwell's Equations, Tensor

$$\nabla \cdot \mathbf{D} = 4\pi\rho \qquad \nabla \times \mathbf{E} = -\frac{1}{c} \frac{\partial \mathbf{B}}{\partial t} \qquad \nabla \cdot \mathbf{B} = 0 \qquad \nabla \times \mathbf{H} = \frac{4\pi}{c} \mathbf{J} + \frac{1}{c} \frac{\partial \mathbf{D}}{\partial t}$$

$$\oiint \mathbf{D} \cdot d\mathbf{a} = 4\pi Q \quad \oint \mathbf{E} \cdot dl = \frac{-1}{c} \iint \frac{\partial \mathbf{B}}{\partial t} \qquad \oiint \mathbf{B} \cdot d\mathbf{a} = 0 \qquad \oint \mathbf{H} \cdot dl = \iint \left(\frac{4\pi}{c} \mathbf{J} + \frac{1}{c} \frac{\partial \mathbf{D}}{\partial t} \right) \cdot d\mathbf{a} = \frac{4\pi}{c} I$$

$$\mathbf{E} = -\nabla\phi - \frac{1}{c} \frac{\partial \mathbf{A}}{\partial t} \qquad \mathbf{B} = \nabla \times \mathbf{A} = -\frac{q}{c} \frac{\mathbf{v} \times \hat{\mathbf{r}}}{r^2} = -\boldsymbol{\beta} \times \mathbf{E}, \, \beta \ll 1 \qquad d\mathbf{B} = \frac{I}{c} d\boldsymbol{\ell} \times \frac{\hat{\mathbf{r}}}{r^2} \quad \left[\text{LL 43.7} \right]$$

$$\nabla^2 \mathbf{A} = \frac{-4\pi}{c} \mathbf{J} \quad \left[\text{LL 43.4 p110} \right] \qquad A(\mathbf{r}) = \frac{1}{c} \int d\mathbf{r}' \frac{\mathbf{J}(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} \quad \left[\text{LL 43.5 p111} \right] \qquad \mathbf{S} = \frac{c}{4\pi} \mathbf{E} \times \mathbf{B}$$

For $B_0 \hat{\mathbf{z}}$: $\qquad \mathbf{A} = r \frac{B_0}{2} \hat{\boldsymbol{\varphi}} \quad or \quad \mathbf{A} = x B_0 \hat{\mathbf{y}} \quad or \quad \mathbf{A} = -y B_0 \hat{\mathbf{x}}$

$$\nabla \cdot \left(\nabla \times \mathbf{B} = \frac{4\pi}{c} \mathbf{J} + \frac{1}{c} \frac{\partial \mathbf{E}}{\partial t} \right) \to 0 = \frac{4\pi}{c} \left(\nabla \cdot \mathbf{J} + \frac{\partial \rho}{\partial t} \right) \qquad \mathbf{F} = e \left(\mathbf{E} + \frac{\mathbf{v}}{c} \times \mathbf{B} \right) = \frac{I}{c} \boldsymbol{\ell} \times \mathbf{B} \quad \frac{\mathbf{F}}{V} = \frac{\mathbf{J} \times \mathbf{B}}{c}$$

Radiation

$$S \equiv |\mathbf{S}| = \frac{1}{2} \sqrt{\frac{\varepsilon}{\mu}} E^2 \qquad \Rightarrow \qquad E = \sqrt{2 \sqrt{\frac{\mu}{\varepsilon}} S} \qquad \left[\text{SI, Jac 7.13 p298} \right]$$

$$\mathbf{S} = \frac{c}{4\pi} \mathbf{E} \times \mathbf{H} = \frac{c}{4\pi} E^2 \mathbf{n} \quad \Rightarrow \qquad E = \sqrt{\frac{4\pi}{c} S} \qquad \left[\text{gaussian, L\&L p120} \right]$$

$$\frac{d\langle P\rangle}{d\Omega} = \frac{d\sigma}{d\Omega}\langle I\rangle = \frac{d\sigma}{d\Omega}\frac{c}{8\pi}E^2 \qquad P = \frac{2}{3}\frac{e^2a^2}{c^3} = \frac{2}{3}\frac{\ddot{p}^2}{c^3} \qquad \text{dielectric: } \frac{d\sigma}{d\Omega} = k^4a^6\left(\frac{\varepsilon-1}{\varepsilon+2}\right)^2\left|\hat{\mathbf{E}}\times\hat{\mathbf{r}}\right|^2$$

Thompson scattering: $\qquad \dfrac{d\langle P\rangle}{d\Omega} = \dfrac{c}{8\pi}\left(\dfrac{\omega}{c}\right)^4 P^2\sin^2\theta \qquad \dfrac{d\sigma}{d\Omega} = \dfrac{e^4}{m^2c^4}\left|\hat{\mathbf{E}}\times\hat{\mathbf{r}}\right|$

$$\mathbf{E}(t) = \left[\frac{e(1-\beta^2)(\hat{\mathbf{r}}-\boldsymbol{\beta})}{(1-\boldsymbol{\beta}\cdot\hat{\mathbf{r}})^3 R^2} + \frac{e}{c}\frac{\hat{\mathbf{r}}\times(\hat{\mathbf{r}}-\boldsymbol{\beta})\times\dot{\boldsymbol{\beta}}}{(1-\boldsymbol{\beta}\cdot\hat{\mathbf{r}})^3 R}\right]_{t'} \qquad \mathbf{B}(t) = \left[\hat{\mathbf{n}}\times\mathbf{E}\right]_{t'}$$

$$t = t' + \frac{|\mathbf{r}-\mathbf{r}_0(t')|}{c} \qquad\qquad \mathbf{R}(t') = \mathbf{r}-\mathbf{r}_0(t')$$

Non-Rel: $\quad \mathbf{E}\approx\left[\dfrac{e}{c}\dfrac{\hat{\mathbf{n}}\times(\hat{\mathbf{n}}-\boldsymbol{\beta})\times\dot{\boldsymbol{\beta}}}{R}\right]_{t'} \quad \dfrac{dP(t)}{d\Omega} = r^2\left(\approx R^2\right)\hat{\mathbf{n}}\cdot\mathbf{S} = \dfrac{c}{4\pi}\dfrac{e^2}{c^4}\left|\hat{\mathbf{n}}\times\dot{\mathbf{v}}\right|^2 = \dfrac{e^2}{4\pi c^3}\left|\dot{\mathbf{v}}\right|^2\sin^2\theta$ [Jac 14.21 p665]

$$P(t) = \int d\Omega\,\frac{dP}{d\Omega} = \frac{2e^2}{3c^3}\left|\dot{\mathbf{v}}\right|^2 \quad \text{[Jac 14.22 p665]}$$

Multipole: $\quad P(t) = \dfrac{2\left|\ddot{\mathbf{P}}\right|^2}{3c^3} + \dfrac{2\left|\dddot{\mathbf{m}}\right|^2}{3c^3} + \dfrac{1}{180c^5}\sum\limits_{\alpha,\beta}\left|\dddot{Q}_{\alpha\beta}\right|^2 \quad$ Relativistic: $P(t) = \dfrac{2e^2\gamma^6}{3c}\left[\dot{\beta}^2 - (\boldsymbol{\beta}\times\dot{\boldsymbol{\beta}})^2\right]$

$$\langle P\rangle = \frac{\omega^4}{3c^3}\left(\left|\mathbf{P}\right|^2 + \left|\mathbf{m}\right|^2\right) \qquad P_L(t) = \frac{2e^2\gamma^6}{3c}\dot{\beta}^2 \qquad P_T = \frac{2e^2}{3m^2c}\gamma^4\left|\dot{\boldsymbol{\beta}}\right|^2, \qquad \text{for fixed }\left|\frac{d\mathbf{p}}{dt}\right|: \quad P_T = \gamma^2 P_L$$

$$\frac{dP(t')}{d\Omega} = \frac{e^2}{4\pi c}\left[\frac{\left\{\hat{\mathbf{n}}\times\left[(\hat{\mathbf{n}}-\boldsymbol{\beta})\times\dot{\boldsymbol{\beta}}\right]\right\}^2}{(1-\boldsymbol{\beta}\cdot\hat{\mathbf{n}})^5}\right]_{t'} \qquad \sin\theta = \hat{\mathbf{v}}\times\hat{\mathbf{n}}$$

$$\frac{dP_L(t')}{d\Omega} = \frac{e^2\dot{v}^2}{4\pi c^3}\frac{\sin^2\theta}{(1-\beta\cos\theta)^5} \overset{\beta\approx1}{\approx} \frac{\theta^2}{1/2\gamma^2 + \theta^2/2} \overset{\theta\approx1/\gamma}{\sim} \gamma^8\,??$$

$$\frac{dP_T(t')}{d\Omega} = \frac{e^2}{2\pi c^3}\frac{\left|\dot{\mathbf{v}}\right|^2}{(1-\beta\cos\theta)^3}\left[1-\frac{\sin^2\theta\cos^2\phi}{\gamma^2(1-\beta\cos\theta)^2}\right], \;\; \phi = \hat{\mathbf{n}}\;to\;\dot{\boldsymbol{\beta}}\;in\;\dot{\boldsymbol{\beta}}-\hat{\mathbf{3}}\,??$$

Reflection/Refraction

$$\left|\mathbf{E}\right| = \left|\mathbf{B}\right| \quad \langle\mathbf{S}\rangle = \frac{c}{8\pi}\operatorname{Re}\mathbf{E}\times\mathbf{B}^* = \frac{c}{8\pi}\operatorname{Re}\mathbf{E}\cdot\mathbf{E}^* = \frac{c}{8\pi}\left|E\right|^2 \quad \langle\mathbf{g}\rangle = \frac{1}{8\pi c}\operatorname{Re}\mathbf{E}\times\mathbf{B}^* = \frac{\left|E\right|^2}{8\pi c}\hat{\mathbf{z}}$$

$\hat{x}\pm i\hat{y}\to(LHCirc+helicity)/(RHCirc-helicity) \qquad E_t, H_t\;cont;\;\;D_n, B_n\;cont$

$$E_\perp: \frac{E'}{E} = \frac{2n_1\cos\theta}{n_1\cos\theta + (\mu_1/\mu_2)\sqrt{n_2^2 - n_1^2\sin^2\theta}} \qquad \frac{E''}{E} = \frac{n_1\cos\theta - (\mu_1/\mu_2)\sqrt{n_2^2 - n_1^2\sin^2\theta}}{n_1\cos\theta + (\mu_1/\mu_2)\sqrt{n_2^2 - n_1^2\sin^2\theta}}$$

$$E_\parallel: \frac{E'}{E} = \frac{2n_1 n_2\cos\theta}{(\mu_1/\mu_2)n_2^2\cos\theta + n_1\sqrt{n_2^2 - n_1^2\sin^2\theta}} \qquad \frac{E''}{E} = \frac{(\mu_1/\mu_2)n_2^2\cos\theta - n_1\sqrt{n_2^2 - n_1^2\sin^2\theta}}{(\mu_1/\mu_2)n_2^2\cos\theta + n_1\sqrt{n_2^2 - n_1^2\sin^2\theta}}$$

$$R \equiv \frac{\langle S_y{}''\rangle}{\langle S_y\rangle} = \frac{E''^2}{E^2} \quad R_{normal} = \left(\frac{n_1 - n_2}{n_1 + n_2}\right)^2, \qquad R_\perp = \left(\frac{\sin(\theta' - \theta)}{\sin(\theta' + \theta)}\right)^2$$

$$R_\parallel = \left(\frac{\tan(\theta' - \theta)}{\tan(\theta' + \theta)}\right)^2, \qquad \theta + \theta' \to \frac{\pi}{2} \Rightarrow R_\parallel \to 0$$

$$n(\omega) = \sqrt{\varepsilon(\omega)\mu(\omega)} \qquad \varepsilon_0 = \varepsilon(\omega \ll 1/\tau) \approx 1 + 4\pi n e^2/m \quad \varepsilon(\omega \gg \gamma) = \varepsilon_0 - \frac{4\pi n f e^2}{m\omega^2},$$

$$\sigma = \frac{f n e^2}{m}\tau \qquad decay: \nabla^2 B = \frac{4\pi\sigma\mu}{c^2}\frac{\partial B}{\partial t}$$

$$\delta = \sqrt{\frac{c^2}{\omega\sigma\mu}} \qquad H(x) = H_0 \exp\left[-\frac{1-i}{\delta}x\right] \qquad \mathbf{E}(x) = \hat{\mathbf{z}}(i-1)\frac{c}{4\pi\sigma\delta}\mathbf{H}(x)$$

$$\frac{\langle Power\rangle}{area} = \frac{1}{2\sigma\delta}\left(\frac{c}{4\pi}H_0\right)^2 \qquad k_{eff} = \frac{c}{4\pi}H_0 \qquad \sigma_{Cu} = 10^{17}\ s^{-1}\ ??$$

Waveguides

$$\mathbf{TM}: E_z(x,y) = \sin\frac{n\pi x}{a}\sin\frac{m\pi y}{b} \quad \mathbf{TE}: B_z(x,y) = \cos\frac{n\pi x}{a}\cos\frac{m\pi y}{b} \quad \gamma^2 = \left(\frac{n\pi}{a}\right)^2 + \left(\frac{m\pi}{b}\right)^2$$

$$\omega_{\min} = \gamma_{\min}c \qquad \frac{\omega^2}{c^2} = k^2 + \gamma^2$$

$$v_{ph} = \frac{\omega}{k} = \sqrt{c^2 + \frac{\gamma^2 c^2}{k^2}} \qquad v_g = \frac{d\omega}{dk} = \frac{kc^2}{\sqrt{\gamma^2 c^2 + k^2 c^2}} \quad v_g v_{ph} = c^2$$

$$E_t = \nabla_t \phi e^{ikz - i\omega t} \qquad \nabla_t^2 \phi = 0 \qquad B_t = \hat{\mathbf{z}} \times \mathbf{E}_t$$

Spherical Image Charges

$$q_{out} = -q_{in}\frac{R}{d_{in}} \qquad d_{out} = \frac{R^2}{d_{in}} \qquad\qquad \phi = -\lambda \ln r^2\ ??$$

Index

The index is not yet developed, so go to the web page on the front cover, and text-search in this document.