Chapter 1

Transmission Lines

1.1 Introduction

In a typical electrical engineering curriculum, students learn circuit theory early in the program. Circuit theory is the application of Kirchoff's voltage and current laws together with component current/voltage relationships to find the currents and voltages in a circuit.

Circuit theory is an approximation to Maxwell's equations of electromagnetism. In a circuit, when the voltage between two nodes on a pair of wires is changed at one location, the voltage between any other pair of nodes on the same wires changes instantaneously. But in reality, the voltage change cannot propagate faster than the speed of light.

Transmission line theory can be viewed as a correction to circuit theory that is needed when wires or other conductors in a system are long enough that this propagation delay along the wires or conductors cannot be neglected. If a system operates at a given frequency, the frequency determines the wavelength of the waves propagating on the system. If the length of the interconnects in the system is small compared to the wavelength, we can usually use circuit theory to model the system. If the interconnects are longer than the wavelength, we often use transmission line theory. This is not a hard boundary, however, as stringent performance requirements sometimes require the use of transmission line theory or a full-wave analysis using Maxwell's equations even when components in the system are small compared to the wavelength.

After developing the basics of transmission line theory, we will turn to an even more powerful and accurate analytical framework based on Maxwell's equations. Transmission line theory is an approximation to the modal solutions that arise from Maxwell's equations. In turn, KVL and KCL are simplifications of two of Maxwell's equations, Ampere's law and Faraday's law. The logical relationships between these analytical frameworks is shown in Figure 1.1.

Appliations of transmission line theory include the design of high speed digital interconnects, very large scale integrated (VLSI) circuit design, microwave and radio frequency (RF) circuits, and any electrical system in which the rate of change of the signals in the system is fast in relation to the propagation time from one location in the system to another. Types of systems that use transmission line theory in the design process include:

- Smart devices, mobile phones, handheld electronic devices
- Cellular communication hardware

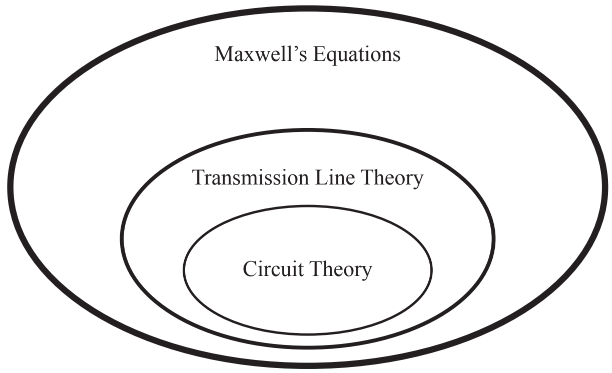


Figure 1.1: Logical relationship between Maxwell's equations, transmission line theory, and circuit theory.

- Radar systems
- Optical communication systems and optical fiber links
- Bluetooth, near field communication (NFC), and other short range wireless hardware
- Display technology
- Integrated circuits, processor chips, and digital links
- Computer motherboards
- Global positioning system receivers
- Satellite communication systems
- Satellite remote sensing
- Electronically scanned phased array antenna systems
- Radiowave propagation between the earth and ionosphere
- Biomedical systems

as well as many others. The range of technological applications that are impacted by transmission line theory is truly enormous.

1.2 Types of Transmission Lines

A transmission line guides energy from one point to another in such a way that the energy does not spread as it propagates. Transmission line examples include

- Coaxial cables (network, television)
- Twisted pair lines (telephone, network)
- Waveguides, optical fibers
- Printed circuit board trace; metalized line on an integrated circuit (microstrip and stripline)
- Power lines
- Earth/ionosphere system

The circuit diagram symbol for a transmission line is two wires with junctions marked as small circles:

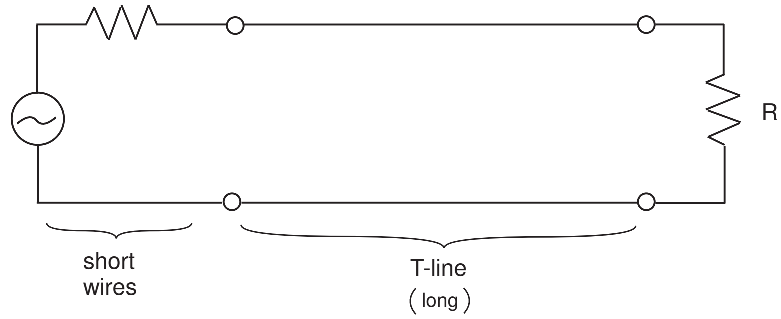


Figure 1.2: Circuit diagram symbol for a transmission line.

One important aspect of transmission line theory is that many different types of lines, including systems that do not appear to be transmission lines at all, such as optical coatings and even empty space, can be treated using the same basic set of equations. We will study the behavior of transmission lines for two types of driving sources: transients or pulses, and sinusoidal or time harmonic excitation.

1.3 Transmission Line Equations

How can we analyze the behavior of currents and voltages on a transmission line? We need a set of equations that govern the currents and voltages at different locations along the line. One way to arrive at these equations is to model the transmission line as a sequence of lumped circuit elements. The inductance and capacitance provide the propagation delay as energy moves along the transmission line, and the resistance represents losses. One can then take the limit as the length of the lumped-element section goes to zero, and arrive at a set of partial differential equations for the current and voltage on the transmission line.

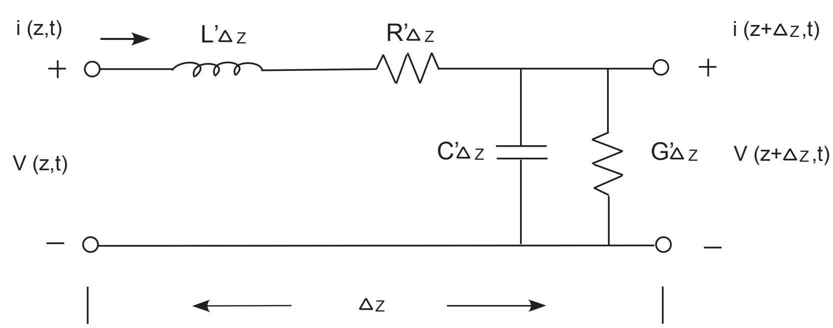


Figure 1.3: Lumped element model for a short section of a transmission line.

In the lumped element model, it is convenient to represent the capacitance, inductance, and resistance as per-unit-length quantities, so that

$$L' = \text{Inductance per unit length (H/m)} \tag{1.1}$$
$$C' = \text{Capacitance per unit length (F/m)} \tag{1.2}$$
$$R' = \text{Resistance per unit length } (\Omega/\text{m}) \tag{1.3}$$
$$G' = \text{Conductance per unit length (S/m)} \tag{1.4}$$

It is helpful at this point to understand these quantities in terms of a particular transmission line example. For a simple pair of parallel wires, $L'\Delta z$ represents energy stored in the magnetic field around the wires for a section of length Δz. $C'\Delta z$ represents the capacitance between the two pieces of wire. $R'\Delta z$ represents the series resistance of the wires, and $G'\Delta z$ represents parallel conductance of the dielectric material around the wires.

Lossless line. If the dielectric material between the conductors is a perfect insulator, then $G' = 0$. If the conductors making up the transmission line are perfect, then $R' = 0$. In this case, the line is said to be lossless. Real transmission lines are lossy, but in many cases the loss is small, so it is very common to approximate a transmission line as lossless.

After we develop the basic theory of transmission lines, we generally treat transmission lines at the undergraduate level as lossless. For a short microstrip transmission line on a PCB operating at 1 GHz, the reduction in signal power due to dielectric and conductive losses over a few centimeters might be a few tenths of a dB. For some applications, this is not important. For other applications, such as high sensitivity receiving antenna systems, minimizing loss is critical, and the transmission line loss must be accounted for in the design. For basic understanding and design of transmission line systems, however, we can analyze and design the system as if the transmission line is lossless, keeping in mind that signals will be attenuated as they propagate down the line.

1.3.1 Telegrapher Equations

Applying Kirchoff's voltage law (KVL) around the loop leads to

$$v(z,t) - i(z,t)R'\Delta z - L'\Delta z\frac{\partial i(z,t)}{\partial t} - v(z+\Delta z,t) = 0 \tag{1.5}$$

$$v(z,t) - v(z+\Delta z,t) = \Delta z\left[i(z,t)R' + L'\frac{\partial i(z,t)}{\partial t}\right] \tag{1.6}$$

$$-\left[\frac{v(z+\Delta z,t) - v(z,t)}{\Delta z}\right] = R'i(z,t) + L'\frac{\partial i(z,t)}{\partial t} \tag{1.7}$$

When we let $\Delta z \to 0$, the left hand side becomes the definition of a derivative, so that

$$-\frac{\partial v(z,t)}{\partial z} = R'i(z,t) + L'\frac{\partial i(z,t)}{\partial t} \tag{1.8}$$

Using Kirchoff's current law at the top left node, we obtain

$$i(z,t) - G'\Delta Z v(z+\Delta z,t) - C'\Delta z\frac{\partial v(z+\Delta z,t)}{\partial t} - i(z+\Delta z,t) = 0 \tag{1.9}$$

Again taking the limit as $\Delta z \to 0$, this equation becomes

$$-\frac{\partial i(z,t)}{\partial z} = G' v(z,t) + C' \frac{\partial v(z,t)}{\partial t} \tag{1.10}$$

Equations (1.8) and (1.10) are known as the Telegrapher equations.

1.3.2 Wave Equation

The Telegrapher equations are coupled first order partial differential equations. Since it is simpler to solve a single second order differential equation, we combine these two equations into a single equation. We first take the derivative of Eq. (1.8) with respect to z, to obtain

$$-\frac{\partial^2 v}{\partial z^2} = R' \frac{\partial i}{\partial z} + L' \frac{\partial^2 i}{\partial t \partial z} \tag{1.11}$$

We can substitute $\partial_t i$ from Eq. (1.10). We also need $\partial_{tz} i$, which we can get by differentiating Eq. (1.10) with respect to t:

$$-\frac{\partial^2 i}{\partial z \partial t} = G' \frac{\partial v}{\partial t} + C' \frac{\partial^2 v}{\partial t^2} \tag{1.12}$$

Substituting Eqs. (1.10) and (1.12) into Eq. (1.11) leads to

$$-\frac{\partial^2 v}{\partial z^2} = -R' \overbrace{\left[G' v + C' \frac{\partial v}{\partial t}\right]}^{-\frac{\partial i}{\partial z}} - L' \overbrace{\left[G' \frac{\partial v}{\partial t} + C' \frac{\partial^2 v}{\partial t^2}\right]}^{-\frac{\partial^2 i}{\partial t\,\partial z}} \tag{1.13}$$

$$\frac{\partial^2 v}{\partial z^2} = R'G' v + (R'C' + L'G') \frac{\partial v}{\partial t} + L'C' \frac{\partial^2 v}{\partial t^2} \tag{1.14}$$

This is a second order partial differential equation, where the only unknown is the voltage $v(z,t)$ on the transmission line. This is called the wave equation. A similar equation for $i(z,t)$ could be derived by eliminating the voltage instead, but once we know the voltage on the line, the current can be found using the transmission line equations.

1.3.3 Wave Solutions

How do we solve the wave equation (1.14)? Most of the time when we use differential equations in engineering, we look up or remember the general form of the solution and solve for the unknowns using initial or boundary conditions. For the wave equation, in the lossless case the general solution consists of two traveling waves of the form

$$v(z,t) = v^+(z - ut) + v^-(z + ut) \qquad \text{(Lossless line)} \tag{1.15}$$

The term $v^+(z-ut)$ represents a pulse or wave traveling to the right ($+z$ direction), and v^- represents a pulse traveling to the left ($-z$ direction). The functions v^+ and v^- depend on the excitation of the transmission line, and the constant u is determined by the coefficients of the wave equation.

Let's look at the first part of the general solution where the excitation produces a square pulse as given by

$$p(x) = \begin{cases} 1 & |x| < 1 \\ 0 & \text{otherwise} \end{cases} \tag{1.16}$$

We will set $v^+(z-ut) = p(z-ut)$. At time $t=0$, the pulse $v^+(z)$ is centered at $z=0$. At the time $t=t_o$ the pulse becomes

$$v^+(z-ut_o) = \begin{cases} 1 & |z-ut_o| < 1 \\ 0 & \text{otherwise} \end{cases} \tag{1.17}$$

The pulse is now centered at the position $z = ut_o$. The pulse has moved in the $+z$ direction. The resulting velocity is given by

$$\text{velocity} = \frac{\Delta z}{\Delta t} = \frac{ut_o}{t_o} = u \tag{1.18}$$

The other part of the general solution $v^-(z+ut)$ travels at the same velocity in the $-z$ direction.

1.3.4 Phase Velocity

To solve for the constant u, we plug either part of the general solution (1.15) into the wave equation. Using the chain rule for the derivative,

$$\frac{\partial^2}{\partial z^2} v^+(z-ut) = v^{+\prime\prime}(z-ut) \tag{1.19}$$

where the primes denote ordinary differentiation. Follow the same process to get the second derivative with respect to time leads to

$$\frac{\partial^2}{\partial t^2} v^+(z-ut) = v^{+\prime\prime}(z-ut)(-u)^2 \tag{1.20}$$

Substituting these two terms into the wave equation gives

$$v^{+\prime\prime}(z-ut) = L'C'\, u^2\, v^{+\prime\prime}(z-ut). \tag{1.21}$$

In order for this equality to hold, we must have that $u^2 LC = 1$, so that

$$u = \frac{1}{\sqrt{L'C'}} \qquad \text{(Phase velocity)} \tag{1.22}$$

This quantity is called the phase velocity of waves on the transmission line. Phase velocity is also represented by some authors with the symbol v_p.

For simple, air- or dielectric-filled transmission lines, the phase velocity is

$$\text{Coaxial Cable}: \quad \frac{1}{\sqrt{\mathrm{L'C'}}} = \left(\sqrt{\frac{2\pi}{\mu \ln(b/a)}}\right)\left(\sqrt{\frac{\ln(b/a)}{2\pi\epsilon}}\right) = \frac{1}{\sqrt{\mu\epsilon}} \tag{1.23}$$

$$\text{Two Wire}: \quad \frac{1}{\sqrt{\mathrm{L'C'}}} = \frac{1}{\sqrt{\mu\epsilon}} \tag{1.24}$$

$$\text{Parallel Plate}: \quad \frac{1}{\sqrt{\mathrm{L'C'}}} = \frac{1}{\sqrt{\mu\epsilon}} \tag{1.25}$$

where μ and ϵ are parameters of the material separating the conductors. For these transmission lines, the phase velocity only depends on the properties of the material around the transmission line, not the geometry.

For transmission lines which consist of a dielectric (insulator) and a pair of conductors, $\mu = \mu_o$, where μ_o is the permeability of free space ($\mu_o = 4\pi \times 10^{-7}$ H/m), and $\epsilon = \epsilon_r \epsilon_o$, where ϵ_o is the permittivity of free space ($\epsilon_o \simeq 8.854 \times 10^{-12}$ F/m) and ϵ_r is the relative permittivity of the dielectric. A typical value for the phase velocity is

$$u = \frac{1}{\sqrt{\mu\epsilon}} = (\frac{1}{\sqrt{\mu_o\ \epsilon_o}})(\frac{1}{\sqrt{\epsilon_r}}) = \frac{c}{\sqrt{\epsilon_r}} \approx \frac{2}{3}c \approx 2 \times 10^8 \text{m/s} \tag{1.26}$$

where $c \simeq 3 \times 10^8$ m/s is the speed of light in a vacuum.

One of the most common types of transmission lines is microstrip, which is a conductive trace on a printed circuit board over a dielectric substrate layer and a ground plane. The dielectric material can be glass fibers impregnated with epoxy as in the common FR4 PCB type, or for better performance, ceramic, teflon, or a composite material can be used. For microstrip, the electromagnetic fields associated with a signal on the trace are in both the air above the PCB and in the dielectric substrate, so the effective permittivity that governs the phase velocity $u = c/\sqrt{\epsilon_{\text{r,eff}}}$ is between the free space value and the permittivity of the dielectric ($\epsilon_r > \epsilon_{\text{r,eff}} > 1$).

If the conductive trace is sandwiched between two ground planes in a multilayer PCB, it is referred to as stripline. For microstrip and stripline, the width of the trace is usually designed so that the characteristic impedance of the transmission line defined in Sec. 1.3.5 is equal to the standard value of $50\,\Omega$.

1.3.5 Characteristic Impedance

The ratio between the current and voltage waves on a transmission line is called the characteristic impedance. We can find the characteristic impedance from the Telegrapher's equations.

For a lossless line, the first of the two Telegrapher's equations is

$$-\frac{\partial v}{\partial z} = L' \, \frac{\partial i}{\partial t} \tag{1.27}$$

Let us consider just the forward traveling wave,

$$-\frac{\partial}{\partial z} v^+(z-ut) = L' \, \frac{\partial}{\partial t} i^+(z-ut) \tag{1.28}$$

Using the chain rule this becomes

$$-v^{+\prime}(z-ut) = L' \, (-u) i^{+\prime}(z-ut) \tag{1.29}$$

Integrating both side with respect to $x = z - ut$ gives

$$\begin{aligned} -\int v^{+\prime}(x)\, dx &= -u \, L' \int i^{+\prime}(x)\, dx && (1.30) \\ -v^+ &= -uL' i^+ + \text{constant} && (1.31) \end{aligned}$$

Since $i^+ = 0$ if $v^+ = 0$, the constant is zero, and we have

$$\frac{v^+}{i^+} = u \, L' = \frac{L'}{\sqrt{L' \, C'}} = \sqrt{\frac{L'}{C'}} \tag{1.32}$$

The constant on the right-hand side has units of Ohms, and we call this the "characteristic impedance" of the line:

$$Z_o = \sqrt{\frac{L'}{C'}} \tag{1.33}$$

If we repeat this derivation for the reverse traveling wave, we get

$$\frac{v^-}{i^-} = -u \, L' = -\sqrt{\frac{L'}{C'}} = -Z_o \tag{1.34}$$

The total current can then be related to the total voltage using

$$\begin{aligned} v(z,t) &= v^+(z-ut) + v^-(z+ut) && (1.35) \\ i(z,t) &= \frac{v^+(z-ut)}{Z_o} - \frac{v^-(z+ut)}{Z_o} && (1.36) \end{aligned}$$

This allows us to find the current on a transmission line if we know the forward and reverse voltage waveforms.

The minus sign in the second equation is important. The characteristic impedance Z_o is not a simple resistance produced by the conductors in the line, because even a transmission line constructed from perfect conductors has a finite, nonzero characteristic impedance. An impedance is the ratio of total voltage to total

Figure 1.4: Relationship between the structural parameters of a transmission line, lumped element model parameters, and the characteristic impedance and phase velocity associated with propagating waves on the transmission line.

current, whereas the characteristic impedance is the ratio of the forward voltage waveform to the associated current and the negative of the ratio of the reverse voltage waveform to the associated current. If Eq. (1.36) had a positive sign instead of a negative sign, then Z_o would be a regular impedance. Can you understand the meaning of this minus sign physically?

The logical flow of transmission line modeling is illustrated in Fig. 1.4. The transmission line structural parameters (length, width, separation distance, dielectric substrate permittivity, etc.) determine the lumped element model parameters, which in turn govern the the characteristic impedance and phase velocity associated with propagating waves on the transmission line.

1.4 Sinusoidal Steady State

Electromagnetics applications can be divided into two broad classes:

- Time-domain: Excitation is not sinusoidal (pulsed, broadband, etc.)
  - Ultrawideband communications
  - Pulsed radar
  - Digital signals
- Time-harmonic: Excitation is sinusoidal or narrowband enough to be approximated as sinusoidal
  - Narrowband communication schemes - amplitude modulation (AM), frequency modulation (FM), phase shift keying (PSK), etc.
  - Continuous wave radar
  - Optical communications

Time-harmonic systems are fundamental to applications of electrical engineering. The concept of sharing a communication channel by using carrier sinusoids with different frequencies together with receivers tuned to discriminate among the carriers dates back to the earliest days of radio communications.

1.4.1 Phasor Notation

In analyzing time-harmonic systems, we assume that the signal of interest is narrowband enough that it can be approximated as a sinusoid. This approximation works very well for many important applications. Due to capacitance, inductance, and propagation delays in a system, the phase of a signal depends on where the signal is measured. For this reason, it takes two parameters to characterize the signal at any point in the system:

$$v(x,y,z,t) = \underbrace{v_o(x,y,z)}_{\text{Amplitude}} \cos\left[\omega t + \underbrace{\phi(x,y,z)}_{\text{Phase}}\right] \tag{1.37}$$

where ω is the time frequency of the signal in radians per meter. It is inconvenient to have the phase $\phi(x,y,z)$ inside the argument of the cosine function. Dealing with time-harmonic signals is much easier if we express these two degrees of freedom in a more symmetric way, as the real and imaginary parts of a complex number, which we call a phasor. The definition of a phasor voltage $\tilde{V}$ is

$$v(x,y,z,t) = \text{Re}\left\{\tilde{V}(x,y,z)e^{j\omega t}\right\} \tag{1.38}$$

At this point, we will drop the x and y dependence, and assume that for a transmission line the voltage only depends on time and the position z along the line.

How does the complex number $\tilde{V}$ relate to the real voltage V? By placing the complex number $\tilde{V}$ in polar form, we can express the voltage as

$$\begin{aligned} v(z,t) &= \text{Re}\left\{|\tilde{V}(z)|e^{j\angle\tilde{V}(z)}e^{j\omega t}\right\} && (1.39)\\ &= |\tilde{V}(z)|\text{Re}\left\{e^{j[\omega t+\angle\tilde{V}(z)]}\right\} && (1.40)\\ &= |\tilde{V}(z)|\cos\left[\omega t+\angle\tilde{V}(z)\right] && (1.41)\end{aligned}$$

By comparing this to Eq. (1.37), we can see that the magnitude of the voltage is equal to the magnitude of the phasor and the phase shift of the voltage relative to ωt is equal to the phase of $\tilde{V}$. Keep in mind that there is no such thing as a complex voltage. The real and imaginary parts of the phasor voltage $\tilde{V}$ simply offer a convenient tool for keeping track of the magnitude and phase in Eq. (1.37) at different locations in a circuit or system.

Another simplification that results from the use of phasor notation is that time derivatives become multiplication by $j\omega$, through the use of the identity

$$\frac{\partial v(z,t)}{\partial t} = \text{Re}\left\{j\omega\tilde{V}(z)e^{j\omega t}\right\} \tag{1.42}$$

The current-voltage relationship for a capacitor, for example, is

$$i(t) = C'\frac{dv(t)}{dt} \tag{1.43}$$

In the phasor domain, this becomes

$$\tilde{I} = j\omega C'\tilde{V} \tag{1.44}$$

The inverse relationship is

$$\tilde{V} = \tilde{I} \quad \underbrace{\frac{1}{j\omega C'}}_{\text{Impedance}} \tag{1.45}$$

The beauty of this result is that the capacitor current-voltage relationship now has the form of Ohm's law, but with an imaginary value in place of resistance. So, we can handle resistors, capacitors, and inductors without having to solve differential equations by using phasor notation.

1.4.2 Time-harmonic Wave Equation

By substituting Eqs. (1.38) and (1.42) into the wave equation (1.14), we obtain the time-harmonic wave equation,

$$\begin{aligned}\frac{d^2\tilde{V}}{dz^2} &= \left[R'G' + j\omega(R'C' + L'G') - \omega^2 L'C'\right]\tilde{V} &(1.46)\\ &= (R' + j\omega L')(G' + j\omega C')\tilde{V} &(1.47)\end{aligned}$$

The constant on the right-hand side is the square of the complex propagation constant, with the symbol γ, so that

$$\gamma^2 = (R' + j\omega L')(G' + j\omega C') \tag{1.48}$$

The solution to the ordinary differential equation (1.47) has the form

$$\tilde{V}(z) = Ae^{mz} + Be^{-mz} \tag{1.49}$$

Substituting this expression into Eq. (1.47) leads to

$$m^2\tilde{V} = \gamma^2\tilde{V} \tag{1.50}$$

so that the general solution can be expressed as

$$\tilde{V}(z) = Ae^{\gamma z} + Be^{-\gamma z} \tag{1.51}$$

The constants A and B are determined by the excitation and boundary conditions on the transmission line.

In general, γ is complex, and can be expressed in terms of its real and imaginary parts as

$$\gamma = \alpha + j\beta \tag{1.52}$$

Using this in the general solution leads to

$$\tilde{V}(z) = Ae^{\alpha z}e^{j\beta z} + Be^{-\alpha z}e^{-j\beta z} \tag{1.53}$$

If we use Eq. (1.38) to find the voltage on the transmission line, we obtain

$$\begin{aligned}v(z,t) &= \mathrm{Re}\left\{Ae^{\alpha z}e^{j\beta z}e^{j\omega t} + Be^{-\alpha z}e^{-j\beta z}e^{j\omega t}\right\} &(1.54)\\ &= \mathrm{Re}\left\{|A|e^{j\phi_A}e^{\alpha z}e^{j\beta z}e^{j\omega t} + |B|e^{j\phi_B}e^{-\alpha z}e^{-j\beta z}e^{j\omega t}\right\} &(1.55)\\ &= \underbrace{|A|e^{\alpha z}\cos\left[\omega t + \beta z + \phi_A\right]}_{\text{Reverse wave}} + \underbrace{|B|e^{-\alpha z}\cos\left[\omega t - \beta z + \phi_B\right]}_{\text{Forward wave}} &(1.56)\end{aligned}$$

By looking at this expression, we can understand the physical meaning of the real and imaginary parts of the complex propagation constant γ. The real part α represents attenuation and has units of Nepers per meter (Np/m). The imaginary part β determines the wavelength of the wave, and is called the wavenumber with units of radians per meter (rad/m). β is also called the spatial frequency or phase constant of the wave. The phase velocity of the wave is $u = \omega/\beta$ and the wavelength is $\lambda = 2\pi/\beta$ (Fig. 1.5).

From Eq. (1.56), we can see that the constant A represents the amplitude of the wave moving in the $-z$ direction, and B represents the wave moving in the $+z$ direction. Because of this, we rename the constants so that $V_o^+ = B$ and $V_o^- = A$, so that (1.51) becomes

$$\tilde{V}(z) = V_o^+e^{-\gamma z} + V_o^-e^{\gamma z} \tag{1.57}$$

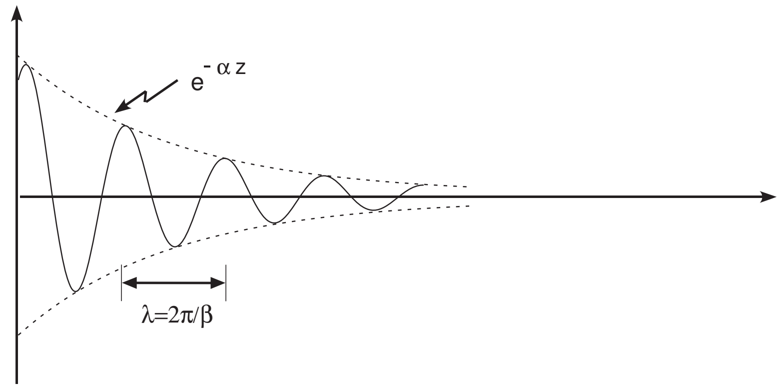


Figure 1.5: Propagating, attenuating forward wave.

1.4.3 Current Solution

To get the current on the line, we use one of the Telegrapher's equations in time-harmonic form:

$$-\frac{d\tilde{V}(z)}{dz} = (R + j\omega L)\tilde{I}(z) \tag{1.58}$$

Substituting the general solution (1.51) leads to

$$\begin{aligned}
\tilde{I}(z) &= -\frac{1}{R + j\omega L}\frac{d}{dz}\left[V_o^+ e^{-\gamma z} + V_o^- e^{\gamma z}\right] && (1.59)\\
&= -\frac{1}{R + j\omega L}\left[-\gamma V_o^+ e^{-\gamma z} + \gamma V_o^- e^{\gamma z}\right] && (1.60)\\
&= \frac{\gamma}{R + j\omega L}\left[V_o^+ e^{-\gamma z} - V_o^- e^{\gamma z}\right] && (1.61)\\
&= \frac{\sqrt{(R + j\omega L)(G + j\omega C)}}{R + j\omega L}\left[V_o^+ e^{-\gamma z} - V_o^- e^{\gamma z}\right] && (1.62)\\
&= \underbrace{\sqrt{\frac{G + j\omega C}{R + j\omega L}}}_{\frac{1}{Z_o}}\left[V_o^+ e^{-\gamma z} - V_o^- e^{\gamma z}\right] && (1.63)\\
&= \underbrace{\frac{V_o^+}{Z_o}}_{I_o^+} e^{-\gamma z} - \underbrace{\frac{V_o^-}{Z_o}}_{I_o^-} e^{\gamma z} && (1.64)
\end{aligned}$$

1.4.4 Lossless Transmission Lines

When analyzing and designing systems with transmission lines, it is common to neglect dielectric and conductive losses and treat the transmission line as if it were lossless. Therefore, transmission line theory for the lossless case is important as a first approximation. In engineering practice, the loss can be included in the analysis using a design software package.

For a lossless line,

$$\gamma = j\omega\sqrt{L'C'} = j\beta \qquad \text{(Lossless line)} \tag{1.65}$$

so that the attenuation constant α is zero, and there is no decay of the amplitude of a wave as it propagates. The general solution for the voltage simplifies to

$$\boxed{\tilde{V}(z) = V_o^+ e^{-j\beta z} + V_o^- e^{j\beta z} \qquad \text{(Voltage on a lossless line)}} \tag{1.66}$$

Because many transmission lines can be approximated as lossless, we often use this expression instead of (1.51). The current solution for a lossless line is

$$\boxed{\tilde{I}(z) = \frac{V_o^+}{Z_o} e^{-j\beta z} - \frac{V_o^-}{Z_o} e^{j\beta z} \qquad \text{(Current on a lossless line)}} \tag{1.67}$$

Equations (1.66) and (1.67) are the basic theoretical results from which arise all of the transmission line analysis techniques we will cover in this section. The first term with phase $e^{-j\beta z}$ represents a wave on the transmission line propagating in the $+z$ direction. The second term with phase $e^{j\beta z}$ propagates in the $-z$ direction. This can be seen by converting the voltage or current from the phasor domain to the time domain with (1.38) and then applying the analysis in Sec. 1.3.3.

1.4.5 Reflection Coefficient

At the load end of a transmission line (Fig. 1.6), we can use boundary conditions to find the ratio of the forward and reflected waves.

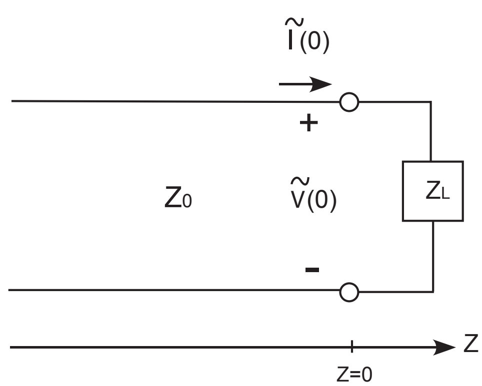


Figure 1.6: Transmission line and load.

For the sinusoidal steady state, it is convenient to shift the coordinate system so that load end is at $z = 0$. The goal is to find

$$\Gamma_L = \frac{V_o^-}{V_o^+} \tag{1.68}$$

The voltage and current boundary conditions at the load end, together with Ohm's law for the load impedance, lead to the following relationship between the total current and voltage at the load end of the transmission line:

$$\tilde{V}(0) = Z_L \tilde{I}_L(0) \tag{1.69}$$

Substituting the general voltage and current solutions leads to

$$V_o^+ + V_o^- = Z_L \left(\frac{V_o^+}{Z_o} - \frac{V_o^-}{Z_o} \right) \tag{1.70}$$

Now we can solve for the reflection coefficient:

$$\boxed{\Gamma_L = \frac{V_o^-}{V_o^+} = \frac{Z_L - Z_o}{Z_L + Z_o} \qquad \text{(Reflection coefficient)}} \tag{1.71}$$

Although this is the same expression as was obtained for the transient case (except that it is a phasor-domain formula and allows for complex load impedances), the meaning of the reflection coefficient is different. For the sinusoidal steady state, the forward and reverse waves exist everywhere on the transmission line. If we know V_o^+, for example, we can find V_o^- using Γ_L, and then we can use Eq. (1.57) or (1.66) to find the voltage anywhere on the transmission line.

Generalized reflection coefficient. In the lossless case, it is also sometimes useful to define a generalized reflection coefficient as the ratio of the forward and reverse waves at any point on the transmission line:

$$\Gamma(z) = \frac{V_o^- e^{j\beta z}}{V_o^+ e^{-j\beta z}} = \Gamma_L e^{j2\beta z} \tag{1.72}$$

The generalized reflection coefficient can be plotted in the complex plane and used as a graphical tool to understand the behavior of a transmission line system, the Smith chart covered in Sec. 1.6.

1.4.6 Standing Waves

When forward and reverse sinusoidal waves are both present on a transmission line, the two propagating waves add to form a standing wave pattern. If we apply a voltmeter to a transmission line instead of an oscilloscope, at high frequencies the voltmeter cannot respond rapidly enough to follow the $\cos \omega t$ time variation, so what we actually measure is the standing wave pattern on the line. Standing waves are also helpful into gaining insight into transmission line phenomena for different types of loads.

Using the load reflection coefficient, the phasor voltage on a transmission line can be written as

$$\tilde{V}(z) = V_o^+ e^{-j\beta z} + \Gamma_L V_o^+ e^{j\beta z} \tag{1.73}$$

In terms of the generalized reflection coefficient in Eq. (1.72), this becomes

$$\tilde{V}(z) = V_o^+ e^{-j\beta z} \left[1 + \Gamma(z)\right] \tag{1.74}$$

To analyze the standing wave pattern, we look at the magnitude of the phasor:

$$\begin{aligned} |\tilde{V}(z)| &= |V_o^+||1 + \Gamma(z)| \\ &= |V_o^+||1 + \Gamma_L e^{j2\beta z}| \\ &= |V_o^+||1 + \underbrace{|\Gamma_L| e^{j\theta_L}}_{\Gamma_L} e^{j2\beta z}| \end{aligned} \tag{1.75}$$

$$\begin{aligned} &= |V_o^+| \left[(1 + |\Gamma_L| e^{j\theta_L + j2\beta z})(1 + |\Gamma_L| e^{-j\theta_L - j2\beta z})\right]^{1/2} \\ &= |V_o^+| \left[1 + 2|\Gamma_L| \cos\left(2\beta z + \theta_L\right) + |\Gamma_L|^2\right]^{1/2} \end{aligned} \tag{1.76}$$

We can understand this function graphically by going back to Eq. (1.75). If we plot the phasor voltage in the complex plane, we find that its value is equal to $|V_o^+|$ on the real axis plus a complex number with magnitude $|\Gamma_L||V_o^+|$ and phase $2\beta z$, as shown in Fig. 1.7. The magnitude of the phasor is equal to the distance from the origin to the sum of the two terms.

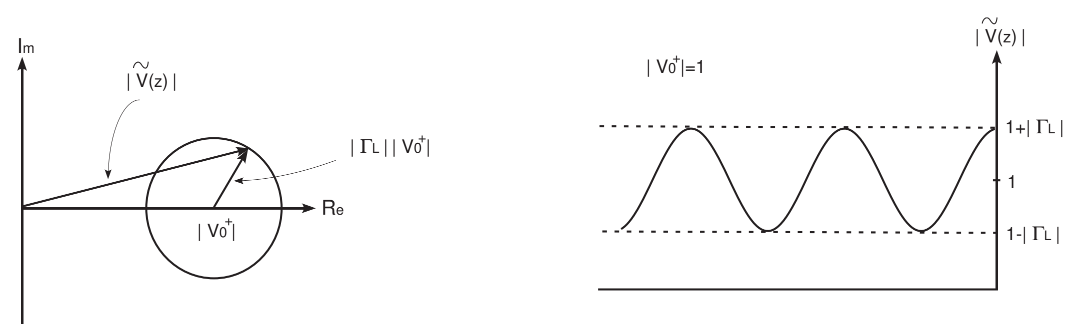


Figure 1.7: (a) Graphical representation of Eq. (1.75). (b) Corresponding standing wave pattern.

It is easy to see graphically that $|\tilde{V}(z)|$ is bounded by

$$|\tilde{V}(z)|_{\text{max}} = |V_o^+|(1 + |\Gamma_L|) \tag{1.77}$$

$$|\tilde{V}(z)|_{\text{min}} = |V_o^+|(1 - |\Gamma_L|) \tag{1.78}$$

The phasor travels around the circle each time z changes by π/β, which is equal to $\lambda/2$ or one half wavelength. Because the magnitude of the phasor voltage is the amplitude of the time-varying voltage in Eq. (1.37), $|\tilde{V}(z)|$ is the envelope of the voltage along the transmission line as it oscillates in time.

VSWR. The voltage standing wave ratio (VSWR) is defined to be

$$S = \frac{|\tilde{V}(z)|_{\text{max}}}{|\tilde{V}(z)|_{\text{min}}} = \frac{1+|\Gamma_L|}{1-|\Gamma_L|} \tag{1.79}$$

This quantity is useful because it is more easily measured on high frequency transmission lines than the time-varying voltage itself. For a matched load, $\Gamma_L = 0$, so that the VSWR is equal to one and there is no standing wave on the transmission line.

1.4.7 Load Examples

Matched load ($\Gamma_L = 0$): Only a forward wave exists on the transmission line, and there is no standing wave.

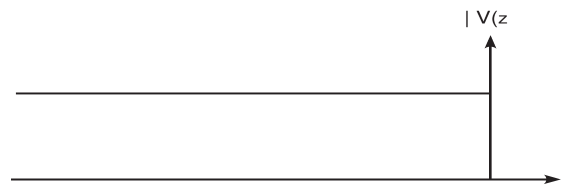


Figure 1.8: Standing wave pattern for a matched load.

Open circuit ($\Gamma_L = 1$): In this case, $|\tilde{V}(z)|_{\text{min}}$ is zero and the VSWR is infinite. The standing wave pattern exhibits nulls spaced one half wavelength apart along the transmission line with a maximum at the load.

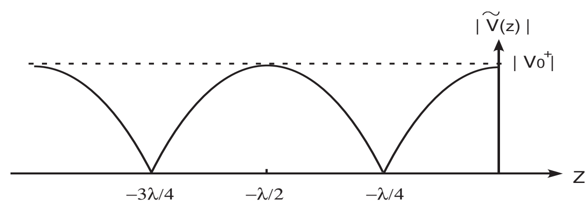


Figure 1.9: Standing wave pattern for an open circuit load.

Short circuit ($\Gamma_L = -1$): $|\tilde{V}(z)|_{\text{min}}$ is zero as with the open circuit load, and the VSWR is also infinite. In this case, however, there is a null instead of a maximum at the short circuit load.

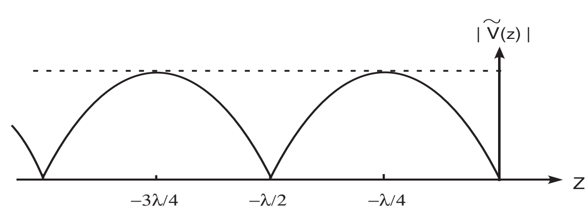


Figure 1.10: Standing wave pattern for a short circuit load.

1.4.8 Input Impedance

We need one more tool in order to analyze a complete sinusoidal steady state transmission system. Unlike the transient case, the impedance looking into the generator end depends on the entire transmission line and the load.

To understand impedance on a transmission line for a time-harmonic excitation, we can define a line impedance that is the ratio of the phasor voltage to the phasor current at a point on the line:

$$\begin{aligned} Z_{\text{in}}(z) &= \frac{\tilde{V}(z)}{\tilde{I}(z)} && (1.80)\\ &= \frac{V_o^+(1+\Gamma_L e^{j2\beta z})}{V_o^+(1-\Gamma_L e^{j2\beta z})/Z_o} && (1.81)\\ &= Z_o\frac{1+\Gamma(z)}{1-\Gamma(z)} && (1.82)\end{aligned}$$

If the generator is located at $z=-\ell$, then the input impedance seen by the source is $Z_{\text{in}} = Z_{\text{in}}(-\ell)$. By substituting Eq. (1.71) for Γ_L and applying trigonometric identities to Eq. (1.81), the input impedance can be placed in an alternate form

$$Z_{\text{in}} = Z_{\text{in}}(-\ell) = Z_o\frac{Z_L + jZ_o\tan\beta\ell}{Z_o + jZ_L\tan\beta\ell} \tag{1.83}$$

This now allows us to analyze a complete time-harmonic transmission line system (Fig. 1.11).

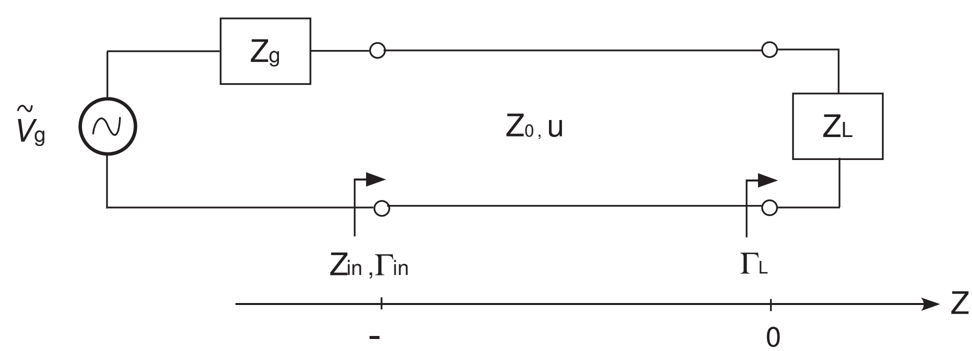


Figure 1.11: Transmission line system with sinusoidal excitation.

In the steady state, as noted above the impedance seen by the source is Z_{in} as given by Eq. (1.83), so the voltage on the transmission line at the input port can be found using a voltage divider:

$$\tilde{V}(-\ell) = \tilde{V}_g\frac{Z_{\text{in}}}{Z_g + Z_{\text{in}}} \tag{1.84}$$

We can then find V_o^+ using Eq. (1.73), so that

$$V_o^+ = \frac{\tilde{V}(-\ell)}{e^{j\beta\ell} + \Gamma_L e^{-j\beta\ell}} \tag{1.85}$$

Once V_o^+ is known, all currents and voltages anywhere on the transmission line can be determined.

Example: Sinusoidal Steady State Transmission Line System

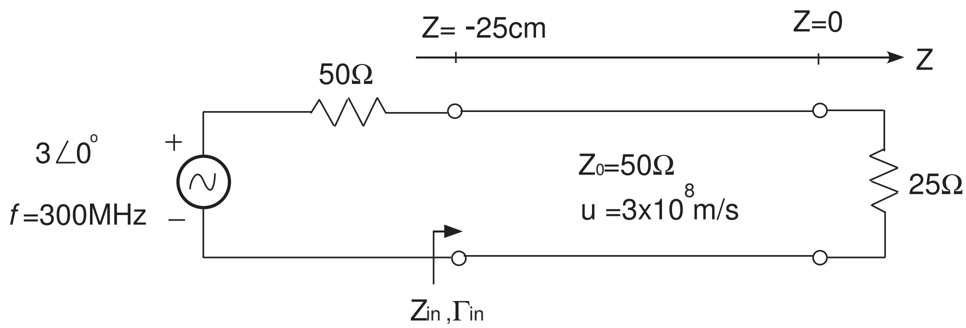


Figure 1.12: Transmission line with source and load.

1. Find Γ_L

2. Find $\Gamma_{\text{in}} = \Gamma(-\ell)$

3. Find Z_{in}

4. Find $\tilde{V}_{\text{in}}$

5. Find V_o^+

6. With these quantities, we can determine whatever else we may want to know about the system. Use the result to find the voltage across the load impedance and the time average power dissipated in the load.

Special Cases

Let's consider a few common loads and line lengths, and for each we will determine the input impedance looking into the generator end of the line using

$$Z_{\text{in}}(-\ell) = Z_o \frac{e^{-j\beta z} + \Gamma_L e^{j\beta z}}{e^{-j\beta z} - \Gamma_L e^{j\beta z}} \tag{1.86}$$

or one of the alternate forms of this expression that we derived previously.

Open circuit ($\Gamma_L = 1$):

$$\begin{aligned} Z_{\text{in}}^{\text{oc}}(-\ell) &= Z_o \frac{e^{j\beta\ell} + e^{-j\beta\ell}}{e^{j\beta\ell} - e^{-j\beta\ell}} \\ &= Z_o \frac{2\cos(\beta\ell)}{2j\sin(\beta\ell)} \\ &= -jZ_o \cot(\beta\ell) \end{aligned} \tag{1.87}$$

Short circuit ($\Gamma_L = -1$):

$$Z_{\text{in}}^{\text{sc}}(-\ell) = jZ_o \tan(\beta\ell) \tag{1.88}$$

Notice that in both the open and short circuit cases, the input impedance is purely imaginary, corresponding to a reactive load, and the lines appear inductive or capacitative. By changing the length ℓ, we can make the line look like a capacitor or inductor of any value. This principle is often used in microwave designs. (Is is possible to realize a reactive impedance corresponding to a very large inductance or capacitance?)

Half-integer line length ($\ell = n\lambda/2$, $n = 1, 2, 3, \ldots$)

$$\begin{aligned} Z_{\text{in}}(-n\lambda/2) &= Z_o \left. \frac{Z_L + jZ_o \tan\beta\ell}{Z_o + jZ_L \tan\beta\ell} \right|_{\tan(\beta n\lambda/2) = \tan(n\pi) = 0} \\ &= Z_L \end{aligned} \tag{1.89}$$

The load impedance repeats along the line each half wavelength.

Quarter-wave transformer ($\ell = n\lambda/2 + \lambda/4$, $n = 0, 1, 2, \ldots$) The input impedance looking into a quarter-wave section (or an integer number of half-wavelengths plus $\lambda/4$) as shown in Fig. 1.13 is

$$Z_{\text{in}}(n\lambda/2 + \lambda/4) = Z_o \left. \frac{Z_L + jZ_o \tan\beta\ell}{Z_o + jZ_L \tan\beta\ell} \right|_{\tan(\beta\lambda/4) = \tan(\pi/2) \to \infty} \tag{1.90}$$

$$= \frac{Z_o^2}{Z_L} \tag{1.91}$$

In order to see a matched load looking into the quarter-wave line, we can set

$$Z_{o1} = \frac{Z_{o2}^2}{Z_L} \quad \Rightarrow \quad Z_{o2} = \sqrt{Z_{o1} Z_L} \tag{1.92}$$

This is called quarter-wave matching.

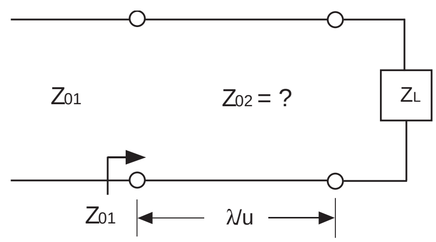


Figure 1.13: Quarter-wave matching transformer. The characteristic impedance of the quarter wave section of transmission line can be designed to match the load resistance Z_L to the impedance Z_{o1} of the input line.

1.5 Power

There are several ways to quantify power for sinusoidal steady state systems:

1. Instantaneous power: $p_i(t) = v(t)i(t)$.
2. Time-average power: $p_{\text{av}} = \frac{1}{T}\int_0^T p^i(t)\,dt$, $T = 2\pi/\omega$.
3. Complex power: $\tilde{P} = \tilde{V}\tilde{I}^*$

The time-average power is related to complex power by

$$p_{\text{av}} = \frac{1}{2}\text{Re}\{\tilde{V}\tilde{I}^*\} \tag{1.93}$$

The imaginary part of $\tilde{V}\tilde{I}^*$ is not associated with dissipated or supplied power, but rather represents changes in the amount of energy stored in inductive and capacitative elements. We will look at each of these for the case of a transmission line system.

1.5.1 Instantaneous Power

Energy is carried along a transmission line by both the forward and reverse waves. The instantaneous power arriving at a load is

$$\begin{aligned} p^+(t) &= v^+(t)i^+(t) \\ &= \text{Re}\left\{V_o^+ e^{j\omega t}\right\}\text{Re}\left\{\frac{V_o^+}{Z_o}e^{j\omega t}\right\} \\ &= \text{Re}\left\{|V_o^+|e^{j\phi^+}e^{j\omega t}\right\}\text{Re}\left\{\frac{|V_o^+|}{Z_o}e^{j\phi^+}e^{j\omega t}\right\} \\ &= \frac{|V_o^+|^2}{Z_o}\cos^2\left(\omega t + \phi^+\right) \end{aligned} \tag{1.94}$$

What this result means is that each half cycle of the forward wave delivers power to the load. If we repeat this derivation for the reverse wave, we find that

$$\begin{aligned} p^-(t) &= -\frac{|V_o^-|^2}{Z_o}\cos^2\left(\omega t + \phi^-\right) \\ &= -|\Gamma_L|^2\frac{|V_o^+|^2}{Z_o}\cos^2\left(\omega t + \phi^+ + \theta_L\right) \end{aligned} \tag{1.95}$$

where $\Gamma_L = |\Gamma_L| e^{j\theta_L}$. The negative sign means that the reverse wave carries power away from the load. The net power delivered to the load is equal to the sum of the incident and reflected power:

$$p(t) = p^+(t) + p^-(t) \tag{1.96}$$

1.5.2 Time-Average Power

The time-average power associated with the forward wave is

$$\begin{aligned} p_{\text{av}}^+ &= \frac{1}{T}\int_0^T \frac{|V_o^+|^2}{Z_o}\cos^2(\omega t + \phi^+)\,dt \\ &= \frac{|V_o^+|^2}{Z_o}\underbrace{\frac{1}{T}\int_0^T \cos^2(\omega t + \phi^+)\,dt}_{1/2} \\ &= \frac{|V_o^+|^2}{2Z_o} \end{aligned} \tag{1.97}$$

The time-average power carried away by the reverse wave can be computed in the same way, but we will use the phasor expression in Eq. (1.93) to illustrate an alternate approach:

$$p_{\text{av}}^- = \frac{1}{2}\text{Re}\left\{\tilde{V}^-(0)\tilde{I}^{-*}(0)\right\} \tag{1.98}$$

$$\begin{aligned} &= \frac{1}{2}\text{Re}\left\{V_o^-\left(-\frac{V_o^-}{Z_o}\right)^*\right\} \\ &= -\frac{|V_o^-|^2}{2Z_o} \\ &= -|\Gamma_L|^2\frac{|V_o^+|^2}{2Z_o} \end{aligned} \tag{1.99}$$

The net time-average power delivered to the load is

$$p_{\text{av}} = p_{\text{av}}^+ + p_{\text{av}}^- = \frac{|V_o^+|^2}{2Z_o}(1 - |\Gamma_L|^2) \tag{1.100}$$

What happens if the load is purely reactive (lossless)?

Another way to arrive at the same result is to compute the power absorbed by the load directly from the total phasor voltage at the load:

$$\begin{aligned} p_{\text{av}} &= \frac{1}{2}\text{Re}\left\{\tilde{V}(0)\tilde{I}^*(0)\right\} \\ &= \frac{1}{2}\text{Re}\left\{V_{\text{o}}^+(1+\Gamma_L)\frac{V_{\text{o}}^{+*}(1-\Gamma_L)^*}{Z_{\text{o}}}\right\} \\ &= \frac{|V_{\text{o}}^+|^2}{2Z_{\text{o}}}(1 - |\Gamma_L|^2) \end{aligned} \tag{1.101}$$

which is identical to (1.100).

1.6 Smith Chart

The Smith chart provides a graphical way to solve the transmission line equations that we have derived. Most high frequency engineering is done using computer aided design packages, so we don't really need the Smith chart as a calculation tool. But it does provide a powerful way to communicate the behavior of a transmission line system visually. In fact, software packages and instruments often present computed or measured values on a Smith chart. Thus, the Smith chart is mainly a tool for gaining insight into transmission line systems.

The Smith chart is a plot of a reflection coefficient (1.72) in the complex plane. Superimposed on the complex plane is a grid of circles that represent the real and imaginary parts of the impedance corresponding to the reflection coefficient.

Unit circle and purely reactive impedances. A passive load cannot reflect more power than is incident on it, so from Eq. (1.99), we must have $|\Gamma_L| \leq 1$. Thus, for most transmission line systems the reflection coefficient is confined to be on or inside the unit circle.

Impedance circles. Now, let's derive equations for the curved grid representing the impedance corresponding to Γ_L. If we solve Eq. (1.71) for the load impedance, we obtain

$$Z_L = Z_o \frac{1+\Gamma_L}{1-\Gamma_L} \tag{1.102}$$

Because we don't want to have to have a different Smith chart for every possible value of the characteristic impedance, we will rearrange this expression and work with normalized impedance, which we will identify with a lower case symbol:

$$z_L = \frac{Z_L}{Z_o} = \frac{1+\Gamma_L}{1-\Gamma_L} \qquad \text{(Normalized impedance)} \tag{1.103}$$

Now, we break both z_L and Γ_L into their real and imaginary parts,

$$\begin{aligned} r_L + jx_L &= \frac{1+\Gamma_{Lr}+j\Gamma_{Li}}{1-\Gamma_{Lr}-j\Gamma_{Li}} \\ &= \frac{1+\Gamma_{Lr}+j\Gamma_{Li}}{1-\Gamma_{Lr}-j\Gamma_{Li}} \frac{1-\Gamma_{Lr}+j\Gamma_{Li}}{1-\Gamma_{Lr}+j\Gamma_{Li}} \\ &= \frac{1-\Gamma_{Lr}^2-\Gamma_{Li}^2}{(1-\Gamma_{Lr})^2+\Gamma_{Li}^2} + j\frac{2\Gamma_{Li}}{(1-\Gamma_{Lr})^2+\Gamma_{Li}^2} \end{aligned}$$

With some algebra, the real and imaginary parts of this equation can be rearranged into the forms

$$\left(\Gamma_{Lr} - \frac{r_L}{1+r_L}\right)^2 + \Gamma_{Li}^2 = \left(\frac{1}{1+r_L}\right)^2 \tag{1.104}$$

$$(\Gamma_{Lr}-1)^2 + \left(\Gamma_{Li} - \frac{1}{x_L}\right)^2 = \left(\frac{1}{x_L}\right)^2 \tag{1.105}$$

Both of these equations represent circles. The first one is centered at $(\Gamma_{Lr}, \Gamma_{li}) = (r_L/(1+r_L), 0)$ and has radius $1/(1+r_L)$. The second circle is centered at $(1, 1/x_L)$ and has radius $1/x_L$. Since the imaginary part of the impedance can be positive or negative, we have to consider two circles for each value of x_L. For a given reflection coefficient in the complex plane, two circles intersect at that point, one given by (1.104) for a particular value of r_L and the other given by (1.105) for a value of x_L. On the Smith chart, the r_L and x_L are labeled, so that the value of $z_L = r_L + jx_L$ can easily be read from the chart.

Before using the Smith chart to analyze transmission lines, it is helpful to learn our way around the chart by considering some important landmarks (Fig. 1.14):

1. The unit circle $|\Gamma_L| = 1$ corresponds to lossless loads (capacitative, inductive, open circuit, short circuit).
2. The real axis corresponds to purely resistive load impedances. The left side of the real axis corresponds to $R_L < Z_o$, and the right side to $R_L > Z_o$.
3. The upper half plane represents inductive loads and the lower half plane represents capacitative loads.
4. The center of the Smith chart ($\Gamma = 0$) corresponds to a matched load ($z_L = 1$).
5. The point $\Gamma = -1$ corresponds to a short circuit.
6. The point $\Gamma = 1$ corresponds to an open circuit or infinite load impedance. Since small capacitances, large inductances, and large resistances all lead to $\Gamma \simeq 1$, all of the impedance circles go through that point.

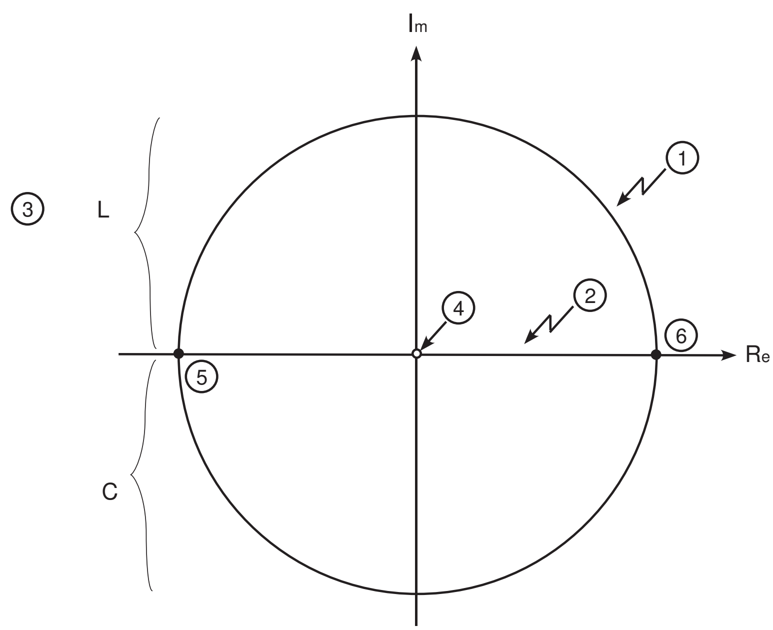


Figure 1.14: Smith chart landmarks. The unit circle represents purely reactive impedances. The real axis represents purely resistive impedances. The upper half plane corresponds to inductive impedances and the lower half plane to capacitative impedances. The center of the Smith chart represents an impedance that is matched to the characteristic impedance of the transmission line. The points -1 and $+1$ on the unit circle represent short circuit and open circuit loads, respectively.

The Smith chart can also be used for admittances, instead of impedances, by rotating the reflection coefficient by 180 degrees around the center of the Smith chart. When the Smith chart is flipped in this way to its admittance personality, all values on the Smith chart must be reinterpreted as admittances.

Generalized reflection coefficient. The next key principle for the Smith chart is that we can also plot the generalized reflection coefficient

$$\Gamma(z) = \frac{V_o^- e^{j\beta z}}{V_o^+ e^{-j\beta z}} = \Gamma_L e^{j2\beta z} \tag{1.106}$$

and the corresponding normalized input impedance $z_{\text{in}}(z) = Z_{\text{in}}(z)/Z_o$ as a function of position on the line. As z moves away from zero, the phase of $\Gamma(z)$ changes such that $\Gamma(z)$ on the Smith chart moves around a circle centered at the origin. The radius of the circle is $|\Gamma_L|$. Increasing z causes the phase angle in Eq. (1.106) to increase, which corresponds to counterclockwise rotation. Since the generator is at $z = -\ell$, moving from the load towards the generator corresponds to clockwise rotation along the circle. If z changes by $\lambda/2$, the generalized reflection coefficient moves once around the circle. If the distance moved along the line is given in wavelengths, so that $L = \ell/\lambda$, then we rotate $2L$ times around the circle.

VSWR. The voltage standing wave ratio on a transmission line is

$$S = \frac{1 + |\Gamma|}{1 - |\Gamma|} \tag{1.107}$$

where Γ can be the load reflection coefficient or the generalized reflection coefficient anywhere on the line. A circle of constant $|\Gamma|$ on the Smith chart is also a circle of constant VSWR. When the VSWR circle crosses the positive real axis, the imaginary part of z_{in} is zero and the real part is equal to some value $r > 1$, so the magnitude of the generalized reflection coefficient is

$$|\Gamma(z)| = \frac{r - 1}{r + 1} \tag{1.108}$$

If we solve this equation for r and compare the resulting expression to (1.107), we find that $S = r$. As we move along a transmission line, the voltage standing wave maxima occur when the generalized reflection coefficient crosses the positive real axis. The minima occur when it crosses the negative real axis.

Admittance. When transmission lines or circuit elements are in parallel, it is convenient to convert from impedances to admittances. There are two ways to work with admittances on a Smith chart. One is to add another grid for admittances in a different color. The other is to shift an impedance point to a new point and then reinterpret the impedance circles on the Smith chart as admittance lines. The load admittance is

$$y_L = \frac{1}{z_L} = \frac{1 - \Gamma_L}{1 + \Gamma_L}$$

By comparing this to

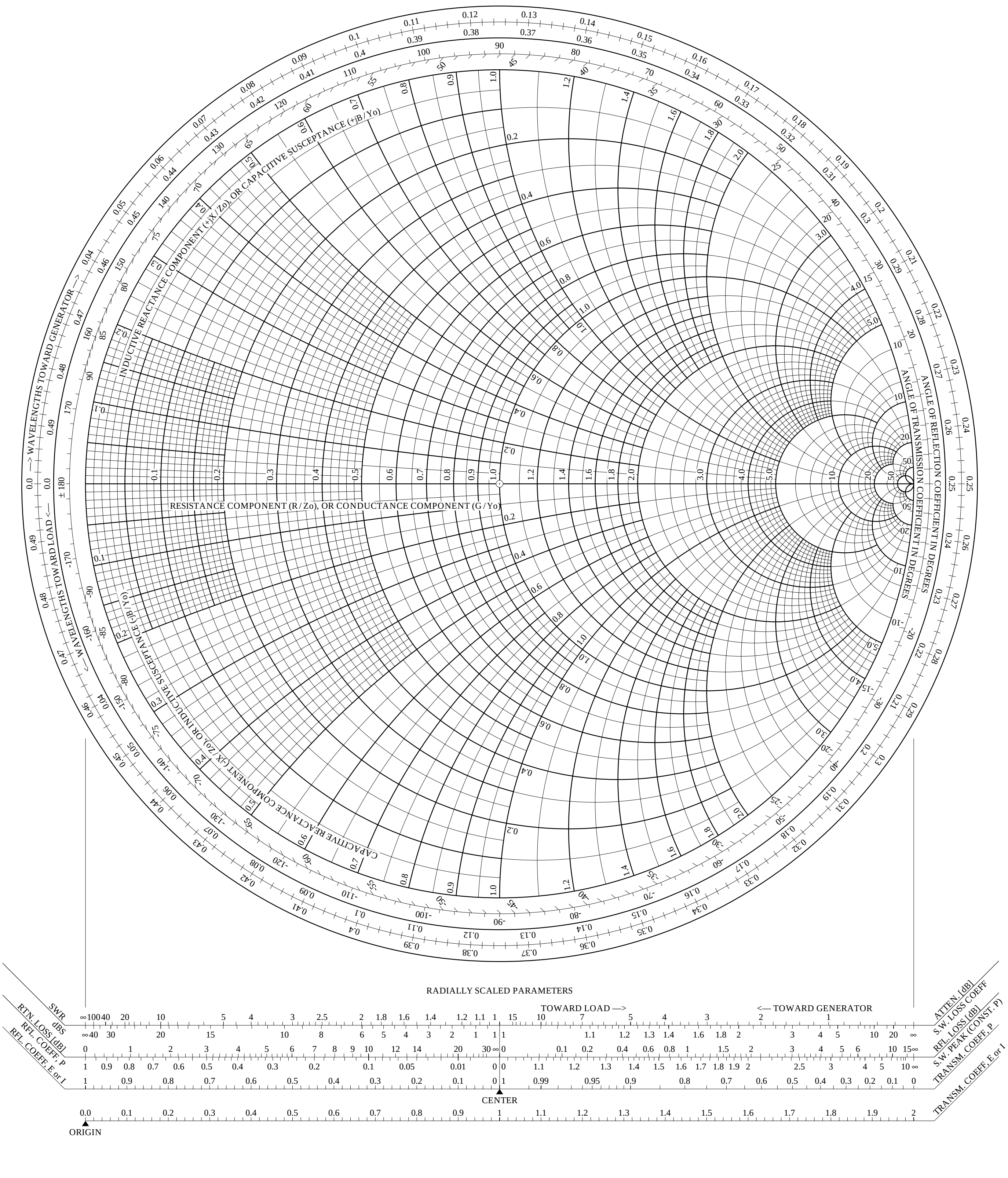
$$\begin{aligned} z_{\text{in}}(z = -\lambda/4) &= \frac{1 + \Gamma_L e^{2j\beta z}}{1 - \Gamma_L e^{2j\beta z}} \\ &= \frac{1 + \Gamma_L e^{j\pi}}{1 - \Gamma_L e^{j\pi}} \\ &= \frac{1 - \Gamma_L}{1 + \Gamma_L} \end{aligned}$$

it can be seen that a $\lambda/4$ rotation on the Smith chart transforms an impedance to an admittance. This corresponds to reflection with respect to the origin. So, if we are given an admittance, we can plot it on a single color Smith chart using its real and imaginary parts as values for the impedance circles, and the resulting reflection coefficient is found by reflecting the point about the origin. Alternatively, a two-color Smith chart can be used, with one set of circles representing admittances, and another set of circles in a different color representing admittances.

1.6.1 Smith Chart Solution Procedure

To solve a lossless transmission line problem with generator and load impedance graphically on a Smith chart, the following steps are involved:

1. Find the normalized load impedance and plot it on the Smith chart.
2. Rotate the load reflection coefficient clockwise around a circle of constant radius on the Smith chart by an angle of $2\beta l$ or $2\ell/\lambda$ times around the circle.
3. Read the normalized input impedance $z_{\text{in}}(-\ell)$ from the Smith chart, and unnormalize to get Z_{in}. This can be used in a voltage divider to get $\tilde{V}(-\ell)$, from which V_o^+ can be found using equations.
4. Other quantities can be read from the Smith chart as well:
   (a) VSWR: The voltage standing wave ratio is equal to the real part of the normalized input impedance when the constant VSWR circle crosses the positive real axis.
   (b) Voltage maxima/minima: The first voltage maximum occurs when the generalized reflection coefficient first crosses the positive real axis. The voltage minima occur when it crosses the negative real axis. The rotation angles at which the extrema occur can be read from the Smith chart and converted to distances along the line.



Example: Smith Chart Solution

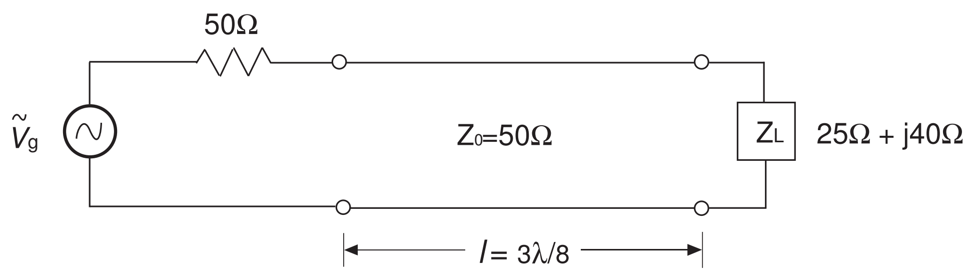


Figure 1.15: Transmission line with source and load.

1. Find the normalized load impedance and plot on Smith chart.

2. Rotate to the generator end.

3. Read off the normalized input impedance z_{in} and unnormalize to get Z_{in}.

4. Read the VSWR from the Smith chart.

5. Find the distance between the load and the first voltage minimum in wavelengths.

6. If the generator voltage is 1 V, solve for the voltage at the line input end.

7. Using a voltage divider with the generator impedance and the line input impedance, find the forward voltage wave amplitude V_o^+.

8. Find the time average power dissipated in the load.

1.7 Matching

A mismatched load ($Z_L \neq Z_o$) means that power is reflected and is not delivered to the load. If this is undesirable, an impedance matching network can be used to make the load appear to be matched. There are many ways to do this:

- Stub tuning
- Quarter-wave matching
- Lumped element matching
- Matching transformers
- Multistage (broadband) matching networks
- Tapered transmission lines
- Isolators (non-reciprocal devices allowing only one-way power flow)

1.7.1 Shunt Single-Stub Matching

One type of matching network is a short section of transmission line placed in parallel with the main line at some distance from the load (Fig. 1.16). The stub is typically terminated with an open or short circuit. Since we are placing two elements in parallel, it is convenient to use admittances for the design procedure, since the admittance of two parallel elements is the sum of the admittances of the elements. The basic principle is to place the stub at a location where the input admittance on the main line is of the form $Y_o + jB$, and then choose the length of the stub so that its input admittance is $-jB$. The parallel combination of the stub and main line then has admittance $Y_{\text{in}} = Y_o$, which represents a match.

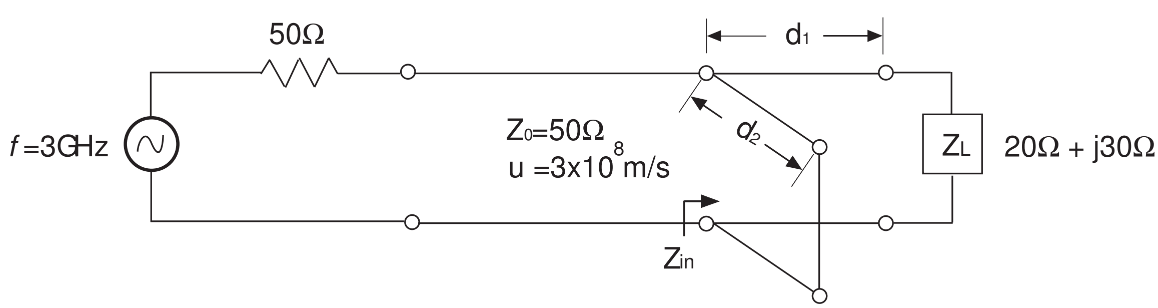


Figure 1.16: Shunt single-stub matching network.

The goal is to find the length of the stub d_2 and the distance of the stub from the load d_1 such that the input impedance looking into the junction of the stub and main line is equal to Z_o. The solution procedure is as follows:

1. Normalize the load impedance.

2. Plot y_L on the Smith chart. On a single-color Smith chart, this is done by plotting z_L and then rotating by $180°$ to y_L.

3. Rotate towards the generator (clockwise) around the constant VSWR circle until the input impedance reaches the $g = 1$ circle. At this point, the normalized input admittance of the line is of the form $y_{d_1} = 1 + jb$. Read y_{d_1} and d_1 in wavelengths from the Smith chart.

4. The input admittance of the sub must be $y_{\text{stub}} = -jb$, so that $y_{\text{in}} = y_{d_1} + y_{\text{stub}} = 1$. From the Smith chart, determine the length of the stub d_2 that gives this input admittance. The generator end of the stub is at the point $(0, -1)$. On a single color Smith chart, we flip this about the origin to the point $(1, 0)$. We then rotate towards the generator end until the admittance is $-jb$, and read off the length of the stub in wavelengths. The lengths can easily be converted to meters by multiplying by λ.

5. What would d_2 be if we want to use an open circuit stub?

There are other solutions for this matching problem. We could continue rotating away from the load until the main line input impedance hits the $g = 1$ circle again.

1.7.2 Series Single-Stub Matching

If the stub is in series, then the design procedure changes, so that we rotate from the load end of the main line to the $r = 1$ circle, and then add a series stub to cancel the imaginary part of the impedance.

Shunt Stub Matching Example 2

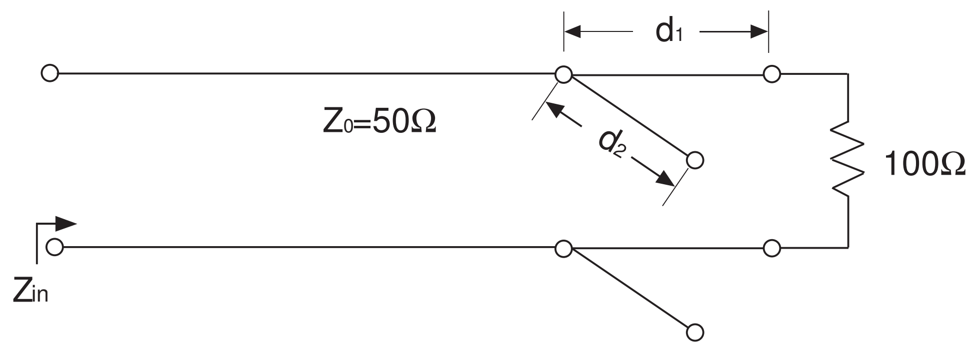


Figure 1.17: Shunt sub matching example.

1. Normalize the load impedance.
2. Plot y_L on the Smith chart.
3. Rotate towards the generator (clockwise) around the constant VSWR circle until the input impedance reaches the $g = 1$ circle. Read y_{d_1} and d_1 in wavelengths from the Smith chart.

4. The input admittance of the sub must be $y_{\text{stub}} = -jb$. For an open circuit stub, we start at $y_{L,\text{stub}} = 0$ and rotate towards the generator end until the admittance is $-jb$, and read off the length of the stub in wavelengths.

5. Find another solution to the problem:

1.8 Transients on Transmission Lines

So far, we have considered sinusoidal excitations on transmission lines. Since many signals are narrowband and can be approximated for the purpose of transmission line design as purely sinusoidal, the phasor domain analysis of earlier sections is powerful and can be applied to many situations. Signals can also be broadband or transient in nature, and in that case a different analysis technique is needed. The most common application for transient analysis is signalling on a digital interconnect between logic gates.

Consider the transmission line circuit in Fig. 1.18.

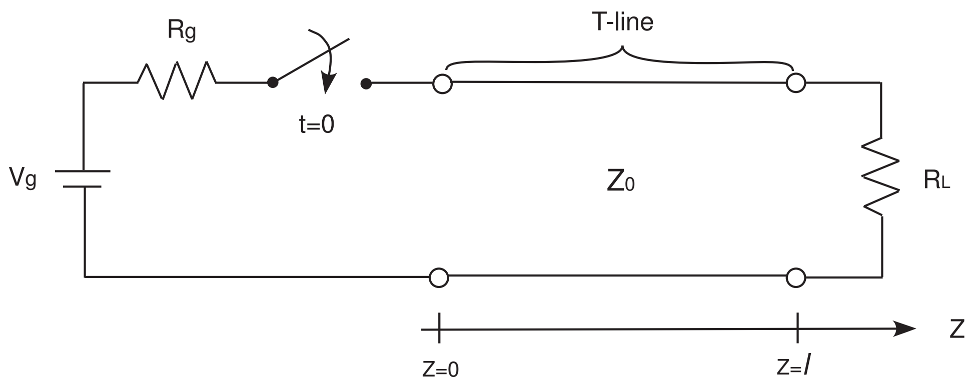


Figure 1.18: Transmission line circuit.

When the switch closes a step voltage appears across the generator end of the line. What is the value of that voltage? The characteristic impedance is not the total impedance of the line but rather the relationship between the forward and reverse traveling voltages and currents. But when the switch first closes there is no reverse wave, because the forward step has not had time to travel down to the end and back. Therefore, the characteristic impedance of the line is the total impedance of the line imediately after the switch closes. This leads to the equivalent circuit shown in Fig. 1.19 when the switch is first closed. From the equivalent

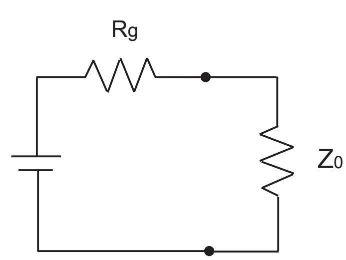


Figure 1.19: Equivalent circuit at $t = 0$.

circuit we can use a voltage divider relationship to calculate the magnitude of the step voltage,

$$v_1^+(z=0,t=0) = \frac{Z_o}{Z_o+R_g}\,V_g \tag{1.109}$$

This initial pulse then starts to travel down the line at speed u. At time $t=\ell/(2u)$, for example the step is halfway down the line. At $t=\ell/u$, the step arrives at the load.

1.8.1 Reflection Coefficient

What happens when the step hits the load? The pulse will reflect, and the v^- term in the wave equation solution will no longer be zero. We need to find the amplitude of the reflected wave. This is easy to do using boundary conditions at the load end of the transmission line. The boundary conditions are

$$v(\ell,T) = v_L(T) \tag{1.110}$$
$$i(\ell,T) = i_L(T) \tag{1.111}$$

where $T=\ell/u$ and $v_L(T)$ and $i_L(T)$ are the voltage across and the current through the load resistor. v_L and i_L are related by Ohm's law: $v_L = i_L R_L$. Putting Ohm's law together with the boundary conditions at the load end of the line, we obtain

$$v(\ell,T) = R_L i(\ell,T) \tag{1.112}$$

Using Eqs. (1.35) and (1.36), this can be rewritten as

$$v_1^+ + v_1^- = R_L\left[\frac{v_1^+}{Z_o} - \frac{v_1^-}{Z_o}\right] \tag{1.113}$$

where the line voltages and currents are all evaluated at $z=\ell$ and $t=T$. Solving this for v_1^- gives

$$v_1^- = \frac{R_L - Z_o}{R_L + Z_o} v_1^+ \tag{1.114}$$

We call the constant in this expression the load reflection coefficient:

$$\Gamma_L = \frac{R_L - Z_o}{R_L + Z_o} \qquad \text{(Load reflection coefficient)} \tag{1.115}$$

From this we can find the magnitude of the reflected wave:

$$v_1^- = \Gamma_L v_1^+ = \frac{R_L - Z_o}{R_L + Z_o}\frac{Z_o}{Z_o + R_g}\,V_g \tag{1.116}$$

The voltage waveform on the transmission line at a time that is after the reflection from the load, and before the reflected pulse arrives at the generator, is shown in Fig. 1.20.

What happens when the reflected wave gets back to the generator? Using the same idea as at the load end, we can show that

$$v_2^+ = \frac{R_g - Z_o}{R_g + Z_o} v_1^- = \Gamma_g v_1^- = \Gamma_g \Gamma_L v_1^+ \tag{1.117}$$

Can you derive this expression on your own?

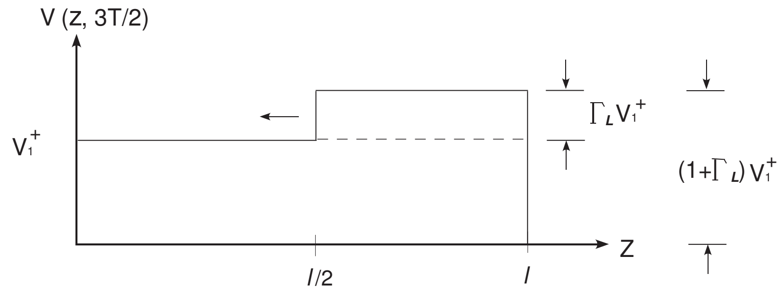


Figure 1.20: Voltage waveform at time $t = 3T/2$.

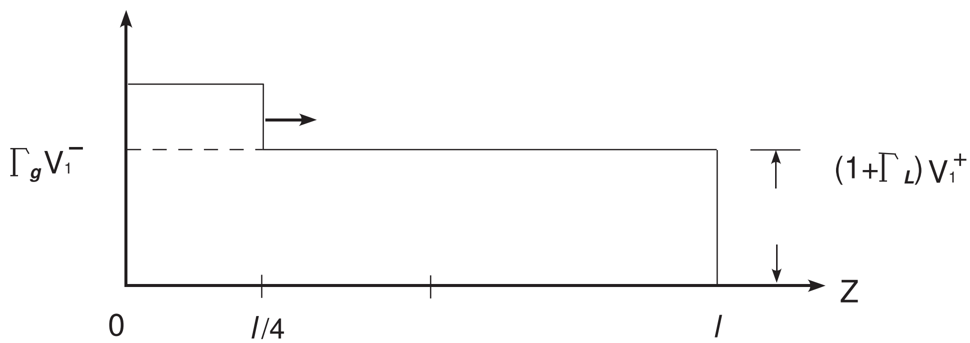


Figure 1.21: Voltage waveform at time $t = 9T/4$.

Figure 1.21 shows the voltage waveform after the reflected pulse has reflected again from the generator end of the transmission line. The total voltage at a point on the transmission line is the sum of all the reflected and forward steps (v_1^+, v_1^-, v_2^+) that have occured up until the current time.

If we look at the results we have found so far, we can see a pattern:

$$
\begin{array}{rcll}
v_1^+ & = & \frac{Z_o}{Z_o+R_g}\,V_g & \text{(Voltage divider at } t=0) \\
v_1^- & = & \Gamma_L v_1^+ & \text{(First reflection at the load)} \\
v_2^+ & = & \Gamma_g \Gamma_L v_1^+ & \text{(Reflection at generator)} \\
v_2^- & = & \Gamma_L \Gamma_g \Gamma_L v_1^+ & \\
v_3^+ & = & \Gamma_g \Gamma_L \Gamma_g \Gamma_L v_1^+ & \\
\vdots & & &
\end{array}
\tag{1.118}
$$

We can write out the total voltage at some point on the line at $t = \infty$ as an infinite series:

$$\begin{aligned} v(z, t=\infty) &= v_1^+ + \Gamma_L v_1^+ + \Gamma_g \Gamma_L v_1^+ + \Gamma_L \Gamma_g \Gamma_L v_1^+ + \Gamma_g \Gamma_L \Gamma_g \Gamma_L v_1^+ + \cdots \\ &= v_1^+ \left[(1+\Gamma_L) + (1+\Gamma_L)\Gamma_L \Gamma_g + (1+\Gamma_L)\Gamma_L^2 \Gamma_g^2 + \cdots \right] \quad (1.119) \\ &= v_1^+ (1+\Gamma_L) \underbrace{\left[1 + \Gamma_L \Gamma_g + \Gamma_L^2 \Gamma_g^2 + \cdots \right]}_{\text{Geometric series}} \quad (1.120) \\ &= v_1^+ (1+\Gamma_L) \frac{1}{1 - \Gamma_L \Gamma_g} \quad (1.121) \end{aligned}$$

If we plug in the definitions of Γ_L and Γ_g, this reduces to

$$v(z, t=\infty) = \frac{R_L}{R_g + R_L} V_g \quad (1.122)$$

This is the steady state voltage on the transmission line. Does this result make sense?

1.8.2 Bounce Diagrams

A convenient tool for understanding transmission line transients is a bounce diagram. One axis of the diagram is the z coordinate, and the other is time. On it we plot the location of the leading edge of the reflected pulse as it propagates between the load and generator terminations, and note the amplitude of each reflection.

The bounce diagram for a single transmission line with load reflection coefficient Γ_L and generator reflection coefficient Γ_g is shown in Fig. 1.22. The voltage at a point on the transmission line at a time t is the sum

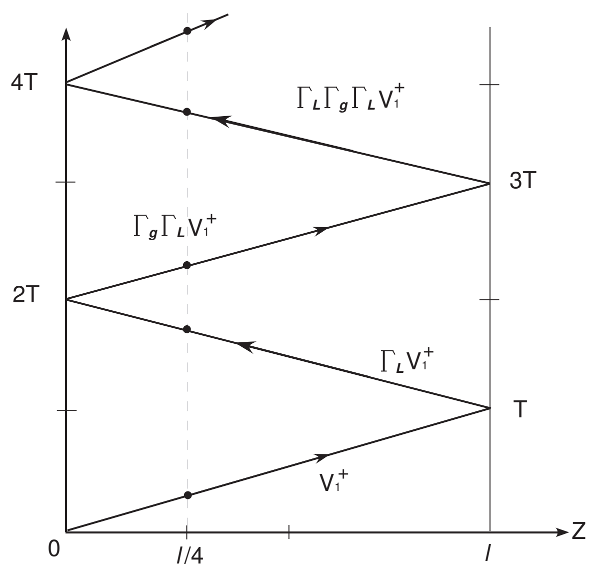


Figure 1.22: Bounce diagram for a single transmission line with load reflection coefficient Γ_L and generator reflection coefficient Γ_g.

of all of the reflections that have occured up to the given time. So, if we want to plot the voltage at a point on the line as a function of time, we draw a vertical line on the bounce diagram at the given location. The reflected pulse passes that point at each time for which the vertical line crosses the line representing the leading edge of the reflected pulse. At a given time, the voltage on the transmission line is the sum of all the reflection amplitudes below the current time (Fig. 1.23).

We can also find the current using

$$i = \frac{v^+}{Z_o} - \frac{v^-}{Z_o} \tag{1.123}$$

If the currents and voltages are evaluated at one of the ends of the transmission line, this can be rearranged to obtain

$$\frac{i^-}{i^+} = \frac{-v^-/Z_o}{v^+/Z_o} = -\frac{v^-}{v^+} = -\Gamma \tag{1.124}$$

where Γ is the voltage reflection coefficient at the load or generator end. So, the current bounce diagram is the same as the voltage bounce diagram, but with the signs of all the reflection coefficients reversed and v_1^+ replaced by i_1^+.

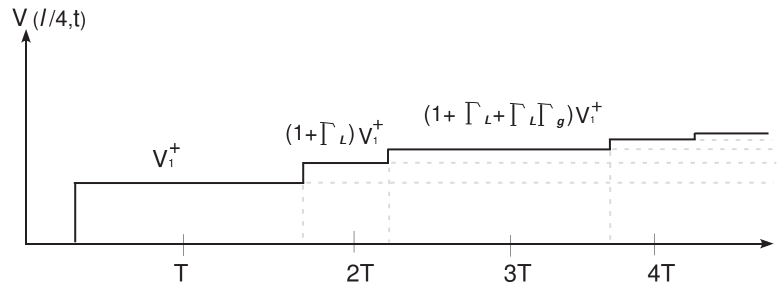


Figure 1.23: Voltage as a function of time at the location $z = \ell/4$ obtained from the bounce diagram.

1.8.3 Multi-section Lines

Transmission lines can be placed in series or in parallel. By proper application of boundary conditions at each junction, any situation can be handled using the techniques we have developed. Let's consider the former case, as shown in Fig. 1.24.

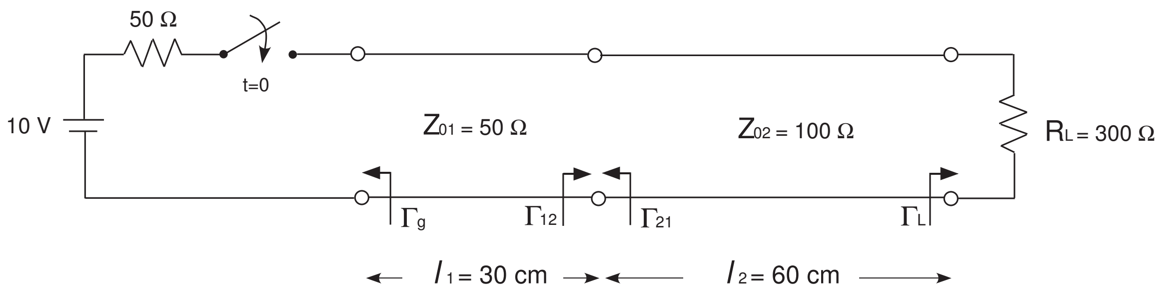


Figure 1.24: Two transmission lines in series.

We first find the initial pulse amplitude

$$v_{11}^{+} = 10\frac{50}{50+50} = 5\,\mathrm{V}$$

and compute the reflection coefficients at each junction:

$$\begin{aligned}
\Gamma_g &= \frac{50-50}{50+50} = 0 \\
\Gamma_{12} &= \frac{100-50}{100+50} = \frac{1}{3} \\
\Gamma_{21} &= \frac{50-100}{50+100} = -\frac{1}{3} \\
\Gamma_L &= \frac{300-100}{300+100} = \frac{1}{2}
\end{aligned}$$

The amplitude of the first reflection from the junction between the two lines is $v_{11}^{-} = 5\Gamma_{12} = 5/3\,\mathrm{V}$. At the time of the first reflection, the total voltage at the right end of line 1 is $v_{11}^{+} + v_{11}^{-} = 5 + 5/3 = 20/3\,\mathrm{V}$.

Applying the voltage boundary condition, this must also be the voltage at the left end of the second line, so that a pulse with amplitude $v_{21}^+ = 20/3\,\text{V}$ is launched down the second line. This pulse reflects from the load with amplitude $v_{21}^- = \Gamma_L v_{21}^+ = 10/3\,V$. The load reflection reaches the junction between the two lines, and reflects to the right with amplitude $v_{22}^+ = \Gamma_{21} v_{21}^- = -10/9$, while at the same time launching a wave down the first transmission line with amplitude $v_{22}^+ + v_{21}^- = 20/9$. Because the generator impedance is matched to the characteristic impedance of line 1, there is no reflection from the generator, but reflections continue between the right and left ends of line 2.

The resulting bounce diagram is shown in Fig. 1.25. Figure 1.26 shows the voltage at $z = 15\,\text{cm}$ as a function of time.

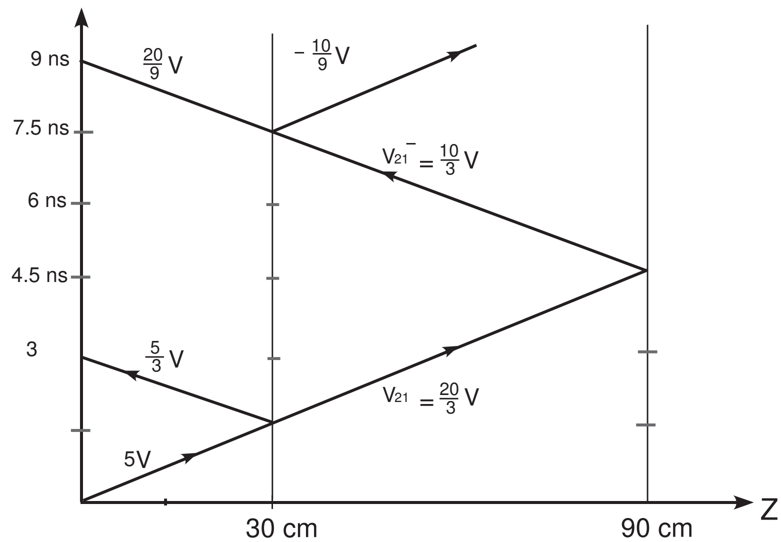


Figure 1.25: Bounce diagram for the series transmission lines in Fig. 1.24.

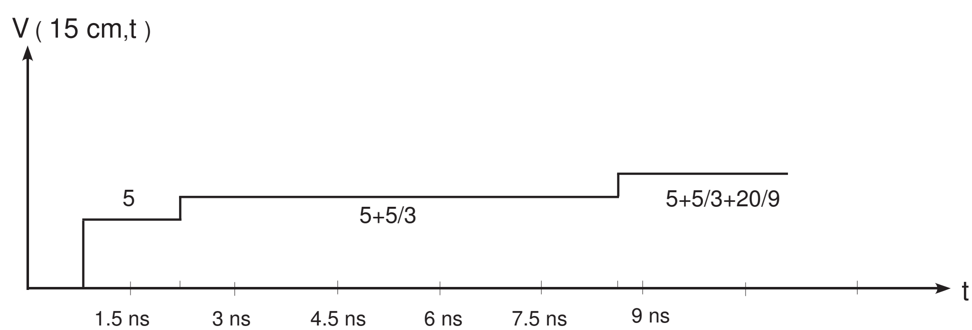


Figure 1.26: Voltage for the series transmission lines in Fig. 1.24 at $z = 15\,\text{cm}$ as a function of time.

We can also handle a branched transmission line. If we replace the second line in Fig. 1.24 with two lines in parallel, the system shown in Fig. 1.27 is obtained. The same forward wave with amplitude $v_{11}^+ = 5\,\text{V}$ is launched at $t = 0$. When the pulse arrives at the branch junction, the effective load impedance seen by the pulse is the parallel combination of the characteristic impedances of lines 2 and 3:

$$R_L = 50\|50 = 25\,\Omega \tag{1.125}$$

The reflection coefficient is then

$$\Gamma_1 = \frac{25-50}{25+50} = -\frac{1}{3}$$

The reflection back down line 1 has amplitude $v_{11}^- = \Gamma_1 v_{11}^+ = -5/3\,\text{V}$. Pulses with amplitude $v_{11}^+ + v_{11}^- = 10/3\,\text{V}$ are launched down lines 2 and 3. Continuing in this way, we can compute the values of all the reflections at each junction and complete the analysis of the branched line system. A bounce diagram can be drawn for the branched lines as shown in Fig. 1.28.

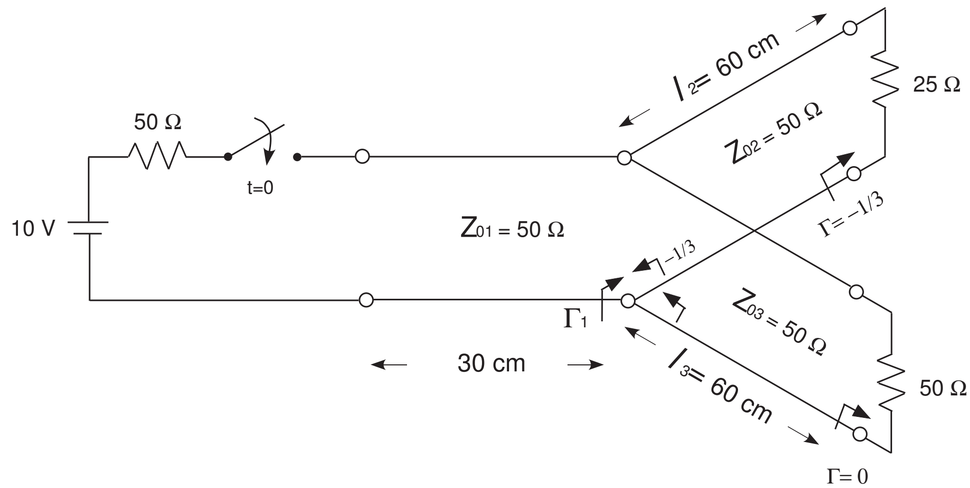


Figure 1.27: Branched transmission lines.

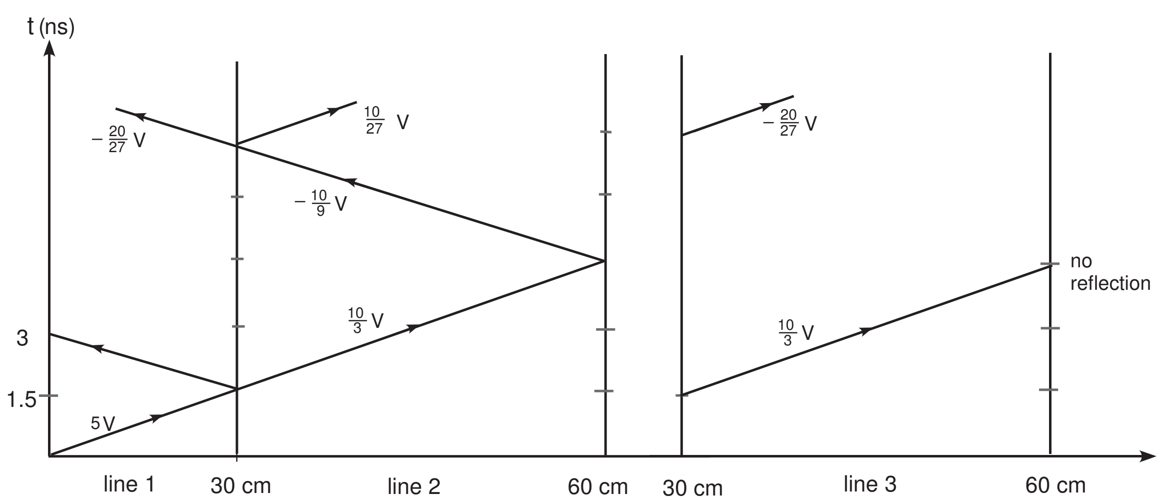


Figure 1.28: Bounce diagram for the branched transmission lines in Fig. 1.27.

1.8.4 Reactive Load

What happens if the load is a capacitor? Using the general principles we have already developed, we can solve this and many other types of new problems.

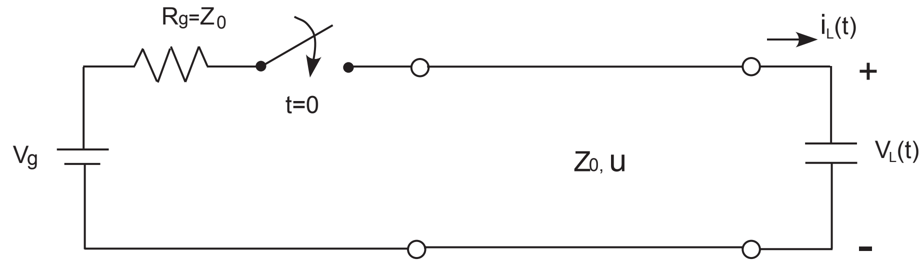


Figure 1.29: Transmission line with a reactive load.

Figure 1.29 shows a transmission line with a reactive load. At $t = 0$, the voltage on the capacitor is assumed to be zero. This remains unchanged until the step launched by the generator arrives at the load at time $T = \ell/u$. At this time, the capacitor begins to charge. The steady state voltage on the capacitor will be V_g, as shown in Fig. 1.30.

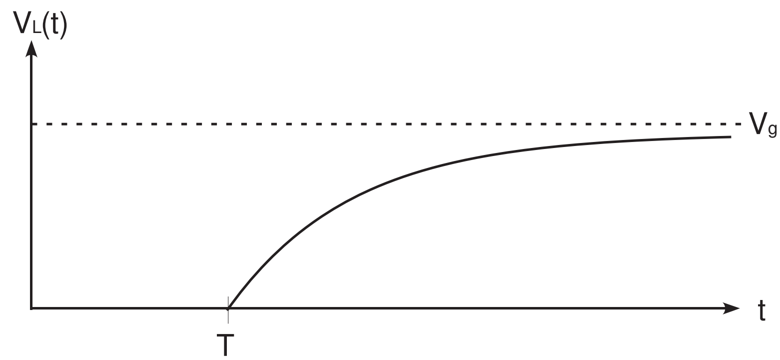


Figure 1.30: Voltage across load capacitor.

We can find the analytical form of the voltage across the capacitor by solving a differential equation. By the current boundary condition at the load end of the transmission line and the capacitor voltage-current relationship, we can say that

$$i^+ + i^- = C\frac{dv_L(t)}{dt}$$

Using the relationship between the current and voltage waves on the transmission line, we can rewrite this as

$$\frac{v^+}{Z_o} - \frac{v^-}{Z_o} = C\frac{dv_L(t)}{dt}$$

There are two unknowns in this equation, but we can reduce that to one by using the voltage boundary condition $v_L = v^+ + v^-$ to eliminate v^-, so that we have

$$\frac{v^+}{Z_o} - \frac{v_L - v^+}{Z_o} = C\frac{dv_L(t)}{dt}$$

By rearranging this into standard form, we obtain the differential equation that we need to solve:

$$C\frac{dv_L(t)}{dt} + \frac{v_L}{Z_o} = \frac{2v^+}{Z_o} \tag{1.126}$$

This can be solved using standard techniques for differential equations. The form of the solution for $t \geq T$ is

$$v_L(t) = A + Be^{-m(t-T)} \tag{1.127}$$

The constants A, B, and m can be found by substituting this into the differential equation (1.126). After doing this, the final solution is found to be

$$v_L(t) = 2v^+ - 2v^+e^{-(t-T)/(Z_oC)} \tag{1.128}$$

The amplitude of the reflected wave at the load end is

$$v^-(t) = v^+(1 - 2e^{-(t-T)/(Z_oC)}) \tag{1.129}$$

This result can be used to find the voltage at, say, the generator end of the transmission line. The voltage waveform given in Eq. (1.129) arrives at the generator at time $t = 2T$. Before that time, the voltage at the generator end is $V_g/2$. At time $t = 2T$, the arriving reflected wave has amplitude $-v^+ = -V_g/2$, so the total voltage at the generator changes to zero. As the reflected wave becomes less negative, the voltage at the generator increases. The steady state voltage is the sum of the initial forward wave $v^+ = V_g/2$ and the limiting value of the reflected wave, which is $v^-(\infty) = v^+ = V_g/2$, so that the final voltage at the generator is V_g. This is shown in Fig. 1.31.

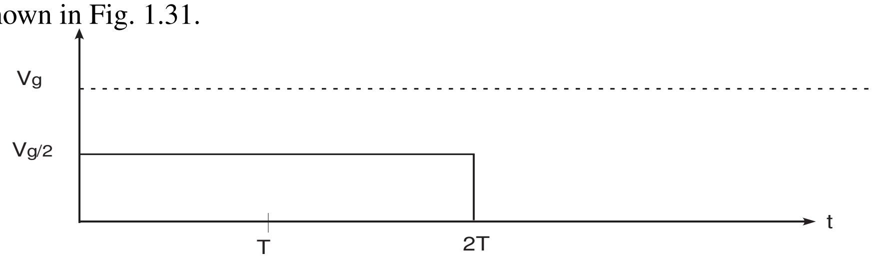


Figure 1.31: Voltage at the generator end of the transmission line.

If we turn off the source at some time t_o, then v^+ goes to zero and the voltage at the generator end changes to $v^- = V_g/2$. The edge of the step down in the v^+ wave propagates to the right. When it arrives at the load, v^+ is zero all the way along the line, and the capacitor is charged to V_g, so the reverse wave changes to $v^- = V_g$. At this time, the capacitor begins discharging and v^- decays to zero. The step up in v^- propagates to the left until it arrives at the generator at time $t = t_o + T$. The voltage at the generator then changes to V_g and decays to zero (Fig. 1.32).

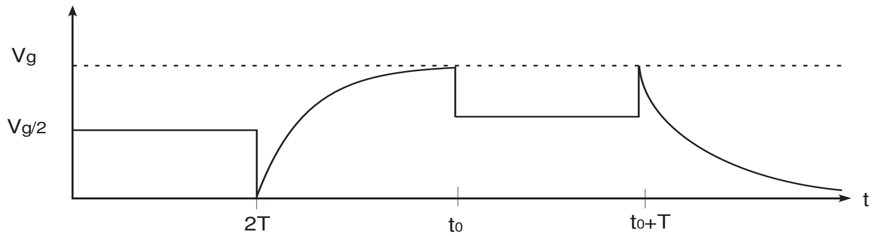


Figure 1.32: Voltage at the generator end of the transmission line if source is turned off.

1.9 Digital Signaling

One important application of transmission line theory is modeling connections carrying digital signals between logic elements. Two of the main issues that must be dealt with are terminations to reduce reflections and cross-talk between closely spaced lines.

1.9.1 Microstrip

One important type of transmission line that is used for both analog and digital systems is the microstrip, which consists of a conductive strip separated from a ground plane by a dielectric layer (Fig. 1.33). An example is a printed circuit board trace with a ground plane on the bottom of the board.

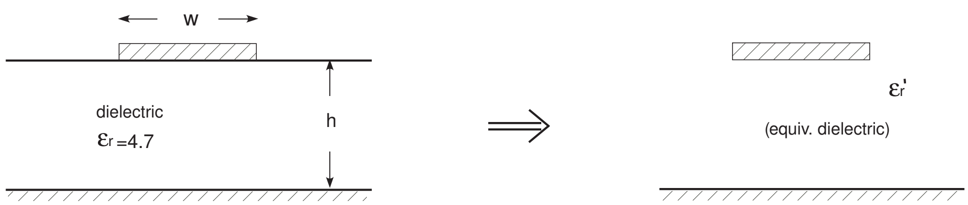


Figure 1.33: (a) Microstrip transmission line. (b) Dielectric and air replaced by an effective medium with relative permittivity ϵ_r'.

The microstrip produces electric and magnetic fields in both the dielectric and the air above the dielectric. To a good approximation, the dielectric and air can be replaced by an effective medium everywhere above the ground plane with relative permittivity

$$\epsilon_r' = \frac{\epsilon_r + 1}{2} + \frac{\epsilon_r - 1}{2} \frac{1}{\sqrt{1 + 10h/w}} \tag{1.130}$$

The inductance per unit length of a microstrip line can be approximated by

$$L \simeq \begin{cases} \frac{60}{c} \ln\left[\frac{8h}{w} + \frac{w}{4h}\right] & \frac{w}{h} \le 1 \\ \frac{120\pi}{c}\left[\frac{w}{h} + 1.393 + 0.667\ln\left(\frac{w}{h} + 1.444\right)\right]^{-1} & \frac{w}{h} \ge 1 \end{cases} \tag{1.131}$$

The capacitance per unit length is

$$C \simeq \begin{cases} \frac{\epsilon_r'}{60c\ln\left[\frac{8h}{w} + \frac{w}{4h}\right]} & \frac{w}{h} \le 1 \\ \frac{\epsilon_r'}{120\pi c}\left[\frac{w}{h} + 1.393 + 0.667\ln\left(\frac{w}{h} + 1.444\right)\right] & \frac{w}{h} \ge 1 \end{cases} \tag{1.132}$$

These formulas allow microstrip transmission lines to be designed for a given characteristic impedance.

1.10 Printed Circuit Board (PCB) Termination

Consider a connection between two digital logic elements, as shown in Fig. 1.34. This can be modeled as a microstrip transmission line. There are several ways to terminate the system to minimize undesirable reflections on the transmission line.

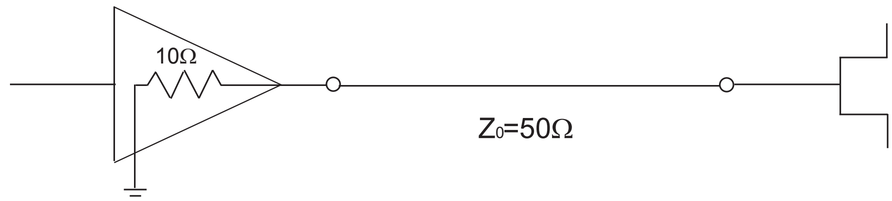


Figure 1.34: Connection between two digital logic elements.

No termination. The FET at the receiver end appears as a high impedance load, which is essentially an open circuit. The load reflection coefficient is $\Gamma_L \simeq 1$. For a driver impedance of $10\,\Omega$ and pulse amplitude 5 V, the initial forward step amplitude is

$$v_1^+ = \frac{50}{10+50} 5\,V = \frac{25}{6}\,V \simeq 4.167\,V$$

This wave propagates until reaches the receiver, at which time it reflects and a reverse pulse propagates towards the driver. The source reflection coefficient is

$$\Gamma_s = \frac{10-50}{10+50} = -\frac{2}{3}$$

Repeated reflections lead to the voltage signal at the receiver end shown in Fig. 1.35.

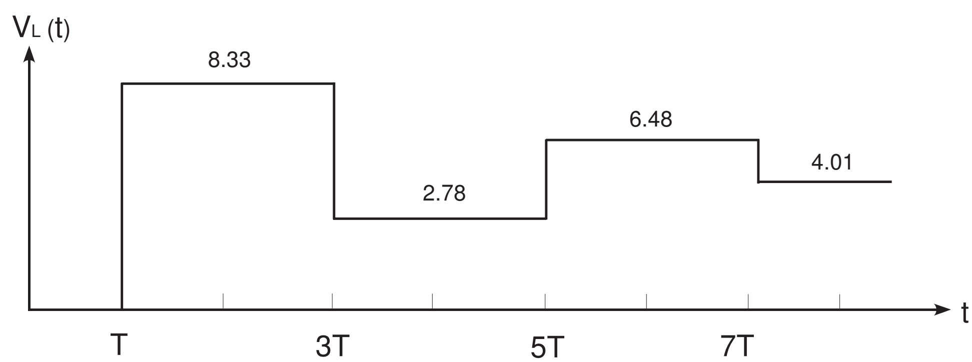


Figure 1.35: Voltage as a function of time at the receiver end.

The change in the voltage decreases by $\Gamma_L \Gamma_s = -2/3$ at each bounce. In order for the voltage at the receiver to settle to within 10%, we must have

$$|\Gamma_L \Gamma_s|^N \leq 0.1 \tag{1.133}$$

which first occurs when $N = 6$. The time required for the voltage to settle to this level is

$$t_s = 2NT \tag{1.134}$$

If the pulse is repeated with a frequency f, we would like to have the voltage settle by at least the middle of the pulse, so that the next pulse is not disturbed too strongly by reflections still occurring from the previous pulse. This means that we must have

$$t_s \le \frac{1}{4f} \tag{1.135}$$

Substituting the definition of phase velocity and using $N = 6$, for the given example we find that $f < u/(48L)$.

If the line is very short, then the delay time T is small, and so the settling time t_s is also small. In this case, termination may be unnecessary. A rule of thumb is that if the pulse rise time is greater than $6T$, termination is not needed. For example, a trace 6.6" long on standard FR4 PC board has a delay of $T = 1$ ns. The pulse rise or fall time must be greater than 6 ns in order to neglect termination.

For a longer line or higher pulse repetition rates, reflections may be intolerable. For $L = 10\,\text{cm}$ and $u = 2 \times 10^8$ m/s, we find for the given example that $f < 42\,\text{MHz}$. Alternately, the length of the line must be less than about $0.1\,\lambda$. For a digital system, this is a very low operating frequency. To do better, we need to add some kind of termination to the line to reduce the reflections.

Load termination. Another way to terminate the connection is to add a matching resistor of value equal to Z_o in parallel with the receiver transistor (Fig. 1.36). This eliminates all reflections. But the power dissipated at the load is

$$P = \frac{V^2}{R} \simeq \frac{4.166^2}{50} \simeq 0.35\,\text{W}$$

For a system with many connections, the total dissipated power would be intolerably large. Moreover, this is more power than can be supplied by a typical driver circuit.

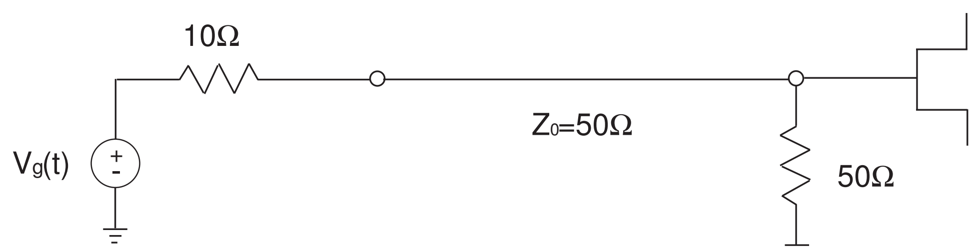


Figure 1.36: Matching resistor at the load.

Source termination. Another termination scheme with less power consumption is to add a resistor at the driver end which increases the effective source impedance to Z_o (Fig. 1.37). The initial forward wave has amplitude $v_1^+ = 2.5\,\text{V}$, and the reflection from the receiver is $v_1^- = 2.5\,\text{V}$. But the reflection coefficient looking into the driver and source termination is zero, so there is only one bounce, and the settling time is faster than in the case of no termination. Because the receiver still appears as a high impedance load, little

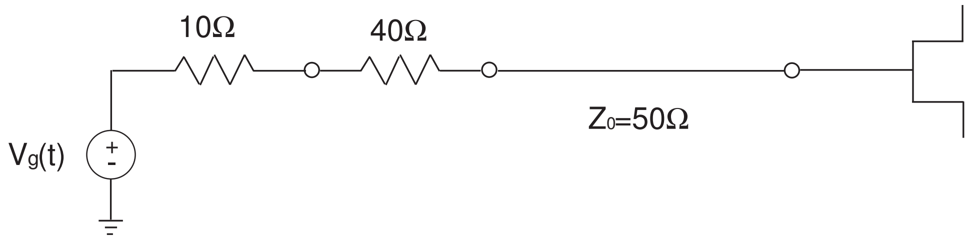


Figure 1.37: Matching resistor at the source.

current flows and the power dissipated is small. Some potential problems with this approach are that the driver impedance may be different depending on whether it is sourcing or sinking current; multiple loads do not see a "high" voltage at the same time; and there is still some power dissipated due to the source resistance.

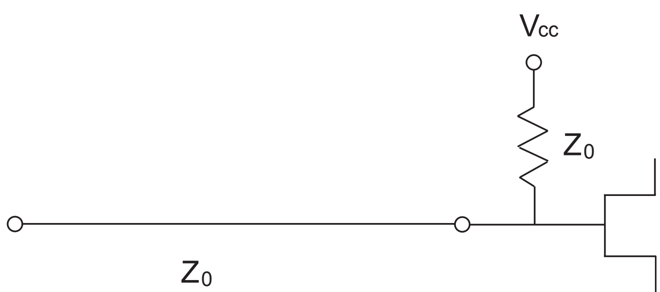


Figure 1.38: Thevenin load termination.

Thevenin termination. Another approach is to use a pullup resistor at the receiver (Fig. 1.38). When the driver switches high and a forward wave travels to the load end, there is no reflection, since the termination has value Z_o. But in this case the voltage across the resistor is small and not very much current flows through it. The current associated with the reflected wave is almost equal and opposite to the incident current. When the driver goes low, the incident current turns off, so there is a net current flow towards the driver. This setup has the advantage that the driver does not have to supply very much current and only has to sink current when the pulse turns off.

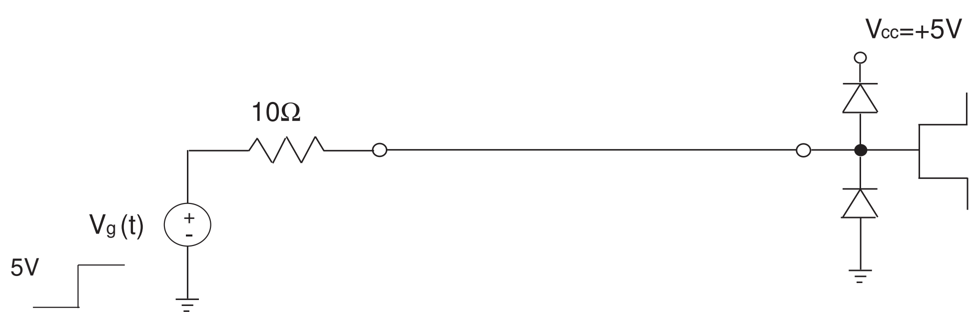


Figure 1.39: Diode clamp termination.

Diode clamp termination. If the diode turn-on voltage is v_d, and the incident voltage is v_1^+, when the incident pulse arrives at the load, the upper diode in Fig. 1.39 will turn on. The voltage at the load will be clamped at $v_{cc} + v_d$, so the amplitude of the reflected wave is only $v_{cc} + v_d - v_1^+$. For the given example, with $v_d = 0.7\,\text{V}$, $v_1^- = 1.53\,\text{V}$, which is smaller than the reflection in the case of no termination.

When further reflected forward pulses arrive at the diode, as long as the total current through the diode is positive, the load appears to be a short, since the total voltage at the diode is clamped and new pulses cannot change the load voltage. At some point, the total current through the diode drops to zero and the diode shuts off, and the reflected wave must be determined by requiring the total current (the sum of all forward and reverse currents) through the diode to be zero. After that, the diode remains off, and the load appears to be an open circuit.

(Time $t = 3T$) For the example circuit, the second forward wave has amplitude $v_2^+ = \Gamma_s v_1^- \simeq -1.022\,\text{V}$. Since the voltage is clamped to 5.7 V, the reflection cannot change the voltage, so $v_2^- = -v_2^+$. To check and make sure the diode is still on, we need to look at the current through the diode, which at the time of the second reflection is $(v_1^+ - v_1^- + v_2^+ - v_2^-)/Z_o \simeq 0.59/Z_o$. Since this is positive, the diode is still on.

(Time $t = 5T$) At the time of the third reflection from the load, the current would go negative if we considered the diode to still be on. So, the diode turns off, and the total load current $(v_1^+ - v_1^- + v_2^+ - v_2^- + v_3^+ - v_3^-)/Z_o$ must be zero. Solving for v_3^- gives a value of $-.0926$. After this time, the pulse bounces between the source and open load. In the steady state, no current flows, and the load voltage approaches 5 V.

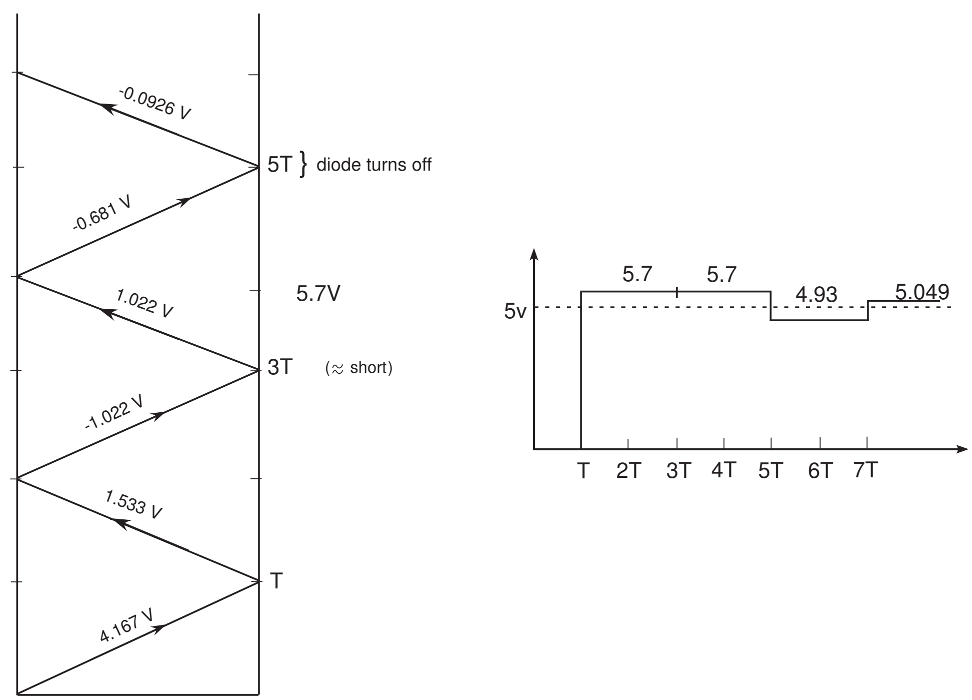


Figure 1.40: (a) Bounce diagram and load voltage plot (b) for diode termination.

The bottom diode will never let the voltage at the load become smaller then $-v_d$. The diodes also protect the other circuit elements from damage due to static discharge.

1.11 Cross-talk

Consider two coupled transmission lines. One line (A) is active, and the other line is quiet (Q). The lines share a common ground. The lines are coupled by mutual capacitance per unit length C_m (F/m) and mutual inductance per unit length L_m (H/m). The goal is to determine the signal induced on the quiet line when the active line is driven by a source.

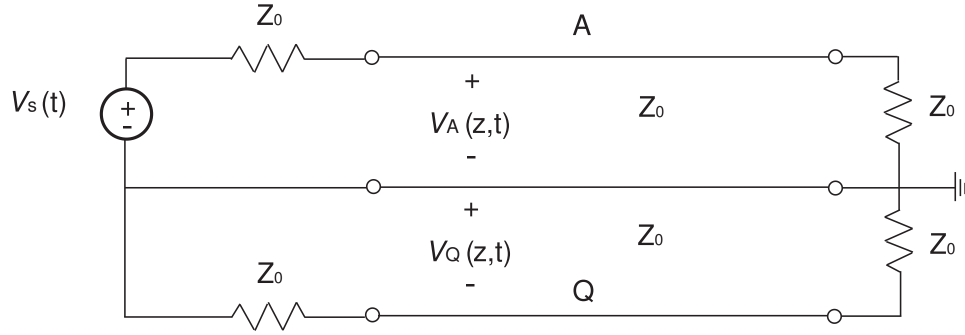


Figure 1.41: Coupled transmission lines. The lines share a common ground.

Capacitative Coupling

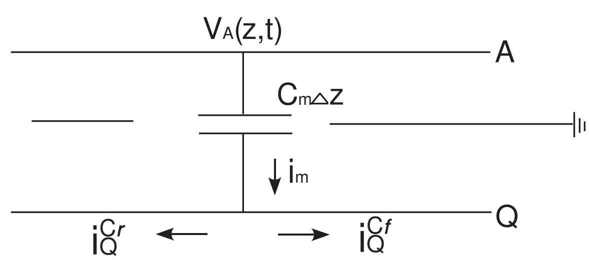


Figure 1.42: Capacitative coupling.

As with the distributed capacitance associated with a single transmission line, the distributed mutual inductance can also be approximated by a lumped element section as shown in in Fig. 1.42. The current and voltage satisfy

$$\Delta i_m(z,t) = C_m \Delta z \frac{\partial v_A(z,t)}{\partial t} \tag{1.136}$$

The current induced on the quiet line splits in half going in each direction, so that

$$\Delta i_Q^{Cr}(z,t) = \Delta i_Q^{Cf}(z,t) = \frac{C_m \Delta z}{2} \frac{\partial v_A(z,t)}{\partial t} \tag{1.137}$$

The Δ means that this is only the contribution from coupling of one small section of the line, and we will later add up the contributions all along the line to get the total induced voltage. The corresponding forward and reverse voltage waveforms are

$$\Delta v_Q^{Cr}(z,t) = \Delta v_Q^{Cf}(z,t) = \frac{Z_o C_m \Delta z}{2} \frac{\partial v_A(z,t)}{\partial t} \tag{1.138}$$

Inductive Coupling

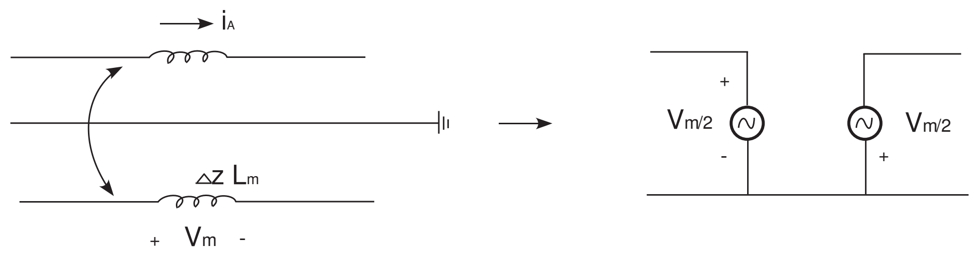


Figure 1.43: Inductive coupling.

For the mutual inductance in Fig. 1.43, a similar analysis leads to

$$\Delta v_Q^{Lr}(z,t) = -\Delta v_Q^{Lf}(z,t) = \frac{L_m \Delta z}{2Z_o}\frac{\partial v_A(z,t)}{\partial t} \tag{1.139}$$

Total Coupling

The combined inductive and capacitative forward and reverse voltage contributions from a section of the line are

$$\Delta v_Q^f = \Delta v_Q^{Cf} + \Delta v_Q^{Lf} = \underbrace{\left(\frac{Z_o C_m}{2} - \frac{L_m}{2Z_o}\right)}_{K_f} \Delta z \frac{\partial v_A(z,t)}{\partial t} \tag{1.140}$$

$$\Delta v_Q^r = \Delta v_Q^{Cr} + \Delta v_Q^{Lr} = \underbrace{\left(\frac{Z_o C_m}{2} + \frac{L_m}{2Z_o}\right)}_{K_r} \Delta z \frac{\partial v_A(z,t)}{\partial t} \tag{1.141}$$

The constants K_f and K_r represent the strength of the coupling between the two lines. Notice that K_r is always positive, whereas K_f can be positive, negative, or zero.

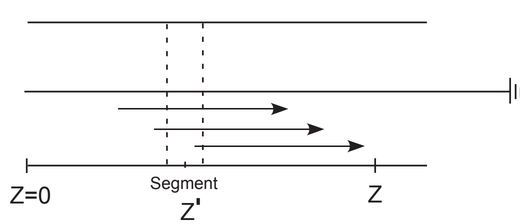


Figure 1.44: Summing up the contributions $\Delta v_Q^f(z',t)$ at each z' between 0 and z to get the total forward wave at z.

Now we want to add up the contributions from the coupling due to each short section of the lines by integrating. For the forward wave, the voltage at a point z on the quiet line is due to the forward coupled wave at all points $z' \leq z$ (see Fig. 1.44). The voltage on the active line at z' is $v_A(z',t) = v_A(t - z'/u)$. Using Eq. (1.140), this induces a voltage at z' of

$$\Delta v_Q^f(z',t) = K_f \Delta z' \frac{\partial v_A(t - z'/u)}{\partial t} \tag{1.142}$$

This voltage signal takes a time $t_o = (z - z')/u$ to arrive at the point z, so that

$$\begin{aligned} \Delta v_Q^f(z,t) &= K_f \Delta z' \frac{\partial v_A(t - (z - z')/u - z'/u)}{\partial t} \\ &= K_f \Delta z' \frac{\partial v_A(t - z/u)}{\partial t} \end{aligned} \tag{1.143}$$

We now integrate the contributions from all z' between the left end of the line and the point z:

$$\begin{aligned} v_Q^f(z,t) &= \lim_{\Delta z' \to 0} \sum K_f \Delta z' \frac{\partial v_A(t - z/u)}{\partial t} \\ &= K_f \int_0^z \frac{\partial v_A(t - z/u)}{\partial t}\, dz' \\ &= K_f z \frac{\partial v_A(t - z/u)}{\partial t} \end{aligned} \tag{1.144}$$

Does this result make sense? Let's look at each term:

K_f: As the coupling capacitance and inductance grow, K_f increases, and the induced voltage also increases, as expected.

z: The farther we go down the line, the bigger z is, and the bigger the induced voltage. In other words, one mile of parallel quiet line picks up a lot more voltage than one inch.

v_A: Because of the time derivative, the induced voltage doesn't depend on the DC voltage on the active line, but on how fast the active signal changes. This makes sense, because it is the transients that couple energy to the quiet line, not the DC voltage.

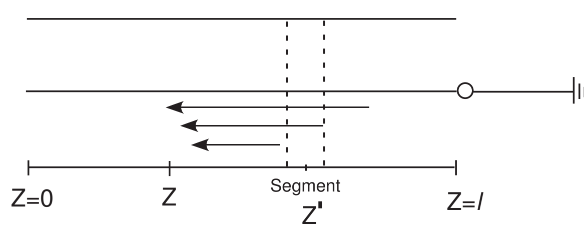


Figure 1.45: Summing up the contributions $\Delta v_Q^r(z',t)$ at each z' between z and ℓ to get the total reverse wave at z.

Now we will follow a similar procedure for the reverse wave. To find the total induced reverse wave at a point z on the quiet line, we have to add up the contributions for all $z' > z$. The induced reverse wave contribution at z' is

$$\Delta v_Q^r(z',t) = K_r \Delta z' \frac{\partial v_A(t - z'/u)}{\partial t} \tag{1.145}$$

As this contribution propagates from z' to z, the time shift is $t_o = (z' - z)/u$, and after that time shift the signal at z is

$$\begin{aligned} \Delta v_Q^r(z,t) &= K_r \Delta z' \frac{\partial v_A(t - (z' - z)/u - z'/u)}{\partial t} \\ &= K_r \Delta z' \frac{\partial v_A(t + z/u - 2z'/u)}{\partial t} \end{aligned} \tag{1.146}$$

Now we integrate the contributions from z to ℓ:

$$v_Q^r(z,t) = K_r \int_z^\ell \frac{\partial v_A(t + z/u - 2z'/u)}{\partial t}\, dz' \tag{1.147}$$

Using the chain rule twice and combining the results,

$$\begin{aligned}
\frac{\partial}{\partial t} v_A(t + z/u - 2z'/u) &= v_A'(t + z/u - 2z'/u) \\
\frac{\partial}{\partial z'} v_A(t + z/u - 2z'/u) &= v_A'(t + z/u - 2z'/u)(-2/u) \\
\frac{\partial}{\partial t} v_A(t + z/u - 2z'/u) &= (-u/2)\frac{\partial}{\partial z'} v_A(t + z/u - 2z'/u)
\end{aligned}$$

Now, we can work the integral in Eq. (1.147) using the fundamental theorem of calculus:

$$\begin{aligned}
v_Q^r(z,t) &= K_r \int_z^\ell (-u/2)\frac{\partial}{\partial z'} v_A(t + z/u - 2z'/u)\, dz' \\
&= (-K_r u/2)\; v_A(t + z/u - 2z'/u)\Big|_z^\ell \\
&= (-K_r u/2)\,[v_A(t + z/u - 2\ell/u) - v_A(t + z/u - 2z/u)] \\
&= (K_r u/2)\,[v_A(t - z/u) - v_A(t + z/u - 2\ell/u)]
\end{aligned} \tag{1.148}$$

Equations (1.144) and (1.148) allow us to find the induced voltage on the quiet line in terms of the voltage on the forward line.

Example

Consider a pair of coupled transmission lines with length $\ell = 1$, coupling constants $K_f = -0.02\,\text{ns/m}$, $K_r = 0.15\,\text{ns/m}$, and phase velocity $u = 0.2\,\text{m/ns}$. We want to plot the forward and reverse voltages at $z = 0$ and $z = \ell$ as functions of time. The active line is driven with the signal shown in Fig. 1.46

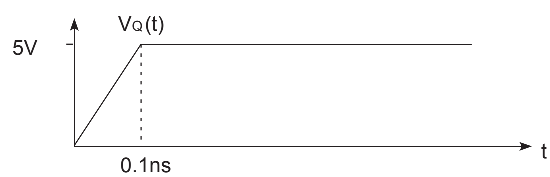


Figure 1.46: Voltage signal on active line.

Forward wave. At $z = 0$, we have $v_Q^f(0,t) = 0$ because of the factor of z in Eq. (1.144). At $z = \ell$,

$$v_Q^f(\ell,t) = (-0.02\,\text{ns/m})(1\,\text{m})\frac{\partial}{\partial t} v_A(t - 5\,\text{ns}) \tag{1.149}$$

The derivative of the active voltage is zero except during the rise time of the pulse, when the derivative is equal to $50\,\text{V/ns}$. The forward voltage on the quiet line at $z = \ell$ is then

$$v_Q^f(\ell,t) = \begin{cases} (-.02\,\text{ns})(50\,\text{V/ns}) = -1\,\text{V} & 5\,\text{ns} \le t \le 5.1\,\text{ns} \\ 0 & \text{otherwise} \end{cases} \tag{1.150}$$

as shown in Fig. 1.47.

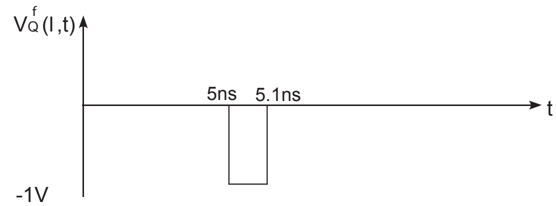


Figure 1.47: Induced forward wave on quiet line at $z = \ell$.

Reverse wave. At $z = 0$, we have

$$v_Q^r(0, t) = 0.015 \, [v_A(t) - v_A(t - 10\,\text{ns})] \tag{1.151}$$

This waveform is shown in Fig. 1.48. At $z = \ell$, the two terms in the reverse wave cancel, so $v_Q^r(\ell, t) = 0$.

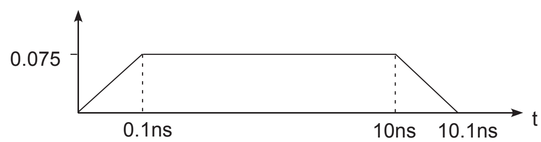


Figure 1.48: Induced reverse wave on quiet line at $z = 0$.

1.12 Review

Transmission line equations

- Lumped element equivalent circuit
- Telegrapher equations
- Wave equation
- Voltage and current solutions, phase velocity, characteristic impedance

Sinusoidal steady state (time-harmonic)

- Phasor notation
- Time-harmonic wave equation
- Voltage solution - forward and reverse waves
- Propagation constant, wavenumber
- Current solution and characteristic impedance
- Reflection coefficient
- Generalized reflection coefficient
- Standing waves, VSWR
- Input impedance
- Matched load, open circuit load, short circuit load, half-integer line, quarter-wave line
- Instantaneous power, complex power, time-average power
- Smith chart
  * Reflection coefficient plane
  * Impedance circles
  * Landmarks: unit circle, real axis, upper/lower half planes, center, short, open
  * Standing waves - VSWR, voltage maxima and minima
  * Admittance
- Matching - shunt or series single stub

Transients

- Voltage solution - forward and reverse waves
- Phase velocity
- Current solution and characteristic impedance
- Reflection coefficient
- Steady state voltage
- Bounce diagrams

Digital signaling

- Matching terminations - load, source, diode clamp

Transmission Line Fundamentals

Sinusoidal steady state

Phasor: $v(z,t) = \text{Re}\left\{\tilde{V}(z)e^{j\omega t}\right\}$

Lossless time-harmonic wave equation: $\dfrac{d^2\tilde{V}(z)}{dz^2} + \beta^2\tilde{V}(z) = 0$, $\beta^2 = \omega^2 LC$

Voltage solution: $\tilde{V}(z) = V_o^+ e^{-j\beta z} + V_o^- e^{j\beta z}$

Current solution: $\tilde{I}(z) = \frac{V_o^+ e^{-j\beta z}}{Z_o} - \frac{V_o^- e^{j\beta z}}{Z_o}$

Reflections: Current and voltage boundary conditions at end of line, $\Gamma_L = \dfrac{V_o^-}{V_o^+} = \dfrac{Z_L - Z_o}{Z_L + Z_o}$

Generalized reflection coefficient: $\Gamma(z) = \dfrac{V_0^- e^{j\beta z}}{V_0^+ e^{-j\beta z}} = \Gamma_L e^{j2\beta z}$

$$\tilde{V}(z) = V_o^+ e^{-j\beta z}[1 + \Gamma(z)]$$

Standing wave pattern: $v(z,t) = |\tilde{V}(z)| \cos[\omega t + \phi(z)]$

Input impedance: $Z_{\text{in}}(z) = \dfrac{\tilde{V}(z)}{\tilde{I}(z)} = Z_o \dfrac{1 + \Gamma(z)}{1 - \Gamma(z)}$

$$Z_{\text{in}}(-\ell) = Z_o \frac{Z_L + jZ_o \tan\beta\ell}{Z_o + jZ_L \tan\beta\ell}$$

Power: $p_{\text{av}} = \frac{1}{2}\text{Re}\left\{\tilde{V}\tilde{I}^*\right\}$

Smith chart: Complex $\Gamma(z)$ plane together with impedance circles from $\dfrac{Z_{\text{in}}(z)}{Z_o} = \dfrac{1 + \Gamma(z)}{1 - \Gamma(z)}$

Shunt stub matching: Find z so that $Y_{\text{in}}(z) = Y_o + jB$, add parallel reactance $-jB$ at z so that the parallel combination is Y_o (matched)

Series stub matching: Find z so that $Z_{\text{in}}(z) = Z_0 + jX$, add series reactance $-jX$ at z so that the series combination is Z_o (matched)

Transients

Lossless wave equation: $\dfrac{\partial^2 v(z,t)}{\partial z^2} = LC\dfrac{\partial^2 v(z,t)}{\partial t^2}$

Voltage solution: $v(z,t) = v^+(z - ut) + v^-(z + ut)$, $u = 1/\sqrt{LC}$

Current solution: $i(z,t) = v^+(z - ut)/Z_o - v^-(z + ut)/Z_o$, $Z_o = \sqrt{L/C}$

Reflections: Current and voltage boundary conditions at end of line, $\Gamma_L = \dfrac{v^-}{v^+} = \dfrac{R_L - Z_o}{R_L + Z_o}$

Chapter 2

Electrostatics

2.1 Vectors

The laws of electromagnetics were originally formulated using a system of many partial differential equations. Today, we use a more compact notation that is much more convenient. But in order to be able to use the notation, one must first understand it.

The fundamental definition of electric field is in terms of force on a test charge. That force has a magnitude and direction. We represent the ratio of that force to the strength of the charge mathematically as a vector:

$$\overline{E} = E_x\hat{x} + E_y\hat{y} + E_z\hat{z} \tag{2.1}$$

where $\hat{x}$ is a unit vector of length one in the $+x$ direction and the other two unit vectors are defined similarly. E_x, E_y, and E_z are real or complex numbers called the components of $\overline{E}$. We can also express the components of the vector using other sets of linearly independent unit vectors. The magnitude of the vector is given by the same character without an accent:

$$E = \|\overline{E}\| = \sqrt{|E_x|^2 + |E_y|^2 + |E_z|^2} \tag{2.2}$$

Vector field. A vector field assigns a vector to each point in space, so the components of the vector are functions of position:

$$\overline{E}(x, y, z) = E_x(x, y, z)\hat{x} + E_y(x, y, z)\hat{y} + E_z(x, y, z)\hat{z} \tag{2.3}$$

The components may also depend on other independent variables such as time or frequency.

Examples

Scalar fields: temperature $T(x, y, z)$, pressure $p(x, y, z)$, electric potential $V(x, y, z)$.

Vector fields: wind velocity $\overline{v}(x, y, z)$, electric field intensity $\overline{E}(x, y, z)$, magnetic field intensity $\overline{H}(x, y, z)$.

Unit vectors. Unit vectors have length one, so that $\|\hat{x}\| = \|\hat{y}\| = \|\hat{z}\| = 1$. We can also come up with a unit vector in the direction of an arbitrary vector $\overline{A}$ using

$$\hat{a} = \frac{\overline{A}}{\|\overline{A}\|} \tag{2.4}$$

Position vector. The position vector is defined by

$$\overline{r} = x\hat{x} + y\hat{y} + z\hat{z} \tag{2.5}$$

This is not really a vector, but is merely a compact way to represent the point (x, y, z).

Dot product. Two vectors can be combined to form a scalar:

$$\overline{A} \cdot \overline{B} = A_x B_x + A_y B_y + A_z B_z \tag{2.6}$$
$$= AB \cos\psi \tag{2.7}$$

where ψ is the angle between the two vectors. If $\overline{A} \cdot \overline{B} = 0$, the vectors are orthogonal. Also, $\overline{A} \cdot \overline{A} = \|A\|^2$.

Cross product. Two vectors can also be combined to form another vector:

$$\overline{A} \times \overline{B} = \hat{n}AB \sin\psi \tag{2.8}$$

where $\hat{n}$ is a unit vector in the direction given by the right hand rule applied to the vectors $\overline{A}$ and $\overline{B}$ and ψ is the angle between the vectors. If we switch the order of $\overline{A}$ and $\overline{B}$, the cross product changes sign. For the rectangular unit vectors,

$$\begin{array}{lll} \hat{x} \times \hat{x} = 0, & \hat{x} \times \hat{y} = \hat{z}, & \hat{x} \times \hat{z} = -\hat{y} \\ \hat{y} \times \hat{x} = -\hat{z}, & \hat{y} \times \hat{y} = 0, & \hat{y} \times \hat{z} = \hat{x} \\ \hat{z} \times \hat{x} = \hat{y}, & \hat{z} \times \hat{y} = -\hat{x}, & \hat{z} \times \hat{z} = 0 \end{array} \tag{2.9}$$

These relationships can be used to express the cross product using components as

$$\overline{A} \times \overline{B} = (A_y B_z - A_z B_y)\hat{x} + (A_z B_x - A_x B_z)\hat{y} + (A_x B_y - A_y B_x)\hat{z} \tag{2.10}$$

Another handy rule for computing the cross product is

$$\overline{A} \times \overline{B} = \begin{vmatrix} \hat{x} & \hat{y} & \hat{z} \\ A_x & A_y & A_z \\ B_x & B_y & B_z \end{vmatrix} \tag{2.11}$$

where the vertical bars denote the matrix determinant operation. An identity that connects the dot and cross products is

$$\overline{A} \cdot (\overline{B} \times \overline{C}) = \overline{B} \cdot (\overline{C} \times \overline{A}) = \overline{C} \cdot (\overline{A} \times \overline{B}) \tag{2.12}$$

Integrals

Vector fields can be integrated over paths and surfaces. Paths and surfaces can be represented using parameterizations. A path is defined by three functions such that the point

$$\overline{r}(t) = (f(t), g(t), h(t)), \quad a \le t \le b \tag{2.13}$$

traces out the path as the parameter t ranges from a to b. A surface is parameterized by functions of two parameters:

$$\overline{r}(s,t) = (f(s,t), g(s,t), h(s,t)), \quad a \le s \le b, c \le t \le d \tag{2.14}$$

Path integrals. A path integral of a vector field is written as

$$\int_P \overline{A} \cdot d\overline{\ell} \tag{2.15}$$

where P represents a path and $d\overline{\ell}$ is a vector tangent to the path with a differential length. Using a parameterization to change the integration variable from a point in the x, y, z plane to the parameter of the path, a path integral can be transformed into a standard scalar integral:

$$\begin{aligned} \int_P \overline{A} \cdot d\overline{\ell} &= \int_a^b \left[A_x\hat{x} + A_y\hat{y} + A_z\hat{z}\right]_{(x,y,z)=(f(t),g(t),h(t))} \cdot d\left[f(t)\hat{x} + g(t)\hat{y} + h(t)\hat{z}\right] \\ &= \int_a^b \left[A_x(t)f'(t) + A_y(t)g'(t) + A_z(t)h'(t)\right] \, dt \end{aligned} \tag{2.16}$$

Example: To integrate $\overline{A} = x\hat{x} + y\hat{y}$ over a straight line from $(0,0)$ to $(1,0)$, we parameterize the path using $(x,y) = (t,0)$, $0 \le t \le 1$. The integral is

$$\begin{aligned} \int_P \overline{A} \cdot d\overline{\ell} &= \int_0^1 (t\hat{x} + 0\hat{y}) \cdot \hat{x} \, dt \\ &= \int_0^1 t \, dt \\ &= \frac{1}{2} \end{aligned}$$

Surface integrals. A surface integral is written as

$$\int_S \overline{A} \cdot d\overline{S} \tag{2.17}$$

where S represents a surface and $d\overline{S}$ is a normal differential area element vector. A surface integral can be evaluated using a parameterization as with a path integral. In many cases, however, the integration surface is simple enough that we can write down $d\overline{S}$ by inspection using

$$d\overline{S} = \hat{n} \, dS \tag{2.18}$$

where $\hat{n}$ is a unit vector normal to the surface and dS is a differential area element. If a vector field represents flow, then the surface integral represents the total amount of flow through the surface.

Example: We want to integrate $\overline{A} = 3(z+1)\hat{z}$ over a square with corners $(0,0,0)$, $(1,0,0)$ $(1,1,0)$, and $(0,1,0)$. Because the surface is confined to the $z = 0$ plane, the differential area element is $dx\,dy$ and the surface normal vector is $\hat{z}$. The integral is

$$\begin{aligned} \int_S \overline{A} \cdot d\overline{S} &= \int_0^1 \int_0^1 3\hat{z} \cdot \hat{z}\, dx\, dy \\ &= \int_0^1 \int_0^1 3\, dx\, dy \\ &= 3 \end{aligned}$$

Derivatives

Because vector fields can change with position, we can measure the amount of change using derivatives of vector fields, much like a scalar derivative gives the rate of change of a scalar function. Vector derivatives are derived in terms of the operator

$$\nabla = \hat{x}\frac{\partial}{\partial x} + \hat{y}\frac{\partial}{\partial y} + \hat{z}\frac{\partial}{\partial z} \tag{2.19}$$

With this mathematical symbol, known as the "del" operator, three first order derivative operations involving vectors and one second order derivative operator can be defined by combining the del operator with vectors using the dot and cross products.

Gradient. The gradient operation transforms a scalar to a vector:

$$\nabla f(x,y,z) = \hat{x}\frac{\partial f(x,y,z)}{\partial x} + \hat{y}\frac{\partial f(x,y,z)}{\partial y} + \hat{z}\frac{\partial f(x,y,z)}{\partial z} \tag{2.20}$$

This vector points in the direction of most rapid increase of the function $f(x,y,z)$.

Example: If $f(x,y) = x^2 + y^2$, then $\nabla f = 2x\hat{x} + 2y\hat{y}$.

Curl. The curl operation transforms a vector to a vector:

$$\nabla \times \overline{A} = \hat{x}\left(\frac{\partial A_z}{\partial y} - \frac{\partial A_y}{\partial z}\right) + \hat{y}\left(\frac{\partial A_x}{\partial z} - \frac{\partial A_z}{\partial x}\right) + \hat{z}\left(\frac{\partial A_y}{\partial x} - \frac{\partial A_x}{\partial y}\right) \tag{2.21}$$

This vector gives the amount of rotation of the vector field $\overline{A}$.

Example: $\nabla \times (y\hat{x} - x\hat{y}) = \hat{z}(-1-1) = -2\hat{z}$.

Divergence. The divergence operation transforms a vector to a scalar:

$$\nabla \cdot \overline{A} = \frac{\partial A_x}{\partial x} + \frac{\partial A_y}{\partial y} + \frac{\partial A_z}{\partial z} \tag{2.22}$$

The value of this scalar is positive at sources of the vector field and negative at sinks.

Example: If $\nabla \cdot (x\hat{x} + y\hat{y}) = 1 + 1 = 2$.

Laplacian. The Laplacian transforms a scalar to a scalar or a vector to a vector:

$$\nabla^2 = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2} \tag{2.23}$$

When this operator acts on a vector field, it does not mix the components of the vector field. The Laplacian of a vector field is a combination of the Laplacian operator applied to each component: $\nabla^2\overline{A} = (\nabla^2 A_x)\hat{x} + (\nabla^2 A_y)\hat{y} + (\nabla^2 A_z)\hat{z}$.

Some important identities are

$$
\begin{aligned}
\nabla \times (\nabla V) &= 0 && (2.24)\\
\nabla \cdot (\nabla \times \overline{A}) &= 0 && (2.25)\\
\nabla \cdot (\nabla V) &= \nabla^2 V && (2.26)\\
-\nabla \times (\nabla \times \overline{A}) + \nabla(\nabla \cdot \overline{A}) &= \nabla^2 \overline{A} && (2.27)
\end{aligned}
$$

2.2 Differential Forms

With vector analysis, there are some operations such as the curl derivative that are difficult to understand physically. We will introduce a notation called the calculus of differential forms that is very similar to vector analysis, but gives us an intuitive way to visualize these operations.

Since it is easy to convert a vector to a differential form and vice versa, we can use either notation to solve a given problem. Mathematically, working with differential forms is very similar to working with vectors, but the pictures that we can draw for fields when we express them as differential forms are quite different and in many cases provide a better way to visualize otherwise complicated phenomena. So, we will use mostly vectors, and convert to differential forms in places where they help in illuminating a difficult point.

2.2.1 What is a Differential Form?

A differential form is a quantity that can be integrated, including the differentials. In the integral below, $3x\,dx$ is a differential form:

$$\int_a^b \underbrace{3x\,dx}_{\text{one-form}}$$

This differential form has degree one because it is integrated over a 1-dimensional region, or path. We call a differential form of degree one a one-form. Differential forms can be added together, with the differentials being linearly independent. The sum of dx and dy, for example, is the one-form $dx + dy$. The sum of $3x\,dx$ and $2\,dx$ is $(3x + 2)\,dx$ since we can combine like differentials.

Two-forms are integrated by double integrals over surfaces. For example, $zx^2 dx \wedge dy$ is a two-form. Two-forms under integral signs are written $zx^2 dxdy$, without the wedge. We want to take these differential forms out from under the integral signs so that we can combine them and take derivatives of them, much like we do with vectors. This requires a special rule for combining forms, called the *wedge* or *exterior product* and represented by the symbol $\wedge$, that allows us to think of two-forms as a combination of one-forms.

Table 2.1 shows differential forms of various degrees. Zero-forms and three-forms correspond to scalars or functions. One-forms and two-forms correspond to vectors. It should be obvious why the one-form and vector given in the table go together. The relationship between the two-form and its associated vector will become clear below when we show how to draw pictures of two-forms. Functions are "integrated" by evaluating them at a point, and a point is zero-dimensional, so we can call functions zero-forms.

Degree	Region of Integration	General Expression	Vector/Scalar
zero-form	"Point"	$f(x, y, z)$	$f(x, y, z)$
one-form	Path	$A_x dx + A_y dy + A_z dz$	$A_x\hat{x} + A_y\hat{y} + A_z\hat{z}$
two-form	Surface	$A_x dy \wedge dz + A_y dz \wedge dx + A_z dx \wedge dy$	$A_x\hat{x} + A_y\hat{y} + A_z\hat{z}$
three-form	Volume	$\rho(x, y, z)\, dx \wedge dy \wedge dz$	$\rho(x, y, z)$

Table 2.1: Differential forms.

We will now discuss each type of form separately, and show how to draw pictures for them. Drawing a picture of a vector is easy: a line with an arrow. But there is only one type of picture for vector analysis. With differential forms, there are three pictures. This is one of the things that makes differential forms helpful in electromagnetic theory.

2.2.2 One-forms

A one-form is drawn as surfaces. The one-form dx has surfaces perpendicular to the x-axis spaces one unit apart, as shown in Fig. 2.1(a). $5\,dx$ also has surfaces perpendicular to the x-axis but they are spaced more closely: five per unit distance. This is because of the way one-forms are integrated. If we integrate dx along a path from the point $(.5, 0, 0)$ to the point $(1.5, 0, 0)$, we get one. If we draw the path, it crosses one of the surfaces of dx. With a larger coefficient, the integral is bigger, so the path has to cross more surfaces.

The one-form dy has surfaces perpendicular to the y-axis. The one-form $2\,dz$ has surfaces perpendicular to the z-axis spaced twice as closely as those for dx, as in Fig. 2.1(b).

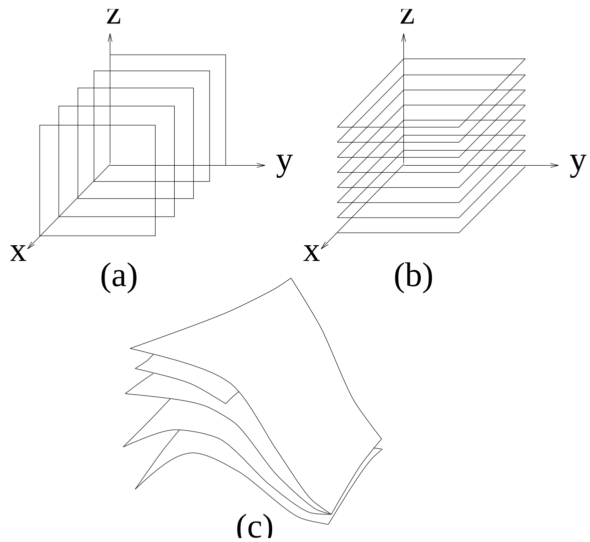


Figure 2.1: (a) The one-form dx. The surfaces of dx are infinite in the z and y directions, and are drawn with edges only for graphical clarity. For the same reason, not all of the surfaces are shown. (b) The one-form $2dz$, with surfaces perpendicular to the z-axis and spaced twice as closely as those of dx. (c) A general one-form, with curved surfaces and surfaces that end or meet each other.

More complicated forms can also be drawn. The one-form $dx + 5dy$ is drawn as slanted surfaces that are perpendicular to the vector $\hat{x} + 5\hat{y}$. The one-form $f dx$ consists of surfaces that are perpendicular to the x-axis but with spacing that gets closer or farther apart depending on the value of the function f. In general, the surfaces of a one-form can twist wildly, end, or meet each other. An example of this is shown in Fig. 2.1(c).

Fig. 2.2 shows how an arbitrary one-form is integrated over a path. The integral of a one-form over a path is the number of surfaces of the one-form pierced by the path. Since we integrate along the path in a particular direction, we have to keep track of the *orientation* of each surface. The orientation of a form is determined by the sign of its coefficients. If a path crosses a surface of the one-form dx in the $+x$ direction, for example, that contributes a positive value to the integral. If the path crosses the surface in the $-x$ direction, that contributes negatively.

In rectangular coordinates, to convert one-forms into vectors and vectors into one-forms, we interchange basis one-forms with basis vectors as below:

$$dx \leftrightarrow \hat{x}, \quad dy \leftrightarrow \hat{y}, \quad dz \leftrightarrow \hat{z}. \tag{2.28}$$

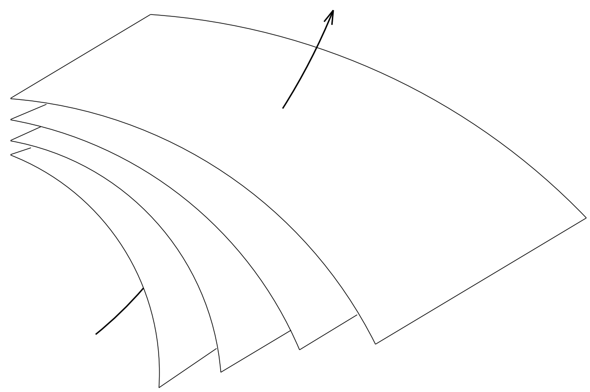


Figure 2.2: Integrating a one-form over a path graphically. Since the path crosses four surfaces, the value of the integral is four.

The vector corresponding to a one-form is sometimes called the one-form's *dual* vector.

2.2.3 Two-forms

A two-form is drawn as two sets of surfaces that intersect to form tubes. To draw $dx \wedge dy$, we superimpose the surfaces of dx and the surfaces of dy as in Fig. 2.3(a). This produces tubes that point in the z direction. This explains the correspondence between the two-form and its dual vector given in Table 2.1.

Graphically, understanding the integral of a two-form over a surface is easy. We just count how many tubes pass through it, as in Fig. 2.3(b). Of course, we have to keep track of the orientation of the two-form (the direction of the tubes) and the orientation of the surface it is being integrated over. A surface is oriented by choosing one of the two normal directions. Tubes crossing in the same direction as the orientation make a positive contribution to the value ot the integral; tubes crossing in the negative directions contribute negatively.

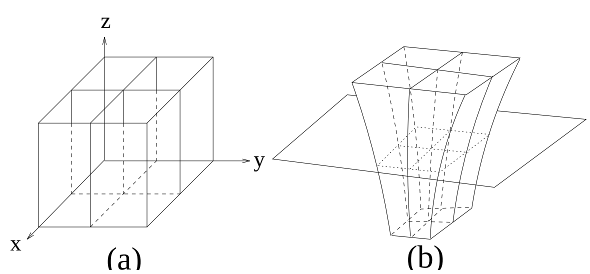


Figure 2.3: (a) The two-form $dx \wedge dy$, with tubes in the z direction. The tubes are infinitely long, but are drawn as finite for clarity. There are also infinitely many tubes in the x and y directions. (b) A two-form is integrated over a surface by counting the number of tubes passing through the surface.

As we noted earlier, between the differentials of a two-form there is a special product, the wedge $\wedge$ or exterior product. Sometimes, the wedge is dropped, so that the two-form $dx\,dy$, for example, is the exterior product of the one-forms dx and dy, or $dx \wedge dy$.

The exterior product is anticommutative, so that switching the order of two differentials changes the sign: $dx \wedge dy = -dy \wedge dx$. One consequence of this property is that $dx \wedge dx = dy \wedge dy = dz \wedge dz = 0$. The anticommutativity of the exterior products allows us to simplify differential forms.

Example 2.1. Exterior product of one-forms.

Let $a = 3dx + dy$ and $b = 2dx + 3dy$. Then

$$\begin{aligned} a \wedge b &= (3dx + dy) \wedge (2dx + 3dy) \\ &= 6dx \wedge dx + 9dx \wedge dy + 2dy \wedge dx + 3dy \wedge dy \\ &= 9dx \wedge dy - 2dx \wedge dy \\ &= 7dx \wedge dy. \end{aligned}$$

This two-form is dual to the cross product $(3\hat{x} + \hat{y}) \times (2\hat{x} + 3\hat{y}) = 7\hat{z}$.

For convenience, we always put differentials of two-forms into the right cyclic orders $dy \wedge dz$, $dz \wedge dx$ and $dx \wedge dy$. Two-forms with differentials in right cyclic order can be converted to vectors by interchanging basis forms and basis vectors as follows:

$$dy \wedge dz \leftrightarrow \hat{x}, \quad dz \wedge dx \leftrightarrow \hat{y}, \quad dx \wedge dy \leftrightarrow \hat{z}. \tag{2.29}$$

There are two types of vectors: those that are dual to one-forms, and those that are dual to two-forms. Usually, vectors dual to two-forms represent flow or flux.

2.2.4 Three-forms

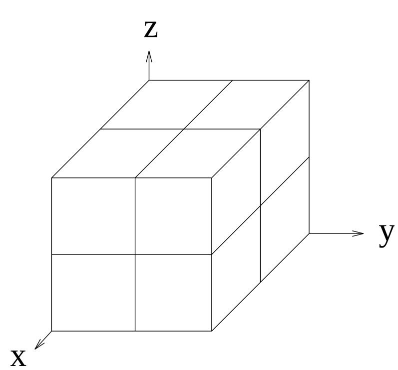


Figure 2.4: The three-form $dx \wedge dy \wedge dz$, with cubes of side equal to one. The cubes fill all space; only a few of them are drawn in the figure.

A three-form has three sets of surfaces that form boxes (Fig. 2.4). The larger the coefficient of the three-form, the smaller and more tightly packed the boxes. The integral of a three-form over a volume is the number of boxes inside the volume, taking into account the sign of the contribution to the integral if the coefficient of the three-form is negative.

We always put the differentials of a three-form in right cyclic order, $dx \wedge dy \wedge dz$. Since any combination of the three differentials dx, dy and dz can be converted to $dx \wedge dy \wedge dz$ using the anticommutativity of the exterior product, any sum of three-forms combines into one term. The coefficient of this term is a scalar. The three-form is dual to this scalar.

A three-form represents a volume density, so the coefficient of a three-form has units of length^{-3}. Some scalars, such as temperature, are not volume densities. The calculus of differential forms lets us keep these two types of quantities separate.

Finally, there are no four-forms, since there are only three differentials, so that any four-form has a repeated differential in each term and so must vanish.

2.3 Integrating Differential Forms

When integrating a vector field over a path or surface, the dot product with a differential vector actually converts the vector field into a differential form. So, it is more natural to integrate a differential form than a vector field.

Consider for example the one-form $\alpha = 2dx + 3xdy$ and a path P which lies along the curve $y = x^2$ from the point $(0, 0)$ to $(1, 1)$. We wish to find

$$\int_P \alpha \tag{2.30}$$

This is done by parameterizing the path P in terms of a new variable t, so that the path becomes $(x = t, y = t^2)$, with t ranging from zero to one. We then substitute these values for x and y into the integral:

$$\begin{aligned}
\int_P \alpha(x, y) &= \int_0^1 \alpha(t, t^2) \\
&= \int_0^1 [2dt + 3td(t^2)] \\
&= \int_0^1 (2 + 6t^2)dt \\
&= 4
\end{aligned}$$

When the symbol d acts on t^2, we use implicit differentiation to obtain $2t\,dt$. Integrating the dual vector $2\hat{x} + 3x\hat{y}$ over the same path gives the same result.

This same approach can be used to evaluate surface integrals of two-forms as well. A parameterization of a surface requires two variables, so that the surface is given by $(x = a(s, t),\ y = b(s, t),\ z = c(s, t))$ where a, b and c are functions of s and t. These functions are substituted into the two-form to be integrated, yielding a two-form with the differentials $ds \wedge dt$, which is then integrated over the appropriate limits in s and t.

2.3.1 Star Operator

The star operator relates zero-forms with three-forms and one-forms with two-forms according to the relationships

$$\star 1 = dx \wedge dy \wedge dz \tag{2.31}$$

and

$$\begin{aligned} \star dx &= dy \wedge dz && (2.32) \\ \star dy &= dz \wedge dx && (2.33) \\ \star dz &= dx \wedge dy && (2.34) \end{aligned}$$

Also, $\star\star = 1$, so that $\star\, dy \wedge dz = dx$, for example. Graphically, for a one-form α, the tubes of the two-form $\star\,\alpha$ are perpendicular to the surfaces of α.

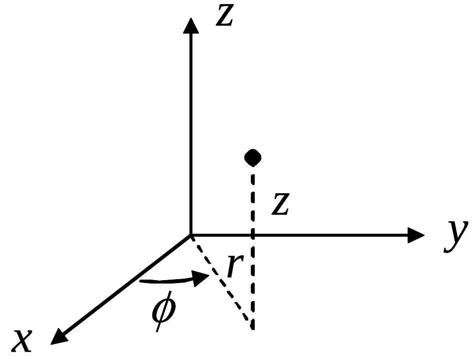
2.3.2 Summary

Differential forms are classified by degree: zero-forms are functions, one-forms are dual to vectors and are drawn as surfaces, two-forms are also dual to vectors but are drawn as tubes, and three-forms are dual to scalars and are drawn as boxes. Differential forms combine using the exterior product to yield differential forms of higher degree. One-forms are integrated over paths, two-forms are integrated over surfaces, and three-forms are integrated over volumes.

2.4 Cylindrical and Spherical Coordinates

This section reviews the cylindrical and spherical coordinate systems, which are convenient when solving problems that involve cylindrical or spherical geometries.

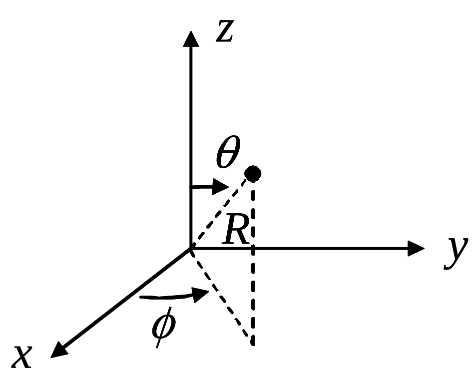
2.4.1 Cylindrical Coordinates



The location of a point is represented by the coordinates (r, ϕ, z)
$r = \sqrt{x^2 + y^2}$
$\phi = \tan^{-1}(y/x)$ = angle from $+x$ axis

Swept Variable	Differential Length	Unit Vector
r	dr	$\hat{r}$
ϕ	$rd\phi$	$\hat{\phi}$
z	dz	$\hat{z}$

2.4.2 Spherical Coordinates



The location of a point is represented by the coordinates (R, θ, ϕ)
$R = \sqrt{x^2 + y^2 + z^2}$
$\phi = \tan^{-1}(y/x)$ = angle from $+x$ axis
$\theta = \cos^{-1}(z/R)$ = angle from $+z$ axis

Swept Variable	Differential Length	Unit Vector
R	dR	$\hat{R}$
θ	$R\,d\theta$	$\hat{\theta}$
ϕ	$R \sin\theta d\phi$	$\hat{\phi}$

2.5 Maxwell's Equations

Electromagnetic behavior can be described using a set of four fundamental relations known as *Maxwell's Equations*. These equations are a mathematical model for observations about electric and magnetic fields discovered by Faraday, Ampére, and others. Maxwell's equations in point form are

$$\nabla \times \overline{E} = -\frac{\partial \overline{B}}{\partial t} \tag{2.35}$$

$$\nabla \times \overline{H} = \overline{J} + \frac{\partial \overline{D}}{\partial t} \tag{2.36}$$

$$\nabla \cdot \overline{D} = \rho_v \tag{2.37}$$

$$\nabla \cdot \overline{B} = 0 \tag{2.38}$$

The first equation is known as Faraday's law, the second is Ampére's law, and the third and fourth are Gauss's laws for the electric and magnetic fields. The field and source quantities are

$\overline{E}$:	Electric field intensity	V/m
$\overline{H}$:	Magnetic field intensity	A/m
$\overline{D}$:	Electric flux density	C/m^2
$\overline{B}$:	Magnetic flux density	Wb/m^2
$\overline{J}$:	Electric current density	A/m^2
ρ_v:	Electric charge density	C/m^3

The first four quantities represent the electromagnetic field. The last two quantities are sources that radiate electromagnetic fields.

In free space (i.e., a vacuum), the electric and magnetic field intensities and flux densities are related by constitutive relations, which are

$$\overline{D} = \epsilon_0 \overline{E} \tag{2.39a}$$

$$\overline{B} = \mu_0 \overline{H} \tag{2.39b}$$

where $\epsilon_0 \simeq 8.854 \times 10^{-12}$ F/m is the permittivity of free space and $\mu_0 = 4\pi \times 10^{-7}$ H/m is the permeability. Air affects the permittivity slightly, but we generally approximate the permittivity of air by the free space value. The effect of dielectrics and magnetic materials on the electromagnetic field can be modeled as a change in the value of ϵ and μ.

Maxwell's equations can also be expressed in integral form:

$$\oint_C \overline{E} \cdot d\overline{\ell} = -\frac{d}{dt} \int_A \overline{B} \cdot d\overline{s} \tag{2.40}$$

$$\oint_C \overline{H} \cdot d\overline{\ell} = \int_A \overline{J} \cdot d\overline{s} + \frac{d}{dt} \int_A \overline{D} \cdot d\overline{s} \tag{2.41}$$

$$\oint_S \overline{D} \cdot d\overline{s} = \int_V \rho_v \, dV = Q \tag{2.42}$$

$$\oint_S \overline{B} \cdot d\overline{s} = 0 \tag{2.43}$$

where S is the closed surface bounding the volume V and C is the closed path bounding the area A. We will show how the integral and point forms of these equations are related a little later.

Suppose that the fields do not change in time (*static fields*). All of the time derivatives go to zero, and so Maxwell's equations become

$$\oint_C \overline{E} \cdot d\bar{\ell} = 0 \quad \text{(Equivalent to KVL)} \tag{2.44}$$

$$\oint_C \overline{H} \cdot d\bar{\ell} = \int_A \overline{J} \cdot d\overline{s} \quad \text{(Equivalent to KCL)} \tag{2.45}$$

$$\oint_S \overline{D} \cdot d\overline{s} = \int_V \rho_v \, dV = Q \tag{2.46}$$

$$\oint_S \overline{B} \cdot d\overline{s} = 0 \tag{2.47}$$

Note that for the case of static fields, the electric and magnetic fields are no longer coupled. Therefore, we can treat them separately. For electric fields, we are dealing with *electrostatics*. For magnetic fields, we are dealing with *magnetostatics*.

2.6 Charge and Current Distributions

We need to remind ourselves about charges. The volume charge density is the amount of charge per unit volume. It can vary with position in the volume. The total charge is the integral of the charge density over the volume V, or

$$Q = \int_V \rho_v \, dV \tag{2.48}$$

where Q is the charge in coulombs (C). When dealing with things like conductors, the charge may be distributed on the *surface* of a material. We therefore are interested in the *surface charge density* ρ_s with units of C/m^2. This is the amount of charge per unit area on the surface. The total charge Q would be the integral of ρ_s over the surface. Finally, we can have a *line charge density* ρ_ℓ with units of C/m which is the amount of charge per unit distance along a line. An example of this might be a very thin wire. The total charge Q would be the integral of ρ_ℓ over the length of the line segment.

$\overline{J}$ is the volume current density, measured in units of A/m^2. It represents the amount of current flowing through a unit surface area. The total current flowing through a surface A is

$$I = \int_A \overline{J} \cdot d\overline{s} \tag{2.49}$$

If the current is confined to the surface of a conducting object, we can define a *surface current density* $\overline{J}_s$ with units of A/m. $\overline{J}_s$ is the amount of current per unit length, where the length represents the "cross-section" of the surface.

2.7 Electric Field Intensity, Electric Flux Density, and Electric Potential

Let's also quickly review electric fields and flux. Coulomb's law states that

1. An isolated charge q induces an electric field $\overline{E}$ at every point in space. At an observation point a distance R from this charge, the electric field is

$$\overline{E} = \hat{R}\frac{q}{4\pi\epsilon R^2} \tag{2.50}$$

where $\hat{R}$ is the unit vector pointing from the charge to the observation point. ϵ is called the *permittivity* of the medium.

2. The force on a charge q' due to an electric field is

$$\overline{F} = q'\overline{E} \tag{2.51}$$

The physical constant in Coulomb's law is the permittivity of the material in which the charges are embedded, and is often given in terms of a relative permittivity with respect to the permittivity of free space:

$\epsilon = \epsilon_0 \epsilon_r$	
$\epsilon_0 \simeq 8.854 \times 10^{-12}$ F/m	permittivity of vacuum (free space)
ϵ_r	relative permittivity or dielectric constant of material

The electric flux density due to an electric field is

$$\overline{D} = \epsilon\overline{E} \tag{2.52}$$

To find the electric potential difference between two points, we integrate the electric field intensity over a path between two points:

$$V_{21} = V_2 - V_1 = -\int_{P_1}^{P_2} \overline{E} \cdot d\bar{\ell} \tag{2.53}$$

For static fields, V_{21} is independent of the path taken. This implies that

$$\oint_C \overline{E} \cdot d\bar{\ell} = 0 \tag{2.54}$$

for a closed path or contour C. This is one of Maxwell's equations for statics, and it is also Kirchhoff's voltage law.

2.8 Electric Properties of Materials

The effect of dielectrics and magnetic materials on electromagnetic fields can be included in Maxwell's equations by changing the value of the permittivity and permeability that appear in the constitutive relationships (2.39). We classify materials based upon their *constitutive parameters*:

ϵ	electrical permittivity	F/m
μ	magnetic permeability	H/m
σ	conductivity	S/m

We will consider the permeability μ later in our discussion of magnetostatics. For now, let's focus on the other two.

Conductors and dielectrics are classified by how well they conduct current. Dielectrics are good insulators, meaning that the electrical resistance is high and can often be neglected. Good conductors on the other hand have low resistance. An ideal dielectric is a perfect insulator, whereas an ideal or perfect electric conductor (PEC) has zero resistance. Semiconductors are somewhere in between, and additionally have the property that they can be doped in such a way that current can be controlled and restricted to a given region.

2.8.1 Conductors

Conductivity relates current density to the electric field which supplies the force to move the charges. This relationship is

$$\overline{J} = \sigma \overline{E} \tag{2.55}$$

which looks a lot like Ohm's law (in fact, Ohm's law is a simplification of this relationships). Given this definition, we see that a perfect insulator has $\sigma = 0$ so that $\overline{J} = 0$.

A perfect conductor, on the other hand, has $\sigma \to \infty$, which implies that to have finite current density we must have $\overline{E} = 0$ since $\overline{E} = \overline{J}/\sigma$. The surface of a perfect conductor is always an equipotential surface, meaning that the electric potential is the same everywhere on the conductor. Perfect conductors are impenetrable to changing electric and magnetic fields. They can support a surface current, but the fields and current do not penetrate the PEC body.

For metals $\sigma \sim 10^7$S/m, which is very large, so one can often set $\overline{E} = 0$ inside a good conductor and to approximate the metal as a PEC. Due to the skin effect, time harmonic electromagnetic fields that oscillate at high frequencies are confined to a thin layer near the surface of a good conductor. The surface current on a PEC object is the idealized limit of the skin effect layer as the conductivity becomes large and the skin depth goes to zero.

It is common in electromagnetic theory to model some current sources as *impressed currents*, meaning that we consider the current to be produced by an external source, and we treat it mathematically as a given source term in Maxwell's equations. In other words, we approximate the current as being independent of the fields we are solving for. In general, we have both induced conduction current and impressed current, so that

$$\begin{aligned} \overline{J} &= \overline{J}_{\text{induced}} + \overline{J}_{\text{impressed}} && (2.56) \\ &= \sigma \overline{E} + \overline{J}_{\text{impressed}} && (2.57) \end{aligned}$$

The impressed current is generally used as a model for external sources of electromagnetic power that are outside the system or region of space that we are modeling. Conduction currents are induced by fields produced by external sources of power.

2.8.2 Dielectrics

To develop a simple model for a dielectric material, we will first assume that the dielectric is perfect, so that $\sigma = 0$ and $\overline{J} = 0$. At the microscopic level, an atom or molecule in the dielectric material consists of a positively charged nucleus and negatively charged electron cloud. Unlike a good conductor, for which the electrons can move freely through the material, the electrons in a dielectric only shift their positions relative to the nuclei as if the center of gravity of the cloud were attached to a spring with a given force constant. In the resting state, the cloud center is coincident with the nucleus center, leading to a neutral charge configuration.

If an external electric field $\overline{E}_{\text{ext}}$ is applied to the material, the center of the electron cloud will be displaced from its equilibrium value. While we still have charge neutrality, we can consider that there will be an electric field emanating from the positively charged nucleus and ending at the negatively charged cloud center. This process of creating electric *dipoles* within the material by applying an electric field is called *polarizing* the material.

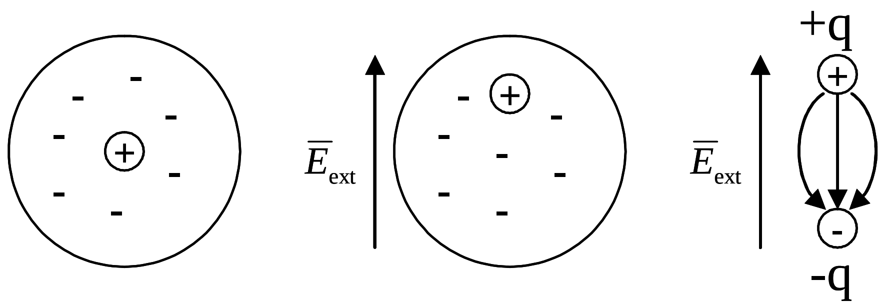


Figure 2.5: Electron cloud displaced by an externally applied electric field.

The induced electric field created by our new dipole is referred to as a *polarization* field, and it is weaker and in the opposite direction to $\overline{E}_{\text{ext}}$. If we think of sliding a dielectric slab in between the plates of a capacitor connected to a voltage source, additional charge must flow from the source onto the capacitor plates to maintain a constant voltage between the plates. The flux density increases to

$$\overline{D} = \epsilon_0 \overline{E} + \overline{P} \tag{2.58}$$

where $\epsilon_0 \simeq 8.854 \times 10^{-12}$ F/m is the permittivity of free space (vacuum) and $\overline{P}$ is referred to as the *polarization vector* in the material. This quantity is proportional to the applied field strength, since the separation of the charges in the dielectric is more pronounced for stronger external fields. We can therefore write

$$\overline{P} = \epsilon_0 \chi_e \overline{E} \tag{2.59}$$

$$\overline{D} = \epsilon_0 \overline{E} + \epsilon_0 \chi_e \overline{E} = \epsilon_0 (1 + \chi_e) \overline{E} = \epsilon \overline{E} \tag{2.60}$$

$$\epsilon = \epsilon_0 \underbrace{(1 + \chi_e)}_{\epsilon_r} \tag{2.61}$$

where ϵ_r is the *relative permittivity* of the dielectric.

Some representative values for relative permittivity are

Material	Relative permittivity ϵ_r
Free space	1
Air	1.006
Polystyrene	2.6
Water	80
Barium titanate	1000 - 10,000

2.9 Gauss's Law and Capacitance

One application of Gauss's law is the analysis of capacitors. The definition of capacitance is

$$C = \frac{Q}{V} \tag{2.62}$$

with units of Farads = Coulombs/Volt, where V is the potential difference between the conductor with charge $+Q$ and the conductor with charge $-Q$. The larger the capacitor, the more charge accumulates on the two conductors that make up the capacitor for a given applied voltage.

To compute the capacitance of a given structure, the basic procedure is to assume a charge Q on the positive conductor, then use Gauss's law to find the electric field and the voltage between the conductors:

1. Assume a charge Q on the conductors
2. Find $\overline{D}$ from Gauss's Law
3. Find $\overline{E} = \overline{D}/\epsilon$.
4. Find V from $\overline{E}$
5. Find the capacitance from $C = Q/V$

2.9.1 Parallel Plate Capacitor

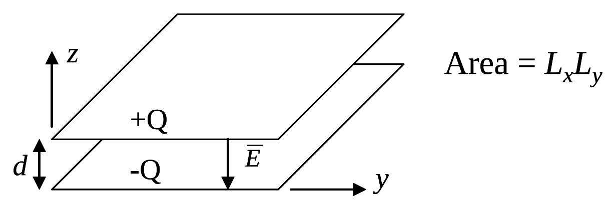


Figure 2.6: Parallel plate capacitor.

1. Assume the charge on the upper plate is Q. We also assume that the charge will evenly distribute itself over the conducting plates, and neglect fringing fields near the edges of the plates.

2. Use Gauss's law to find $\overline{D}$. The Gaussian surface is a cube that contains the top plate. Since we are neglecting fringing fields outside the plates, the only contribution to the surface integral is from the part of the Gaussian surface that is between the plates:

$$\begin{aligned}
\oint \overline{D} \cdot d\overline{s} &= Q \\
\int_0^{L_y} \int_0^{L_x} D_o dx dy &= D_o L_x L_y \\
D_o &= \frac{Q}{L_x L_y} = \frac{Q}{A} \\
\overline{D} &= -\frac{Q}{A}\hat{z}
\end{aligned}$$

3. The electric field intensity is

$$\overline{E} = -\frac{Q}{\epsilon A}\hat{z}$$

4. The potential between the two plates is

$$V = -\int_0^d -\frac{Q}{\epsilon A}\hat{z} \cdot \hat{z} dz = \frac{Qd}{\epsilon A}$$

5. Finally, the capacitance is

$$C = \frac{Q}{V} = \frac{\epsilon A}{d} \tag{2.63}$$

Question: Why does larger ϵ increase the capacitance?

Differential Forms

The electric flux density two-form is $D = -D_o\, dx \wedge dy$, so

$$\begin{aligned} \oint D &= Q \\ \int_0^{L_y} \int_0^{L_x} D_o\, dx \wedge dy &= D_o L_x L_y \\ D_o &= \frac{Q}{L_x L_y} = \frac{Q}{A} \\ D &= -\frac{Q}{A} dx \wedge dy \end{aligned}$$

The electric field intensity one-form is

$$E = \frac{1}{\epsilon} \star D = -\frac{Q}{\epsilon A} dz$$

The potential between the two plates is

$$V = -\int_0^d -\frac{Q}{\epsilon A} dz = \frac{Qd}{\epsilon A}$$

and we obtain the same result, $C = \epsilon A/d$, for the capacitance.

2.9.2 Energy

Electromagnetic fields store energy. For electric fields, the energy density is

$$w_e = \frac{1}{2}\overline{E} \cdot \overline{D} \tag{2.64}$$

which has units of Joules/m^3. With differential forms, this becomes a three-form, $(1/2)E \wedge D$. The total energy in a volume V is

$$W_e = \frac{1}{2}\int_V E \wedge D \tag{2.65}$$

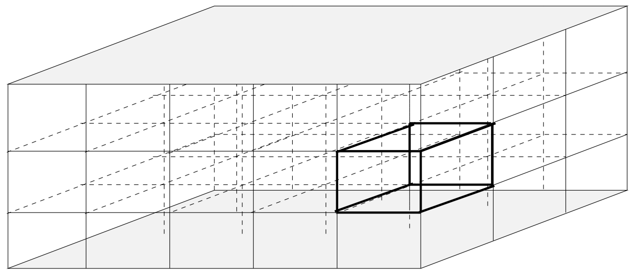


Figure 2.7: Tubes of the two-form D intersect with surfaces of the one-form E to produce boxes, representing energy stored in the capacitor.

For the parallel plate capacitor,

$$\begin{aligned} W_e &= \frac{1}{2}\int_0^d \int_0^{L_y} \int_0^{L_x} \frac{Q^2}{\epsilon A^2} dx \wedge dy \wedge dz \\ &= \frac{1}{2}\frac{Q^2}{\epsilon A^2} Ad \\ &= \frac{1}{2}\frac{C^2V^2d}{\epsilon A} \\ &= \frac{1}{2}CV^2 \end{aligned} \tag{2.66}$$

2.9.3 Spherical Shell Capacitor

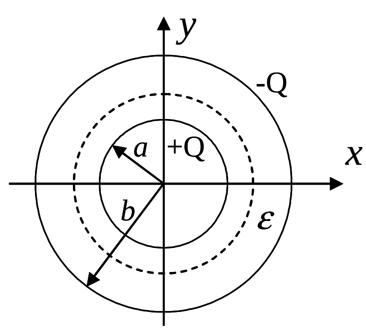


Figure 2.8: Spherical shell capacitor geometry. Solid lines represent conducting spheres of radius a and b.

The charge will evenly distribute itself over the conducting spherical shells.

$$\text{Gauss's law:} \quad \oint \overline{D} \cdot d\overline{s} = Q \tag{2.67}$$

$$\int_0^{2\pi} \int_0^{\pi} D_o R^2 \sin\theta \, d\theta d\phi = Q \tag{2.68}$$

$$4\pi R^2 D_o = Q \tag{2.69}$$

$$D_o = \frac{Q}{4\pi R^2} \tag{2.70}$$

$$\text{Flux density:} \quad \overline{D} = \frac{Q}{4\pi R^2}\hat{R} \tag{2.71}$$

$$\text{Field intensity:} \quad \overline{E} = \frac{Q}{4\pi\epsilon R^2}\hat{R} \tag{2.72}$$

$$\text{Potential:} \quad V_{ab} = V_a - V_b = -\int_b^a \frac{Q}{4\pi\epsilon R^2}\hat{R} \cdot \hat{R} dR \tag{2.73}$$

$$= \left.\frac{Q}{4\pi\epsilon R}\right|_b^a = \frac{Q}{4\pi\epsilon}\left(\frac{1}{a} - \frac{1}{b}\right) \tag{2.74}$$

$$\text{Capacitance:} \quad C = \frac{Q}{V_{ab}} = \frac{4\pi\epsilon}{\frac{1}{a} - \frac{1}{b}} \tag{2.75}$$

2.9.4 Electric Field Boundary Conditions

At a boundary between two different dielectrics, the tangential components of $\overline{E}$ have to be equal on either side of the interface:

$$E_{1t} = E_{2t} \tag{2.76}$$

If this were not the case, then the closed path integral of $\overline{E}$ would not be zero around a small closed path at the boundary. Similarly, the normal components of $\overline{D}$ have to be equal on either side of the boundary, or else Gauss's law would not give zero charge inside a small closed surface at the boundary:

$$D_{1n} = D_{2n} \tag{2.77}$$

If there were charge stored on the boundary, then this changes to $D_{1n} - D_{2n} = \rho_s$, where ρ_s is the surface charge density. (A dielectrics will in general have a nonzero surface charge at the boundary due to polarization by an external electric field, but we have already taken this into account when we changed the value of the dielectric constant ϵ in Eq. (2.60), so we do not include it on the right-hand side of Gauss's law.)

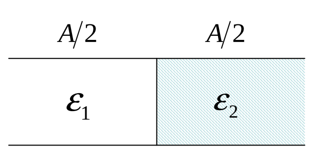


Figure 2.9: Parallel plate capacitor with two different dielectrics.

We can use this to analyze a parallel plate capacitor filled with two different dielectrics, with constants ϵ_1 and ϵ_2, each occupying half of the area of the plates. Because the electric field intensity $\overline{E}$ is tangential to the interface between the dielectrics, $\overline{E}$ must be the same in both dielectrics. So, the flux density is of the form

$$\overline{D} = \begin{cases} -\epsilon_1 E_o \hat{z} & \text{region 1} \\ -\epsilon_2 E_o \hat{z} & \text{region 2} \end{cases} \tag{2.78}$$

Applying Gauss's law,

$$\int_{A/2} \epsilon_1 E_o \hat{z} \cdot dxdy\hat{z} + \int_{A/2} \epsilon_2 E_o \hat{z} \cdot dxdy\hat{z} = \epsilon_1 E_o A/2 + \epsilon_2 E_o A/2 = Q$$

Solving for E_o,

$$E_o = \frac{2Q}{A(\epsilon_1 + \epsilon_2)}$$

The potential is

$$\begin{aligned} V &= -\int \overline{E} \cdot d\bar{\ell} \\ &= -\int_0^d \frac{-2Q}{A(\epsilon_1 + \epsilon_2)} \hat{z} \cdot dz\hat{z} \\ &= \frac{2Qd}{A(\epsilon_1 + \epsilon_2)} \end{aligned}$$

and the capacitance is

$$C = \frac{A(\epsilon_1 + \epsilon_2)}{2d} \tag{2.79}$$

If the dielectric layers were layered one on top of the other instead of side by side, then the flux density would have to be the same in both materials, and the solution would change correspondingly.

Chapter 3

Magnetostatics

In electrostatics, we considered the electric field and flux density arising from charges. In magnetostatics, we are concerned with the magnetic field intensity and flux density arising from currents. The basic equations of interest are Ampere's Law and Gauss' Law for magnetostatics:

Point form:

$$\nabla \cdot \overline{B} = 0 \tag{3.1}$$

$$\nabla \times \overline{H} = \overline{J} \tag{3.2}$$

Integral form:

$$\oint_S \overline{B} \cdot d\overline{s} = 0 \tag{3.3}$$

$$\oint_C \overline{H} \cdot d\overline{\ell} = \int_S \overline{J} \cdot d\overline{s} \tag{3.4}$$

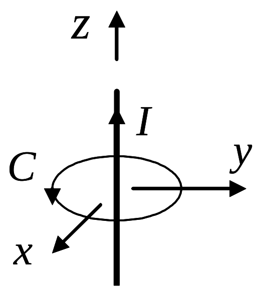
We will begin by using the integral form of Ampere's Law to determine magnetic fields much like we used Gauss' Law to determine electric fields in the prior section.

3.1 Ampere's Law

Suppose we are given a static or DC current distribution, which means we have charges that move and the velocity is constant in time. We can use Ampere's Law to determine the magnetic field that emanates from the current. Let's try a few examples:

3.1.1 Line Current

Suppose we have a current of I Amperes moving through a thin wire. The current flows in the $\hat{z}$ direction. Recall that to find the direction of the magnetic field, we use the right hand rule, which means that we put the thumb of the right hand in the direction of the current and the fingers indicate the direction of the magnetic field. The magnetic field will loop around the wire. By the symmetry of the problem, $\overline{H}$ is of the form $H_o\hat{\phi}$, and we find the constant H_o using Ampere's law:



$$\begin{aligned}
\oint_C \overline{H}\cdot d\overline{\ell} &= \int_S \overline{J}\cdot d\overline{s} = I && (3.5)\\
\int_0^{2\pi} H_o\hat{\phi}\cdot\hat{\phi} r\,d\phi &= I\\
2\pi H_o r &= I\\
H_o &= \frac{I}{2\pi r}\\
\overline{H} &= \frac{I}{2\pi r}\hat{\phi} && (3.6)
\end{aligned}$$

The right hand side of Ampere's law evaluates to the total current passing through the loop C. The resulting magnetic flux density is

$$\overline{B} = \mu\overline{H} = \frac{\mu I}{2\pi r}\hat{\phi} \qquad (3.7)$$

The intensity of the magnetic field decreases with distance, but the integral of the field along a closed path that encircles the wire is constant, regardless of the radius of the path.

Differential Forms

If we draw the magnetic field intensity vector field in Eq. (3.6), the picture is not very intuitive, because the magnetic field appears to "curl" everywhere in space, even though by Eq. (3.2) the quantity $\nabla\times\overline{H} = \overline{J}$ really is nonzero everywhere except on the z axis where the current is flowing.

The differential form picture for the line current adds some nice intuition to this. Let's derive the magnetic field intensity one-form for the line current. By symmetry, H is of the form $H_1 d\phi$, and we can find the constant H_1 using Ampere's law:

$$\begin{aligned}
\oint_C H &= \int_S J = I && (3.8)\\
\int_0^{2\pi} H_1 d\phi &= I\\
2\pi H_1 r &= I\\
H_1 &= \frac{I}{2\pi}\\
H &= \frac{I}{2\pi}d\phi && (3.9)
\end{aligned}$$

If we draw this differential form, we get I surfaces extending away from the line current (one surface for each Ampere of current), as shown in Fig. 3.1. Recall that J is a two-form, so it is represented by tubes running along the wire. The same number of surfaces of H extend away from the current as there are tubes of J. This is exactly what Ampere's law for magnetostatics in Eq. (3.4) means graphically.

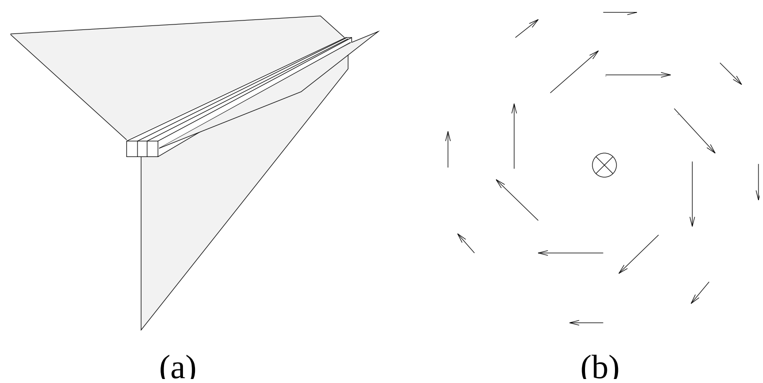
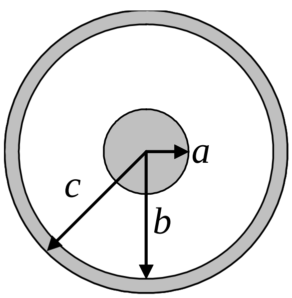


Figure 3.1: Magnetic field due to a line current. (a) Differential forms: surfaces of H are produced by tubes of J. (b) Vectors: the magnetic field intensity vector points in the right-hand rule direction around the line current.

3.1.2 Coaxial Cable

On the conductors, the current density can be written as



$$\overline{J} = \begin{cases} \frac{I}{\pi a^2}\hat{z} & \text{inner conductor} \\ \frac{-I}{\pi(c^2-b^2)}\hat{z} & \text{outer conductor} \end{cases} \tag{3.10}$$

We can compute the magnetic field in four regions:

Inside the inner conductor ($r < a$):

$$\int_0^{2\pi} H_o\hat{\phi}\cdot\hat{\phi}rd\phi = \int_0^{2\pi}\int_0^r \frac{I}{\pi a^2}\hat{z}\cdot\hat{z}r'dr'd\phi \tag{3.11}$$

$$2\pi r H_o = \frac{I}{\pi a^2}2\pi\frac{r^2}{2} = I\frac{r^2}{a^2} \tag{3.12}$$

$$H_o = \frac{Ir}{2\pi a^2} \tag{3.13}$$

$$\overline{H} = \frac{Ir}{2\pi a^2}\hat{\phi} \tag{3.14}$$

Between the conductors ($a < r < b$): Note that the integral on the left hand side is the same for all regions, except that H_o takes on different values.

$$2\pi r H_o = \int_0^{2\pi}\int_0^a \frac{I}{\pi a^2}\hat{z}\cdot\hat{z}rdrd\phi = \frac{I}{\pi a^2}2\pi\frac{a^2}{2} = I \tag{3.15}$$

$$H_o = \frac{I}{2\pi r} \tag{3.16}$$

$$\overline{H} = \frac{I}{2\pi r}\hat{\phi} \tag{3.17}$$

Inside the outer conductor ($b < r < c$):

$$
\begin{aligned}
2\pi r H_o &= I - \int_0^{2\pi}\int_b^r \frac{I}{\pi(c^2-b^2)} r' dr' d\phi && (3.18)\\
&= I - \frac{I}{\pi(c^2-b^2)} 2\pi \left(\frac{r^2}{2} - \frac{b^2}{2}\right) && (3.19)\\
&= I - I\left(\frac{r^2-b^2}{c^2-b^2}\right) && (3.20)\\
&= I\left[\frac{c^2-b^2-r^2+b^2}{c^2-b^2}\right] && (3.21)\\
&= I\left(\frac{c^2-r^2}{c^2-b^2}\right) && (3.22)\\
\overline{H} &= \frac{I}{2\pi r}\left(\frac{c^2-r^2}{c^2-b^2}\right)\hat{\phi} && (3.23)
\end{aligned}
$$

Outside the coaxial cable ($r > c$): On the right hand side of (3.22), we simply replace r with the outer limit of integration c (the left hand side remains the same). The right hand side is zero, so that

$$\overline{H} = 0 \qquad (3.24)$$

This makes sense, since there the *net* current enclosed is zero.

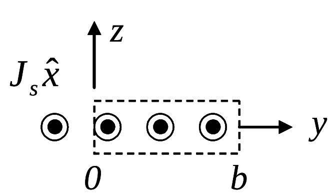
3.1.3 Infinite Current Sheet

An infinite current sheet can be hard to visualize, but it is an idealized current source that produces a very simple magnetic field. The current sheet can be thought of as many closely spaced parallel wires in a plane carrying current in the same direction.

Consider a surface current in the x-y plane flowing in the $\hat{x}$ direction. By the symmetry of the source, the magnetic field will be

$$\overline{H} = \begin{cases} -H_o\hat{y} & z > 0 \\ H_o\hat{y} & z < 0 \end{cases} \qquad (3.25)$$

where the constant H_o is unknown. Let's consider doing the integration over a square path of side length b:



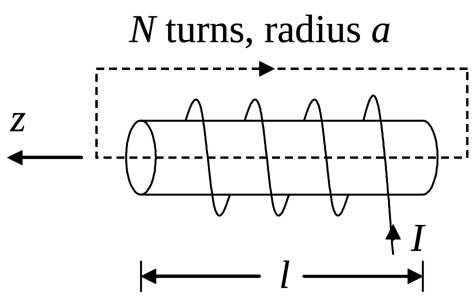
$$
\begin{aligned}
\int_0^b H_o dy - \int_b^0 H_o dy &= \int_0^b J_s dy && (3.26)\\
2H_o b &= J_s b && (3.27)\\
H_o &= \frac{J_s}{2} && (3.28)\\
\overline{H} &= \pm\frac{J_s}{2}\hat{y} && (3.29)
\end{aligned}
$$

Note that we only integrated the surface current density over the line from 0 to b in y. We did not integrate in z since the surface current density is in units of A/m (i.e., we only need to integrate in one dimension to get the total current enclosed by the square path).

3.1.4 Solenoid (coil)

Inside a solenoid carrying DC current, the magnetic field is concentrated and in the direction of the axis of the coil. Outside the solenoid, magnetic field lines spread out, and the field can be approximated as zero.

Inside a z-directed solenoid, $\overline{H} = H_o\hat{z}$. To find H_o, we choose as the closed path in Ampere's law a square path with one side in the solenoid and the other outside. The total current passing through the path is NI, where N is the number of turns of wire and I is the current flowing in the wire. We let the side of the path outside of the solenoid go off to infinity where the fields are zero. The main contribution to the line integral then comes from the part of the path that is within the solenoid of length ℓ:



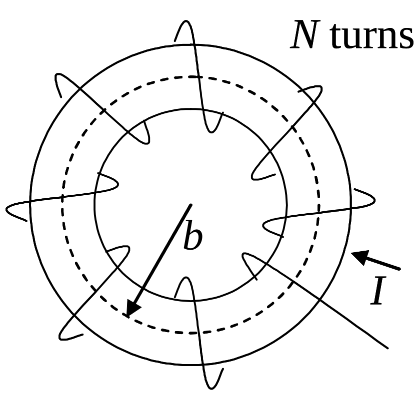
$$\int_0^\ell H_o\hat{z}\cdot\hat{z}dz = NI \tag{3.30}$$

$$H_o = \frac{NI}{\ell} \tag{3.31}$$

$$\overline{H} = \frac{NI}{\ell}\hat{z} \tag{3.32}$$

$$\overline{B} = \mu\frac{NI}{\ell}\hat{z} \tag{3.33}$$

3.1.5 Toroid



$$\int_0^{2\pi} H_o\hat{\phi}\cdot\hat{\phi}rd\phi = NI \tag{3.34}$$

$$2\pi rH_o = NI \tag{3.35}$$

$$H_o = \frac{NI}{2\pi r} \tag{3.36}$$

$$\overline{H} = \frac{NI}{2\pi r}\hat{\phi} \tag{3.37}$$

$$\overline{B} = \mu\frac{NI}{2\pi r}\hat{\phi} \tag{3.38}$$

3.2 Magnetic Properties of Materials

The magnetic permeability μ is similar to the permittivity of a dielectric material, except that it represents polarization of a magnetic material such as iron or another type of ferrous metal.

Using a classical description of matter, all atoms have electrons which orbit the nucleus. This orbiting charge represents a current loop which creates a magnetic moment. In most materials, called diamagnetic materials, these atoms are randomly aligned so that there is no net magnetic effect.

The electrons also have a property referred to as "spin" which creates another contribution to the magnetic moment. In atoms with even numbers of electrons, there are always two spins that are equal but opposite, resulting in zero net magnetic moment from spin. For an odd number of electrons, there is a net magnetic effect from the single unpaired electron.

When a material is exposed to a magnetic field $\overline{H}$, we can express the magnetic flux density as

$$\overline{B} = \mu_0\overline{H} + \mu_0\overline{M} = \mu_0(\overline{H} + \overline{M}) \tag{3.39}$$

where $\overline{M}$ is called the magnetization vector of a material. This vector represents the vector sum of the magnetic dipole moments of the atoms. Physically, the magnetic field is aligning the atomic magnetic dipoles. The degree to which these dipoles can be aligned is represented by the magnitude of $\overline{M}$. Much like we did in electrostatics, we define a magnetic susceptibility χ_m and write

$$\overline{M} = \chi_m\overline{H} \tag{3.40}$$

$$\overline{B} = \mu_0(\overline{H} + \chi_m\overline{H}) = \mu_0 \underbrace{(1+\chi_m)}_{\mu_r} \overline{H} = \mu\overline{H} \tag{3.41}$$

The units of μ are H/m. For most materials, χ_m is so small that we can write $\mu_r = 1$. Ferromagnetic materials, however, which are suceptible to magnetic alignment, can have high values of μ_r. For example, pure iron has $\mu_r = 2 \times 10^5$.

Types of magnetic materials:

1. Diamagnetic: $\chi_m < 0$ ($\chi_m \sim -10^{-5}$, $\mu_r \sim 1$)
2. Paramagnetic: $\chi_m > 0$ ($\chi_m \sim 10^{-5}$, $\mu_r \sim 1$)
3. Ferromagnetic: $|\chi_m| \gg 1$, $\mu_r \gg 1$

3.3 Inductance

The physical behavior of currents and magnetic fields is interesting. We know that currents create magnetic fields. However, a static magnetic field (created by a magnet) which cuts through a loop of wire will not create a current unless the loop or magnet are moving.

We will see that time-varying magnetic fields can induce currents. If we have two separate loops, and drive a time-varying current through one, we will observe a current in the other. This magnetic linking is mutual inductance. If the current in a solenoid changes, the changing magnetic field in the solenoid generates an electromotive force (EMF) in the loops in the solenoid. This is self inductance.

Despite the fact that we need time-varying currents/fields to have this coupling, we can compute the strength of the coupling using static analysis, based on the energy stored in the magnetic field produced by the solenoid. The strength of the coupling is the inductance of the system:

$$\text{Inductance: } L = \frac{\Lambda}{I} \tag{3.42}$$

where

Λ = net magnetic flux linkage = net flux linking the loops in the solenoid
I = current producing flux linkage

3.3.1 Steps for Computing Inductance

1. Assume a current and a form for $\overline{H}$
2. Calculate $\overline{H}$ using Ampere's Law
3. Calculate $\overline{B} = \mu\overline{H}$
4. Calculate $\Lambda = \int_s \overline{B} \cdot d\overline{s}$
5. Calculate $L = \Lambda/I$

3.3.2 Solenoid

1.-3. The magnetic flux density for a current I was computed previously in Eq. (3.33):

$$\overline{B} = \mu\frac{NI}{\ell}\hat{z} \tag{3.43}$$

4. For a single loop, the flux passing through the loop is

$$\begin{aligned}
\Lambda_1 &= \int_0^{2\pi}\int_0^a \overline{B} \cdot d\overline{s} && (3.44)\\
&= \int_0^{2\pi}\int_0^a \mu\frac{NI}{\ell}\hat{z} \cdot \hat{z} r dr d\phi && (3.45)\\
&= \mu 2\pi \frac{a^2}{2}\frac{NI}{\ell} && (3.46)\\
&= \mu\pi a^2\frac{NI}{\ell} && (3.47)
\end{aligned}$$

For N loops, the flux linkage is

$$\Lambda = N\Lambda_1 = \mu\pi a^2\frac{N^2 I}{\ell} \tag{3.48}$$

5. The inductance is

$$L = \frac{\Lambda}{I} = \frac{\mu\pi a^2 N^2}{\ell} \tag{3.49}$$

You may wonder why we use $\Lambda = N\Lambda_1$. If we pass a different time-varying current $i(t)$ through the solenoid, then the voltage across one loop would be $v(t) = L_1 di/dt = (\Lambda_1/I)di/dt$. Across all the loops in the inductor, the voltages add in series, so that $v(t) = (N\Lambda_1/I)di(t)/dt$.

3.3.3 Toroid

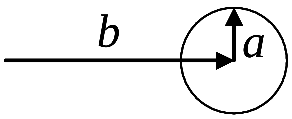
Based on our prior work, steps 1-3 yield

$$\overline{B} = \mu\frac{NI}{2\pi r}\hat{\phi} \tag{3.50}$$

The integral in step 4 is relatively difficult to perform, so let's make an approximation. If the radius a of the core of the toroid is small compared to the radius of the toroid (*i.e.*, $b \gg a$), then

$$\overline{B} \approx \mu\frac{NI}{2\pi b}\hat{\phi} \tag{3.51}$$

To obtain the net flux, we integrate over the area of the toroid cross-section:



$$\begin{aligned}
\Lambda_1 &= \int\int \mu \frac{NI}{2\pi b}\hat{\phi}\cdot\hat{\phi}drdz && (3.52)\\
&= \mu\frac{NI}{2\pi b}\pi a^2 = \mu\frac{NIa^2}{2b} && (3.53)\\
\Lambda &= N\Lambda_1 = \mu\frac{N^2Ia^2}{2b} && (3.54)
\end{aligned}$$

The inductance is therefore:

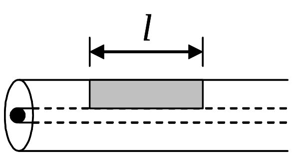
$$L = \frac{\Lambda}{I} = \mu\frac{N^2a^2}{2b} \tag{3.55}$$

3.3.4 Coax

We have already computed

$$\overline{B} = \frac{\mu I}{2\pi r}\hat{\phi} \tag{3.56}$$

The flux linking the two conductors is now the flux passing through the area between the conductors:



$$\begin{aligned}
\Lambda &= \int_a^b\int_0^\ell \frac{\mu I}{2\pi r}\hat{\phi}\cdot\hat{\phi}dzdr && (3.57)\\
&= \frac{\mu I\ell}{2\pi}\ln r|_a^b && (3.58)\\
&= \frac{\mu I\ell}{2\pi}\ln\left(\frac{b}{a}\right) && (3.59)
\end{aligned}$$

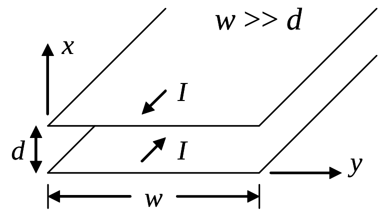
The inductance is

$$\begin{aligned}
L &= \frac{\mu\ell}{2\pi}\ln\left(\frac{b}{a}\right) && (3.60)\\
\frac{L}{\ell} &= \frac{\mu}{2\pi}\ln\left(\frac{b}{a}\right) && (3.61)
\end{aligned}$$

3.3.5 Parallel Flat Conductors

For parallel, flat conductors, the inductance per unit length can be found from this analysis:

$\overline{H} \approx H_o \hat{y}$



$$\int_0^w H_o \hat{y} \cdot \hat{y} dy = H_o w = I \tag{3.62}$$

$$\overline{H} = \frac{I}{w} \hat{y} \tag{3.63}$$

$$\overline{B} = \frac{\mu I}{w} \hat{y} \tag{3.64}$$

$$\Lambda = \int_0^d \int_0^\ell \frac{\mu I}{w} dz dx = \frac{\mu I}{w} \ell d \tag{3.65}$$

$$\frac{L}{\ell} = \mu \frac{d}{w} \tag{3.66}$$

Chapter 4

Dynamic Fields

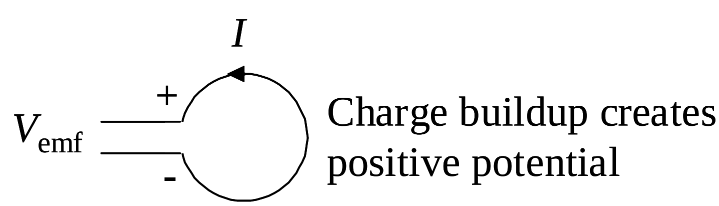
We are now ready to look at the behavior of fields that vary in time. To begin, let's examine a very interesting relationship between changing magnetic flux and electric field intensity, Faraday's Law. We will then look at Ampere's law and the rest of Maxwell's equations.

4.1 Faraday's Law

A time-varying magnetic flux through a loop will cause a current to flow in the loop of wire. Faraday's Law describes this effect. To begin, consider a wire loop as shown. A magnetic field supplied by an external source passes through the loop. The flux through the loop is defined as

$$\Lambda = \int_S \overline{B} \cdot d\overline{s} \tag{4.1}$$

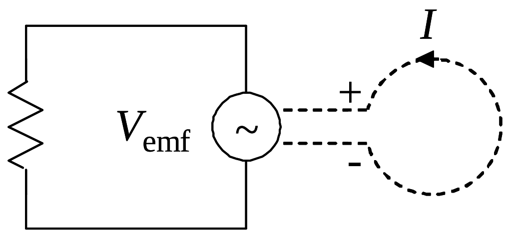
with units of Webers (Wb). When this flux changes in time, a current will flow in the loop. This means that a voltage has been created across the loop terminals called the *electromotive force* (EMF):



$$V_{\text{emf}} = -\frac{d\Lambda}{dt} = -\frac{d}{dt}\int_S \overline{B} \cdot d\overline{s} \tag{4.2}$$

V_{emf} represents the voltage available to a load circuit attached to the loop. Note that if $\overline{B}$ does not change in time, $V_{\text{emf}} = 0$.

The negative sign comes from Lenz's Law, which states that the induced current will oppose the change in flux. To ensure the correct polarity of V_{emf}, we use the right hand rule with the thumb in the direction of $d\overline{s}$, and the fingers give the direction of the $+$ terminal to the $-$ terminal. For the above loop, let $d\overline{s}$ be out of the page. If Λ decreases, V_{emf} is positive.



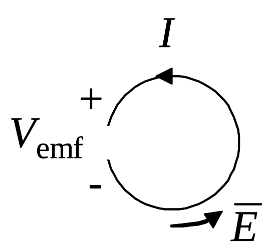
Notice also the following. Since $V_{\text{emf}} \neq 0$ in this system, $\overline{E} \neq 0$. If we integrate $\overline{E}$ around the loop, we will obtain the voltage V_{emf}:

$$V_{\text{emf}} = \oint_C \overline{E} \cdot d\overline{\ell} = -\frac{d}{dt} \int_S \overline{B} \cdot d\overline{s} \tag{4.3}$$

This is **Faraday's Law** in integral form. The direction of the path integral is given by the right hand rule with respect to $d\overline{s}$. We think of the gap as being small enough that we can consider the integral of $\overline{E}$ to be around a closed path. Since this integral is nonzero, the electric field near a time varying magnetic field is nonconservative.

The definition of electric potential used in electrostatics includes a minus sign, because we were finding the potential along a path between two electrodes. With induced EMF, the situation is different: the electric field is pushing the charge to make one terminal more positive than the other, so we do not have the minus sign when finding V_{emf}.

For the above example, the right hand side of Eq. (4.3) has a positive value, so the electric field is in the same direction as $d\overline{\ell}$. The direction of $d\overline{\ell}$ is given by the right hand rule with the thumb in the direction of $d\overline{s}$, which is out of the page, so $d\overline{\ell}$ is in the counterclockwise direction. The counterclockwise electric field pushes charge so that the $+$ side becomes positive, making V_{emf} positive. If the flux were increasing, then the electric field would reverse, and the $+$ side would become negatively charged, making V_{emf} negative.

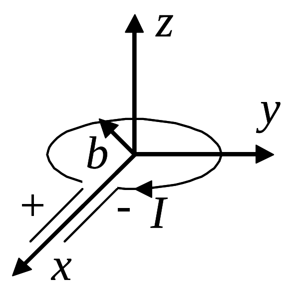


There are two ways to obtain this effect. The magnetic flux through a loop can be changed by varying the magnetic field or by varying the location or orientation of the loop in a fixed magnetic field. These two effects are referred to as:

- Transformer EMF: A time varying magnetic field linking a stationary loop.
- Motional EMF: A moving loop with a time varying orientation relative to the direction of $\overline{B}$.

4.1.1 Transformer Action

If the applied $\overline{B}$ through a loop changes in time, the induced potential at the loop terminals is called the transformer EMF.



Consider the loop shown with

$$\overline{B} = B_o t\hat{z} \tag{4.4}$$

$$\Lambda = \int_0^{2\pi}\int_0^b B_o t\, r dr d\phi = B_o t\frac{b^2}{2}2\pi = B_o t\pi b^2 \tag{4.5}$$

$$V_{\text{emf}} = -\frac{d\Lambda}{dt} = -B_o\pi b^2 \tag{4.6}$$

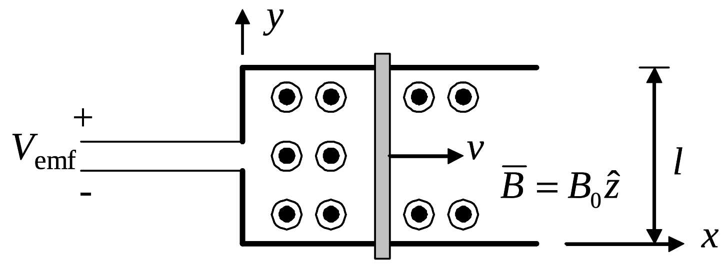
$\overline{B}$ is increasing in time, so I is induced as shown to oppose the change. V_{emf} is therefore negative.

This effect is used in a transformer. A time varying voltage or alternating current (AC) applied to the transformer's primary coil creates a time varying magnetic field. Transformers generally have a ferrite core with high magnetic permeability to increase the flux density. The time varying magnetic flux induces a voltage in the secondary winding. The number of turns in the primary and secondary coils can be adjusted to step up or step down the amplitude of the voltage across the terminals of the secondary coil relative to the amplitude of the voltage applied to the primary coil.

4.1.2 Generator Action

Generator action occurs when a loop is mechanically altered while the flux density remains constant. A sliding bar on two wire rails leads to a simple integration of the flux density over the loop area. A rotating loop is a more practical case and is the basis for most electric generators.

Sliding Bar



$$\Lambda(t) = \int_0^{\ell}\int_0^{vt} B_o dx dy = B_o \ell v t \tag{4.7}$$

$$V_{\text{emf}} = -\frac{d}{dt}\Lambda = -B_o \ell v \tag{4.8}$$

Rotating Loop

A loop of length ℓ and width w is rotating with an angular velocity of ω within a constant magnetic field given by

$$\overline{B} = \hat{z}B_o \tag{4.9}$$

The magnetic flux through the loop is

$$\Lambda = \int_S \overline{B} \cdot d\overline{s} \tag{4.10}$$

$$= \int_S \hat{z}B_o \cdot \hat{n} ds \tag{4.11}$$

where $\hat{n} = \cos(\omega t)\,\hat{z} + \sin(\omega t)\,\hat{y}$ and ω is the rotation rate in rad/sec of the loop. So,

$$\Lambda = \int_0^{\ell}\int_0^{w} B_o \cos(\omega t)\, ds \tag{4.12}$$

$$= B_o w \ell \cos(\omega t) \tag{4.13}$$

The EMF is

$$V_{\text{emf}} = B_o A \omega \sin(\omega t) \tag{4.14}$$

where $A = w\ell$ is the area of the loop.

In a generator, the magnetic field produced by the induced current in the loop creates an opposing force, so that external mechanical torque must be used to rotate the loop. In this way, mechanical energy is converted to electrical energy.

4.1.3 Inductor Law

The voltage-current relationship for an inductor is a simplification of Faraday's law. The voltage induced across the terminals of a solenoid carrying a current $i(t)$ is

$$
\begin{aligned}
v(t) &= \oint_C \overline{E} \cdot d\overline{\ell} && (4.15)\\
&= -\frac{d}{dt}\int_S \overline{B} \cdot d\overline{s} && (4.16)\\
&= -\frac{d}{dt}N\int_S -\frac{\mu N i(t)}{\ell}\, ds && (4.17)\\
&= \frac{d}{dt}N\frac{\mu N i(t)}{\ell}\pi a^2 && (4.18)\\
&= \underbrace{\frac{\mu\pi a^2 N^2}{\ell}}_{\text{Inductance } L}\ \frac{di(t)}{dt} && (4.19)\\
&= L\frac{di(t)}{dt} && (4.20)
\end{aligned}
$$

The factor of N in Eq. (4.17) is because the surface S is really N disks bounded by each turn of the coil. The extra minus sign in (4.17) arises because the direction of $d\overline{s}$ is opposite to the direction of the magnetic field produced by $i(t)$ if the current flows from the $+$ reference to the $-$ reference.

4.2 Ampere's Law and Displacement Current

The right-hand side of Ampere's Law includes two terms with units of current (A):

$$\oint \overline{H} \cdot d\overline{\ell} = \frac{d}{dt}\int_S \overline{D} \cdot d\overline{s} + \int_S \overline{J} \cdot d\overline{s} = \frac{d}{dt}\int_S \overline{D} \cdot d\overline{s} + I_c \tag{4.21}$$

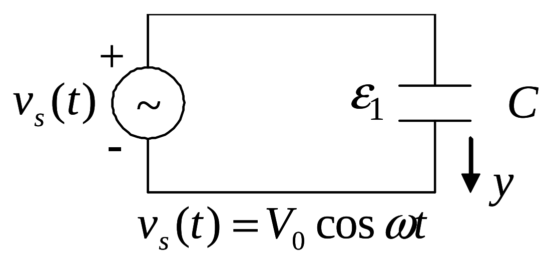
where I_c represents the conduction current. Although the term

$$I_d = \frac{d}{dt}\int_S \overline{D} \cdot d\overline{s} \tag{4.22}$$

does not represent the physical motion of charge, it has units of Amperes. We call this the displacement current.

Consider a parallel plate capacitor as shown. In the wire, $\overline{E} = \overline{D} = 0$, so there is no displacement current in the wire. The conduction current in the wire is

$$I_c \quad = \quad C\frac{dv_s}{dt} = -CV_o\omega \sin \omega t \tag{4.23}$$



In the capacitor, $\overline{J} = 0$ so

$$I_d = \frac{d}{dt}\int_S \overline{D} \cdot d\overline{s} \tag{4.24}$$

Since $\overline{E} = (V_o/d)\cos \omega t\ \hat{y}$, the electric flux density is $\overline{D} = (V_o\epsilon_1/d)\cos \omega t\ \hat{y}$, and the displacement current term in Ampere's law inside the capacitor is

$$I_d = \frac{d}{dt}\int_S \frac{V_o\epsilon_1}{d}\cos(\omega t)\ dzdx = -\underbrace{\frac{\epsilon_1 A}{d}}_{C} V_o\omega \sin \omega t = -CV_o\omega \sin \omega t \tag{4.25}$$

We can see that $I_d = I_c$, so that the displacement current allows continuity of current.

Notice that Eq. (4.25) can also be written as $I_d = Cdv/dt$, so the capacitor voltage-current relationship comes from Ampere's law. Also, by Ampere's law in Eq. (4.21), the changing electric field between the capacitor plates produces a magnetic field around the capacitor, just as the wire produces a magnetic field.

4.3 Boundary Conditions

Now that we have considered each of Maxwell's equations, we need to look at boundary conditions for all of the field quantities. This is the tool we use to deal with electromagnetic problems involving materials. Our goal is to understand how boundaries in materials impact electric and magnetic fields.

Boundary conditions are powerful tools that are routinely used in analytical and numerical methods for solving electromagnetic boundary value problems. When we studied transmission lines, we introduced boundary conditions at the junctions between the transmission line and the circuit elements or source attached to it. Using these boundary conditions allowed us to solve for the unknown forward and reverse wave amplitudes on the transmission line. Similarly, boundary conditions for the electromagnetic field allow us to solve for unknown wave amplitudes at the junction between different materials. With software tools for electromagnetic modeling, a computational domain with finite size is used to model an infinite region of space using an absorbing boundary condition.

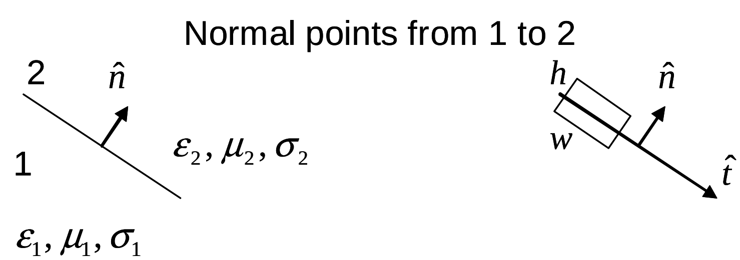
We first break the electric field and magnetic field into components that are either tangential or normal to the boundary, as given by

$$\begin{aligned}
\overline{E}_1 &= E_{1t}\,\hat{t} + E_{1n}\,\hat{n} && (4.26)\\
\overline{E}_2 &= E_{2t}\,\hat{t} + E_{2n}\,\hat{n} && (4.27)\\
\overline{H}_1 &= H_{1t}\,\hat{t} + H_{1n}\,\hat{n} && (4.28)\\
\overline{H}_2 &= H_{2t}\,\hat{t} + H_{2n}\,\hat{n}. && (4.29)
\end{aligned}$$

The goal is to find the relationship between the tangential or normal component on either side of a boundary surface.

Electric field intensity. We will begin with electric fields. Consider a material boundary as shown, with a closed contour that crosses the boundary. Faraday's law is

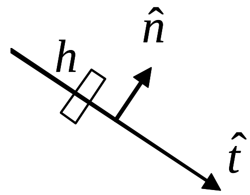
$$\oint_C \overline{E}\cdot d\overline{\ell} = -\frac{d}{dt}\int_S \overline{B}\cdot d\overline{s} \tag{4.30}$$



To apply this to the contour, let the contour shrink to zero area so that the right hand side of Faraday's law goes to zero. If we let $w \to 0$, the left hand side becomes

$$\int_{-h/2}^{0} E_{n1}\hat{n}\cdot\hat{n}d\ell + \int_{-h/2}^{0} E_{n1}\hat{n}\cdot(-\hat{n})d\ell + \int_{0}^{h/2} E_{n2}\hat{n}\cdot\hat{n}d\ell + \int_{0}^{h/2} E_{n2}\hat{n}\cdot(-\hat{n})d\ell = 0 \tag{4.31}$$

In other words, $0 = 0$. This isn't very useful.



If we instead let $h \to 0$, we get

$$\int_0^w E_{1t}\hat{t}\cdot(-\hat{t})d\ell + \int_0^w E_{2t}\hat{t}\cdot\hat{t}d\ell = 0 \tag{4.32}$$

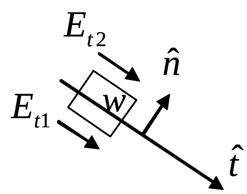
$$\int_0^w (E_{2t} - E_{1t})d\ell = 0 \tag{4.33}$$

$$E_{2t} - E_{1t} = 0 \tag{4.34}$$

So, the tangential component of the electric field is **continuous** across the boundary. Note that we can also write

$$\hat{n} \times (\overline{E}_2 - \overline{E}_1) = 0 \tag{4.35}$$

since $\hat{n} \times \overline{E}$ is the tangential electric field.



Magnetic field intensity. If we apply the same technique to Ampere's law, we get the same result with one important difference:

$$\oint_C \overline{H}\cdot d\overline{\ell} = \frac{d}{dt}\int_S \overline{D}\cdot d\overline{s} + \int_S \overline{J}\cdot d\overline{s} \tag{4.36}$$

Suppose that $\overline{J}$ represents a surface current that flows in the $\hat{s} = \hat{n} \times \hat{t}$ direction. Note that $\hat{s}$ is tangent to the surface but points normal to the surface of our integration contour. Then we have

$$\int_0^w H_{1t}\hat{t}\cdot(-\hat{t})d\ell + \int_0^w H_{2t}\hat{t}\cdot\hat{t}d\ell = \int_0^w J_s d\ell \tag{4.37}$$

$$\int_0^w (H_{2t} - H_{1t})d\ell = \int_0^w J_s d\ell \tag{4.38}$$

$$H_{2t} - H_{1t} = J_s \tag{4.39}$$

where J_s represents surface current and has units of A/m. This can also be expressed as

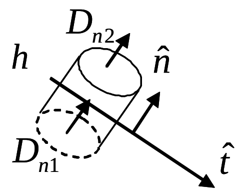
$$\hat{n} \times (\overline{H}_2 - \overline{H}_1) = \overline{J}_s \tag{4.40}$$

So, the tangential $\overline{H}$ can be discontinuous across the interface if a surface current exists.

Some authors label the two regions differently, so that the surface normal points from region 2 to region 1. In this case, $\overline{H}_1$ and $\overline{H}_2$ must be swapped in (4.40).

Electric flux intensity. Let's try Gauss' law. We use a "pillbox" for the integration surface:

$$\oint_S \overline{D} \cdot d\overline{s} = \int_V \rho_v dV \tag{4.41}$$



If $h \to 0$, the right hand side will go to zero unless there is a surface charge density. Then,

$$\int_A D_{n2}\hat{n} \cdot \hat{n} ds + \int_A D_{n1}\hat{n} \cdot (-\hat{n}) ds = \int_A \rho_s ds \tag{4.42}$$

$$D_{n2} - D_{n1} = \rho_s \tag{4.43}$$

$$\hat{n} \cdot (\overline{D}_2 - \overline{D}_1) = \rho_s \tag{4.44}$$

So normal $\overline{D}$ is discontinous by the surface charge density.

Magnetic flux density. For magnetic fields, Gauss's law implies that

$$B_{n2} - B_{n1} = 0 \tag{4.45}$$

$$\hat{n} \cdot (\overline{B}_2 - \overline{B}_1) = 0 \tag{4.46}$$

so that normal $\overline{B}$ is continous.

4.3.1 PEC

One very important ideal special case of the boundary conditions is that of the perfect electric conductor (PEC). The electric field must be identically zero inside a PEC. Static magnetic fields can exist inside a PEC body, but they cannot change in time, or else by Faraday's law an electric field would be induced in the PEC. If the boundary lies along the surface of a PEC body, then because $E_{1t} = 0$ inside the PEC, we must also have that $E_{2t} = 0$. So, we always have the condition $\hat{n} \times \overline{E} = 0$ on the surface of a PEC body.

To summarize all of the boundary conditions:

Field	Two Dielectrics	Dielectric-Conductor	PEC
Tangential $\overline{E}$	Continuous	Continuous: $\hat{n} \times (\overline{E}_2 - \overline{E}_1) = 0$	$\hat{n} \times \overline{E} = 0$
Tangential $\overline{H}$	Continuous	Discontinuous: $\hat{n} \times (\overline{H}_2 - \overline{H}_1) = \overline{J}_s$	$\hat{n} \times \overline{H} = \overline{J}_s$
Normal $\overline{D}$	Continuous	Discontinuous: $\hat{n} \cdot (\overline{D}_2 - \overline{D}_1) = \rho_s$	$\hat{n} \cdot \overline{D} = \rho_s$
Normal $\overline{B}$	Continuous	Continuous: $\hat{n} \cdot (\overline{B}_2 - \overline{B}_1) = 0$	$\hat{n} \cdot \overline{B} = 0$

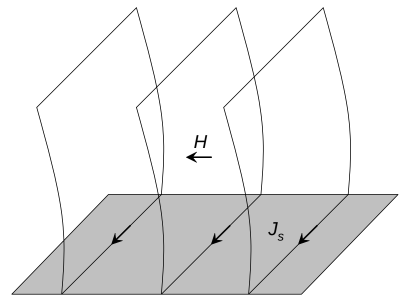


Figure 4.1: At a boundary, surfaces of H are either continuous across the boundary or they must end on lines of surface current J_s. The current flows along the lines of the surface current density one-form.

Differential Forms

Differential forms provide nice pictures for the boundary conditions. For the magnetic field intensity one-form, if $H_1 = 0$ below a boundary, then the surfaces of H_2 above the boundary must be normal to the boundary, and they end on lines representing the surface current density one-form J_s (Fig. 4.1). Unlike other one-forms like E and H, for which the physical direction of the field is normal to the one-forms surfaces, the direction of flow of the surface current is *along* the one-form lines on the boundary.

For the electric flux density two-form, if $D_1 = 0$ below the boundary, then the tubes of D_2 are normal to the boundary and end on squares of the surface charge density two-form $\rho_s ds$.

Tubes of B and surfaces of E must always be continuous across a boundary surface.

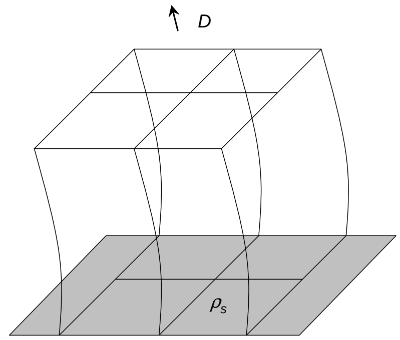


Figure 4.2: At a boundary, tubes of D are either continuous across the boundary or they must end on squares of surface charge.

4.4 Relating Maxwell's Equations in Point and Integral Form

We need to recall once again the vector derivative operations:

Gradient:

$$\nabla f(x,y,z) = \frac{\partial f}{\partial x}\hat{x} + \frac{\partial f}{\partial y}\hat{y} + \frac{\partial f}{\partial z}\hat{z} \tag{4.47}$$

Curl:

$$\nabla \times \overline{F}(x,y,z) = \hat{x}\left(\frac{\partial F_z}{\partial y} - \frac{\partial F_y}{\partial z}\right) + \hat{y}\left(\frac{\partial F_x}{\partial z} - \frac{\partial F_z}{\partial x}\right) + \hat{z}\left(\frac{\partial F_y}{\partial x} - \frac{\partial F_x}{\partial y}\right) \tag{4.48}$$

Divergence:

$$\nabla \cdot \overline{F}(x,y,z) = \frac{\partial F_x}{\partial x} + \frac{\partial F_y}{\partial y} + \frac{\partial F_z}{\partial z} \tag{4.49}$$

We also have two vector identities:

$$\nabla \times \nabla f = 0 \tag{4.50}$$

$$\nabla \cdot (\nabla \times \overline{F}) = 0 \tag{4.51}$$

The derivative operators and path and surface integrals are related by two theorems from vector calculus:

Stokes' Theorem:

$$\int_S \nabla \times \overline{F} \cdot d\overline{s} = \oint_C \overline{F} \cdot d\overline{\ell} \tag{4.52}$$

Divergence Theorem:

$$\int_V \nabla \cdot \overline{F} dV = \oint_S \overline{F} \cdot d\overline{s} \tag{4.53}$$

These are generalizations of the fundamental theorem of calculus to two and three dimensions. The fundamental theorem of calculus relates the integral of a derivative of a function over an interval to the value of the function at the endpoints or boundary of the interval. Similarly, Stokes' theorem relates the integral of the derivative of a vector field over a surface to the integral of the vector field over the boundary of the surface.

Let's use these theorems to derive Maxwell's equations in point form from the equations in integral form:

Faraday's Law:

$$\begin{aligned}
\oint_C \overline{E}\cdot d\bar{\ell} &= -\frac{d}{dt}\int_S \overline{B}\cdot d\overline{s} \\
\int_S \nabla\times\overline{E}\cdot d\overline{s} &= -\frac{d}{dt}\int_S \overline{B}\cdot d\overline{s} \\
\nabla\times\overline{E} &= -\frac{\partial}{\partial t}\overline{B} \qquad (4.54)
\end{aligned}$$

The last step in the derivation follows because the integrals in the second relationship are equal for any surface S.

Ampere's Law:

$$\begin{aligned}
\oint_C \overline{H}\cdot d\bar{\ell} &= \frac{d}{dt}\int_S \overline{D}\cdot d\overline{s} + \int_S \overline{J}\cdot d\overline{s} \\
\int_S \nabla\times\overline{H}\cdot d\overline{s} &= \frac{d}{dt}\int_S \overline{D}\cdot d\overline{s} + \int_S \overline{J}\cdot d\overline{s} \\
\nabla\times\overline{H} &= \frac{\partial}{\partial t}\overline{D} + \overline{J} \qquad (4.55)
\end{aligned}$$

Gauss' Laws:

$$\begin{aligned}
\oint_S \overline{D}\cdot d\overline{s} &= \int_V \rho_v dV \\
\int_V \nabla\cdot\overline{D} dV &= \int_V \rho_v dV \\
\nabla\cdot\overline{D} &= \rho_v \qquad (4.56)
\end{aligned}$$

$$\begin{aligned}
\oint_S \overline{B}\cdot d\overline{s} &= 0 \\
\nabla\cdot\overline{B} &= 0 \qquad (4.57)
\end{aligned}$$

4.5 Continuity of Charge

Consider a volume V containing a charge density ρ_v and total charge Q. The only way for Q to change is by charge entering/leaving the surface S bounding V. If I is the net current flowing across S out of V, then

$$I = -\frac{dQ}{dt} = -\frac{d}{dt}\int_V \rho_v dV \qquad (4.58)$$

From the definition of the total current in terms of the volume current density,

$$I = \oint_S \overline{J} \cdot d\overline{s} \tag{4.59}$$

Using this together with Eq. (4.58) leads to

$$\oint_S \overline{J} \cdot d\overline{s} = -\frac{d}{dt}\int_V \rho_v dV \tag{4.60}$$

This equation represents the integral form of the law of continuity of charge. If we now use the divergence theorem, we obtain

$$\oint_S \overline{J} \cdot d\overline{s} \quad = \quad \int_V \nabla \cdot \overline{J} dV \tag{4.61}$$

By combining the two previous equations, it follows that

$$\nabla \cdot \overline{J} \quad = \quad -\frac{\partial}{\partial t}\rho_v \tag{4.62}$$

This is the continuity of charge law in point form.

4.6 Differential Forms

Let's redo this using differential forms notation. The exterior derivative operator is

$$d = \left(\frac{\partial}{\partial x}dx + \frac{\partial}{\partial y}dy + \frac{\partial}{\partial z}dz\right)\wedge \tag{4.63}$$

This is analogous to the ∇ operator in vector notation, except that we combine it with differential forms using the wedge operation $\wedge$ instead of the dot or cross products:

$$\begin{aligned}
\text{Gradient} = d(\text{zero-form}) &= \left(\frac{\partial}{\partial x}dx + \frac{\partial}{\partial y}dy + \frac{\partial}{\partial z}dz\right)\wedge f \\
&= \frac{\partial f}{\partial x}dx + \frac{\partial f}{\partial y}dy + \frac{\partial f}{\partial z}dz \\
&= \text{one-form}
\end{aligned}$$

$$\begin{aligned}
\text{Curl} = d(\text{one-form}) &= \left(\frac{\partial}{\partial x}dx + \frac{\partial}{\partial y}dy + \frac{\partial}{\partial z}dz\right)\wedge (F_x dx + F_y dy + F_z dz) \\
&= \left(\frac{\partial E_z}{\partial y} - \frac{\partial E_y}{\partial z}\right) dy\wedge dz + \left(\frac{\partial E_x}{\partial z} - \frac{\partial E_z}{\partial x}\right) dz\wedge dx + \left(\frac{\partial E_y}{\partial x} - \frac{\partial E_x}{\partial y}\right) dx\wedge dy \\
&= \text{two-form}
\end{aligned}$$

$$\begin{aligned}
\text{Divergence} = d(\text{two-form}) &= \left(\frac{\partial}{\partial x}dx + \frac{\partial}{\partial y}dy + \frac{\partial}{\partial z}dz\right)\wedge (F_x dy\wedge dz + F_y dz\wedge dx + F_z dx\wedge dy) \\
&= \left(\frac{\partial F_x}{\partial x} + \frac{\partial F_y}{\partial y} + \frac{\partial F_z}{\partial z}\right) dx\wedge dy\wedge dz \\
&= \text{three-form}
\end{aligned}$$

The exterior derivative of a p-form is a $(p+1)$-form. The exterior derivative of a three-form is a four-form, which must have a repeated differential and so is zero.

If we apply the exterior derivative to any differential form twice, we get zero ($dd = 0$).

The exterior derivative also satisfies a product rule analogous to the product rule for the partial derivative, $d(\alpha\wedge\beta) = d\alpha\wedge\beta + (-1)^p\alpha\wedge d\beta$ where p is the degree of α.

The generalized Stokes theorem is

$$\int_M d\omega = \oint_{bd\,M} \omega \tag{4.64}$$

where ω is a p-form and M is a $(p+1)$–dimensional region of integration with boundary $bd\,M$. If $p = 0$, this is the fundamental theorem of calculus. If $p = 1$, this is analogous to the vector Stokes theorem. For $p = 2$, this is the divergence theorem.

Using differential forms, Maxwell's equations in point form are

$$dE = -\frac{\partial B}{\partial t} \tag{4.65}$$

$$dH = \frac{\partial D}{\partial t} + J \tag{4.66}$$

$$dD = \rho \tag{4.67}$$

$$dB = 0 \tag{4.68}$$

The continuity equation is $dJ = -\partial\rho/\partial t$.

Graphically,

> The exterior derivative of a function is a one-form with surfaces that are level sets for the function.
>
> The exterior derivative of a one-form is a two-form with tubes along the edges of the surfaces of the one-form wherever they end.
>
> The exterior derivative of a two-form is a three-form with boxes wherever tubes of the two-form end or begin.

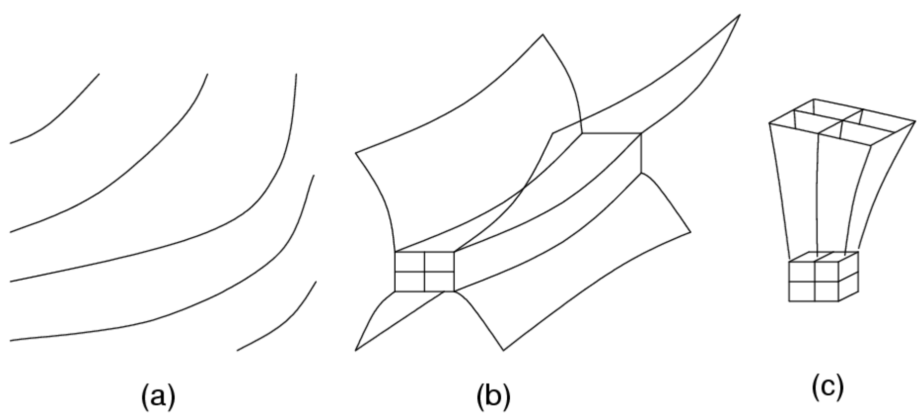


Figure 4.3: (a) Exterior derivative of a zero-form is a one-form drawn as the level sets of the function. (b) Exterior derivative of a one-form is a two-form with tubes where the one-form surfaces end. (c) Exterior derivative of a two-form is a three-form with boxes where the two-form tubes end or begin.

4.7 Maxwell's Equations in Time-Harmonic Form

As with transmission line systems, it is very common with Maxwell's equations to have a single frequency or time-harmonic excitation. After a time-harmonic excitation has been active for a certain amount of time, the transient part of the partial differential equation solution dies out or propagates outside of a region of interest. At this point, the differential equations simplify to what is referred to as the sinusoidal steady state.

To analyze field solutions in the sinusoidal steady state, we assume a time variation of the form $\cos \omega t$. Phasor notation is a very convenient way to work with sinusoidal waveforms. Recall that the definition of a phasor is

$$v(t) = \text{Re}\left\{\tilde{V} e^{j\omega t}\right\} \tag{4.69}$$

where the phasor $\tilde{V}$ is a complex number. What we want to do is to express the components of the electric and magnetic fields as phasors.

Now, we often suppress the coordinate dependence of the fields, but all of the fields are functions of space and time:

$$\begin{aligned} \overline{E} &= \overline{E}(x,y,z,t) = \overline{E}(\overline{R},t) && (4.70)\\ \overline{E}(\overline{R},t) &= \text{Re}\left\{\tilde{\overline{E}}(\overline{R}) e^{j\omega t}\right\} && (4.71) \end{aligned}$$

In the phasor domain, time derivatives become multiplication by $j\omega$:

$$\frac{\partial \overline{E}(\overline{R},t)}{\partial t} = \text{Re}\left\{\tilde{\overline{E}}(\overline{R}) j\omega e^{j\omega t}\right\}$$

Therefore, Maxwell's equations in time-harmonic (phasor) form are

$$\begin{aligned} \oint \tilde{\overline{E}} \cdot d\bar{\ell} &= -j\omega \int \tilde{\overline{B}} \cdot d\bar{s} && (4.72)\\ \oint \tilde{\overline{H}} \cdot d\bar{\ell} &= j\omega \int \tilde{\overline{D}} \cdot d\bar{s} + \int \tilde{\overline{J}} \cdot d\bar{s} && (4.73)\\ \oint_S \tilde{\overline{D}} \cdot d\bar{s} &= \int \tilde{\rho}_v \, dV && (4.74)\\ \oint_S \tilde{\overline{B}} \cdot d\bar{s} &= 0 && (4.75)\\ & && (4.76) \end{aligned}$$

In point form,

$$\begin{aligned} \nabla \times \tilde{\overline{E}} &= -j\omega \tilde{\overline{B}} && (4.77)\\ \nabla \times \tilde{\overline{H}} &= j\omega \tilde{\overline{D}} + \tilde{\overline{J}} && (4.78)\\ \nabla \cdot \tilde{\overline{D}} &= \tilde{\rho}_v && (4.79)\\ \nabla \cdot \tilde{\overline{B}} &= 0 && (4.80) \end{aligned}$$

In later material, we will use the phasor notation exclusively, and so we will drop the tilde on the phasor field and source variables. It should be clear from the context whether the field and source variables represent time domain or phasor quantities.

4.8 Review

Vectors

- Unit vectors, components, magnitude
- Vector fields, scalar fields
- Dot product, cross product
- Path and surface integrals (integration limits, $d\bar{\ell}$ and $d\bar{s}$)

Differential forms

- Converting between vectors and one-forms and two-forms
- One-form surfaces, two-form tubes, three-form boxes

Electrostatics

- Point charge, line charge, surface charge, volume charge density
- Cylindrical and spherical coordinate systems
- Gauss's law, examples: point, line, plane, spherical charges
- Electric potential, path integrals of $\overline{E}$, conservative fields
- Conductors and dielectrics
- Capacitance, examples: parallel plate, cylindrical, spherical
- Energy

Magnetostatics

- Ampere's law, examples: line current, coaxial, sheet current, solenoid, toroid
- Magnetic materials
- Inductance, examples

Dynamic fields

- Faraday's law, Lenz's law, right hand rule reference direction
- Nonconservative electric field (nonzero potential around closed loop)
- Transformer emf, motational emf
- Ampere's law, displacement current
- Boundary conditions, tangential/normal conditions, PEC special case
- Maxwell's equations in point form,
- Gradient, curl, divergence
- Stokes theorem, divergence theorem
- Maxwell's equations in time harmonic form

Fundamentals

Vector analysis

Dot product: $\overline{A} \cdot \overline{B} = A_x B_x + A_y B_y + A_z B_z$, cross product: $\hat{x} \times \hat{y} = -\hat{y} \times \hat{x} = \hat{z}$, etc.

Path integrals: $d\bar{\ell} = \hat{t}dt$, surface integrals: $d\bar{s} = \hat{n}ds$

Cylindrical coordinates: (r, ϕ, z), dr, $rd\phi$, dz, spherical coordinates: (R, θ, ϕ), dR, $Rd\theta$, $R\sin\theta d\phi$

Electrostatics

E surfaces, tubes of D, boxes of charge ρ_v

Gauss's law: $\oint_S \overline{D} \cdot d\bar{s} = \int_V \rho_v \, dV = Q_{enclosed}$

Electric potential: $V_2 - V_1 = -\int_{P_1}^{P_2} \overline{E} \cdot d\bar{\ell}$, $\overline{E} = -\nabla V$

Conductors: $\overline{J} = \sigma \overline{E}$ (point form of Ohm's law)

Dielectrics: $\overline{D} = \epsilon_0 \overline{E} + \overline{P} = \epsilon \overline{E} = \epsilon_r \epsilon_0 \overline{E}$

Capacitance: $C = Q/V$

Magnetostatics

H surfaces, tubes of J, tubes of B

Ampere's law for static fields: $\oint_C \overline{H} \cdot d\bar{\ell} = \int_A \overline{J} \cdot d\bar{s}$

Magnetic materials: $\overline{B} = \mu_0 \overline{H} + \mu_0 \overline{M} = \mu \overline{H} = \mu_r \mu_0 \overline{H}$

Inductance: $L = \Lambda/I$, $\Lambda =$ total linked flux, $\Lambda_1 = \int_A \overline{B} \cdot d\bar{s}$

Dynamics

Faraday's law: $V_{\text{emf}} = \oint_C \overline{E} \cdot d\bar{\ell} = -\frac{d}{dt} \int_A \overline{B} \cdot d\bar{s}$

Ampere's law: $\oint_C \overline{H} \cdot d\bar{\ell} = \int_A \overline{J} \cdot d\bar{s} + \frac{d}{dt} \int_A \overline{D} \cdot d\bar{s}$

Boundary conditions: $\hat{n} \times (\overline{E}_2 - \overline{E}_1) = 0$, $\hat{n} \times (\overline{H}_2 - \overline{H}_1) = \overline{J}_s$, $\hat{n} \cdot (\overline{D}_2 - \overline{D}_1) = \rho_s$, $\hat{n} \cdot (\overline{B}_2 - \overline{B}_1) = 0$

PEC: $\hat{n} \times \overline{E} = 0$, $\hat{n} \cdot \overline{B} = 0$

Gradient: $\nabla f(x,y,z) = \frac{\partial f}{\partial x}\hat{x} + \frac{\partial f}{\partial y}\hat{y} + \frac{\partial f}{\partial z}\hat{z}$

Curl: $\nabla \times \overline{F}(x,y,z) = \hat{x}\left(\frac{\partial F_z}{\partial y} - \frac{\partial F_y}{\partial z}\right) + \hat{y}\left(\frac{\partial F_x}{\partial z} - \frac{\partial F_z}{\partial x}\right) + \hat{z}\left(\frac{\partial F_y}{\partial x} - \frac{\partial F_x}{\partial y}\right)$

Divergence: $\nabla \cdot \overline{F}(x,y,z) = \frac{\partial F_x}{\partial x} + \frac{\partial F_y}{\partial y} + \frac{\partial F_z}{\partial z}$

Stokes' theorem: $\int_S \nabla \times \overline{F} \cdot d\bar{s} = \oint_C \overline{F} \cdot d\bar{\ell}$, divergence theorem: $\int_V \nabla \cdot \overline{F} dV = \oint_S \overline{F} \cdot d\bar{s}$

Maxwell's equations in point form: $\nabla \times \overline{E} = -\frac{\partial}{\partial t}\overline{B}$, $\nabla \times \overline{H} = \frac{\partial}{\partial t}\overline{D} + \overline{J}$, $\nabla \cdot \overline{D} = \rho_v$, $\nabla \cdot \overline{B} = 0$

Time-harmonic case: $\overline{E}(x,y,z,t) = \text{Re}\left\{\tilde{\overline{E}}(x,y,z)e^{j\omega t}\right\}$, $\frac{\partial}{\partial t} \to j\omega$

Chapter 5

Plane Waves

We are now ready to look at the simplest form of electromagnetic waves. From this point on, we will assume that field quantities are phasors unless otherwise stated.

5.1 Wave Equation

Instead of solving Maxwell's equations directly to obtain wave solutions, we will transform the system of first order partial differential equations (PDEs) into a single second order PDE that is easier to solve. We start with Maxwell's equations in time harmonic or phasor form, with the constitutive relations used to eliminate the flux densities $\overline{D}$ and $\overline{B}$:

$$\nabla \times \overline{E} = -j\omega\mu\overline{H} \tag{5.1}$$

$$\nabla \times \overline{H} = j\omega\epsilon\overline{E} + \overline{J} \tag{5.2}$$

$$\nabla \cdot \epsilon\overline{E} = \rho_v \tag{5.3}$$

$$\nabla \cdot \mu\overline{H} = 0 \tag{5.4}$$

The goal is to eliminate all of the field quantities to get an equation for one field only.

Conducting Media. In order to handle lossy materials (conductors), we first rewrite Ampere's Law. If we have a medium which has free charge allowing current flow, then $\overline{J} = \sigma\overline{E}$, and

$$\begin{aligned}\nabla \times \overline{H} = j\omega\epsilon\overline{E} + \sigma\overline{E} &= j\omega[\epsilon + \sigma/j\omega]\overline{E} \\ &= j\omega\underbrace{[\epsilon - j\sigma/\omega]}_{\epsilon_c}\overline{E}\end{aligned} \tag{5.5}$$

This shows that in the phasor domain, the conductivity can be lumped together with the permittivity to produce a new effective complex permittivity:

$$\epsilon_c = \epsilon - j\sigma/\omega = \epsilon_0\left[\epsilon_r - j\frac{\sigma}{\omega\epsilon_0}\right] = \epsilon_0\epsilon_{cr} \tag{5.6}$$

We also sometimes use the notation

$$\epsilon_c = \epsilon' - j\epsilon'' \tag{5.7}$$

for the real and imaginary parts of the complex permittivity. This reduces Ampere's law for a conducting material into the form

$$\nabla \times \overline{H} = j\omega\epsilon_c\overline{E} \tag{5.8}$$

where ϵ_c is complex.

We will assume that there are no impressed sources in our region of interest (i.e., no sources inside the region of interest that are produced by external forces). We can still have charges that move in response to fields, or induced currents, but we have already taken those into account when we made ϵ_c into a complex number.

If we take the curl of Faraday's law, we obtain

$$\nabla \times \nabla \times \overline{E} = -j\omega\mu\nabla \times \overline{H}$$

On the right hand side, we can replace $\nabla \times \overline{H}$ with the electric field using Ampere's Law. We now have a second-order partial differential equation with only one field variable, $\overline{E}$:

$$\nabla \times \nabla \times \overline{E} = -j\omega\mu(j\omega\epsilon_c\overline{E}) = \omega^2\mu\epsilon_c\overline{E}$$

Because of the complexity of the double curl operation, this equation is difficult to solve. We can replace the double curl with a simpler operator using the vector identity

$$\nabla \times \nabla \times \overline{E} = \nabla(\nabla \cdot \overline{E}) - \nabla^2\overline{E} \tag{5.9}$$

where $\nabla^2\overline{E}$ is the Laplacian. In Cartesian coordinates:

$$\nabla^2\overline{E} = \frac{\partial^2\overline{E}}{\partial x^2} + \frac{\partial^2\overline{E}}{\partial y^2} + \frac{\partial^2\overline{E}}{\partial z^2} \tag{5.10}$$

This leads to

$$\nabla(\nabla \cdot \overline{E}) - \nabla^2\overline{E} = \omega^2\mu\epsilon_c\overline{E} \tag{5.11}$$

In a region where the charge density $\rho_v = 0$, from Gauss's law we have $\nabla \cdot \epsilon_c\overline{E} = \epsilon_c\nabla \cdot \overline{E} = 0$, so this equation simplifies to the homogeneous wave equation

$$\nabla^2\overline{E} + \omega^2\mu\epsilon_c\overline{E} = 0 \tag{5.12}$$

Using the notation $k = \omega^2\mu\epsilon_c$ for the propagation constant in the wave equation, this can be simplified to

$$(\nabla^2 + k^2)\overline{E} = 0 \tag{5.13}$$

This PDE is sometimes called the Helmholtz equation.

For lossy materials with $\sigma \neq 0$, ϵ_c is complex, and it is convenient to define the propagation constant in the Helmholtz equation differently. In the lossy case, we use $\gamma^2 = -\omega^2\mu\epsilon_c$ and the Helmholtz equation becomes

$$\nabla^2\overline{E} - \gamma^2\overline{E} = 0 \tag{5.14}$$

5.1.1 Lossless Media

The relationship $\overline{J} = \sigma\overline{E}$ is related to Ohm's law written in the form $I = GV = V/R$, where $G =$ conductance. So, $\sigma > 0$ means energy will be dissipated (loss). If $\sigma = 0$, we call the material a *lossless medium*. In the lossless case, $\omega^2\mu\epsilon$ is a real number. The propagation constant

$$k = \omega\sqrt{\mu\epsilon} \tag{5.15}$$

is the wavenumber. The units of k are radians/meter. This is analogous to the transmission line quantity $\beta = \omega\sqrt{L'C'}$, except that μ (H/m) and ϵ (F/m) are associated with a three-dimensional medium rather than a transmission line structure.

5.2 Plane Wave Solutions

We have shown that in a source-free region and lossless medium, the electric field intensity is governed by the partial differential equation $\nabla^2\overline{E} + k^2\overline{E} = 0$. One type of solution to this PDE is the uniform plane wave. A uniform plane wave is a wave for which there is no variation of the fields over a given plane.

5.2.1 Plane waves propagating in the $\pm z$ direction

To understand plane wave solutions, we will write out the PDE in Cartesian coordinates. If $\overline{E} = \hat{x}E_x + \hat{y}E_y + \hat{z}E_z$, then we have

$$\left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2}\right)(\hat{x}E_x + \hat{y}E_y + \hat{z}E_z) + k^2(\hat{x}E_x + \hat{y}E_y + \hat{z}E_z) = 0 \tag{5.16}$$

Let's focus first on the $\hat{x}$ component of this vector equation:

$$\left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2} + k^2\right)E_x = 0 \tag{5.17}$$

For our first plane wave solution, we will choose the plane of no variation to be the x-y plane. Then

$$\frac{\partial\overline{E}}{\partial x} = \frac{\partial\overline{E}}{\partial y} = \frac{\partial\overline{H}}{\partial x} = \frac{\partial\overline{H}}{\partial y} = 0 \tag{5.18}$$

This similifies Eq. (5.17) to

$$\frac{d^2E_x}{dz^2} + k^2E_x = 0 \tag{5.19}$$

There are similar equations for E_y, H_x, and H_y.

What about E_z and H_z? Using Ampere's law with only the $\hat{z}$ component:

$$\nabla\times\overline{H} = j\omega\epsilon\overline{E} \tag{5.20}$$

$$\hat{z}\left(\frac{\partial H_y}{\partial x} - \frac{\partial H_x}{\partial y}\right) = \hat{z}j\omega\epsilon E_z \tag{5.21}$$

But since

$$\frac{\partial H_y}{\partial x} = \frac{\partial H_x}{\partial y} = 0 \tag{5.22}$$

we have that $E_z = 0$. Using Faraday's law in the same way shows that $H_z = 0$ as well. This shows that $\overline{E}$ and $\overline{H}$ are orthogonal to the x-y plane. Later, we will see that this represents an important property of plane waves, that the fields are orthogonal to the direction of propagation of the wave, so that electromagnetic waves in free space are transverse waves.

Now, all we have to do is solve the differential equation (5.19) for E_x. Assuming that the solution is of the form

$$E_x = Ae^{mz} \tag{5.23}$$

and putting this into the differential equation leads to

$$\begin{aligned} Am^2e^{mz} + k^2Ae^{mz} &= 0 \\ m^2 + k^2 &= 0 \\ m &= \pm jk \end{aligned}$$

The general solution to the differential equation is

$$E_x(z) = E_{xo}^{+}e^{-jkz} + E_{xo}^{-}e^{jkz} \tag{5.24}$$

where the first term is a forward traveling wave and the second is a reverse traveling wave. The coefficients E_{xo}^{+} and E_{xo}^{-} are phasors that determine the intensity of the forward and reverse waves, which in turn depends on the strength and location of the source that produced the wave.

Let's assume that only the forward traveling wave exists, so that $E_x(z) = E_{xo}^{+}e^{-jkz}$. We can then find the magnetic field using Faraday's law:

$$\begin{aligned} \overline{E}(z) &= \hat{x}E_{xo}^{+}e^{-jkz} \\ \nabla \times \overline{E} &= \begin{vmatrix} \hat{x} & \hat{y} & \hat{z} \\ \frac{\partial}{\partial x} & \frac{\partial}{\partial y} & \frac{\partial}{\partial z} \\ E_x & 0 & 0 \end{vmatrix} = \hat{y}\frac{\partial E_x}{\partial z} - \hat{z}\frac{\partial E_x}{\partial y} = \hat{y}\frac{\partial E_x}{\partial z} = -j\omega\mu\overline{H} \\ \hat{y}E_{xo}^{+}e^{-jkz}(-jk) &= -j\omega\mu\hat{y}H_y \\ H_y &= E_{xo}^{+}\frac{k}{\omega\mu}e^{-jkz} = E_{xo}^{+}\frac{\omega\sqrt{\mu\epsilon}}{\omega\mu}e^{-jkz} = \frac{E_{xo}^{+}}{\sqrt{\mu/\epsilon}}e^{-jkz} = H_{yo}^{+}e^{-jkz} \end{aligned} \tag{5.25}$$

From this derivation it follows that $H_{yo}^{+} = E_{xo}^{+}/\eta$, where

$$\eta = \sqrt{\mu/\epsilon} \tag{5.26}$$

η has units of ohms. We call it the intrinsic impedance of the propagation medium. It is analogous to the characteristic impedance of a transmission line, and represents the ratio of the electric to magnetic fields.

The electric and magnetic fields and the direction of propagation have directions that satisfy precise relationships, known as orthogonality relationships. For the $+z$ directed plane wave, we have

$+z$ traveling wave
$+x$ directed electric field
$+y$ directed magnetic field

We can see that $\overline{E} \times \overline{H}$ is a vector that lies in the direction of propagation ($+z$). For this wave,

$$\overline{E}(z) = \hat{x}E_{xo}^{+}e^{-jkz} \tag{5.27}$$

$$\overline{H}(z) = \hat{y}\frac{E_{xo}^{+}}{\eta}e^{-jkz} \tag{5.28}$$

$\overline{E}$ and $\overline{H}$ are perpendicular to each other and are both perpendicular to the direction of propagation.

For the $-z$ traveling wave,

$$\hat{y}E_{xo}^{-}e^{jkz}(jk) = -j\omega\mu\hat{y}H_y \tag{5.29}$$

$$H_y = -E_{xo}^{-}\frac{k}{\omega\mu}e^{jkz} = -\frac{E_{xo}^{-}}{\eta}e^{jkz} = H_{yo}^{-}e^{jkz} \tag{5.30}$$

$$\frac{E_{xo}^{-}}{H_{yo}^{-}} = -\eta \tag{5.31}$$

The field and propagation directions are

$-z$ traveling wave
$+x$ directed electric field
$-y$ directed magnetic field

$\overline{E} \times \overline{H}$ points in the $-\hat{z}$ direction (the direction of propagation).

5.2.2 Plane waves propagating in arbitrary directions

If the plane wave is traveling in an arbitrary direction the solution for the electric field becomes

$$\overline{E} = \overline{E}_o e^{-j\,(k_x x + k_y y + k_z z)} \tag{5.32}$$

where $\overline{E}_o$ is a complex vector constant. For a convenient shorthand we define a wavevector

$$\overline{k} = k_x\hat{x} + k_y\hat{y} + k_z\hat{z}. \tag{5.33}$$

The wavevector points in the direction of propagation and has a magnitude of

$$k = |\overline{k}| = \sqrt{k_x^2 + k_y^2 + k_z^2} = \omega\sqrt{\mu\epsilon} = \frac{2\pi}{\lambda} \tag{5.34}$$

This is known as the plane wave dispersion relation. Now we can write $\overline{E}$ as

$$\overline{E} = \overline{E}_o e^{-j\overline{k}\cdot\overline{r}} \tag{5.35}$$

since

$$\overline{k}\cdot\overline{r} = (k_x\hat{x} + k_y\hat{y} + k_z\hat{z})\cdot(x\hat{x} + y\hat{y} + z\hat{z}) = k_x x + k_y y + k_z z \tag{5.36}$$

If we use the notation $\hat{k}$ to designate the direction of propagation ($\hat{k} = \hat{z}$ for the example considered earlier), then the conditions $\overline{E} \perp \overline{H}$, $\overline{E} \perp \hat{k}$, and $\overline{H} \perp \hat{k}$ indicate that the wave is a **Transverse Electromagnetic (TEM) Wave**.

Let's look at the time domain fields. If $E_{xo}^{+} = |E_{xo}^{+}|e^{j\phi^{+}}$,

$$\begin{aligned}\overline{E}(z,t) = \ &\mathrm{Re}\{\tilde{\overline{E}}(z)e^{j\omega t}\} = \hat{x}\ \mathrm{Re}\{|E_{xo}^{+}|e^{j(\omega t - kz + \phi^{+})}\}\\ &= \hat{x}|E_{xo}^{+}|\cos(\omega t - kz + \phi^{+})\end{aligned} \tag{5.37}$$

$$\overline{H}(z,t) = \ \mathrm{Re}\{\tilde{\overline{H}}(z)e^{j\omega t}\} = \hat{y}\frac{|E_{xo}^{+}|}{\eta}\cos(\omega t - kz + \phi^{+}) \tag{5.38}$$

We can see that $\overline{E}(z,t)$ and $\overline{H}(z,t)$ are in phase for this case.

Let's explore the properties of these waves like we did with transmission lines.

1. **Phase Velocity:** Recall that this is how fast we need to travel if we want to stay at the same point on the wave.

$$\begin{aligned} \xi &= \omega t - kz + \phi^+ = \text{constant} \\ z &= \frac{-\xi + \omega t + \phi^+}{k} \\ u_p &= \frac{dz}{dt} = \frac{\omega}{k} = \frac{\omega}{\omega\sqrt{\mu\epsilon}} = \frac{1}{\sqrt{\mu\epsilon}} \end{aligned} \tag{5.39}$$

2. **Properties in Vacuum:** In a vacuum (free space), $\mu = \mu_0 = 4\pi \times 10^{-7}$ H/m and $\epsilon = \epsilon_0 = 8.854 \times 10^{-12}$ F/m.

$$u_p = \frac{1}{\sqrt{\mu_0 \epsilon_0}} = c \approx 3 \times 10^8 \text{ m/s (speed of light in vacuum)} \tag{5.40}$$

$$\eta = \eta_0 = \sqrt{\frac{\mu_0}{\epsilon_0}} \approx 377\ \Omega \approx 120\pi\ \Omega \text{ (intrinsic impedance of vacuum)} \tag{5.41}$$

3. **Wavelength:** Wavelength is the distance in z necessary to go one complete cycle of the sinusoid.

$$\begin{aligned} kz|_{z=\lambda} &= k\lambda = 2\pi \\ \lambda &= \frac{2\pi}{k} = \frac{2\pi}{\omega/u_p} = \frac{u_p}{f} \\ u_p &= f\lambda \\ k &= \frac{2\pi}{\lambda} = \omega\sqrt{\mu\epsilon} = \frac{\omega}{u_p} \end{aligned} \tag{5.42}$$

5.2.3 Relation between $\overline{E}$ and $\overline{H}$

Consider the ∇ operator acting on an arbitrary plane wave. For example, the gradient is

$$\nabla e^{-j\overline{k}\cdot\overline{r}} = \left(\hat{x}\frac{\partial}{\partial x} + \hat{y}\frac{\partial}{\partial y} + \hat{z}\frac{\partial}{\partial z}\right) e^{-jk_x x - jk_y y - jk_z z} \tag{5.43}$$

$$= \left[\hat{x}(-jk_x) + \hat{y}(-jk_y) + \hat{z}(-jk_z)\right] e^{-j\overline{k}\cdot\overline{r}} \tag{5.44}$$

$$= -j\left[k_x\hat{x} + k_y\hat{y} + k_z\hat{z}\right] e^{-j\overline{k}\cdot\overline{r}} \tag{5.45}$$

$$= -j\overline{k}\, e^{-j\overline{k}\cdot\overline{r}} \tag{5.46}$$

Thus, $\nabla \to -j\overline{k}$ for plane waves. Maxwell's equations become:

$$\begin{array}{ccc} \textbf{Time Harmonic Equation} & & \textbf{Plane Wave} \\ \nabla \times \overline{E} = -j\omega\mu\overline{H} & \Rightarrow & -j\overline{k} \times \overline{E} = -j\omega\mu\overline{H} \\ \nabla \times \overline{H} = j\omega\epsilon\overline{E} & \Rightarrow & -j\overline{k} \times \overline{H} = j\omega\epsilon\overline{E} \\ \nabla \cdot \left(\epsilon\overline{E}\right) & \Rightarrow & -j\overline{k} \cdot \epsilon\overline{E} = 0 \\ \nabla \cdot \left(\mu\overline{H}\right) & \Rightarrow & -j\overline{k} \cdot \mu\overline{H} = 0 \end{array} \tag{5.47}$$

To simplify these expressions, we let $\hat{k}$ be the unit vector in the direction of propagation ($\overline{k}$ direction), so that

$$\hat{k} = \frac{\overline{k}}{\|\overline{k}\|} = \frac{\overline{k}}{\omega\sqrt{\mu\epsilon}}. \tag{5.48}$$

Then Maxwell's equations for a single plane wave become

$$\begin{array}{rcll} \omega\sqrt{\mu\epsilon}\,\hat{k}\times\overline{E} = \omega\mu\overline{H} & \Rightarrow & \hat{k}\times\overline{E} = \eta\overline{H} & \\ \omega\sqrt{\mu\epsilon}\,\hat{k}\times\overline{H} = -\epsilon\overline{E} & \Rightarrow & \hat{k}\times\overline{H} = -\dfrac{\overline{E}}{\eta} & \\ -j\omega\sqrt{\mu\epsilon}\epsilon\,\hat{k}\cdot\overline{E} = 0 & \Rightarrow & \hat{k}\cdot\overline{E} = 0 & \hat{k}\perp\overline{E} \\ -j\omega\sqrt{\mu\epsilon}\mu\,\hat{k}\cdot\overline{H} = 0 & \Rightarrow & \hat{k}\cdot\overline{H} = 0 & \hat{k}\perp\overline{H} \end{array} \tag{5.49}$$

A convenient way to find $\overline{H}$ given $\overline{E}$ is to use the second equation to find the magnitude, $\|\overline{H}\| = \|\overline{E}\|/\eta$, and then choose the direction of $\overline{H}$ so that $\overline{E}\times\overline{H}$ is in the direction of propagation ($\overline{k}$ direction).

5.2.4 Differential Forms

It is interesting to look at plane waves using differential form pictures. For the example given above,

$$E(z,t) = |E_{xo}^{+}|\cos(\omega t - kz + \phi^{+})\,dx \tag{5.50}$$

$$H(z,t) = \frac{|E_{xo}^{+}|}{\eta}\cos(\omega t - kz + \phi^{+})\,dy \tag{5.51}$$

One cycle of the plane wave is shown in Fig. 5.1. As time increases, the picture moves in the direction of propagation. The surfaces of E and H intersect to form tubes in the z direction, representing the flow of energy carried by the wave.

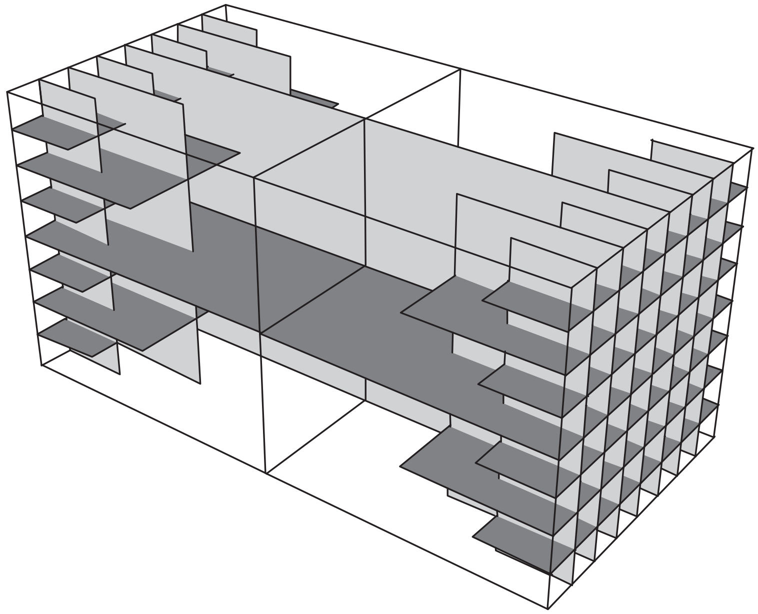


Figure 5.1: Electric and magnetic field intensity one-forms for a plane wave.

5.3 Plane Wave Polarization

Polarization of a wave is the shape the tip of the electric field vector $\overline{E}$ traces as a function of time at a point in space.

In general, this shape is an ellipse, so we say the wave has *elliptical polarization*. Special cases are *circular polarization* and *linear polarization*.

We will assume a forward traveling wave and suppress the '+' superscript.

$$\overline{E} = \hat{x}E_x(z) + \hat{y}E_y(z) = \hat{x}E_{xo}e^{-jkz} + \hat{y}E_{yo}e^{-jkz} = E_{xo}\left(\hat{x} + \hat{y}\frac{E_{yo}}{E_{xo}}\right)e^{-jkz} \tag{5.52}$$

5.3.1 Linear Polarization

If $\alpha = E_{yo}/E_{xo}$ is real, then the wave is *linearly polarized*. In this case,

$$\begin{aligned} \overline{E}(z) &= E_{xo}\,(\hat{x} + \alpha\hat{y})\,e^{-jkz} \\ \overline{\mathcal{E}}(z,t) &= |E_{xo}|\,\{\hat{x}\cos(\omega t - kz + \phi_x) + \hat{y}\alpha\cos(\omega t - kz + \phi_x)\} \\ &= |E_{xo}|(\hat{x} + \alpha\hat{y})\cos(\omega t - kz + \phi_x) \end{aligned} \tag{5.53}$$

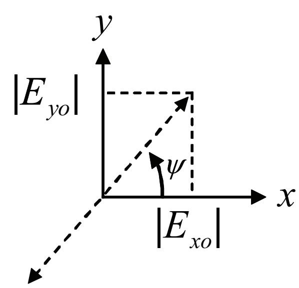
At a point in space (let's choose $z = 0$)

$$\overline{\mathcal{E}}(0,t) = |E_{xo}|(\hat{x} + \alpha\hat{y})\cos(\omega t + \phi_x) \tag{5.54}$$

From this expression, we can see that the vector oscillates in time but always along a line with a fixed angle from the x-axis. The angle of the line is

$$\psi = \tan^{-1}\alpha = \tan^{-1}\left(\frac{E_{yo}}{E_{xo}}\right) \tag{5.55}$$

The components E_{yo} and E_{xo} of the plane wave can be complex, but if the ratio is real, then the polarization is linear.



Linear polarization example: $\overline{E} = [(2 + j2)\hat{x} + \hat{y}(-5 - j5)]\, e^{-jkz}$

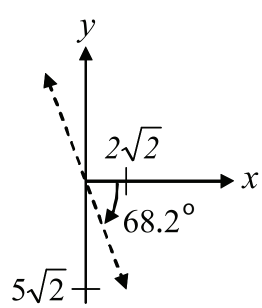
For this plane wave, $E_{xo} = 2 + j2$, $E_{yo} = -5 - j5$. The ratio of the x and y components is

$$\alpha = \frac{E_{yo}}{E_{xo}} = \frac{-5 - j5}{2 + j2} = -\frac{5}{2}\frac{1 + j1}{1 + j1} = -\frac{5}{2} = -2.5$$

The angle of the line traced by the tip of the electric field intensity vector is

$$\psi = \tan^{-1}\left(\frac{E_{yo}}{E_{xo}}\right) = -68.2^\circ \tag{5.56}$$

The magnitudes of the x and y components of the electric field are $|E_{xo}| = 2\sqrt{2}$ and $|E_{yo}| = 5\sqrt{2}$, from which we can plot the line traced by the electric field in the figure below.



The polarization of this wave is linear at an angle of -68.2° from the x axis.

5.3.2 Circular Polarization

If α is imaginary, so that

$$\alpha = \frac{E_{yo}}{E_{xo}} = \pm j \tag{5.57}$$

then the wave is circularly polarized. This means that the components of the electric field are equal in magnitude and 90° out of phase.

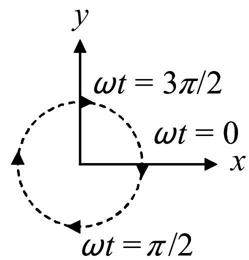
Let's consider an example for which $\alpha = +j$:

$$\begin{aligned}\overline{E}(z) &= E_{xo}(\hat{x} + j\hat{y})e^{-jkz} = E_{xo}(\hat{x} + \hat{y}e^{j\pi/2})e^{-jkz} \\ \overline{\mathcal{E}}(z,t) &= |E_{xo}|\left\{\hat{x}\cos(\omega t - kz + \phi_x) + \hat{y}\cos(\omega t - kz + \phi_x + \pi/2)\right\} \\ &= |E_{xo}|\left\{\hat{x}\cos(\omega t - kz + \phi_x) - \hat{y}\sin(\omega t - kz + \phi_x)\right\}\end{aligned} \tag{5.58}$$

At $z = 0$ (we will assume $\phi_x = 0$ for simplicity) the field is

$$\overline{\mathcal{E}}(0,t) = |E_{xo}|\left\{\hat{x}\cos(\omega t) - \hat{y}\sin(\omega t)\right\} \tag{5.59}$$

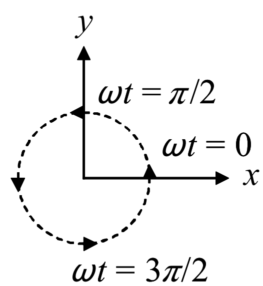
As time increases, the tip of the vector $\overline{\mathcal{E}}(0,t)$ traces out a circle. We assign a sense to the circle. Put your thumb in the direction of propagation, and your fingers in the direction of rotation of $\overline{\mathcal{E}}(0,t)$. In this case, this works out for the left-hand. We therefore call it *Left-Hand Circular Polarization (LHCP)*.



If $\alpha = -j$,

$$\begin{aligned}\overline{E}(z) &= E_{xo}(\hat{x} - \hat{y}e^{j\pi/2})e^{-jkz} \\ \overline{\mathcal{E}}(z,t) &= |E_{xo}|\left\{\hat{x}\cos(\omega t - kz + \phi_x) + \hat{y}\sin(\omega t - kz + \phi_x)\right\} \\ \overline{\mathcal{E}}(0,t) &= |E_{xo}|\left\{\hat{x}\cos(\omega t) + \hat{y}\sin(\omega t)\right\} \text{ for } \phi_x = 0\end{aligned} \tag{5.60}$$

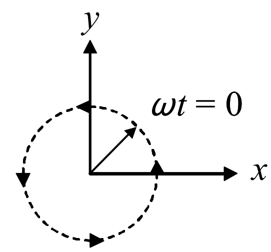
The rule works out for the right hand. Therefore, we call this *Right-Hand Circular Polarization*.



Example: $E_{xo} = 2 + j2$, $E_{yo} = 2 - j2$

$$\begin{aligned}\frac{E_{yo}}{E_{xo}} &= \frac{2 - j2}{2 + j2} = \frac{(2 - j2)(2 - j2)}{8} = -\frac{j8}{8} = -j \\ |E_{xo}| &= |E_{yo}| = 2\sqrt{2} \\ E_{xo} &= 2\sqrt{2}e^{j\pi/4}\end{aligned}$$

So, this will be RHCP with a radius of $2\sqrt{2}$. Because $\phi_x = \pi/4$, at $z = 0$ and $t = 0$, the vector will point at $45°$.



Example: $E_{xo} = 2 - j2$, $E_{yo} = 2 + j2$

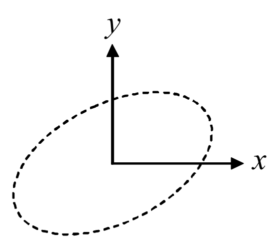
$$\frac{E_{yo}}{E_{xo}} = +j \tag{5.61}$$

$$E_{xo} = 2\sqrt{2}e^{-j\pi/4} \tag{5.62}$$

So, this will be LHCP with a radius of $2\sqrt{2}$. Because $\phi_x = -\pi/4$, at $z = 0$ and $t = 0$, the vector will point at $-45°$.

5.3.3 Elliptical Polarization

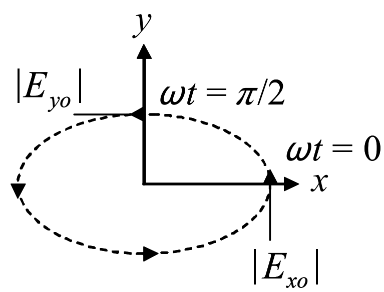
This is the most general case. The tip of the electric field will trace out an ellipse. We can in general have an ellipse rotated from the x-y axes as shown.



As an example of an elliptically polarized wave, let's consider a case with

$$\begin{aligned}
\alpha &= \frac{E_{yo}}{E_{xo}} = -j\frac{|E_{yo}|}{|E_{xo}|} \\
\overline{E}(z) &= E_{xo}\left(\hat{x} + \hat{y}\frac{|E_{yo}|}{|E_{xo}|}e^{-j\pi/2}\right)e^{-jkz} \\
\overline{\mathcal{E}}(z,t) &= |E_{xo}|\left\{\hat{x}\cos(\omega t - kz + \phi_x) + \hat{y}\frac{|E_{yo}|}{|E_{xo}|}\sin(\omega t - kz + \phi_x)\right\} \\
&= \hat{x}|E_{xo}|\cos(\omega t - kz + \phi_x) + \hat{y}|E_{yo}|\sin(\omega t - kz + \phi_x)
\end{aligned} \tag{5.63}$$

In this case for $z = 0$ and $\phi_x = 0$, we get the picture below. Notice that the ellipse has a major (longer) axis and a minor (shorter) axis. The ratio of the length of the major axis to the length of the minor axis is call the *Axial Ratio* ($1 \leq$ axial ratio $< \infty$).



An axial ratio of one represents circular polarization and an infinite axial ratio represents linear polarization. The axial ratio for the wave radiated by an antenna is often expressed in dB. For a circularly polarized antenna, we want an axial ratio as close to 0 dB as possible. For a linearly polarized antenna, the polarization is never perfect, but we hope to have an axial ratio that is as large as possible.

5.4 Electromagnetic Power Density

5.4.1 Poynting Vector

Consider Maxwell's equations in the time domain modified as follows:

$$\overline{\mathcal{H}} \cdot \nabla \times \overline{\mathcal{E}} = -\overline{\mathcal{H}} \cdot \frac{\partial \overline{\mathcal{B}}}{\partial t} \tag{5.64}$$

$$\overline{\mathcal{E}} \cdot \nabla \times \overline{\mathcal{H}} = \overline{\mathcal{E}} \cdot \frac{\partial \overline{\mathcal{D}}}{\partial t} + \overline{\mathcal{E}} \cdot \overline{\mathcal{J}} \tag{5.65}$$

The divergence of a cross product can be expanded using the identity $\nabla \cdot (\overline{\mathcal{E}} \times \overline{\mathcal{H}}) = \overline{\mathcal{H}} \cdot \nabla \times \overline{\mathcal{E}} - \overline{\mathcal{E}} \cdot \nabla \times \overline{\mathcal{H}}$, which is analogous to the product rule for the scalar derivative. Therefore,

$$\nabla \cdot (\overline{\mathcal{E}} \times \overline{\mathcal{H}}) = -\overline{\mathcal{H}} \cdot \frac{\partial \overline{\mathcal{B}}}{\partial t} - \overline{\mathcal{E}} \cdot \frac{\partial \overline{\mathcal{D}}}{\partial t} - \overline{\mathcal{E}} \cdot \overline{\mathcal{J}} \tag{5.66}$$

To further simplify this expression, we use

$$\frac{\partial}{\partial t}\left(\frac{1}{2}\mu\|\overline{\mathcal{H}}\|^2\right) = \frac{\partial}{\partial t}\left(\frac{1}{2}\mu\overline{\mathcal{H}} \cdot \overline{\mathcal{H}}\right) = \frac{1}{2}\mu\left(\frac{\partial \overline{\mathcal{H}}}{\partial t} \cdot \overline{\mathcal{H}} + \overline{\mathcal{H}} \cdot \frac{\partial \overline{\mathcal{H}}}{\partial t}\right) = \overline{\mathcal{H}} \cdot \frac{\partial(\mu\overline{\mathcal{H}})}{\partial t} = \overline{\mathcal{H}} \cdot \frac{\partial \overline{\mathcal{B}}}{\partial t}$$

We will also divide the current into induced and impressed parts, using $\overline{\mathcal{J}} = \sigma\overline{\mathcal{E}} + \overline{\mathcal{J}}_{\rm imp}$. This leads to the differential or point form of Poynting's theorem,

$$\nabla \cdot (\overline{\mathcal{E}} \times \overline{\mathcal{H}}) + \frac{\partial}{\partial t}\left(\frac{1}{2}\mu\|\overline{\mathcal{H}}\|^2\right) + \frac{\partial}{\partial t}\left(\frac{1}{2}\epsilon\|\overline{\mathcal{E}}\|^2\right) + \sigma\overline{\mathcal{E}} \cdot \overline{\mathcal{E}} = -\overline{\mathcal{E}} \cdot \overline{\mathcal{J}}_{\rm imp}$$

Let's integrate over a volume V and apply the divergence theorem to the first term:

$$\boxed{\oint_S (\overline{\mathcal{E}} \times \overline{\mathcal{H}}) \cdot d\overline{s} + \frac{\partial}{\partial t}\int_V \left[\frac{1}{2}\mu\|\overline{\mathcal{H}}\|^2 + \frac{1}{2}\epsilon\|\overline{\mathcal{E}}\|^2\right] dV + \int_V \sigma\|\overline{\mathcal{E}}\|^2 dV = -\int_V \overline{\mathcal{E}} \cdot \overline{\mathcal{J}}_{\rm imp} dV} \tag{5.67}$$

This is the integral form of **Poynting's Theorem** and represents a power balance or conservation of energy for electromagnetic fields. The units of each term (after integration) is Watts. We identify each term as:

$$\oint_S (\overline{\mathcal{E}} \times \overline{\mathcal{H}}) \cdot d\overline{s} = \text{ total power leaving the volume through the surface } S \tag{5.68a}$$

$$\frac{1}{2}\mu\|\overline{\mathcal{H}}\|^2 = \text{ stored magnetic energy density inside } V \tag{5.68b}$$

$$\frac{1}{2}\epsilon\|\overline{\mathcal{E}}\|^2 = \text{ stored electric energy density inside } V \tag{5.68c}$$

$$\frac{\partial}{\partial t}\int_V \left[\frac{1}{2}\mu\|\overline{\mathcal{H}}\|^2 + \frac{1}{2}\epsilon\|\overline{\mathcal{E}}\|^2\right] dV = \text{ rate of } \textbf{increase} \text{ of stored energy inside } V \tag{5.68d}$$

$$\int_V \sigma\|\overline{\mathcal{E}}\|^2 dV = \text{ power lost to heat inside } V \tag{5.68e}$$

$$-\int_V \overline{\mathcal{E}} \cdot \overline{\mathcal{J}}_{\rm imp} dV = \text{ power radiated by impressed current sources} \tag{5.68f}$$

Poynting's theorem states that *the power leaving the volume + the rate of increase in the stored energy + the power going into heat = power radiated by impressed sources.*

The vector field $\overline{\mathcal{E}} \times \overline{\mathcal{H}}$ has units of W/m^2. It represents the density of power carried by electromagnetic waves across the surface S. We call it the **Poynting Vector**:

$$\overline{\mathcal{S}} = \overline{\mathcal{E}} \times \overline{\mathcal{H}} \tag{5.69}$$

This quantity is analogous to instantaneous power $p(t) = v(t)i(t)$, except that $\overline{\mathcal{S}}$ is a power density and has units W/m^2.

In the phasor domain, the average power delivered to a load in a circuit is

$$P = \frac{1}{2}\,\mathrm{Re}\{\tilde{V}\tilde{I}^*\}$$

Similarly, the time-average Poynting vector indicates the average real power density of a time-harmonic wave:

$$\overline{S}_{\mathrm{av}} = \frac{1}{2}\,\mathrm{Re}\{\overline{E} \times \overline{H}^*\} \tag{5.70}$$

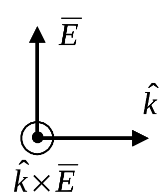
where $\overline{S} = \overline{E} \times \overline{H}^*$ is the complex Poynting vector. The real part of the complex Poynting vector represents the flow of power. The imaginary part represents power that is exchanged between energy stored in the electric and magnetic fields without dissipating or propagating away. This is a field version of the imaginary power represented by $\tilde{V}\tilde{I}^*$ in an LC circuit with no resistance or loss.

So, there are three power quantities associated with electromagnetic fields:

1. Instantaneous Poynting vector (time-domain): $\overline{\mathcal{S}} = \overline{\mathcal{E}} \times \overline{\mathcal{H}}$.
2. Time-average Poynting vector (time-harmonic fields): $\overline{S}_{\mathrm{av}} = \frac{1}{T}\int_0^T \overline{\mathcal{S}}\,dt$, $T = 2\pi/\omega$.
3. Complex Poynting vector (phasor domain): $\overline{S} = \overline{E} \times \overline{H}^*$

The time-average power can be obtained from the complex Poynting vector using $\overline{S}_{\mathrm{av}} = \frac{1}{2}\,\mathrm{Re}\{\overline{E} \times \overline{H}^*\}$.

5.4.2 Poynting Vector for Plane Waves



We know that $\overline{H}(z) = \hat{k} \times \overline{E}(z)/\eta_c$. Therefore

$$\overline{S} = \overline{E} \times \overline{H}^* = \frac{1}{\eta_c^*}\overline{E} \times (\hat{k} \times \overline{E})^* = \frac{1}{\eta_c^*}|\overline{E}|^2\hat{k}$$

The time-average power flux is

$$\overline{S}_{\mathrm{av}} = \frac{1}{2}\,\mathrm{Re}\{\overline{E} \times \overline{H}^*\} = \frac{|\overline{E}|^2}{2}\,\mathrm{Re}\left\{\frac{1}{\eta_c^*}\right\}\hat{k}$$

For lossless media,

$$\overline{S}_{\text{av}} = \frac{|\overline{E}|^2}{2\eta}\hat{k} \tag{5.71}$$

So, the power flow is in the direction of propagation of the wave ($\hat{k}$) and the coefficient is analogous to V^2/R.

Example: Power in a Wireless Communication Link

A Wi-Fi transmitter radiates 50 mW total power. Ignoring the details of the radiation pattern of the transmitting antenna, the radiated power is spread roughly evenly over a sphere as the energy radiates away from the transmitter. A receiving antenna is ten meters from the transmitter. At this distance, the power density incident on the receiver is approximately

$$S_{\text{inc}} \simeq \frac{P_{\text{rad}}}{4\pi R^2} \simeq 40\ \mu\text{W/m}^2 \tag{5.72}$$

Near the receiver, we can approximate the incident field as a plane wave. The intensity of the plane wave can be found by using (5.71) in reverse. The electric field intensity of the incident field at the receiver is

$$E^{\text{inc}} = \sqrt{2\eta S_{\text{inc}}} \simeq 0.17\ \text{V/m} \tag{5.73}$$

Example: Solar Illumination

$|\overline{S}_{\text{av}}|$ = 1 kW/m^2 at the Earth's surface due to sun radiation.

R_e = 6380 km = earth radius

$R_s = 1.5 \times 10^8$ km = radius of earth's orbit around sun

1. Find the total power radiated by the sun

$$P_{\text{sun}} = S_{\text{av}}(4\pi R_s^2) = 2.8 \times 10^{26} W$$

2. Find the total power intercepted by the earth. The earth's cross sectional area is $A_e = \pi R_e^2$.

$$P_{\text{earth}} = S_{\text{av}}\pi R_e^2 = 1.28 \times 10^{17} W$$

3. Find the electric field strength at the earth (assuming single frequency)

$$S_{\text{av}} = \frac{|E_0|^2}{2\eta_0} \to |E_0| = \sqrt{2\eta_0 S_{\text{av}}} = 870\ \text{V/m}$$

5.5 Lossy Media

Let us now consider the wave equation for lossly media. In this case, $\sigma > 0$ and the wave equation is

$$\nabla^2 \overline{E} - \gamma^2 \overline{E} = 0 \tag{5.74}$$

$$\gamma^2 = -\omega^2 \mu \epsilon_c = -\omega^2 \mu(\epsilon - j\sigma/\omega) = -\omega^2 \mu(\epsilon' - j\epsilon'') \tag{5.75}$$

The complex propagation constant γ governs the behavior of the wave. We can find the propagation constant using

$$\boxed{\gamma = \sqrt{-\omega^2 \mu(\epsilon - j\sigma/\omega)}} \tag{5.76}$$

This formula can be used to solve plane wave problems for lossy materials. Once we have γ, we can understand how the plane wave behaves in any lossy material, ranging from good conductors (high conductivity) to good dielectrics (low conductivity).

The propagation constant has real and imaginary parts,

$$\gamma = \sqrt{-\omega^2 \mu(\epsilon - j\sigma/\omega)} = \alpha + j\beta \tag{5.77}$$

The imaginary part, β, governs the phase velocity and wavelength and is similar to the wavenumber k that is used for propagation in lossless materials. The real part, α, governs the decay of the wave due to absorption of the power in the wave and dissipation as thermal energy or heating of the material. A low loss material has a small value for α and the wave propagates a long distance into the material. For a good conductor, α is large, power in a wave is rapidly absorbed by the material, and the penetration depth of a wave into the material is very small.

5.5.1 Plane Waves

Now, if we simplify the wave equation for a uniform plane wave just like we did in the lossless case, we obtain

$$\frac{d^2 E_x}{dz^2} - \gamma^2 E_x = 0 \tag{5.78}$$

The solution to this differential equation is

$$\begin{aligned} E_x(z) &= E_{xo}^+ e^{-\gamma z} + E_{xo}^- e^{\gamma z} \\ &= E_{xo}^+ e^{-\alpha z} e^{-j\beta z} + E_{xo}^- e^{\alpha z} e^{j\beta z} \end{aligned} \tag{5.79}$$

So, the wave decays as it propagates. This means we take $\alpha > 0$, $\beta > 0$ when we take the square root of γ^2.

To determine $\overline{H}$, we use Faraday's law, $\nabla \times \overline{E} = -j\omega\mu\overline{H}$, so that

$$\overline{H} = \frac{1}{\eta_c} \hat{k} \times \overline{E} \tag{5.80}$$

$$\eta_c = \sqrt{\frac{\mu}{\epsilon_c}} \tag{5.81}$$

where η_c is the intrinsic impedance of the lossy medium. Since η_c is a complex number, $\overline{E}$ and $\overline{H}$ are no longer in phase.

5.5.2 Skin Depth

For a $+z$-traveling wave, the magnitude of the electric field is

$$|E_x(z)| = |E_{xo}^{+}E^{-\alpha z}e^{-j\beta z}| = |E_{xo}^{+}|e^{-\alpha z} \tag{5.82}$$

The propagation distance required to attenuate the wave by a factor of e^{-1} is called the *skin depth* δ_s:

$$|E_x(z = \delta_s)| = |E_{xo}^{+}|e^{-1} \rightarrow \delta_s = \frac{1}{\alpha} \tag{5.83}$$

Extremes:	Perfect Conductor	$\sigma = \infty$	$\alpha = \infty$	$\delta_s = 0$
	Dielectric	$\sigma = 0$	$\alpha = 0$	$\delta_s = \infty$

When an AC current flows in a conductor, since $\overline{E}$ decays rapidly, the current $\overline{J} = \sigma\overline{E}$ is concentrated near the conductor surface. In a perfect conductor, the current becomes a surface current density. This is called the skin effect.

5.5.3 Loss Tangent

The loss tangent is simply a commonly-used parameter to describe the loss of a medium. It is defined as:

$$\text{Loss Tangent} = \tan\delta = \frac{\epsilon''}{\epsilon'} \tag{5.84}$$

Often, materials are specified by ϵ' and $\tan\delta$ at a certain frequency:

Polystyrene Foam:	$\epsilon' = 1.03\epsilon_0$	$\tan\delta$ = 0.3×10^{-4}	$f = 3$GHz
Fresh Snow:	$\epsilon' = 1.20\epsilon_0$	$\tan\delta$ = 3×10^{-4}	$f = 3$GHz
Round Steak:	$\epsilon' = 40\epsilon_0$	$\tan\delta$ = 0.3	$f = 3$GHz

Let's put the round steak in the microwave oven (not my favorite way to prepare steak). The complex permittivity is

$$\begin{aligned}\epsilon_c = \epsilon' - j\epsilon'' = \epsilon'\left(1 - j\frac{\epsilon''}{\epsilon'}\right) &= \epsilon'\left(1 - j\tan\delta\right) \\ &= 40\left(1 - j0.3\right)\epsilon_0\end{aligned}$$

$$\gamma = j\omega\sqrt{\mu_0\epsilon_c} = j\omega\sqrt{\mu_0\epsilon_0}\sqrt{40(1-j0.3)} = j\frac{2\pi}{\lambda_0}\sqrt{40(1-j0.3)}$$

At f = 3GHz, $\lambda_0 = 10\,\text{cm}$ = 0.1m:

$$\gamma = \alpha + j\beta = 59 + j402\text{m}^{-1}$$

$$\delta_s = \frac{1}{\alpha} = 0.017\text{m} = 1.7\text{cm}$$

So, the microwave oven heats the surface more rapidly that it heats the center (contrary to popular belief). However, it is true that a microwave immediately starts heating the center (not all heat arrives at the center through heat conduction). For polystyrene foam:

$$\epsilon_c = 1.03\left(1 - j0.3 \times 10^{-4}\right)\epsilon_0$$

$$\gamma = 9.6 \times 10^{-4} + j63.8\text{m}^{-1}$$

Since α is so small, very little wave attenuation (and therefore heating) occurs. This is why you can reheat your meat in a styrofoam box in the microwave without the box getting hot.

5.5.4 Approximations for Good Dielectrics and Good Conductors

We now want to look at approximations for α and β for two limiting cases: low loss dielectrics and good conductors. This helps us to have a more qualitative understanding of how the various material properties (σ, f, ϵ_r) parameters affect plane wave propagation in different types of materials.

Separate formulas for the real and imaginary parts of γ can be found from

$$\gamma^2 = (\alpha + j\beta)^2 = -\omega^2\mu\epsilon' + j\omega^2\mu\epsilon'' \tag{5.85}$$

$$\alpha^2 - \beta^2 + j2\alpha\beta = -\omega^2\mu\epsilon' + j\omega^2\mu\epsilon'' \tag{5.86}$$

Equating real and imaginary parts and solving, we obtain

$$\alpha = \omega\left\{\frac{\mu\epsilon'}{2}\left[\sqrt{1+\left(\frac{\epsilon''}{\epsilon'}\right)^2} - 1\right]\right\}^{1/2} \quad \text{Np/m} \tag{5.87}$$

$$\beta = \omega\left\{\frac{\mu\epsilon'}{2}\left[\sqrt{1+\left(\frac{\epsilon''}{\epsilon'}\right)^2} + 1\right]\right\}^{1/2} \quad \text{rad/m} \tag{5.88}$$

We will look at two cases: (1) if $\epsilon'' \ll \epsilon'$ the material is a low loss medium, and (2) if $\epsilon'' \gg \epsilon'$ the material is a good conductor.

Low loss dielectrics. For a low-loss dielectric, the expression for γ can be put in the form

$$\gamma = j\omega\sqrt{\mu\epsilon'}\left(1 - j\frac{\epsilon''}{\epsilon'}\right)^{1/2} \tag{5.89}$$

The second square root can be approximated using the first two terms of the binomial expansion $\sqrt{1+\Delta} \simeq 1 + \Delta/2$. This results in

$$\gamma \simeq j\omega\sqrt{\mu\epsilon'}\left(1 - j\frac{\epsilon''}{2\epsilon'}\right) \tag{5.90}$$

The real and imaginary part are

$$\alpha \simeq \frac{\omega\epsilon''}{2}\sqrt{\frac{\mu}{\epsilon'}} = \frac{\sigma}{2}\sqrt{\frac{\mu}{\epsilon}} \tag{5.91}$$

$$\beta \simeq \omega\sqrt{\mu\epsilon'} = \omega\sqrt{\mu\epsilon} \tag{5.92}$$

This expression shows that β is the same as in the lossless case, so the plane propagation behavior for a low-loss medium is the same with the addition of a decay term.

The intrinsic impedance is also approximated using the binomial expansion as given by

$$\begin{aligned}\eta &\simeq \sqrt{\frac{\mu}{\epsilon'}}\left(1 + j\frac{\epsilon''}{2\epsilon'}\right) \\ &\simeq \sqrt{\frac{\mu}{\epsilon}}\end{aligned} \tag{5.93}$$

which is the same as it was for the lossless case.

Good conductors. For the case $\epsilon'' \gg \epsilon'$,

$$\gamma = \sqrt{-\omega^2\mu(\epsilon' - j\epsilon'')} \tag{5.94}$$
$$\simeq \sqrt{j\omega^2\mu\epsilon''} \tag{5.95}$$

We now substitute $\epsilon'' = \sigma/\omega$ and $\sqrt{j} = (1+j)/\sqrt{2}$ to get

$$\gamma = \sqrt{\frac{\omega\mu\sigma}{2}}\,(1+j) \tag{5.96}$$

resulting in

$$\alpha = \beta = \sqrt{\frac{\omega\mu\sigma}{2}} = \sqrt{\pi f\mu\sigma}. \tag{5.97}$$

In this case the propagation and decay constants are equal and change with frequency. The approximation for the intrinsic impedance follows a similar derivation, resulting in

$$\eta_c = \sqrt{j\frac{\mu}{\epsilon''}} = (1+j)\sqrt{\frac{\pi f\mu}{\sigma}}. \tag{5.98}$$

With a complex η the electric and magnetic fields are no longer in phase.

Is it valid to assume that dielectrics are low-loss and metals are good conductors?

$$\begin{array}{cc}
\text{Dielectric} & \text{Conductor} \\
1 \gg \frac{\epsilon''}{\epsilon'} & 1 \ll \frac{\epsilon''}{\epsilon'} \\
\frac{1}{100} > \frac{\epsilon''}{\epsilon'} & 100 < \frac{\epsilon''}{\epsilon'} \\
\frac{1}{100} > \frac{\sigma}{\omega\epsilon_r\epsilon_o} & 100 < \frac{\sigma}{\omega\epsilon_r\epsilon_o} \\
\omega > \frac{100\sigma}{\epsilon_r\epsilon_o} & \omega < \frac{\sigma}{100\epsilon_r\epsilon_o} \\
\omega > \frac{100\times10^{-12}}{(4)(8.854\times10^{-12})} & \omega < \frac{10^6}{(100)(8.854\times10^{-12})} \\
\omega > 2.8 \text{ rad/s} & \omega < 10^{12} \text{ rad/s}
\end{array} \tag{5.99}$$

This shows that the approximations derived in this section are accurate for dielectrics and conductors over a very wide frequency range.

Example: Sea Water

Let's look at plane wave propagation through sea water. The material parameters are

$\epsilon_r = 72$–80 (We will use $\epsilon_r = 80$)

$\sigma = 4$

What is the range for the good conductor approximation?

$$\frac{\epsilon''}{\epsilon'} > 100 \tag{5.100}$$
$$\frac{\sigma}{\omega\epsilon_r\epsilon} > 100 \tag{5.101}$$
$$\omega < 56 \text{ MHz} \tag{5.102}$$

What is the range for the low-loss dielectric approximation?

$$\frac{\epsilon''}{\epsilon'} < \frac{1}{100} \tag{5.103}$$
$$\frac{\sigma}{\omega\epsilon_r\epsilon} < \frac{1}{100} \tag{5.104}$$
$$\omega > 565 \text{ GHz} \tag{5.105}$$

At 1 KHz, the decay constant is

$$\alpha\,(1 \text{ kHz}) = \sqrt{\pi \cdot 10^3 \cdot 4 \cdot 4\pi \times 10^{-7}} = 0.126 \text{ Np/m} \tag{5.106}$$

and the skin depth is about 8 meters.

At microwave frequencies, neither approximation is valid, so we have to return to the original expression for γ to find the skin depth. So, for 1 GHz,

$$\begin{aligned}\gamma = j\omega\sqrt{\mu_0\epsilon_c} &= j2\pi 10^9\sqrt{4\pi \times 10^{-7} 8.854 \times 10^{-12}(80 - j4/(2\pi \times 10^9))} \\ &\simeq 78 \text{ Np/m} + j203 \text{ rad/m}\end{aligned}$$

The skin depth is about 1.3 cm.

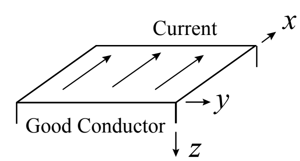
5.5.5 Current Flow in Good Conductors (Skin Effect)

If we have a DC current, the current will be uniformly distributed across the conductor cross section. However, in the AC case, due to the skin effect, the current is concentrated near the conductor surface.

Consider a semi-infinite slab of conducting material carrying a time-harmonic current. The current flows in the x direction and the amplitude of the current decays exponentially in the z direction away from the top surface of the conducting slab.

The fields at the surface of the conducting slab are

$$\overline{E}(z = 0^+) = \hat{x}E_0 \tag{5.107}$$
$$\overline{H}(z = 0^+) = \hat{y}\frac{E_0}{\eta_c} \tag{5.108}$$



Adding the z dependence for a wave in a lossy medium gives us the dependence of the fields on the depth into the slab:

$$\overline{E}(z) = \hat{x}E_0 e^{-\alpha z}e^{-j\beta z} \tag{5.109}$$
$$\overline{H}(z) = \hat{y}\frac{E_0}{\eta_c}e^{-\alpha z}e^{-j\beta z} \tag{5.110}$$

The current density associated with the wave is $\overline{J} = \sigma\overline{E} = \hat{x}\sigma E_0 e^{-\alpha z}e^{-j\beta z} = \hat{x}J_0 e^{-\alpha z}e^{-j\beta z}$. If $\epsilon'' \gg \epsilon'$ (good conductor), then $\alpha = \beta = 1/\delta_s$. The current density becomes

$$\overline{J} = \hat{x}J_0 e^{-(1+j)z/\delta_s} \tag{5.111}$$

Now, we explore the amount of current flowing through the region $0 \le y \le w$ and $0 \le z < \infty$ by integrating the current density:

$$\begin{aligned} I = \int_0^w \int_0^\infty J_0 e^{-(1+j)z/\delta_s} dz dy &= -J_0 w \frac{\delta_s}{1+j} \left[e^{-(1+j)\infty/\delta_s} - e^0 \right] \\ &= J_0 w \frac{\delta_s}{1+j} \end{aligned} \tag{5.112}$$

Let's suppose that instead we integrate in z only over a finite depth. The error in the integrated current density over a finite depth relative to the full integral in (5.112) becomes small as the depth becomes more than a few multiples of the skin depth:

Integral in z over	**Error in calculating I is**
$0 \le z \le 3\delta_s$	5%
$0 \le z \le 5\delta_s$	1%

Therefore, we can treat the conductor as infinitely thick as long as the thickness is larger than about $5\delta_s$. The basic principle is that the majority of the current flows within a few skin depths of the surface. For example, for copper we have at the frequency f = 1 GHz:

$$\begin{aligned} \sigma_c &= 5.8 \times 10^7 \text{ S/m} \\ \delta_s &= \frac{1}{\sqrt{\pi f \mu \sigma_c}} = 2.1\mu\text{m} \end{aligned}$$

This shows that 99% of the current flows within a layer at the surface of the copper that is only 10μm thick.

Resistance

The impedance is the voltage divided by the total current. The voltage along a path of length l in the x direction along the conducting slab is

$$\begin{aligned} V &= -\int \overline{E} \cdot dl \\ &= E_0 l \end{aligned}$$

The impedance is then given by

$$\begin{aligned} Z &= \frac{V}{I} \\ &= (E_0 l) \left(\frac{1+j}{\sigma E_0 w \delta_s} \right) \\ &= \underbrace{\frac{1+j}{\sigma \delta_s}}_{Z_s} \frac{l}{w} \end{aligned} \tag{5.113}$$

where the quantity Z_s is called surface impedance. The surface resistance is the real part of Z_s,

$$R_s = \frac{1}{\sigma \delta_s} = \sqrt{\frac{\pi f \mu}{\sigma}} \tag{5.114}$$

The resistance of the conducting slab is

$$R = R_s \frac{l}{w} \tag{5.115}$$

where l is the length in the direction of current flow and w is the width. The longer the slab, the higher the resistance. The wider the slab, the more surface is available to conduct current, and the lower the resistance.

As a further example, let's look at the resistance per unit length of a coaxial transmission line. The width of the inner conductor is $w_{\text{inner}} = 2\pi a$ and of the outer conductor is $w_{\text{outer}} = 2\pi b$. The resulting resistance per unit length is then

$$\begin{aligned} R' &= \sqrt{\frac{\pi f \mu}{\sigma}} \left(\frac{1}{w_{\text{inner}}} + \frac{1}{w_{\text{outer}}} \right) \\ &= \sqrt{\frac{\pi f \mu}{\sigma}} \left(\frac{1}{2\pi} \right) \left(\frac{1}{a} + \frac{1}{b} \right) \end{aligned} \tag{5.116}$$

This result takes into account the skin effect.

Chapter 6

Reflection and Refraction

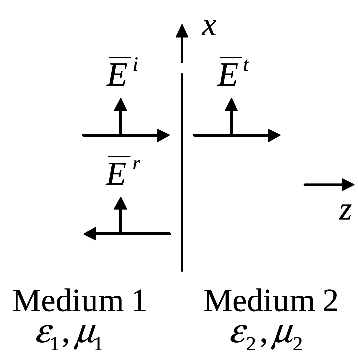
We have analyzed plane wave propagation in a homogeneous medium. We are now interested in exploring what happens when a plane wave traveling in one medium encounters an interface (boundary) with another medium. Understanding this phenomenon allows us to understand things like:

1. How an optical or microwave lens works
2. Why glass causes glare how to design anti-glare windows
3. How buildings and walls influence cellular phone signals
4. Why light bends when it enters water
5. Optical fibers

Many other devices and physical systems can be modeled and designed using the theory of plane wave reflection and transmission.

6.1 Normal Incidence

We will begin by looking at a wave normally incident on the interface between lossless media. The frequency ω is the same in all regions, but the wavenumbers $k_1 = \omega\sqrt{\mu_1\epsilon_1}$ and $k_2 = \omega\sqrt{\mu_2\epsilon_2}$ are different.



6.1.1 Fields

In general, we must allow for three different waves in the system:

1. Incident wave

$$\begin{aligned}\overline{E}^i(z) &= \hat{x}E_o^i e^{-jk_1 z}\\ \overline{H}^i(z) &= \frac{1}{\eta_1}\hat{k}^i\times\overline{E}^i = \hat{y}\frac{E_o^i}{\eta_1}e^{-jk_1 z} \qquad (\hat{k}^i=\hat{z})\\ k_1 &= \omega\sqrt{\mu_1\epsilon_1} \qquad \eta_1=\sqrt{\frac{\mu_1}{\epsilon_1}}\end{aligned}$$

2. Reflected wave

$$\begin{aligned}\overline{E}^r(z) &= \hat{x}E_o^r e^{+jk_1 z}\\ \overline{H}^r(z) &= \frac{1}{\eta_1}\hat{k}^r\times\overline{E}^r = -\hat{y}\frac{E_o^r}{\eta_1}e^{+jk_1 z} \qquad (\hat{k}^r=-\hat{z})\end{aligned}$$

3. Transmitted wave

$$\begin{aligned}\overline{E}^t(z) &= \hat{x}E_o^t e^{-jk_2 z}\\ \overline{H}^t(z) &= \frac{1}{\eta_2}\hat{k}^t\times\overline{E}^t = \hat{y}\frac{E_o^t}{\eta_2}e^{-jk_2 z} \qquad (\hat{k}^t=\hat{z})\\ k_2 &= \omega\sqrt{\mu_2\epsilon_2} \qquad \eta_2=\sqrt{\frac{\mu_2}{\epsilon_2}}\end{aligned}$$

We assume we know E_o^i. We must determine E_o^r and E_o^t. In these wave solutions there are two unknowns, so two equations are required. These equations are obtained from the electromagnetic boundary conditions that hold at the interface between the two materials.

6.1.2 Boundary Conditions

To find a relationship between the incident, reflected, and transmitted waves, we use boundary conditions. Since there is no surface current on the boundary between two dielectrics, the $\overline{E}$ and $\overline{H}$ boundary conditions are

$$\begin{aligned}\hat{n}=\hat{z}: \qquad & \hat{z}\times(\overline{E}_2-\overline{E}_1)\big|_{z=0}=0\\ & \hat{z}\times(\overline{H}_2-\overline{H}_1)\big|_{z=0}=0\end{aligned}$$

At $z=0$ we have:

$$\begin{aligned}\overline{E}_1 &= \overline{E}^i(0)+\overline{E}^r(0) = \hat{x}(E_o^i+E_o^r)\\ \overline{H}_1 &= \overline{H}^i(0)+\overline{H}^r(0) = \hat{y}\frac{(E_o^i-E_o^r)}{\eta_1}\\ \overline{E}_2 &= \overline{E}^t(0) = \hat{x}E_o^t\\ \overline{H}_2 &= \overline{H}^t(0) = \hat{y}\frac{E_o^t}{\eta_2}\end{aligned}$$

Therefore, application of the boundary conditions gives two equations:

$$\hat{z} \times \left[\hat{x}E_o^t - \hat{x}(E_o^i + E_o^r)\right] = 0 \quad \rightarrow \quad E_o^t - (E_o^i + E_o^r) = 0 \tag{6.1}$$

$$\hat{z} \times \left[\hat{y}\frac{E_o^t}{\eta_2} - \hat{y}\frac{(E_o^i - E_o^r)}{\eta_1}\right] = 0 \quad \rightarrow \quad \frac{E_o^t}{\eta_2} - \frac{(E_o^i - E_o^r)}{\eta_1} = 0 \tag{6.2}$$

Substitution of the first equation, $E_o^t = (E_o^i + E_o^r)$, into (6.2) gives

$$\frac{(E_o^i + E_o^r)}{\eta_2} - \frac{(E_o^i - E_o^r)}{\eta_1} = 0 \quad \rightarrow \quad E_o^r\left(\frac{1}{\eta_1} + \frac{1}{\eta_2}\right) = E_o^i\left(\frac{1}{\eta_1} - \frac{1}{\eta_2}\right)$$

$$E_o^r = \left(\frac{\eta_2 - \eta_1}{\eta_2 + \eta_1}\right) E_o^i = \Gamma E_o^i$$

$$E_o^t = E_o^i + \left(\frac{\eta_2 - \eta_1}{\eta_2 + \eta_1}\right) E_o^i = \left(\frac{\eta_2 + \eta_1 + \eta_2 - \eta_1}{\eta_2 + \eta_1}\right) E_o^i$$

$$E_o^t = \left(\frac{2\eta_2}{\eta_2 + \eta_1}\right) E_o^i = \tau E_o^i$$

So, for **Normal Incidence**:

$$\Gamma = \frac{\eta_2 - \eta_1}{\eta_2 + \eta_1} = \text{reflection coefficient} \tag{6.3}$$

$$\tau = \frac{2\eta_2}{\eta_2 + \eta_1} = \text{transmission coefficient} \tag{6.4}$$

Notice that we can write $E_o^t = E_o^i + \Gamma E_o^i = (1 + \Gamma)E_o^i$ as well as $E_o^t = \tau E_o^i$. Therefore, for normal incidence

$$\tau = 1 + \Gamma$$

We also observe that if region 2 is a perfect conductor, $\sigma_2 \rightarrow \infty$, $\epsilon_{c2} \rightarrow \infty$, and $\eta_2 = 0$. This results in $\Gamma = -1$, which is consistent with the boundary condition that the tangential electric field must go to zero at the boundary of the conductor (short circuit).

6.1.3 Dielectric Materials and Index of Refraction

For an interface between dielectric materials with relative permittivities ϵ_{r1} and ϵ_{r2}, the reflection and transmission coefficients can be simplified to

$$\Gamma = \frac{\sqrt{\mu_0/\epsilon_2} - \sqrt{\mu_0/\epsilon_1}}{\sqrt{\mu_0/\epsilon_2} + \sqrt{\mu_0/\epsilon_1}} = \frac{1/\sqrt{\epsilon_{r2}} - 1/\sqrt{\epsilon_{r1}}}{1/\sqrt{\epsilon_{r2}} + 1/\sqrt{\epsilon_{r1}}} \tag{6.5}$$

$$\tau = \frac{2/\sqrt{\epsilon_{r2}}}{1/\sqrt{\epsilon_{r2}} + 1/\sqrt{\epsilon_{r1}}} \tag{6.6}$$

Because the square root of the dielectric coefficient appears in reflection and transmission formulas, we often use the index of refraction

$$n = \sqrt{\epsilon_r} \tag{6.7}$$

in these formulas. In terms of the index of refraction, the reflection and transmission coefficients for a wave traveling from air to a dielectric half space are

$$\Gamma = \frac{1/n_2 - 1/n_1}{1/n_2 + 1/n_1} = \frac{n_1 - n_2}{n_1 + n_2} \tag{6.8}$$

$$\tau = \frac{2/n_2}{1/n_2 + 1/n_1} = \frac{2n_1}{n_1 + n_2} \tag{6.9}$$

6.1.4 Transmission Line Analogy

The expression for Γ is very similar to what we found in transmission lines, with the intrinsic impedance replacing the characteristic impedance. In fact, there is a direct correspondence between the voltage and current on a transmission line and the electric and magnetic fields at a dielectric interface:

Electromagnetic Wave	Electrical Transmission Line
$\overline{E}$	V
$\overline{H}$	I
η	Z_o
Γ	Γ_L
$\overline{E}_1 = \hat{x}\overline{E}_\text{o}^i \left(e^{-jk_1 z} + \Gamma e^{jk_1 z}\right)$	$V = V_\text{o}^+ \left(e^{-jkz} + \Gamma e^{jkz}\right)$

So, all the results of transmission line theory can be used to understand reflection and transmission. For example, the electric field has a standing wave pattern,

$$\begin{aligned} |\overline{E}_1| &= |E_\text{o}^i| \left|1 + \Gamma e^{j2k_1 z}\right| \\ |\overline{E}_1|_\text{max} &= |E_\text{o}^i| \left(1 + |\Gamma|\right) \\ |\overline{E}_1|_\text{min} &= |E_\text{o}^i| \left(1 - |\Gamma|\right) \end{aligned}$$

The standing wave ratio (SWR) is

$$S = \frac{|\overline{E}_1|_\text{max}}{|\overline{E}_1|_\text{min}} = \frac{1 + |\Gamma|}{1 - |\Gamma|}$$

This is exactly the same expression we obtained for transmission lines. As in transmission lines, the standing wave pattern is periodic with period $\lambda_1/2$:

$$e^{j2k_1\ell} = e^{j2\pi} \quad \rightarrow \quad k_1\ell = \pi \quad \rightarrow \quad \ell = \frac{\pi}{k_1} = \frac{\lambda_1}{2}$$

Again we see that the distance between adjacent maxima (or minima) is a half wavelength.

Continuing with the transmission line analogy, the intrinsic impedance of air

$$\eta_\text{air} = \sqrt{\frac{\mu_\text{o}}{\epsilon_\text{o}}} \simeq 377\ \Omega \tag{6.10}$$

can be viewed as the characteristic impedance of a transmission line. The characteristic impedance for a dielectric is

$$\eta = \sqrt{\frac{\mu_\text{o}}{\epsilon_r\ \epsilon_\text{o}}} = \frac{377}{n}\ \Omega, \tag{6.11}$$

The characteristic impedance of a conducting medium is

$$\eta \simeq (1+j)\sqrt{\frac{\pi f \mu}{\sigma}} \simeq 0 \tag{6.12}$$

if σ is large, so a conductor acts as a short circuit.

6.1.5 Power Flow in Region 1

In region 1, we have the fields

$$\begin{aligned}\overline{E}_1 &= \hat{x}E_{\mathrm{o}}^i\left(e^{-jk_1z} + \Gamma e^{jk_1z}\right)\\ \overline{H}_1 &= \hat{y}\frac{E_{\mathrm{o}}^i}{\eta_1}\left(e^{-jk_1z} - \Gamma e^{jk_1z}\right)\end{aligned}$$

The time-average Poynting vector is:

$$\begin{aligned}\overline{S}_{av,1} = \frac{1}{2}\,\mathrm{Re}\left\{\overline{E}_1 \times \overline{H}_1^*\right\} &= \frac{1}{2}\,\mathrm{Re}\left\{E_{\mathrm{o}}^i\left(e^{-jk_1z} + \Gamma e^{jk_1z}\right)\frac{E_{\mathrm{o}}^{i*}}{\eta_1}\left(e^{jk_1z} - \Gamma^* e^{-jk_1z}\right)\hat{x}\times\hat{y}\right\}\\ &= \frac{1}{2}\,\mathrm{Re}\left\{\frac{|E_{\mathrm{o}}^i|^2}{\eta_1}\left[1 - \Gamma^* e^{-j2k_1z} + \Gamma e^{j2k_1z} - |\Gamma|^2\right]\right\}\hat{z}\end{aligned}$$

Recall that $A - A^* = (a+jb) - (a-jb) = j2b = 2\,\mathrm{Im}\{A\}$, so that this can be simplified to

$$\begin{aligned}\overline{S}_{av,1} &= \frac{1}{2}\,\mathrm{Re}\left\{\frac{|E_{\mathrm{o}}^i|^2}{\eta_1}\left[1 - |\Gamma|^2 + 2\,\mathrm{Im}\left\{\Gamma e^{j2k_1z}\right\}\right]\right\}\hat{z}\\ &= \hat{z}\frac{|E_{\mathrm{o}}^i|^2}{2\eta_1}\left[1 - |\Gamma|^2\right]\end{aligned} \tag{6.13}$$

Note that

$$\begin{aligned}\overline{S}_{av}^i &= \hat{z}\frac{|E_{\mathrm{o}}^i|^2}{2\eta_1}\\ \overline{S}_{av}^r &= -\hat{z}\frac{|E_{\mathrm{o}}^i|^2}{2\eta_1}|\Gamma|^2 = -|\Gamma|^2\overline{S}_{av}^i\end{aligned}$$

So, $\overline{S}_{av,1}$ is simply the sum of the power densities in the incident and reflected waves. The average power density in region 2 is

$$\overline{S}_{av,2} = \hat{z}\frac{|\overline{E}_2|^2}{2\eta_2} = \hat{z}|\tau|^2\frac{|E_{\mathrm{o}}^i|^2}{2\eta_2} \tag{6.14}$$

For lossless media, Γ and τ are real, so that $|\Gamma|^2 = \Gamma^2$ and $|\tau|^2 = \tau^2$. Therefore,

$$\begin{aligned}\frac{1-\Gamma^2}{\eta_1} = \frac{1 - \left(\frac{\eta_2-\eta_1}{\eta_2+\eta_1}\right)^2}{\eta_1} &= \frac{(\eta_2+\eta_1)^2 - (\eta_2-\eta_1)^2}{\eta_1(\eta_2-\eta_1)^2}\\ &= \frac{\eta_2^2 + \eta_1^2 + 2\eta_1\eta_2 - \eta_2^2 - \eta_1^2 + 2\eta_1\eta_2}{\eta_1(\eta_2-\eta_1)^2}\\ &= \frac{4\eta_2}{(\eta_2-\eta_1)^2}\\ &= \frac{1}{\eta_2}\left[\frac{2\eta_2}{\eta_2-\eta_1}\right]^2 = \frac{\tau^2}{\eta_2}\end{aligned}$$

By using this result, we can see that the power in medium 1, Eq. (6.13), is equal to the power density in medium 2, Eq. (6.14), as we would expect based on conservation of energy.

6.1.6 Lossy Media

The results obtained above can be used for the lossless case by simply using complex intrinsic impedances in the formulas for reflection and transmission coefficients:

$$\begin{aligned}\eta_1 &\rightarrow \eta_{c1}\\ \eta_2 &\rightarrow \eta_{c2}\end{aligned}$$

6.2 Oblique (Non-normal) Incidence

We need to first become familiar with the idea of waves traveling in a direction other than $\pm z$. In order for the phase progression to occur along the $\hat{k}$ direction, we need to use the component of the position vector $\overline{r}$ that points along the $\hat{k}$ direction. So, our distance variable is

$$\begin{aligned}\xi^i = \hat{k}^i \cdot \overline{r} &= (\hat{x}\sin\theta_i + \hat{z}\cos\theta_i)\cdot(\hat{x}x + \hat{z}z)\\ &= x\sin\theta_i + z\cos\theta_i\end{aligned}$$

The plane wave exponential is

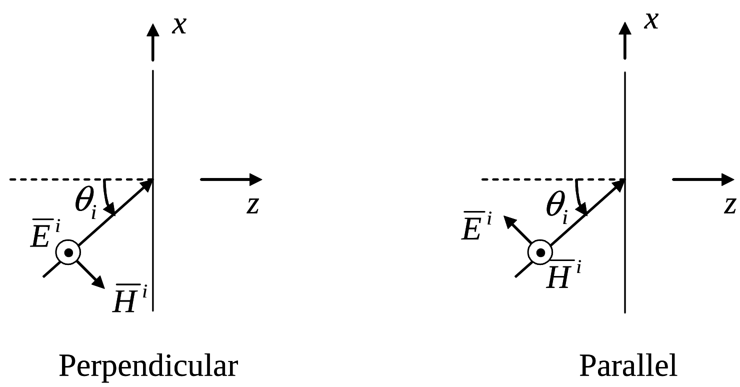
$$e^{-jk_1\hat{k}^i\cdot\overline{r}} = e^{-jk_1(x\sin\theta_i + z\cos\theta_i)}$$

This is the phasor associated with a plane wave traveling at an angle of θ_i with respect to the z axis.

For normal incidence, the polarization of the incident wave is always parallel to the interface between the two materials and the reflection and transmission coefficients are independent of the polarization. This is no longer true for oblique incidence.

For an incident wave at an oblique angle, the reflection and transmission coefficients depend on the polarization. For an arbitrarily polarized incident wave, we represent the wave as a sum of two waves with orthogonal polarizations. We can then find the overall reflection and transmission coefficients for the arbitrarily polarized wave by combining the reflection and transmission coefficients for the two orthogonal polarizations. The two orthogonal polarization cases we will cover are referred to as perpendicular and parallel. In different technical fields, other names are used.

The plane containing the incident wave vector and the surface normal at the interface is the plane of incidence. If the electric field is orthogonal to the plane of incidence, that is perpendicular polarization. If the electric field is in the plane of incidence, that is parallel polarization.



For the case of perpendicular polarization, for example, the electric and magnetic fields are

$$\begin{aligned}\overline{E}^i &= \hat{y}E_o^i e^{-jk_1(x\sin\theta_i + z\cos\theta_i)}\\ \overline{H}^i &= \frac{1}{\eta_1}\hat{k}^i \times \overline{E}^i\\ &= \frac{1}{\eta_1}(\hat{x}\sin\theta_i + \hat{z}\cos\theta_i)\times\hat{y}E_o^i e^{-jk_1(x\sin\theta_i + z\cos\theta_i)}\\ &= \frac{E_o^i}{\eta_1}(-\hat{x}\cos\theta_i + \hat{z}\sin\theta_i)e^{-jk_1(x\sin\theta_i + z\cos\theta_i)}\end{aligned}$$

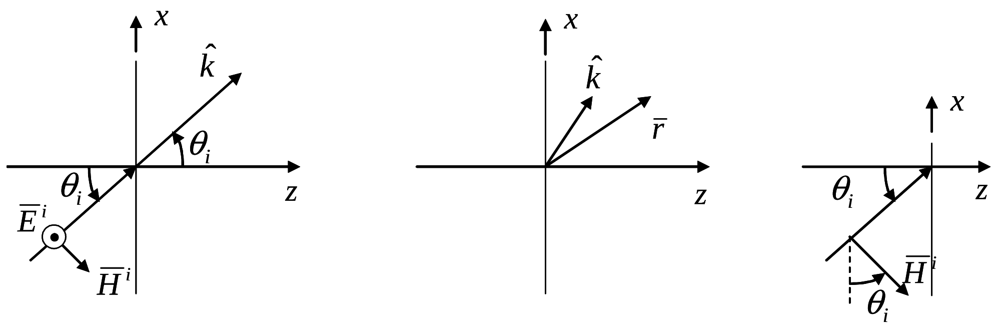


Figure 6.1: Field vectors and the plane of incidence for perpendicular polarization.

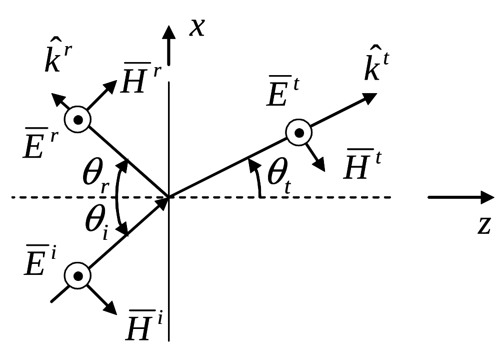
The magnetic field can also be found from the electric field using trigonometry. The field vectors are shown in Fig. 6.1.

To summarize, we define these terms for oblique incidence:

Angle of Incidence:	Angle of the incident wavevector with respect to the normal to the boundary.
Plane of Incidence:	Plane containing normal to boundary and $\hat{k}$ vector
Perpendicular Polarization:	$\overline{E}$ perpendicular to the plane of incidence. Also called transverse electric (TE) polarization, horizontal polarization (h-pol), or s-polarization (from the German word *senkrecht* for perpendicular).
Parallel Polarization:	$\overline{E}$ parallel to the plane of incidence. Also called the transverse magnetic (TM) polarization, vertical polarization (v-pol), or p-polarization (for parallel).

6.2.1 Perpendicular (TE) Polarization

We will now solve for the reflection and transmission coefficients for the case of perpendicular polarization.



The field vectors for this polarization are

1. Incident wave:

$$\begin{aligned}\overline{E}^i &= \hat{y}E_o^i e^{-jk_1(x\sin\theta_i+z\cos\theta_i)}\\ \overline{H}^i &= (-\hat{x}\cos\theta_i+\hat{z}\sin\theta_i)\frac{E_o^i}{\eta_1}e^{-jk_1(x\sin\theta_i+z\cos\theta_i)}\end{aligned}$$

2. Reflected wave:

$$\begin{aligned}\overline{E}^r &= \hat{y}E_o^r e^{-jk_1(x\sin\theta_r-z\cos\theta_r)}\\ \overline{H}^r &= (\hat{x}\cos\theta_r+\hat{z}\sin\theta_r)\frac{E_o^r}{\eta_1}e^{-jk_1(x\sin\theta_r-z\cos\theta_r)}\end{aligned}$$

3. Transmitted wave:

$$\begin{aligned}\overline{E}^t &= \hat{y}E_o^t e^{-jk_2(x\sin\theta_t+z\cos\theta_t)}\\ \overline{H}^t &= (-\hat{x}\cos\theta_t+\hat{z}\sin\theta_t)\frac{E_o^t}{\eta_2}e^{-jk_2(x\sin\theta_t+z\cos\theta_t)}\end{aligned}$$

We now have 4 unknowns: E_o^r, E_o^t, θ_r, θ_t.

Boundary Conditions

$$\begin{array}{ll}\text{Tangential } \overline{E} & \hat{n}\times(\overline{E}^i+\overline{E}^r)\Big|_{z=0}=\hat{n}\times\overline{E}^t\Big|_{z=0}\\ & E_o^i e^{-jk_1x\sin\theta_i}+E_o^r e^{-jk_1x\sin\theta_r}=E_o^t e^{-jk_2x\sin\theta_t}\\ \text{Tangential } \overline{H} & \hat{n}\times(\overline{H}^i+\overline{H}^r)\Big|_{z=0}=\hat{n}\times\overline{H}^t\Big|_{z=0}\\ & -\dfrac{E_o^i}{\eta_1}\cos\theta_i e^{-jk_1x\sin\theta_i}+\dfrac{E_o^r}{\eta_1}\cos\theta_r e^{-jk_1x\sin\theta_r}=-\dfrac{E_o^t}{\eta_2}\cos\theta_t e^{-jk_2x\sin\theta_t}\end{array}$$

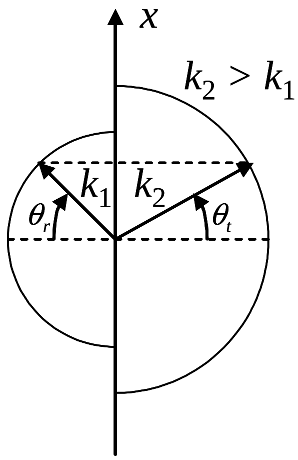
In order for the two equations to be satisfied for all x, the arguments of the exponentials must be equal. We call this the *phase matching condition*:

$$k_1\sin\theta_i=k_1\sin\theta_r=k_2\sin\theta_t$$

This produces Snell's Laws:

$\theta_i=\theta_r$	Snell's Law of Reflection
$k_1\sin\theta_i=k_2\sin\theta_t$	Snell's Law of Refraction

This phase matching condition determines the directions of propagation of the reflected and transmitted waves. A nice picture can be drawn for this in terms of wavevectors:



$k_1 \sin\theta_i = k_1 \sin\theta_r = k_2 \sin\theta_t$
$\Rightarrow$ Tangential component of $\overline{k}_1$ equals tangential component of $\overline{k}_2$.
This makes the phase continuous along the boundary, which is required for the boundary conditions to be satisfied.

Because of the phase matching conditions, the complex exponentials drop out of the boundary conditions:

$$\begin{aligned} E_o^i + E_o^r &= E_o^t \\ \frac{E_o^i}{\eta_1}\cos\theta_i - \frac{E_o^r}{\eta_1}\cos\theta_i &= \frac{E_o^t}{\eta_2}\cos\theta_t \\ \frac{\cos\theta_i}{\eta_1}(E_o^i - E_o^r) &= \frac{\cos\theta_t}{\eta_2}(E_o^i + E_o^r) \end{aligned}$$

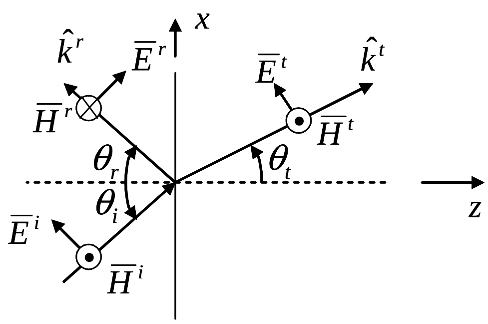
Solving this pair of equations for the unknown amplitudes E_o^r and E_o^t leads to

$$E_o^r = \frac{\eta_2/\cos\theta_t - \eta_1/\cos\theta_i}{\eta_2/\cos\theta_t + \eta_1/\cos\theta_i} E_o^i = \Gamma_\perp E_o^i \tag{6.15}$$

$$E_o^t = \frac{2\eta_2/\cos\theta_t}{\eta_2/\cos\theta_t + \eta_1/\cos\theta_i} E_o^i = \tau_\perp E_o^i = (1+\Gamma_\perp)E_o^i \tag{6.16}$$

6.2.2 Parallel (TM) Polarization

We also need to solve for the reflection and transmission coefficients for parallel polarization.



The field vectors for this polarization are

1. Incident wave

$$\begin{aligned}\overline{E}^i &= (\hat{x}\cos\theta_i - \hat{z}\sin\theta_i)E_o^i e^{-jk_1(x\sin\theta_i + z\cos\theta_i)} \\ \overline{H}^i &= \hat{y}\frac{E_o^i}{\eta_1} e^{-jk_1(x\sin\theta_i + z\cos\theta_i)}\end{aligned}$$

2. Reflected wave

$$\begin{aligned}\overline{E}^r &= (\hat{x}\cos\theta_r + \hat{z}\sin\theta_r)E_o^r e^{-jk_1(x\sin\theta_r - z\cos\theta_r)} \\ \overline{H}^r &= -\hat{y}\frac{E_o^r}{\eta_1} e^{-jk_1(x\sin\theta_r - z\cos\theta_r)}\end{aligned}$$

3. Transmitted wave

$$\begin{aligned}\overline{E}^t &= (\hat{x}\cos\theta_t - \hat{z}\sin\theta_t)E_o^t e^{-jk_2(x\sin\theta_t + z\cos\theta_t)} \\ \overline{H}^t &= \hat{y}\frac{E_o^t}{\eta_2} e^{-jk_2(x\sin\theta_t + z\cos\theta_t)}\end{aligned}$$

Boundary Conditions

$$\begin{array}{ll}\text{Tangential } \overline{E} & \hat{n}\times(\overline{E}^i + \overline{E}^r)\Big|_{z=0} = \hat{n}\times\overline{E}^t\Big|_{z=0} \\ & E_o^i\cos\theta_i e^{-jk_1 x\sin\theta_i} + E_o^r\cos\theta_r e^{-jk_1 x\sin\theta_r} = E_o^t\cos\theta_t e^{-jk_2 x\sin\theta_t} \\ \text{Tangential } \overline{H} & \hat{n}\times(\overline{H}^i + \overline{H}^r)\Big|_{z=0} = \hat{n}\times\overline{H}^t\Big|_{z=0} \\ & \dfrac{E_o^i}{\eta_1}e^{-jk_1 x\sin\theta_i} - \dfrac{E_o^r}{\eta_1}e^{-jk_1 x\sin\theta_r} = \dfrac{E_o^t}{\eta_2}e^{-jk_2 x\sin\theta_t}\end{array}$$

Phase matching again produces Snell's Laws. Then, application of the boundary conditions leads to the reflection and transmission coefficients for the parallel polarization case:

$$E_o^r = \frac{\eta_2\cos\theta_t - \eta_1\cos\theta_i}{\eta_2\cos\theta_t + \eta_1\cos\theta_i}E_o^i = \Gamma_\parallel E_o^i \tag{6.17}$$

$$E_o^t = \frac{2\eta_2\cos\theta_i}{\eta_2\cos\theta_t + \eta_1\cos\theta_i}E_o^i = \tau_\parallel E_o^i \tag{6.18}$$

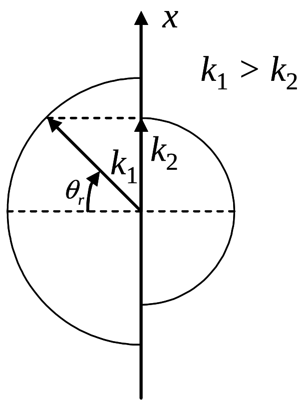
6.2.3 Computing Reflection and Transmission Coefficients

To compute the reflection and transmission coefficients at an oblique angle for a wave that is perpendicularly or parallel polarized, Snell's law is first used to find the transmitted angle. The reflection and transmission coefficient formulas derived above are then used to find the reflection and transmission coefficients for the desired polarization.

For an arbitrarily polarized wave, we break up the incident field into perpendicular and parallel components and use the respective reflection and transmission coefficients for each component.

6.3 Total Internal Reflection

Suppose we have the case where $k_1 > k_2$ (generally $\epsilon_1 > \epsilon_2$). Then, our k-vector diagram becomes



If $\theta_i = \theta_r$ gets too large, we have the case where $k_2 \sin\theta_t$ can't be large enough to equal $k_1 \sin\theta_i$.

When the incidence angle θ_i is such that

$$k_1 \sin\theta_i = k_2 \text{ (so } \theta_t = 90°)$$

we call $\theta_i = \theta_c$ the *critical angle*

In this case, the transmitted wave travels parallel to the interface so that the constant phase planes are parallel to the y-z plane.

If $\theta_i > \theta_c$, then

$$k_2 \sin\theta_t = k_1 \sin\theta_i \to \sin\theta_t = \frac{k_1}{k_2}\sin\theta_i > 1$$

A real angle θ_t cannot satisfy this equation. In fact, θ_t becomes a complex number. What this means is that the z component of the transmitted wavevector k_{2z} becomes imaginary and the transmitted wave decays exponentially away from the boundary. Rather than determining θ_t, we can determine $\sin\theta_t$ and $\cos\theta_t$. $\sin\theta_t$ is determined by Snell's Law as above, which can be rewritten as

$$\sin\theta_t = \frac{k_1}{k_2}\sin\theta_i = \frac{\beta}{k_2}$$

Then

$$\cos\theta_t = \sqrt{1-\sin^2\theta_t} = \sqrt{1-\left(\frac{k_1}{k_2}\right)^2\sin^2\theta_i} = \pm j\sqrt{\left(\frac{k_1}{k_2}\right)^2\sin^2\theta_i - 1} = \pm j\frac{\alpha}{k_2}$$

We will chose the negative root for reasons to be seen below.

The propagation term in the argument of the exponential for the transmitted wave becomes

$$k_2(x\sin\theta_t + z\cos\theta_t) = k_2\left(x\frac{\beta}{k_2} - jz\frac{\alpha}{k_2}\right) = \beta x - j\alpha z$$

Therefore, we can write the transmitted field as

$$\overline{E}^t = \hat{y}E_o^t e^{-jk_2(x\sin\theta_t + z\cos\theta_t)} = \hat{y}E_o^t e^{-\alpha z}e^{-j\beta x}$$

The wave decays in z, but phase progression (propagation) occurs in x. We call such a wave an **evanescent wave**. This evanescent wave decays exponentially and does not carry real power in the z direction.

At the critical angle $\theta_i = \theta_c$, $\cos\theta_t = 0$ so $\alpha = 0$ (no decay).

The reflection coefficient is

$$\begin{aligned}\Gamma_\perp &= \frac{\eta_2/\cos\theta_t - \eta_1/\cos\theta_i}{\eta_2/\cos\theta_t + \eta_1/\cos\theta_i} = \frac{j\eta_2/\alpha - \eta_1/\cos\theta_i}{j\eta_2/\alpha + \eta_1/\cos\theta_i} = -\frac{A^*}{A} \\ |\Gamma_\perp| &= \frac{|A^*|}{|A|} = 1\end{aligned}$$

The same is true for $\Gamma_\parallel$. Because the magnitude of the reflection coefficient is one, *no real power is transferred across the interface*. All power is reflected. We therefore call this *total internal reflection*.

To find the Poynting vector in medium 2, we need the magnetic field, which is

$$\begin{aligned}\overline{H}^t &= (-\hat{x}\cos\theta_t + \hat{z}\sin\theta_t)\frac{E_o^t}{\eta_2}e^{-jk_2(x\sin\theta_t + z\cos\theta_t)} \\ &= (\hat{x}j\sqrt{\sin^2\theta_t - 1} + \hat{z}\sin\theta_t)\frac{E_o^t}{\eta_2}e^{-\alpha z}e^{-j\beta x}\end{aligned}$$

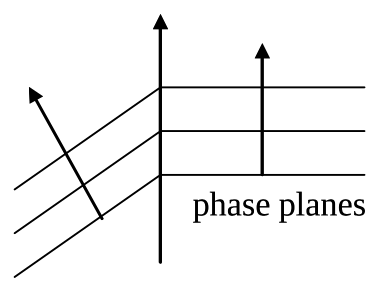
The complex Poynting vector is

$$\begin{aligned}\overline{S}^t &= \overline{E}^t \times \overline{H}^{t*} \\ &= E_o^t e^{-\alpha z}e^{-j\beta x}\frac{E_o^{t*}}{\eta_2}e^{-\alpha z}e^{j\beta x}\left[\hat{z}j\sqrt{\sin^2\theta_t - 1} + \hat{x}\sin\theta_t\right] \\ &= \frac{|E_o^t|^2}{\eta_2}e^{-2\alpha z}\left[\hat{x}\sin\theta_t + \hat{z}j\sqrt{\sin^2\theta_t - 1}\right]\end{aligned}$$

Since the z component of the Poynting vector is imaginary, the power traveling in the $\hat{z}$ direction is reactive (stored energy). The real power is

$$\overline{S}_{av}^t = \frac{1}{2}\frac{|E_o^t|^2}{\eta_2}e^{-2\alpha z}\hat{x}\sin\theta_t$$

which flows in the $\hat{x}$ direction (along the interface).



6.4 Total Transmission (Brewster's Angle)

Suppose we want $\Gamma = 0$ (no reflection), so that

$$\Gamma_\perp = \frac{\eta_2/\cos\theta_t - \eta_1/\cos\theta_i}{\eta_2/\cos\theta_t + \eta_1/\cos\theta_i} = 0$$

For this to occur, we must have

$$\begin{aligned}\eta_2\cos\theta_i &= \eta_1\cos\theta_t \\ \frac{\mu_2}{\epsilon_2}\left[1 - \sin^2\theta_i\right] &= \frac{\mu_1}{\epsilon_1}\left[1 - \sin^2\theta_t\right] = \frac{\mu_1}{\epsilon_1}\left[1 - \left(\frac{k_1}{k_2}\right)^2\sin^2\theta_i\right] = \frac{\mu_1}{\epsilon_1}\left[1 - \frac{\mu_1\epsilon_1}{\mu_2\epsilon_2}\sin^2\theta_i\right]\end{aligned}$$

We can solve for the incidence angle to obtain

$$\sin\theta_i = \sqrt{\frac{1 - (\mu_1\epsilon_2/\mu_2\epsilon_1)}{1 - (\mu_1/\mu_2)^2}} \tag{6.19}$$

If $\mu_1 = \mu_2$, $\sin\theta_i = \infty$ which is impossible. So, we can't have total transmission for the perpendicular polarization in non-magnetic media.

For the parallel polarization,

$$\begin{aligned}\Gamma_\parallel &= \frac{\eta_2\cos\theta_t - \eta_1\cos\theta_i}{\eta_2\cos\theta_t + \eta_1\cos\theta_i} = 0 \\ \eta_2\cos\theta_t &= \eta_1\cos\theta_i\end{aligned}$$

The solution for this case (following an analogous procedure) is

$$\sin\theta_i = \sqrt{\frac{1-(\epsilon_1\mu_2/\epsilon_2\mu_1)}{1-(\epsilon_1/\epsilon_2)^2}}$$

Here, if $\mu_1 = \mu_2$, we obtain

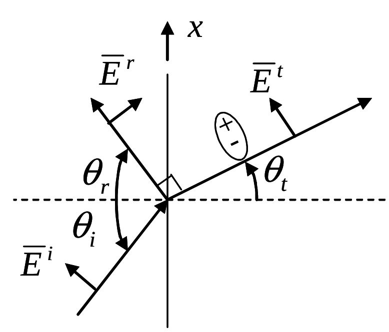
$$\begin{aligned}\sin\theta_B &= \sqrt{\frac{1-\epsilon_1/\epsilon_2}{1-(\epsilon_1/\epsilon_2)^2}} = \sqrt{\frac{1}{1+\epsilon_1/\epsilon_2}} \\ \cos\theta_B &= \sqrt{1-\frac{1}{1+\epsilon_1/\epsilon_2}} = \sqrt{\frac{\epsilon_1/\epsilon_2}{1+\epsilon_1/\epsilon_2}} \\ \tan\theta_B &= \frac{\sin\theta_B}{\cos\theta_B} = \sqrt{\frac{\epsilon_2}{\epsilon_1}} \\ \theta_B &= \tan^{-1}\left(\sqrt{\frac{\epsilon_2}{\epsilon_1}}\right)\end{aligned}$$

We call θ_B *Brewster's Angle*.

To explain Brewster's angle physically, we have to revisit the permittivity. Recall that materials are made of atoms whose charge will shift slightly under the influence of an electric field (the atoms or molecules will polarize, with positive and negative charge separating a bit to made a dipole). The permittivity ϵ represents the susceptibility of the medium to this alignment. When a time-varying field hits the medium, these dipoles oscillate. This charge motion represents a current which causes re-radiation of another electromagnetic wave. This explains why a reflected field is created.

We will later see that dipoles do not radiate along their polar axis. The oscillating dipole in material 2 won't radiate in the direction of the reflection. Therefore, $E^r = 0$. It can be shown that

$$\theta_B + \theta_t = 90^\circ$$



For the perpendicular polarization, the dipole oscillation is normal to the page, so we will always have radiation in the reflected direction. Therefore, there is no Brewster angle for this case (unless one of the materials is magnetic).

6.4.1 Power

For the perpendicular polarization, the power densities associated with the incident, reflected, and transmitted waves are

$$\begin{aligned}
\overline{S}^i &= \overline{E}^i \times \overline{H}^{i*} \\
&= \hat{y}E_{\text{o}}^i e^{-j\overline{k}^i\cdot\overline{r}} \times (-\hat{x}\cos\theta_i + \hat{z}\sin\theta_i)\frac{E_{\text{o}}^{i*}}{\eta_1^*}e^{+j\overline{k}^i\cdot\overline{r}} \\
&= \frac{|E_{\text{o}}^i|^2}{\eta_1}(\hat{z}\cos\theta_i + \hat{x}\sin\theta_i) \qquad (6.20)
\end{aligned}$$

$$\begin{aligned}
\overline{S}^r &= \overline{E}^r \times \overline{H}^{r*} \\
&= \hat{y}\Gamma_\perp E_{\text{o}}^i e^{-j\overline{k}^r\cdot\overline{r}} \times (\hat{x}\cos\theta_r + \hat{z}\sin\theta_r)\frac{\Gamma_\perp^* E_{\text{o}}^{i*}}{\eta_1^*}e^{+j\overline{k}^r\cdot\overline{r}} \\
&= \frac{|\Gamma_\perp E_{\text{o}}^i|^2}{\eta_1}(-\hat{z}\cos\theta_i + \hat{x}\sin\theta_i) \qquad (6.21)
\end{aligned}$$

$$\begin{aligned}
\overline{S}^t &= \overline{E}^t \times \overline{H}^{t*} \\
&= \hat{y}\tau_\perp E_{\text{o}}^i e^{-j\overline{k}^t\cdot\overline{r}} \times (-\hat{x}\cos\theta_t + \hat{z}\sin\theta_t)\frac{\tau_\perp^* E_{\text{o}}^{i*}}{\eta_2^*}e^{+j\overline{k}^t\cdot\overline{r}} \\
&= \frac{|\tau_\perp E_{\text{o}}^i|^2}{\eta_2}(\hat{z}\cos\theta_t + \hat{x}\sin\theta_t) \qquad (6.22)
\end{aligned}$$

If the materials are lossy or θ_t is complex due to total internal reflection, we have to be a little more careful with the complex conjugations.

The ratios of the reflected and transmitted power flux densities to the incident power density are

$$\frac{S_z^r}{S_z^i} = |\Gamma_\perp|^2 \qquad (6.23)$$

$$\frac{S_z^t}{S_z^i} = \frac{|\tau_\perp E_0|^2\cos\theta_t/\eta_2}{|E_0|^2\cos\theta_i/\eta_1} = \frac{\eta_1\cos\theta_t}{\eta_2\cos\theta_i}|\tau_\perp|^2 \qquad (6.24)$$

where we consider the z component of the power flow (i.e., into medium 2). Similar results hold for parallel polarization. It is possible to have $|\tau| > 1$, but by the principle of conservation of energy, the ratio of transmitted power to incident power is always less than or equal to one.

Example:

Consider a wave normally incident on window glass. We can model the glass interface as a half-space dielectric medium with index of refraction 1.5. The reflection coefficient for this case is

$$\Gamma = \frac{\eta_2 - \eta_1}{\eta_2 + \eta_1} = \frac{\sqrt{\mu_2/\epsilon_2} - \sqrt{\mu_1/\epsilon_1}}{\sqrt{\mu_2/\epsilon_2} + \sqrt{\mu_1/\epsilon_1}} = \frac{1/n_2 - 1}{1/n_2 + 1} = -0.2$$

In this expression, n_2 is the index of refraction of the glass, which is $\sqrt{\epsilon_{r2}}$ (see Sec. 6.1.3). This is different from the characteristic impedance $\eta_2 = \sqrt{\mu_2/\epsilon_2}$ associated with the glass. The permeabilities cancel out of the formula because both materials (air and glass) are nonmagnetic and $\mu_1 = \mu_2$. Because magnetic materials are less common than dielectrics, reflection and transmission formulas are often given for convenience in terms of the index of refraction.

The transmission coefficient is

$$\tau = \frac{2\eta_2}{\eta_2 + \eta_1} = \frac{2/n_2}{1/n_2 + 1} = 0.8$$

From (6.24), the fraction of transmitted power is

$$\frac{S^t}{S^r} = 1.5|0.8|^2 = 0.96$$

From this analysis, we see that at normal incidence 96% of the incident power is transmitted into the glass.

Now, we change the incidence angle to $45°$. From Snell's law, the transmitted angle is $\theta_t \simeq 28°$. If the wave is perpendicularly polarized, the reflection coefficient is

$$\Gamma_\perp = \frac{\eta_2/\cos\theta_t - \eta_1/\cos\theta_i}{\eta_2/\cos\theta_t + \eta_1/\cos\theta_i} = \frac{1/(n_2\cos\theta_t) - 1/\cos\theta_i}{1/(n_2\cos\theta_t) + 1/\cos\theta_i} \simeq -0.3$$

and the transmission coefficient is

$$\tau_\perp = \frac{2\eta_2/\cos\theta_t}{\eta_2/\cos\theta_t + \eta_1/\cos\theta_i} \simeq 0.7$$

At this incidence angle, 9% of the power is reflected and 91% percent of the power in the incident wave is transmitted into the glass.

Chapter 7

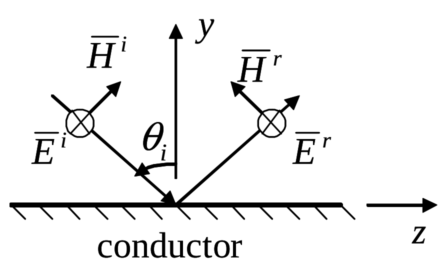
Waveguides

Waveguides are designed to transport electromagnetic waves over a specified path. Some examples of waveguides are coaxial cable, parallel conducting plates, rectangular waveguides, circular waveguides, coaxial cable, biconical, sectoral, microstrip, dielectric slab, and optical fibers. We will look at several of these.

The fundamental principle we will encounter with waveguides is that of *modes*. In free space, we can have any number of plane waves traveling in all directions. There are an uncountable number of possible plane directions. Which plane wave or distribution of plane waves that we actually get depends on the location and shape of the source that produces the waves. Inside a waveguide, only a countable number of combinations of plane wave solutions satisfy the boundary conditions at the waveguide walls. Each of these is called a mode. Which mode or combination of modes that actually exists in the waveguide depends on the source driving the waveguide. Also, each mode can be modeled as a transmission line with a different characteristic impedance and phase velocity.

7.1 Parallel Plates

We first consider a plane wave reflecting off a conducting surface. Note that we will change our coordinate directions from what we used previously for reflection and transmission.



7.1.1 TE (Transverse Electric)

$$\begin{aligned}
\overline{E}^i &= \hat{x}E_o^i e^{-jk_0(-y\cos\theta_i + z\sin\theta_i)} \\
\overline{H}^i &= (\hat{y}\sin\theta_i + \hat{z}\cos\theta_i)\frac{E_o^i}{\eta_0} e^{-jk_0(-y\cos\theta_i + z\sin\theta_i)} \\
\overline{E}^r &= \hat{x}\Gamma_\perp E_o^i e^{-jk_0(y\cos\theta_i + z\sin\theta_i)} \\
\overline{H}^r &= (\hat{y}\sin\theta_i - \hat{z}\cos\theta_i)\Gamma_\perp \frac{E_o^i}{\eta_0} e^{-jk_0(y\cos\theta_i + z\sin\theta_i)}
\end{aligned}$$

where we have used that $\theta_r = \theta_i$. Since $\Gamma_\perp = -1$, the total fields are:

$$\begin{aligned}
\overline{E} = \overline{E}^i + \overline{E}^r &= \hat{x}E_o^i \left(e^{jk_0 y\cos\theta_i} - e^{-jk_0 y\cos\theta_i}\right) e^{-jk_0 z\sin\theta_i} \\
&= \hat{x}j2E_o^i \sin\left(k_0 y\cos\theta_i\right) e^{-jk_0 z\sin\theta_i} \\
\overline{H} = \overline{H}^i + \overline{H}^r &= \frac{E_o^i}{\eta_0}\Big[\hat{y}\sin\theta_i \left(e^{jk_0 y\cos\theta_i} - e^{-jk_0 y\cos\theta_i}\right) \\
&\quad + \hat{z}\cos\theta_i \left(e^{jk_0 y\cos\theta_i} + e^{-jk_0 y\cos\theta_i}\right)\Big] e^{-jk_0 z\sin\theta_i} \\
&= 2\frac{E_o^i}{\eta_0}\left[\hat{y}j\sin\theta_i \sin\left(k_0 y\cos\theta_i\right) + \hat{z}\cos\theta_i\cos\left(k_0 y\cos\theta_i\right)\right] e^{-jk_0 z\sin\theta_i}
\end{aligned}$$

Note that $E_x = H_y = 0$ at $k_0 y\cos\theta_i = m\pi$, $m = 1, 2, \ldots$. Since tangential $\overline{E}$ and normal $\overline{H}$ are zero here, we could slide a conductor in at these points without disturbing the field. If we could get the field into this structure, it would have the same form as we have observed here.

If we look at this another way, if we put a plate at $y = d$, then this field will exist in the structure. The value of θ_i at which the wave will travel will be given by:

$$k_0 d\cos\theta_i = m\pi, \qquad m = 1, 2, 3, \ldots$$

Let's generalize this a little more. Now that we see what the waves look like, we recognize that the angle θ_i is not very important in general. Let's take our wavenumber k_0 and make it into a vector.

$$\begin{aligned}
\overline{k} &= k_0\hat{k}^i = \hat{y}k_y + \hat{z}k_z \\
k_y &= k_0\cos\theta_i \\
k_z &= k_0\sin\theta_i
\end{aligned}$$

$$k_y^2 + k_z^2 = k_0^2$$

Now

$$\begin{aligned}
\overline{E} &= \hat{x}j2E_o\sin(k_y y)e^{-jk_z z} \\
\overline{H} &= \frac{2E_o}{\eta_0 k_0}\left[\hat{y}jk_z\sin(k_y y) + \hat{z}k_y\cos(k_y y)\right] e^{-jk_z z}
\end{aligned}$$

The Poynting vector is:

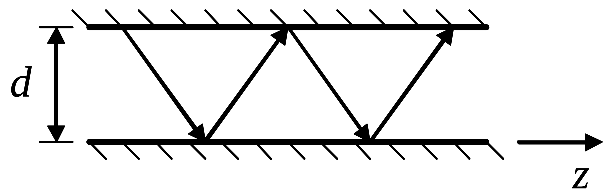
$$\begin{aligned}
\overline{S} = \overline{E}\times\overline{H}^* &= \left[j2E_o\sin(k_y y)e^{-jk_z z}\right]\left[\frac{2E_o^*}{\eta_0 k_0}e^{jk_z z}\right]\left[-\hat{z}jk_z\sin(k_y y) - \hat{y}k_y\cos(k_y y)\right] \\
&= \frac{4|E_o|^2}{\eta_0 k_0}\sin(k_y y)\left[\hat{z}k_z\sin(k_y y) - \hat{y}jk_y\cos(k_y y)\right]
\end{aligned}$$

Notice that all real power flows in the z direction!! So, our structure is guiding power along the z direction.

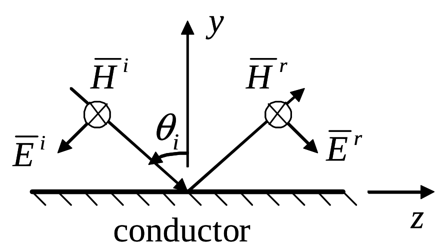
So, the physical picture is that we have two plates which confine the electromagnetic wave. Because we know that

$$k_y d = m\pi$$

there is a countably infinite set of field configurations possible. We call each possible configuration a **mode**. In this case, we call the modes TE$_m$ modes. Note that if $m = 0$, $\overline{E} = 0$, which means no field exists. Therefore, m ranges from 1 to ∞.



7.1.2 TM (Transverse Magnetic)



$$
\begin{aligned}
\overline{H}^i &= \hat{x}\frac{E_o}{\eta_0}e^{-j(-k_y y + k_z z)} \\
\overline{E}^i &= (-\hat{y}k_z - \hat{z}k_y)\frac{E_o}{k_0}e^{-j(-k_y y + k_z z)} \\
\overline{H}^r &= \hat{x}\frac{E_o}{\eta_0}e^{-j(k_y y + k_z z)} \\
\overline{E}^r &= (-\hat{y}k_z + \hat{z}k_y)\frac{E_o}{k_0}e^{-j(k_y y + k_z z)}
\end{aligned}
$$

$$
\begin{aligned}
\overline{H} = \overline{H}^i + \overline{H}^r &= \hat{x}\frac{E_o}{\eta_0}\left(e^{jk_y y} + e^{-jk_y y}\right)e^{-jk_z z} \\
&= \hat{x}\frac{2E_o}{\eta_0}\cos(k_y y)e^{-jk_z z} \\
\overline{E} = \overline{E}^i + \overline{E}^r &= \frac{E_o}{k_0}\left[-\hat{y}k_z\left(e^{jk_y y} + e^{-jk_y y}\right) - \hat{z}k_y\left(e^{jk_y y} - e^{-jk_y y}\right)\right]e^{-jk_z z} \\
&= \frac{2E_o}{k_0}\left[-\hat{y}k_z\cos(k_y y) - \hat{z}jk_y\sin(k_y y)\right]e^{-jk_z z}
\end{aligned}
$$

Once again, if $k_y d = m\pi$ (i.e. plate placed at $y = d$), there will be no impact on the fields since $E_z(y = d) = 0$.

In this case, if $k_y = m\pi/d$ and $m = 0$ we have that $k_y = 0$. Since $k_y^2 + k_z^2 = k_0^2$, then $k_z = k_0$.

$$
\begin{aligned}
\overline{E} &= -\hat{y}\frac{2E_o}{k_0}k_z e^{-jk_z z} = \hat{y}2E_o e^{-jk_0 z} \\
\overline{H} &= \hat{x}\frac{2E_o}{\eta_0}e^{-jk_0 z}
\end{aligned}
$$

This is just a plane wave traveling in the $+z$ direction.

Therefore, we have TM$_m$ modes for $m = 0, 1, 2, \ldots$. The TM$_0$ mode is a TEM (transverse electromagnetic) mode.

7.1.3 Cutoff

For both TE and TM cases:

$$k_y = \frac{m\pi}{d}$$

Since:

$$
\begin{aligned}
k_y^2 + k_z^2 &= k_0^2 \\
k_z &= \sqrt{k_0^2 - k_y^2} = \sqrt{\omega^2\mu_0\epsilon_0 - (m\pi/d)^2}
\end{aligned}
$$

Since the wave travels as $e^{-jk_z z}$, if $k_z \to 0$, there is no propagation. We call this the *cutoff condition* for the waveguide. The frequency ω_m at which the mth mode is cutoff is constructed from:

$$
\begin{aligned}
\omega_m^2\mu_0\epsilon_0 &= \left(\frac{m\pi}{d}\right)^2 \\
\omega_m &= \frac{m\pi}{d}\frac{1}{\sqrt{\mu_0\epsilon_0}} = \frac{m\pi}{d}c_0
\end{aligned}
$$

We call ω_m the *cutoff frequency* for the mth mode. If the waveguide is filled with a material, we replace μ_0, ϵ_0, and c_0 by the values for that material.

Note that:

$$
\begin{aligned}
k_z &= \sqrt{\left(\frac{2\pi f}{c_0}\right)^2 - \left(\frac{2\pi f_m}{c_0}\right)^2} \\
&= \frac{2\pi f}{c_0}\sqrt{1 - \left(\frac{f_m}{f}\right)^2} = k_0\sqrt{1 - \left(\frac{f_m}{f}\right)^2}
\end{aligned}
$$

We can now define:

Guide Wavelength: $$\lambda_g = \lambda_z = \frac{2\pi}{k_z} = \frac{2\pi}{k_0\sqrt{1 - \left(\frac{f_m}{f}\right)^2}} = \frac{\lambda_0}{\sqrt{1 - \left(\frac{f_m}{f}\right)^2}}$$

Phase Velocity: $$u_p = \frac{\omega}{k_z} = \frac{\omega}{k_0\sqrt{1 - \left(\frac{f_m}{f}\right)^2}} = \frac{c_0}{\sqrt{1 - \left(\frac{f_m}{f}\right)^2}}$$

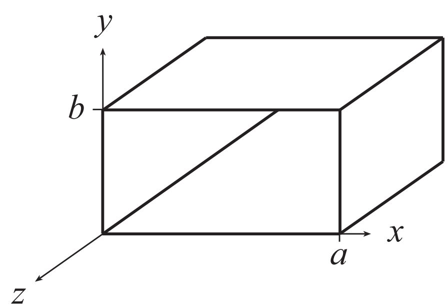
If $\omega < \omega_m$, we are below cutoff and k_z becomes imaginary. Then:

$$\overline{E}_{\text{TE}} = j2\hat{x}E_o \sin\left(\frac{m\pi y}{d}\right) e^{-\alpha z}$$

We have sinusoidal variation in y due to the $+y$ and $-y$ traveling waves. However, the wave decays in z. This is another example of an evanescent wave. Looking back at our wave propagation concept, if $k_z = 0$, $k_y = k_0$ which means that $\theta_i = 0$ (the "wave" is traveling vertically between the plates, but there is no propagation in z).

7.2 Rectangular Metallic Waveguide

A rectangular waveguide has four metallic sides. In this analysis we assume that the metallic walls are perfect electric conductors (PEC). The dimension of the waveguide is $a \times b$. By convention, $a > b$ and we orient the long side in the x direction.



Our goal is to solve Maxwell's equations and apply boundary conditions. The waveguide confines the wave propagation in both the x and y directions. There are six vector field components (E_x, E_y, E_z, H_x, H_y, H_z) that we need to find.

The goal is to use Maxwell's equations to relate the vector quantities. This will simplify the number of differential equations that we have to solve.

If we expand Faraday's Law $\nabla \times \overline{E} = -j\omega\mu\overline{H}$ in components, we get

$$\frac{\partial E_z}{\partial y} + jk_z E_y = -j\omega\mu H_x \tag{7.1}$$

$$-jk_z E_x - \frac{\partial E_z}{\partial x} = -j\omega\mu H_y \tag{7.2}$$

$$\frac{\partial E_y}{\partial x} - \frac{\partial E_x}{\partial y} = -j\omega\mu H_z \tag{7.3}$$

Now use Ampere's Law $\nabla \times \overline{H} = j\omega\epsilon\overline{E}$ to obtain

$$\frac{\partial H_z}{\partial y} + jk_z H_y = jw\epsilon E_x \tag{7.4}$$

$$-jk_zH_x - \frac{\partial H_z}{\partial x} = j\omega\epsilon E_y \tag{7.5}$$

$$\frac{\partial H_y}{\partial x} - \frac{\partial H_x}{\partial y} = j\omega\epsilon E_z \tag{7.6}$$

We are going to combine these equations to relate various components in terms of other components. Substitute Eq. (7.5) into Eq. (7.1),

$$\begin{aligned}
-j\omega\mu H_x &= \frac{\partial E_z}{\partial y} + jk_z\left[\frac{1}{j\omega\epsilon}\left(-jk_zH_x - \frac{\partial H_z}{\partial x}\right)\right] && (7.7)\\
&= \frac{\partial E_z}{\partial y} + \frac{k_z}{\omega\epsilon}(-jk_z)H_x - \frac{k_z}{\omega\epsilon}\frac{\partial H_z}{\partial x} && (7.8)
\end{aligned}$$

Simplifying,

$$\left(-j\omega^2\mu\epsilon + jk_z^2\right)H_x = \omega\epsilon\frac{\partial E_z}{\partial y} - k_z\frac{\partial H_z}{\partial x} \tag{7.9}$$

$$H_x = \frac{j}{k^2 - k_z^2}\left[\omega\epsilon\frac{\partial E_z}{\partial y} - k_z\frac{\partial H_z}{\partial x}\right] \tag{7.10}$$

Thus, if we have E_z and H_z we can find H_x. By making similar substitutions, we can find the relationships

$$H_y = \frac{-j}{k^2 - k_z^2}\left[\omega\epsilon\frac{\partial E_z}{\partial x} + k_z\frac{\partial H_z}{\partial y}\right] \tag{7.11}$$

$$E_x = \frac{-j}{k^2 - k_z^2}\left[k_z\frac{\partial E_z}{\partial x} + \omega\mu\frac{\partial H_z}{\partial y}\right] \tag{7.12}$$

$$E_y = \frac{-j}{k^2 - k_z^2}\left[k_z\frac{\partial E_z}{\partial y} - \omega\mu\frac{\partial H_z}{\partial x}\right] \tag{7.13}$$

These four equations show that if we can find E_z and H_z, we can find all of the other field components (E_x, E_y, H_x, H_y). We will solve for E_z and H_z separately. A fully general solution can be obtained by combining the two sets of results (similar to parallel and perpendicular polarization from reflection and transmission).

In a rectangular waveguide the wave is not simply bouncing up and down in the y-z plane like a parallel plate waveguide. Certain waveguide modes (solutions satisfying the boundary conditions) can be thought of as bouncing off of all of the walls. This would look like a spiral. Therefore, the electric and magnetic fields can have components in all three coordinate directions. Based on Eqs. (7.11)-(7.13), however, the total vector fields can be calculated if we determine the E_z and H_z components. The various modes are divided into two different cases for which either E_z or H_z are equal to zero:

Transverse Electric (TE): The electric field is perpendicular to the propagation direction ($E_z = 0$).

Transverse Magnetic (TM): The magnetic field is perpendicular to the propagation direction ($H_z = 0$).

7.2.1 Transverse Magnetic (TM)

The magnetic field is perpendicular to the propagation direction: $(E_z \neq 0, H_z = 0)$. We only have to solve for E_z, and we can get all the other field components from that. The wave equation for this component is

$$\left(\nabla^2 + k^2\right) E_z = 0 \tag{7.14}$$

We solved this differential equation when we first studied plane waves. We assume that E_z is of a form in which you can separate the variables, resulting in

$$E_z = f(x)\; g(y)\; h(z), \tag{7.15}$$

where f, g, and h are unknown functions. This is a realistic assumption and is commonly used to solve some partial differential equations. We plug this into the wave equation to get

$$\begin{aligned} \left(\nabla^2 + k^2\right) E_z &= \left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2} + k^2\right) f(x)g(y)h(z) = 0 && (7.16) \\ &= f''gh + fg''h + fgh'' + k^2 fgh = 0 && (7.17) \end{aligned}$$

We divide by fgh to get

$$\frac{f''}{f} + \frac{g''}{g} + \frac{h''}{h} + k^2 = 0 \tag{7.18}$$

For the term containing f, all the other terms are all just constants, because they depend on other variables. So, Eq. (7.16) can be broken up into three different equations as given by

$$\begin{aligned} \frac{f''}{f} &= -k_x^2 && (7.19) \\ \frac{g''}{g} &= -k_y^2 && (7.20) \\ \frac{h''}{h} &= -k_z^2. && (7.21) \end{aligned}$$

Each of these equations is now just a second order differential equation with the solutions

$$\begin{aligned} f(x) &= Ae^{jk_x x} + Be^{-jk_x x} && (7.22) \\ g(x) &= Ce^{jk_y y} + De^{-jk_y y} && (7.23) \\ h(x) &= Ee^{jk_z z} + Fe^{-jk_z z} && (7.24) \end{aligned}$$

The solution for E_z is then

$$\begin{aligned} E_z(x,y,z) &= fgh && (7.25) \\ &= \left(Ae^{jk_x x} + Be^{-jk_x x}\right)\left(Ce^{jk_y y} + De^{-jk_y y}\right)\left(Ee^{jk_z z} + Fe^{-jk_z z}\right) && (7.26) \end{aligned}$$

Now we need to determine all of the unknowns. The waveguide is infinite in z resulting in no backwards propagating wave, so we may assume $E = 0$.

The waves in the x and y direction are standing waves rather than traveling waves so we will convert the exponential form into sinusoidal form as given by

$$E_z(x,y,z) = \left[A\sin\left(k_x x\right) + B\cos\left(k_x x\right)\right]\left[C\sin\left(k_y y\right) + D\cos\left(k_y y\right)\right]e^{-jk_z z}. \tag{7.27}$$

We need to find the unknowns A, B, C, D, k_x, k_y, and k_z.

The electric field component is in the z direction. Therefore, it is tangential at each waveguide edge, resulting in boundary conditions given by

$$E_z\left(x=0\right) = E_z\left(x=a\right) = E_z\left(y=0\right) = E_z\left(y=b\right) = 0 \tag{7.28}$$

The condition at $x = 0$ results in

$$B\left[C\sin\left(k_y y\right) + D\cos\left(k_y y\right)\right]e^{-jk_z z} = 0 \tag{7.29}$$

For all values of y and z, this requires $B = 0$.

The condition at $y = 0$ results in

$$\left[A\sin\left(k_x x\right) + B\cos\left(k_x x\right)\right]De^{-jk_z z} = 0 \tag{7.30}$$

For all values of x and z, this requires $D = 0$.

The electric field is now given by

$$E_z(x, y, z) = A\sin\left(k_x x\right)C\sin\left(k_y y\right)e^{-jk_z z} = 0. \tag{7.31}$$

The constants can be combined resulting in

$$E_z(x, y, z) = E_o\sin\left(k_x x\right)\sin\left(k_y y\right)e^{-jk_z z} = 0, \tag{7.32}$$

where E_o is just related to the power of the waveguide mode.

Now use the other two boundaries to determine the propagation constants. For $x = a$,

$$E_z(a, y, z) = E_o\sin\left(k_x a\right)\sin\left(k_y y\right)e^{-jk_z z} = 0. \tag{7.33}$$

In order for this to be equal to zero for all y and z, we must have

$$\sin\left(k_x a\right) = 0, \tag{7.34}$$

which results in

$$k_x a = m\pi \tag{7.35}$$

$$k_x = \frac{m\pi}{a}. \tag{7.36}$$

At the $y = b$ boundary,

$$E_z(x, b, z) = E_o\sin\left(k_x x\right)\sin\left(k_y b\right)e^{-jk_z z} = 0, \tag{7.37}$$

resulting in

$$k_y = \frac{n\pi}{b}. \tag{7.38}$$

The remaining unknown k_z is determined by using the magnitude of the wavevector as given by

$$k_z = \sqrt{k^2 - \left(\frac{m\pi}{a}\right)^2 - \left(\frac{n\pi}{b}\right)^2} \tag{7.39}$$

The resulting solution for the TM_{mn} mode is

$$E_z(x,y,z) = E_o \sin\left(\frac{m\pi}{a}x\right)\sin\left(\frac{n\pi}{b}x\right)e^{-jk_z z}. \tag{7.40}$$

If we need to find the other components of the electric and magnetic fields, we use Eqs. (7.10)-(7.13) to obtain

$$E_x = E_{xo}\cos(k_x x)\sin(k_y y)\,e^{-jk_z z} \tag{7.41}$$
$$E_y = E_{yo}\sin(k_x x)\cos(k_y y)\,e^{-jk_z z} \tag{7.42}$$
$$H_x = H_{xo}\sin(k_x x)\cos(k_y y)\,e^{-jk_z z} \tag{7.43}$$
$$H_y = H_{yo}\cos(k_x x)\sin(k_y y)\,e^{-jk_z z} \tag{7.44}$$

What is the lowest order mode? If either $m = 0$ or $n = 0$, then $E_z = 0$. If both $E_z = 0$ and $H_z = 0$, then the total power is zero, therefore $m \geq 1$ and $n \geq 1$. So the lowest order mode is TM_{11}.

7.2.2 Transverse Electric (TE)

Now let's solve the case for which the electric field is perpendicular to the propagation direction ($E_z = 0$ and $H_z \neq 0$). $\overline{H}$ satisfies the same wave equations as $\overline{E}$, so H_z has the same general form as E_z,

$$H_z(x,y,z) = [A\sin(k_x x) + B\cos(k_x x)]\,[C\sin(k_y y) + D\cos(k_y y)]\,e^{-jk_z z}. \tag{7.45}$$

When we apply the boundary conditions H_z is tangential at the boundary. Since $H_t = J_s$, this boundary condition does not help specify the mode. Therefore, we need to calculate the electric field from the magnetic field. Using Eq. (7.12),

$$E_x = \frac{-j\omega\mu}{k^2 - k_z^2}\frac{\partial H_z}{\partial y} \tag{7.46}$$
$$= \frac{-j\omega\mu}{k^2 - k_z^2}[A\sin(k_x x) + B\cos(k_x x)]\,[Ck_y\cos(k_y y) - Dk_y\sin(k_y y)]\,e^{-jk_z z}. \tag{7.47}$$

The boundary condition at $y = 0$ is

$$E_x(y=0) = 0 = \frac{-j\omega\mu}{k^2 - k_z^2}[A\sin(k_x x) + B\cos(k_x x)]\,[Ck_y]\,e^{-jk_z z} \tag{7.48}$$

so that $C = 0$. We then plug in $C = 0$ and apply the boundary condition at $y = b$ to get

$$E_x(y=b) = 0 = \frac{-j\omega\mu}{k^2 - k_z^2}[A\sin(k_x x) + B\cos(k_x x)]\,[-Dk_y\sin(k_y b)]\,e^{-jk_z z}. \tag{7.49}$$

$$\Rightarrow \quad k_y = \frac{n\pi}{b} \tag{7.50}$$

We now follow a similar process with E_y. Using Eq. (7.13),

$$E_y = \frac{j\omega\mu}{k^2 - k_z^2}\frac{\partial H_z}{\partial x} \tag{7.51}$$
$$= \frac{j\omega\mu}{k^2 - k_z^2}[Ak_x\cos(k_x x) - Bk_x\sin(k_x x)]\,[Ck_y\cos(k_y y) - Dk_y\sin(k_y y)]\,e^{-jk_z z}. \tag{7.52}$$

At $x = 0$,

$$E_y(x=0) = 0 = \frac{j\omega\mu}{k^2 - k_z^2}\left[Ak_x\right]\left[Ck_y\cos\left(k_y y\right) - Dk_y\sin\left(k_y y\right)\right]e^{-jk_z z}. \tag{7.53}$$

$$\Rightarrow \quad A = 0 \tag{7.54}$$

We then plug in $A = 0$ and apply the boundary condition at $x = a$ to get

$$E_y(x=a) = 0 = \frac{j\omega\mu}{k^2 - k_z^2}\left[-Bk_x\sin\left(k_x a\right)\right]\left[Ck_y\cos\left(k_y y\right) - Dk_y\sin\left(k_y y\right)\right]e^{-jk_z z}. \tag{7.55}$$

$$\Rightarrow \quad k_x = \frac{m\pi}{a} \tag{7.56}$$

The resulting solution for the TE$_{mn}$ mode is given by

$$H_z = H_{zo}\cos\left(\frac{m\pi}{a}x\right)\cos\left(\frac{n\pi}{b}y\right)e^{-jk_z z} \tag{7.57}$$

$$k_z = \sqrt{k^2 - \left(\frac{m\pi}{a}\right)^2 - \left(\frac{n\pi}{b}\right)^2} \tag{7.58}$$

The other fields have the form

$$E_x = E_{xo}\cos(k_x x)\sin(k_y y)e^{-jk_z z} \tag{7.59}$$

$$E_y = E_{yo}\sin(k_x x)\cos(k_y y)e^{-jk_z z} \tag{7.60}$$

$$H_x = H_{xo}\sin(k_x x)\cos(k_y y)e^{-jk_z z} \tag{7.61}$$

$$H_y = H_{yo}\cos(k_x x)\sin(k_y y)e^{-jk_z z} \tag{7.62}$$

What is the lowest order mode? H_z can have a non-zero amplitude even if $m = 0$ and $n = 0$. However, if both $m = 0$ and $n = 0$ then $E_x = E_y = 0$. In combination with the fact that $E_z = 0$, the total electric field is zero resulting in no power. Thus, the lowest order mode is either TE$_{10}$ or TE$_{01}$.

7.2.3 Cutoff

If we look at the z propagation constant,

$$k_z = \sqrt{\omega^2\mu\epsilon - \left(\frac{m\pi}{a}\right)^2 - \left(\frac{n\pi}{b}\right)^2} \tag{7.63}$$

we can see that if the frequency is small, k_z is imaginary, and the mode decays instead of propagating. We say that the mode is in cutoff. As we increase the frequency, the imaginary part of k_z becomes smaller until $k_z = 0$. This is the **cutoff frequency** of the mode:

$$0 = (2\pi f_{c,mn})^2\mu\epsilon - \left(\frac{m\pi}{a}\right)^2 - \left(\frac{n\pi}{b}\right)^2$$

$$f_{c,mn} = \frac{c}{2}\sqrt{\left(\frac{m}{a}\right)^2 + \left(\frac{n}{b}\right)^2} \tag{7.64}$$

If the frequency is larger than $f_{c,mn}$, then k_z is real and the mode propagates down the waveguide.

Other quantities of interest in the waveguide are the phase velocity, $v_p = \omega/k_z$, and the guide wavelength $\lambda_g = 2\pi/k_z$. If the frequency is near cutoff ($f \simeq f_{c,mn}$), the phase velocity is very large. The group velocity at which we can send information by modulating the waveguide mode satisfies the relationship $v_g v_p = c^2$ so that near cutoff, information-bearing signals travel very slowly down the waveguide.

Each mode has its own cutoff frequency. Two different modes that have the same cutoff frequency are called degenerate modes. If we list the modes for increasing values of the mode numbers, we get

$$\begin{array}{cll}
\text{TM}_{00} & \text{(trivial)} & \\
\text{TE}_{00} & \text{(trivial)} & \\
\text{TM}_{10} & \text{(trivial)} & \\
\text{TM}_{01} & \text{(trivial)} & \\
\textbf{TE}_{\mathbf{10}} & \mathbf{f_{c,10} = \frac{c}{2a}} & \Rightarrow \quad \textbf{Dominant mode} \\
\text{TE}_{01} & f_{c,01} = \frac{c}{2b} & \\
\text{TE}_{11},\ \text{TM}_{11} & f_{c,01} = \frac{c}{2}\sqrt{\frac{1}{a^2} + \frac{1}{b^2}} & \\
\text{TE}_{20},\ \text{TM}_{20} & f_{c,20} = \frac{c}{a} & \\
\vdots & \vdots &
\end{array}$$

In order to have single-mode operation, the operating frequency must be between $f_{c,10}$ and the next larger cutoff frequency.

Can a time-harmonic current source excite a given mode if $f < f_{c,mn}$?

Some other types of waveguides have two separate conductors (parallel wires, coaxial cable, microstrip, etc.). If these waveguides are analyzed mathematically in a similar way to what we have done here for the rectangular waveguide, we find that the dominant mode, or the mode with lowest cutoff frequency, has a cutoff frequency of zero. We call this the TEM mode, because both $\overline{E}$ and $\overline{H}$ are transverse to the axis of the waveguide. So, circuit theory deals with the TEM mode of two-conductor waveguides at low operating frequencies.

7.2.4 Dominant Mode (TE_{10})

The electric and magnetic fields for the TE_{10} mode are

$$H_z = H_{z10} \cos{(\pi x/a)} e^{-jk_z z} \tag{7.65}$$

$$H_x = H_{x10} \sin{(\pi x/a)} e^{-jk_z z} \tag{7.66}$$

$$E_y = E_{y10} \sin{(\pi x/a)} e^{-jk_z z} \tag{7.67}$$

All other field components are zero. The electric field is zero at the right and left walls of the waveguide, and in between varies according to one half-cycle of the sine function. Do these fields satisfy the required boundary conditions?

We can get some insight into the physical behavior of the mode using plane waves. Using Euler's theorem, we can expand the sine function in E_y into two complex exponentials, so that

$$E_y = \underbrace{\frac{E_{y10}}{2j} e^{-j(-\pi x/a + k_z z)}}_{\text{Plane wave}} - \underbrace{\frac{E_{y10}}{2j} e^{-j(\pi x/a + k_z z)}}_{\text{Plane wave}} \tag{7.68}$$

So, the TE_{10} mode can be viewed as a wave that reflects from the left and right sides of the waveguide and "zigzags" back and forth as it propagates in the z direction. As the operating frequency decreases towards the cutoff frequency $f_{c,10}$, the angle of the wave becomes steeper until it becomes 90° at $f = f_{c,10}$ and the wave no longer makes forward progress. For $f < f_{c,10}$, k_z is imaginary, so the fields decay exponentially away from a source and the mode is in cutoff.

It is also interesting to notice that right at the cutoff frequency of the dominant mode ($f = f_{c,10} = c/(2a)$), the free space wavelength $\lambda = c/f$ is equal to $2a$. This gives an important piece of intuition into mode behavior for any (single-conductor) waveguide: in order to have any propagation at all, the free space operating wavelength must be less than about twice the longest dimension of the waveguide cross section.

7.2.5 Mode Patterns

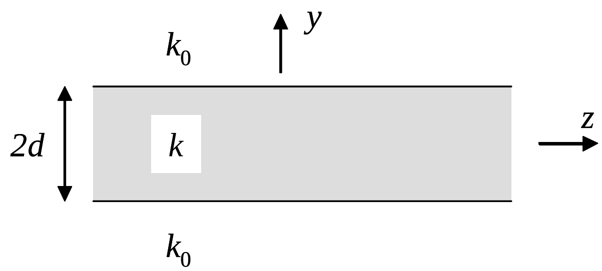
Higher order modes have more variations of the field across the cross section of the waveguide. For the TE_{20} mode, for example, the x dependence of E_y is of the form $\sin{(2\pi x/a)}$, so the field varies according to one full cycle of the sine function. The field has maxima and minima at $x = a/4$ and $x = 3a/4$ and a node or zero at $x = a/2$.

Finally, it is important to observe that the field inside the waveguide must be equal to a sum over all possible modes (even modes that are in cutoff). When we put a time-harmonic current source in the waveguide, its location, orientation, and strength determine the amplitudes of each mode. If the source is at a location where the field pattern of a particular mode is zero, then that mode is not excited and the amplitude of that mode is zero. If the orientation of the fields produced by the source are orthogonal to the fields of a mode, the amplitude of the mode is also zero. In practice, it is hard to design a source that doesn't couple into many modes, so that is why choosing a single-mode operating frequency is important (so the undesired modes decay away).

The two basic types of sources are a straight probe and a loop. The probe produces an electric field aligned with the probe, and the loop produces a magnetic field through the loop. Where would we put sources in order to excite the TE_{10} mode? The TE_{20} mode?

7.3 Dielectric Slab Waveguide

We will now examine the waveguide properties of a "slab" of dielectric material.



Two key concepts concerning dielectric waveguides deserve attention. The first is that, due to the symmetry of the geometry, the fields will either be symmetric or anti-symmetric about the x-z plane. The second is that in order for the field to be guided by the high-permittivity dielectric slab, the fields outside the slab must be evanescent, *i.e.* they decay in the y direction. What we will have inside the slab is a plane wave that bounces back and forth due to total internal reflection. We will use these observations in the formulations that follow.

7.3.1 TE Modes

The electric field for the TE modes must satisfy the homogeneous wave equation. Our experience with the parallel plate waveguide tells us what the solutions must be (before application of the boundary conditions). However, our argument about symmetry makes it so that within the slab, the variation will either be $\sin k_y y$ or $\cos k_y y$ (recall that in general, the field variation can be a combination of these two). Therefore, the electric field can be written as

$$\overline{E} = \hat{x} \begin{cases} E_1 e^{-\alpha y - jk_z z} & y \geq d \\ E_0 \left\{ \begin{array}{c} \sin k_y y \\ \cos k_y y \end{array} \right\} e^{-jk_z z} & |y| \leq d \\ \left\{ \begin{array}{c} - \\ + \end{array} \right\} E_1 e^{+\alpha y - jk_z z} & y \leq -d \end{cases} \tag{7.69}$$

where the top and bottom lines in the braces refer to the antisymmetric and symmetric modes, respectively. Using Faraday's law, we can now compute the magnetic fields

$$\overline{H} = \frac{-1}{j\omega\mu} \nabla \times \overline{E} = \begin{cases} \frac{E_1}{\omega\mu_0} \left(\hat{y} k_z + \hat{z} j\alpha \right) e^{-\alpha y - jk_z z} & y \geq d \\ \frac{E_0}{\omega\mu} \left(\hat{y} k_z \left\{ \begin{array}{c} \sin k_y y \\ \cos k_y y \end{array} \right\} + \hat{z} j k_y \left\{ \begin{array}{r} -\cos k_y y \\ \sin k_y y \end{array} \right\} \right) e^{-jk_z z} & |y| \leq d \\ \left\{ \begin{array}{c} - \\ + \end{array} \right\} \frac{E_1}{\omega\mu_0} \left(\hat{y} k_z - \hat{z} j\alpha \right) e^{+\alpha y - jk_z z} & y \leq -d \end{cases} \tag{7.70}$$

Note that in these field expressions, we have used that the z variation is $e^{-jk_z z}$ both inside and outside the slab. How do we know that this propagation constant is the same in both regions? Note also that we have four unknowns: E_1/E_0, k_y, α, and k_z.

Since these fields must obey the wave equation (with $\partial^2/\partial x^2 = 0$), we know that:

$$k_y^2 + k_z^2 = k_1^2 = \omega^2 \mu \epsilon_1 \tag{7.71}$$

$$-\alpha^2 + k_z^2 = k_2^2 = \omega^2 \mu_0 \epsilon_2 \tag{7.72}$$

which gives us two constraints for determining our unknowns. We need two additional constraints in order to find all four unknowns.

Let's start by enforcing continuity of tangential electric fields at the dielectric-air interface. We will first consider the symmetric modes. Therefore, at $y = d$

$$E_0 \cos(k_y d) e^{-jk_z z} = E_1 e^{-\alpha d} e^{-jk_z z} \qquad \rightarrow \qquad \cos(k_y d) E_0 = e^{-\alpha d} E_1 \tag{7.73}$$

Note that applying continuity at $y = -d$ results in an identical equation, so this does not help us. This stems from the symmetry of the problem, and in reality we have already used this symmetry to break the problem into symmetric and asymmetric modes.

Since we need one more equation, we will apply continuity of tangential (*i.e.* $\hat{z}$ component) magnetic field at the boundary. At $y = d$ we have

$$jk_y \frac{E_0}{\omega \mu} \sin(k_y d) e^{-jk_z z} = j\alpha \frac{E_1}{\omega \mu_0} e^{-\alpha d} e^{-jk_z z} \qquad \rightarrow \qquad \frac{k_y}{\mu} \sin(k_y d) E_0 = \frac{\alpha}{\mu_0} e^{-\alpha d} E_1 \tag{7.74}$$

and again, we get the exact same equation at $y = -d$. The easiest thing to do is to divide these two equations by each other. This yields

$$\frac{\frac{k_y}{\mu} \sin(k_y d) E_o}{\cos(k_y d) E_o} = \frac{\frac{\alpha}{\mu_0} e^{-\alpha d} E_1}{e^{-\alpha d} E_1} \tag{7.75}$$

$$\frac{\alpha}{\mu_0} = \frac{k_y}{\mu} \tan(k_y d). \tag{7.76}$$

which can be re-written as

$$\boxed{(\alpha d) = \frac{\mu_0}{\mu} (k_y d) \tan(k_y d) \qquad \text{Symmetric TE modes} \qquad (7.77)}$$

Combining (7.71) and (7.72) leads to

$$(k_y d)^2 + (\alpha d)^2 = \omega^2 \mu_0 \epsilon_0 (n_1^2 - n_2^2) d^2 \tag{7.78}$$

Finally, we can combine these two equations to the form

$$\tan(k_y d) = \sqrt{\frac{\omega^2 \mu_0 \epsilon_0 (n_1^2 - n_2^2) d^2}{(k_y d)^2} - 1} \tag{7.79}$$

1. Solutions in the range $(m-1)\pi/2 \leq k_y d \leq m\pi/2$, $m = 1, 3, 5, \ldots$ we will call TE$_m$ modes. These correspond to the symmetric TE modes.

2. Cutoff occurs when the mode is no longer guided, which occurs as soon as α becomes imaginary. So, we define cutoff as the frequency at which $\alpha = 0$. Using (7.77), this implies that $\tan(k_y d) = 0$ such that $k_y d = (m-1)\pi/2$, $m = 1, 3, 5, \ldots$. Using (7.78) with $\alpha = 0$ leads to

   $$f_m = \frac{m-1}{4d} \frac{1}{\sqrt{\mu_0 \epsilon_0 (\mu_r \epsilon_r - 1)}} \tag{7.80}$$

   Note that $f_1 = 0$, so the lowest order mode propagates at any frequency. Furthermore, since at cutoff $k_z = k_0$ and $k_z^2 + k_y^2 = k^2$, the angle of incidence of the wave on the dielectric boundary can be expressed as

   $$\theta_i = \sin^{-1} \frac{k_z}{\sqrt{k_z^2 + k_y^2}} = \sin^{-1} \frac{k_0}{k} = \sin^{-1} \sqrt{\frac{\mu_0 \epsilon_0}{\mu \epsilon}} = \theta_c \tag{7.81}$$

   which you may recognize at the critical angle. So, cutoff occurs when the angle of incidence on the boundary is smaller than the critical angle. Makes sense, doesn't it?

   Observe also that the cutoff condition of $k_z = k_0$ means that the propagation constant becomes that of the surrounding medium. We will revisit this below in optical fibers.

3. Note that k_y is frequency dependent, unlike in the parallel plate waveguide.

4. As the frequency gets larger, $\alpha \to \infty$ which means that the field decays very rapidly outside the dielectric. The behavior of the mode becomes like that of a parallel plate waveguide filled with a dielectric.

Note that we could repeat the entire procedure for the antisymmetric TE modes. The dispersion relation (7.78) remains the same. The guidance condition becomes

$$\boxed{(\alpha d) = -\frac{\mu_0}{\mu} (k_y d) \cot(k_y d) \qquad \text{Antisymmetric TE modes} \qquad (7.82)}$$

Again, cutoff occurs for $k_y d = (m-1)\pi/2$, $m = 2, 4, 6, \ldots$. These are therefore the even order TE modes.

We can solve these nonlinear transcendental equations using a nonlinear solver on a computer or calculator. We can also solve these equations graphically. We will plot each equation separately with plot axes of $k_y d$ and αd. Think of $k_y d \equiv x$ and $\alpha d \equiv y$. The various equations then become

$$(\alpha d) = \frac{\mu_0}{\mu} (k_y d) \tan(k_y d) \qquad \text{Symmetric TE modes} \tag{7.83}$$

$$y = x \tan(x), \tag{7.84}$$

$$(\alpha d) = -\frac{\mu_0}{\mu} (k_y d) \cot(k_y d) \qquad \text{Antisymmetric TE modes} \tag{7.85}$$

$$y = -x \cot(x) \tag{7.86}$$

and

$$(k_y d)^2 + (\alpha d)^2 = \omega^2 \mu_o \epsilon_o \left(n_1^2 - n_2^2\right) d^2 \tag{7.87}$$

$$x^2 + y^2 = (k_o d)^2 \left(n_1^2 - n_2^2\right) \tag{7.88}$$

7.3.2 TM Modes

We can repeat the whole process for TM modes. In this case, we have

$$\overline{H} = \hat{x} \begin{cases} H_1 e^{-\alpha y - jk_z z} & y \geq d \\ H_0 \left\{ \begin{array}{c} \sin k_y y \\ \cos k_y y \end{array} \right\} e^{-jk_z z} & |y| \leq d \\ \left\{ \begin{array}{c} - \\ + \end{array} \right\} H_1 e^{+\alpha y - jk_z z} & y \leq -d \end{cases} \tag{7.89}$$

where the top and bottom lines in the braces refer to the antisymmetric and symmetric modes, respectively. Using Ampere's law, we can now compute the electric fields

$$\overline{E} = \frac{1}{j\omega\epsilon} \nabla \times \overline{H} = \begin{cases} \frac{H_1}{\omega\epsilon_0} \left(-\hat{y}k_z - \hat{z}j\alpha\right) e^{-\alpha y - jk_z z} & y \geq d \\ \frac{H_0}{\omega\epsilon} \left(-\hat{y}k_z \left\{ \begin{array}{c} \sin k_y y \\ \cos k_y y \end{array} \right\} - \hat{z}jk_y \left\{ \begin{array}{r} -\cos k_y y \\ \sin k_y y \end{array} \right\} \right) e^{-jk_z z} & |y| \leq d \\ \left\{ \begin{array}{c} - \\ + \end{array} \right\} \frac{H_1}{\omega\epsilon_0} \left(-\hat{y}k_z + \hat{z}j\alpha\right) e^{+\alpha y - jk_z z} & y \leq -d \end{cases} \tag{7.90}$$

We go through the exact same sequence of steps for this case. The dispersion relations remain the same. The guidance conditions become

$$(\alpha d) = \frac{\epsilon_0}{\epsilon}(k_y d)\tan(k_y d) \qquad \text{Symmetric TM modes} \tag{7.91}$$

$$(\alpha d) = -\frac{\epsilon_0}{\epsilon}(k_y d)\cot(k_y d) \qquad \text{Antisymmetric TM modes} \tag{7.92}$$

Dielectric Waveguide Example

How many guided modes exist in a dielectric waveguide that has the following parameters? Index of refraction of the core $n_1 = 1.6$, index of refraction of the cladding $n_2 = 1.5$, wavelength $\lambda = 1.0\,\mu$m, and waveguide core thickness $2d = 4\,\mu$m. What are the values of k_z and α for the dominant TE mode? What can we change the thickness to for single mode operation?

1. **Guided modes.** We could simply use Eq. (7.80), but it is more instructive think about the meaning of what it means for a mode to be guided. The equations for guided modes are

$$\begin{aligned} \alpha d &= k_y d \tan(k_y d) \quad \text{(Symmetric modes, TE)} && (7.93) \\ \alpha d &= -k_y d \cot(k_y d) \quad \text{(Antisymmetric modes, TE)} && (7.94) \\ (k_y d)^2 + (\alpha d)^2 &= (k_o d)^2 \left(n_1^2 - n_2^2\right) && (7.95) \end{aligned}$$

   In order to have a guided mode, α has to be real, so from the third equation, we need to have

$$k_y d < k_o d \sqrt{n_1^2 - n_2^2} = \frac{2\pi}{1.0}\frac{4}{2}\sqrt{n_1^2 - n_2^2} \simeq 7.0$$

   Cutoff occurs where α is zero, which from the zeros of the tangent and cotangent functions in the first two equations corresponds to

$$k_y d = \frac{\pi}{2}(m-1)$$

   For the mth mode to be guided, $k_y d$ has to be larger than this value. So, the biggest value m can have is 5. There are three guided symmetric modes and two guided antisymmetric modes for each polarization: TE_1, TM_1, TE_2, TM_2, TE_3, TM_3, TE_4, TM_4, TE_5, TM_5.

2. **Graphical solution for guided mode parameters.** Using $x = k_y d$ and $y = \alpha d$, the three equations become

$$\begin{aligned} y &= x \tan x && (7.96) \\ y &= -x \cot x && (7.97) \\ x^2 + y^2 &= (k_o d)^2 \left(n_1^2 - n_2^2\right) && (7.98) \end{aligned}$$

   The radius of the circle defined by the third equation is the same as the limiting value computed above for $k_y d$, so $r \simeq 7.0$. To plot the other two functions, we observe that $x \tan x$ is equal to zero when $x = 0, \pi, 2\pi, 3\pi, \ldots$ and is infinite when $x = \pi/2, 3\pi/2, 5\pi/2, \ldots$. The function $-x \cot x$ is equal to zero when $x = \pi/2, 3\pi/2, 5\pi/2, \ldots$ and is infinite when $x = \pi, 2\pi, 3\pi, \ldots$. Also, when $x = 0$, $-x \cot x = -1$. For the TM modes, all that changes is that we get an extra factor of n_2^2/n_1^2 in Eqs. (7.96) and (7.97).

   The guided modes correspond to points where the circle Eq. (7.98) intersects the lines defined by Eqs. (7.96) and (7.97).

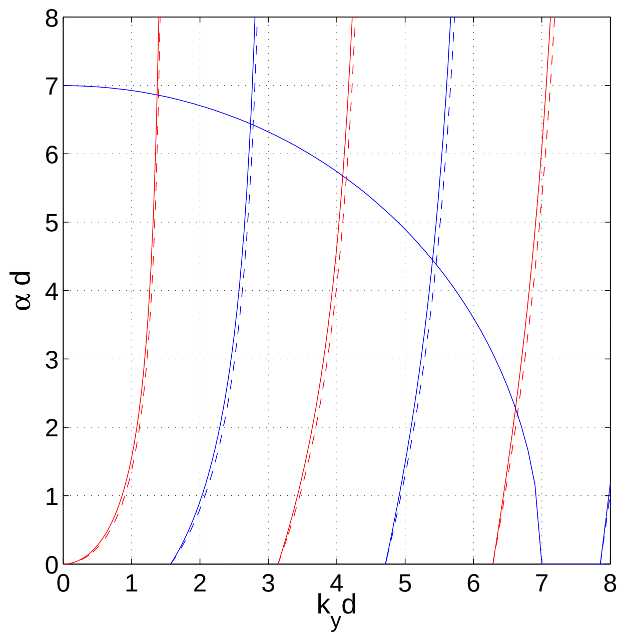


Figure 7.1: Graphical solution method for a dielectric slab waveguide. Solid lines represent TE modes and dashed lines represent TM modes.

The dominant TE mode is TE_1. From the Fig. 7.1, it can be seen that the intersection of the circle and the first solid red line is at about $k_y d = 1.4$ and $\alpha d = 6.8$, so that

$$\begin{aligned} k_y &\simeq \frac{1.4}{2\,\mu\text{m}} = 0.7\,\text{rad}/\mu\text{m} \\ \alpha &\simeq \frac{6.8}{2\,\mu\text{m}} = 3.4\,\text{Np}/\mu\text{m} \\ k_z &= \sqrt{k_2^2 + \alpha^2} \simeq 10.0\,\text{rad}/\mu\text{m} \end{aligned}$$

From the value for α, it can be seen that the wave is mostly confined to within a micrometer or so of the slab. Also, by comparing the wavelength $\lambda = 2\pi/k_z$ in the slab to the free space wavelength, we find that the mode is behaving as if it were in a homogeneous medium with an effective index of refraction of 1.592, which is between n_1 and n_2.

3. For single mode operation, we need

$$\begin{aligned} r &< 0.5\pi \\ \frac{2\pi}{1.0} d\sqrt{1.6^2 - 1.5^2} &< \frac{\pi}{2} \\ \Rightarrow \quad d &< 0.449\,\mu\text{m} \end{aligned}$$

7.4 Review

Plane waves

- Wave equation, dispersion relation
- Plane wave solutions
- Wavenumber, wavelength, phase velocity
- Wavevector
- TEM waves, orthogonality relationships, intrinsic impedance
- Polarization - linear, RH circular, LH circular, elliptical
- Lossy media - good conductors, good insulators, skin depth
- Power and energy - Poynting theorem, Poynting vector

Reflection and refraction

- Normal incidence, transmission line analogy
- Oblique incidence - perpendicular/parallel polarization
- Phase matching condition, k-diagrams, Snell's law
- Total internal reflection, total transmission

Waveguides

- Rectangular metallic waveguide
  TE/TM modes
  Modal solutions, boundary conditions, dispersion relation
  Cutoff
  Dominant mode
  Mode patterns, sources
- Dielectric slab waveguide
  Modal solutions, symmetric/antisymmetric modes
  Cutoff
  Graphical solution method

Fundamentals

Plane waves

$\overline{E} = \overline{E}_0 e^{-j\overline{k}\cdot\overline{r}}$

Wavevector: $\overline{k} = k_x\hat{x} + k_y\hat{y} + k_z\hat{z}$

Dispersion relation: $k^2 = k_x^2 + k_y^2 + k_z^2 = \omega^2\mu\epsilon$

Wavelength: $\lambda = 2\pi/k$, phase velocity: $u_p = \omega/k$

TEM waves: $\hat{k}\cdot\overline{E} = \hat{k}\cdot\overline{H} = \overline{E}\cdot\overline{H} = 0$, $\overline{H} = \hat{k}\times\overline{E}/\eta$, intrinsic impedance: $\eta = \sqrt{\mu/\epsilon}$

Polarization: $\overline{\mathcal{E}}(x,y,z,t) = \text{Re}\left\{\overline{E}(x,y,z)e^{j\omega t}\right\}$

Lossy media: $\epsilon_c = \epsilon - j\sigma/\omega$, $\sqrt{-\omega^2\mu\epsilon_c} = \alpha + j\beta$

Poynting theorem: $\oint_S(\overline{\mathcal{E}}\times\overline{\mathcal{H}})\cdot d\overline{s} + \frac{\partial}{\partial t}\int_V\left[\frac{1}{2}\mu\|\overline{\mathcal{H}}\|^2 + \frac{1}{2}\epsilon\|\overline{\mathcal{E}}\|^2\right]dV + \int_V\sigma\|\overline{\mathcal{E}}\|^2 dV = 0$

Time-average Poynting vector: $\overline{S}_{av} = \frac{1}{2}\text{Re}\left\{\overline{E}\times\overline{H}^*\right\}$ (W/m^2)

Reflection and refraction

Plane waves, boundary conditions, plane of incidence, polarization

Total internal reflection: k_{2z} imaginary (evanescent wave)

Total transmission: $\Gamma = 0$

Waveguides

Rectangular waveguide: plane waves/standing waves, boundary conditions, dispersion relation

Cutoff: $k_z = 0$

$f < f_c \Rightarrow k_z$ imaginary (evanescent wave), $f > f_c \Rightarrow k_z$ real

Dielectric slab waveguide: plane waves/standing waves inside slab, evanescent waves outside, dispersion relations, boundary conditions

Cutoff: $\alpha = 0$

$f < f_c \Rightarrow \alpha$ imaginary (non-guided mode), $f > f_c \Rightarrow \alpha$ real (guided mode)

Graphical solution method

Chapter 8

Antennas

An antenna operating as a transmitter transforms energy from a transmission line or a waveguide to a wave propagating in free space. A receiving antenna transforms energy in a wave in free space to a signal on a transmission line. Examples of antenna types are

1. Dipole: two wires driven by a source at a gap between them.
2. Monopole: single wire above a ground plane.
3. Loop: radiates by producing a magnetic field.
4. Patch antenna: fabricated on a printed circuit board surface.
5. Helical antenna: radiates circular polarization.
6. Yagi: multiple dipole type arms to produce a directive radiation pattern.
7. Aperture antennas: open-ended waveguides, horns, slotted waveguides.
8. Reflector antennas: feed antenna with reflector surface to focus the radiation pattern.
9. Array antennas: multiple antennas weighted to shape radiation pattern electronically.

Antennas are used in thousands of applications and can be designed for a wide variety of purposes. They represent a core area of technology that helps to power satellite communication systems, remote sensing instruments such as weather monitoring satellites, navigational aids, radars and other types of sensors, cellular networks, wireless internet services, and other airborne and mobile communications systems of many kinds.

8.1 Radiation Pattern and Input Impedance

The most basic properties of an antenna are the **radiation pattern** and the **input impedance**. The radiation pattern governs where the energy radiated by the antenna when operated as a transmitter is directed in space. As a receiver, the antenna responds preferentially to signals arriving from angles where the radiation pattern is large. The input impedance represents the load that the antenna presents to its connected transmission line. A dipole antenna with its driving transmission line and radiated electric fields is shown in Fig. 8.1.

If an antenna is small relative to the operating wavelength, the radiation pattern is very broad (close to isotropic). If the antenna is much larger than the operating wavelength, the radiation pattern can be highly directive, so that most of the energy radiates in a narrow solid angle. The radiation pattern is typically parameterized using **directivity** or **gain**.

From the perspective of the transmission line connected to the antenna, the circuit model for an antenna

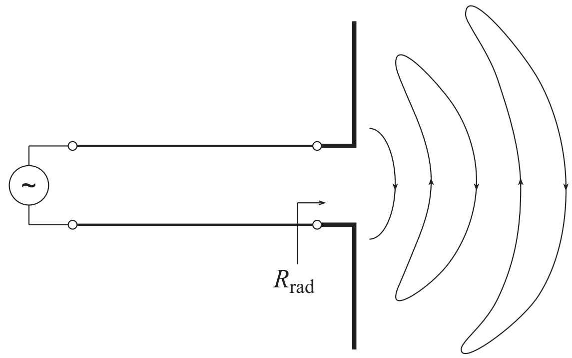


Figure 8.1: Dipole antenna.

includes the loss resistance and the reactance of the antenna, as well as the power radiated by the antenna into space, as a transmission line load impedance. The circuit model is shown in Fig. 8.7. The input impedance of the antenna is

$$Z_{\text{in}} = R_{\text{loss}} + R_{\text{rad}} + jX_{\text{in}} \tag{8.1}$$

R_{loss} represents the losses in the dielectric and conductive structures that make up the antenna. X_{in} represents the reactance of the antenna. R_{rad} represents an equivalent resistance corresponding to the power radiated by the antenna. The radiation resistance is usually the largest contribution to the effective impedance of the antenna at its input port as seen by the transmission line driving it.

The antenna input impedance is typically a strong function of frequency. The reactive part of the input impedance is zero if the antenna is operated at its resonance frequency. We want the reactance to be zero or at least small, so that the input impedance of the antenna is small enough that we can have a reasonably large amount of current flowing into the antenna without requiring a large source voltage. Away from resonance, the reactive part of the impedance may be capacitative or inductive. Antennas that are designed to have a small reactance over a wide range of frequencies are called broadband or ultrawideband antennas. The loss resistance and the reactance require a detailed model or numerical simulation of the near fields at the location of the antenna structure. By the principle of conservation of energy, the radiation resistance can be found from the fields radiated by the antenna.

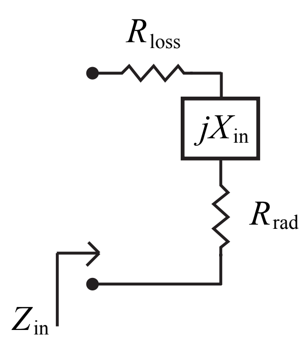


Figure 8.2: Equivalent circuit model of an antenna as a load on the connected transmission line.

To determine these properties for a given antenna, we need to be able to solve Maxwell's equations to get the electromagnetic fields radiated by the antenna. The basic approach is to model the antenna as an equivalent current source and then find the fields produced by the current.

8.2 Calculating Radiated Fields

We have talked a lot about waves: how they propagation, how they behave at interfaces, and how to guide them. We now need to learn how to generate these waves. The basic principle is: *currents radiate fields.*

The question is: how do we determine the fields if we know the currents? Let's go back to our wave equation for $\overline{E}$, but this time allow for the presence of currents:

$$\nabla \times (\nabla \times \overline{E}) = -j\omega\mu\nabla \times \overline{H} = -j\omega\mu(j\omega\epsilon\overline{E} + \overline{J}) = \omega^2\mu\epsilon\overline{E} - j\omega\mu\overline{J}$$

To simplify the derivative operators, we use the identity

$$\nabla \times \nabla \times \overline{E} = \nabla(\nabla \cdot \overline{E}) - \nabla^2\overline{E}$$

Normally, we let $\nabla \cdot \overline{E} = 0$ since $\nabla \cdot \overline{D} = 0$ in a charge-free region. However, we typically create radiation by injecting current into a metal structure. Charge can build up on the structure, resulting in an excess of charge, as shown in Fig. 8.3, so that $\nabla \cdot \overline{D} \neq 0$. This makes the wave equation too difficult to solve directly. We therefore resort to an alternate procedure.

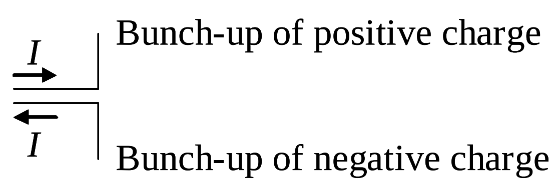


Figure 8.3: Dipole antenna with charge build up on the dipole arms.

Another issue with the Helmholtz equation that we used to derive plane wave solutions is that waves of the form $\overline{E} = \hat{z}e^{-jkz}$, called longitudinal waves, are solutions of the Helmholtz equation, but are not valid solutions of Maxwell's equations. If we were to use the second order equation for the electric field directly to analyze radiation, we would arrive at an incorrect solution for the radiated field. We need a different approach, that will eliminate longitudinal waves from the radiated field solution.

8.2.1 Magnetic Vector Potential and the Radiation Integral

We recall that we used the magnetic vector potential $\overline{A}$ (units Wb/m) in magnetostatics to simplify the analysis. The same approach can be used here. The magnetic vector potential is defined by

$$\overline{B} = \nabla \times \overline{A} \tag{8.2}$$

which naturally satisfies $\nabla \cdot \overline{B} = 0$ since $\nabla \cdot (\nabla \times \overline{A}) = 0$ for any arbitrary vector $\overline{A}$ (i.e., any vector field with zero divergence can be represented as the curl of some other vector field). Let's go through this for dynamic fields. Using Faraday's law,

$$\nabla \times \overline{E} = -j\omega\overline{B} = -j\omega\nabla \times \overline{A}$$

so that

$$\nabla \times \overline{E} + j\omega\nabla \times \overline{A} = \nabla \times (\overline{E} + j\omega\overline{A}) = 0$$

We now define

$$\overline{E} + j\omega\overline{A} = -\nabla\phi \tag{8.3}$$

where ϕ is the scalar electric potential for dynamics. We make this definition because there is an identity $\nabla \times \nabla\phi = 0$ for a scalar function ϕ (in other words, any vector field with zero curl can be represented as the gradient of some scalar function). The minus sign is simply there for consistency with the static electric potential.

Now, using Ampere's law

$$\nabla \times \overline{H} = j\omega\epsilon\overline{E} + \overline{J}$$

together with

$$\overline{H} = \frac{\overline{B}}{\mu} = \frac{1}{\mu}\nabla \times \overline{A}$$

we have that

$$\frac{1}{\mu}\nabla \times \nabla \times \overline{A} = \overline{J} + j\omega\epsilon\overline{E} = \overline{J} + j\omega\epsilon(-j\omega\overline{A} - \nabla\phi)$$
$$\nabla(\nabla \cdot \overline{A}) - \nabla^2\overline{A} = \mu\overline{J} + \omega^2\mu\epsilon\overline{A} - j\omega\mu\epsilon\nabla\phi$$

Now, so far we have only specified $\nabla \times \overline{A}$, which does not uniquely specify $\overline{A}$. We need to specify $\nabla \cdot \overline{A}$ as well. A good choice is

$$\nabla \cdot \overline{A} = -j\omega\mu\epsilon\phi \tag{8.4}$$

We call this the Lorenz gauge. We can take the gradient to obtain $\nabla(\nabla \cdot \overline{A}) = -j\omega\mu\epsilon\nabla\phi$. Then,

$$-\nabla^2\overline{A} = \mu\overline{J} + \omega^2\mu\epsilon\overline{A} = \mu\overline{J} + k^2\overline{A}$$

or

$$(\nabla^2 + k^2)\overline{A} = -\mu\overline{J} \tag{8.5}$$

We have already looked at the homogeneous solutions of this differential equation, which are plane waves. Now, we need to look at the inhomogeneous solutions for a nonzero right-hand side (forcing function). The resulting solution will provide the magnetic vector potential $\overline{A}$ given a current distribution $\overline{J}$. Once we know $\overline{A}$, we can obtain $\overline{H}$ and $\overline{E}$.

The solution process for the PDE (8.5) is a graduate level topic, and is a bit too complicated for this class. In order to solve the PDE, a boundary condition is required. The boundary condition that we will choose is that of free space (or empty space) with a radiation boundary condition at infinity. This means that as we get far away from sources, waves are outgoing and decay as they move away from the source. This is related to the physics principle of causality. With this boundary condition, the solution to the differential equation if the right-hand side is a delta function located at the origin has the form $\overline{A}(\overline{R}) = e^{-jkR}/(4\pi R)$. We can find the solution to the differential equation for any current source by convolving this function with the current source.

The end result of the analysis is that for a current density $\overline{J}(\overline{r}')$, the magnetic vector potential is

$$\overline{A}(\overline{R}) = \frac{\mu}{4\pi}\int_V \overline{J}(\overline{R}')\frac{e^{-jk|\overline{R}-\overline{R}'|}}{|\overline{R}-\overline{R}'|}dV' \tag{8.6}$$

Remember that $\overline{R}$ when used as the argument of a function is simply a surrogate for the position. The $\overline{R}'$ is just a dummy variable in the integration over the source $\overline{J}$. For example, in Cartesian coordinates we could write $\overline{J}(\overline{R}') = \overline{J}(x', y', z')$. So, we have two sets of coordinates:

$\overline{R}'$ represents integration or *source* coordinates, since that is where the current source $\overline{J}$ is evaluated.

$\overline{R}$ represents *observation* or *field* coordinates, because that is where the field $\overline{A}$ is evaluated.

If you have had ECEn 380, you can see that the integral in (8.6) is a convolution, where $\overline{R}$ is like t and $\overline{R}'$ is like τ, so that $\overline{A}(\overline{R}) = \overline{J}(\overline{R}) * g(\overline{R})$. The function $g(\overline{R})$ is a spherical wave radiating away from the source point $\overline{r}'$. This function plays the role of the impulse response of free space, and is called a Green's function.

This result provides a procedure for analyzing antenna radiation:

1. Model the antenna as a time-harmonic current source $\overline{J}$.
2. Compute $\overline{A}$ radiated by the current using Eq. (8.6).
3. Find $\overline{B}$ using $\overline{B} = \nabla \times \overline{A}$ and $\overline{E}$ using Ampere's law or Eq. (8.4) together with (8.3).

8.2.2 Radiation Integral for the Electric Field

Using the results from the previous section, the electric field intensity radiated by a current source in free space is

$$\overline{E}(\overline{r}) = -j\omega\mu \left[1 + \frac{1}{k^2}\nabla\nabla\cdot\right] \int g(\overline{r}, \overline{r}')\overline{J}(\overline{r}')\, d\overline{r}' \tag{8.7}$$

This is the free space radiation integral for the electric field in terms of the scalar Green's function and the electric current density. We could also find the electric field by using the definition of the magnetic vector potential to find the magnetic field and then using Ampere's law, but that would require computing two curl operations, which is more mathematically challenging than (8.7).

This expression is similar to (8.6) for the magnetic vector potential, which shows the close connection between $\overline{E}$ and $\overline{A}$. The component of $\overline{E}$ that arises from the unity term in the square brackets is identical to (8.6) with the exception of an additional scale factor of $-j\omega$. The key difference is that the "$\nabla\nabla$" term in (8.7) removes the longitudinal wave component of the magnetic vector potential, making $\overline{E}$ a valid solution to Maxwell's equations.

8.2.3 Far Field Radiation Integral

Often in antenna analysis we are only interested in the fields far from the antenna. In this case, we can simplify the radiation integral considerably. If the current source is near the origin and the field observation point $\overline{r}$ is far from the origin, then we can use the approximation

$$|\overline{r} - \overline{r}'| \simeq r - \hat{r}\cdot\overline{r}' \tag{8.8}$$

in the radiation integral.

The next step is to use the far field expansion (8.8) in the scalar Green's function. The distance $|\overline{r} = \overline{r}'|$ appears twice in the scalar Green's function, once in the phase and again in the denominator, but we do not need to use approximations of the same accuracy in both places. In the phase term, a small offset matters even if the wave has propagated a long distance, so we must use both terms of (8.8). The denominator does not vary as much if the distance $|\overline{r} = \overline{r}'|$ changes slightly, so there we only need the leading term of the far field expansion. Making these substitutions into the scalar Green's function leads to

$$g(\overline{r}, \overline{r}') \simeq \frac{e^{-jkr}}{4\pi r} e^{jk\hat{r}\cdot\overline{r}'} \tag{8.9}$$

in the far field.

The derivative operators in (8.7) can also be simplified when the observation point is far from the source. The terms in the gradient of the scalar Green's function in the spherical coordinate system can be divided into terms of order $1/r$ and higher order terms according to

$$\nabla \frac{e^{-jkr}}{r} \simeq -jk\hat{r} \frac{e^{-jkr}}{r}$$

This result suggests that the ∇ operator can be replaced with $-jk\hat{r}$ when r is large.

Using this approximation in the radiation integral (8.7) leads to the far field radiation integral

$$\overline{E}(\overline{r}) = -j\omega\mu(1 - \hat{r}\hat{r}\cdot) \frac{e^{-jkr}}{4\pi r} \int e^{jk\hat{r}\cdot\overline{r}'} \overline{J}(\overline{r}')\, d\overline{r}' \tag{8.10}$$

Each term in this expression has a physical meaning. The leading constant adjusts the units of the electric field intensity. The $\hat{r}\hat{r}$ term subtracts out waves with electric field in the r direction, since these are longitudinal waves and are not valid solutions of Maxwell's equation. The term e^{-jkr} represents the phase of a spherical wave as it moves away from the origin, and the factor of $1/r$ accounts for spreading of energy radiated by the source over a sphere of radius r.

8.3 Hertzian Dipole

Let's show how to apply this procedure to find the fields radiated by an antenna. We start with the simplest case. Suppose we have a very short dipole of length ℓ. If we assume that the dipole length is short compared to the wavelength, the current is approximately constant over the length of the antenna. The current density can be written as

$$\overline{J}(\overline{r}') = \overline{J}(x', y', z') = I_o \hat{z}\delta(x')\delta(y') \qquad -\ell/2 \leq z' \leq \ell/2.$$

When ℓ is very short, we call this the **Hertzian dipole** model.

We can find the fields far away from the antenna using (8.10) directly, but we will find the near field radiated by the antenna first, and then we will apply the far field approximation (in the future, we will usually go directly to the far field approximation, since that is what is needed for most antenna problems).

Plugging the current into the equation for the magnetic vector potential yields

$$\overline{A}(\overline{R}) = \frac{\mu_0}{4\pi} \int_{-\ell/2}^{\ell/2} I_o \hat{z} \frac{e^{-jk|\overline{R}-\overline{R}'|}}{|\overline{R}-\overline{R}'|} dz'$$

The field and observation points for the magnetic vector potential integration are shown in Fig. 8.4.

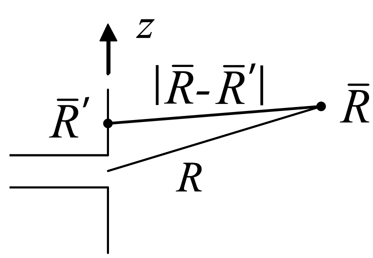


Figure 8.4: Hertzian dipole with field and observation points.

For a very short dipole, $|\overline{R}-\overline{R}'| \approx R$. So,

$$\overline{A}(\overline{R}) = \frac{\mu_0}{4\pi} I_o \hat{z} \frac{e^{-jkR}}{R} \int_{-\ell/2}^{\ell/2} dz' = \frac{\mu_0}{4\pi} I_o \ell \hat{z} \frac{e^{-jkR}}{R}$$

Typically, we are interested in determining the field a fixed distance away from the origin. So, instead of using Cartesian coordinates, it's easier to use spherical coordinates:

$$\hat{z} = \hat{R}\cos\theta - \hat{\theta}\sin\theta$$

$$\overline{A} = (\hat{R}\cos\theta - \hat{\theta}\sin\theta)\frac{\mu_0}{4\pi} I_o \ell \frac{e^{-jkR}}{R}$$

Now, we can find the magnetic and electric fields:

$$\begin{aligned}
\overline{H} &= \frac{1}{\mu_0}\nabla \times \overline{A} \\
&= \hat{\phi}\frac{I_o\ell}{4\pi}\left[jk + \frac{1}{R}\right]\sin\theta\frac{e^{-jkR}}{R} \\
\overline{E} &= \frac{1}{j\omega\epsilon_0}\nabla \times \overline{H} \\
&= \hat{R}\frac{2\eta_0 I_o\ell}{4\pi}\left[1 - \frac{j}{kR}\right]\cos\theta\frac{e^{-jkR}}{R^2} + \hat{\theta}\frac{\eta_0 I_o\ell}{4\pi}\left[jk + \frac{1}{R} - \frac{j}{kR^2}\right]\sin\theta\frac{e^{-jkR}}{R}
\end{aligned}$$

The power density radiating away from the antenna is the $\hat{R}$ component of $\overline{S}$,

$$\begin{aligned}
S_R &= \left(\overline{E} \times \overline{H}^*\right)\cdot\hat{R} \\
&= \left\{\hat{\theta}\frac{\eta_0 I_o\ell}{4\pi}\left[jk + \frac{1}{R} - \frac{j}{kR^2}\right]\sin\theta\frac{e^{-jkR}}{R}\right\} \times \left\{\hat{\phi}\frac{I_o\ell}{4\pi}\left[jk + \frac{1}{R}\right]\sin\theta\frac{e^{-jkR}}{R}\right\}^* \cdot\hat{R} \\
&= \eta_0\left|\frac{I_o\ell}{4\pi}\right|^2\frac{\sin^2\theta}{R^2}\left(k^2 + \frac{jk}{R} - \frac{jk}{R} + \frac{1}{R^2} - \frac{1}{R^2} - \frac{j}{kR^3}\right) \\
&= \eta_0\left|\frac{I_o\ell}{4\pi R}\right|^2\sin^2\theta\left(k^2 - \frac{j}{kR^3}\right)
\end{aligned}$$

So, the real power decays as $1/R^2$ as expected. There is some reactive power stored near the dipole which decays as $1/R^5$.

We are generally interested in field behaviors far from the dipole. So, we only keep the dominant terms for large R. This means we neglect any field terms that decay as $1/R^2$, $1/R^3$. Far away from the antenna, the radiated electric and magnetic fields simplify to

$$\begin{aligned}
\overline{E}_{\text{ff}} &= \hat{\theta}\frac{\eta_0}{4\pi}I_o\ell jk\frac{e^{-jkR}}{R}\sin\theta \\
\overline{H}_{\text{ff}} &= \hat{\phi}\frac{1}{4\pi}I_o\ell jk\frac{e^{-jkR}}{R}\sin\theta
\end{aligned}$$

where the subscript ff stands for far field. We see that

$$\overline{H}_{\text{ff}} = \frac{1}{\eta_0}\hat{k}\times\overline{E}_{\text{ff}}$$

just like for plane waves. This confirms that a spherical wave behaves like a plane wave when the sphere radius is large.

The far field power density is

$$\overline{S}_{\text{av, ff}} = \hat{R}\frac{\eta_0}{2}\left|\frac{kI_o\ell}{4\pi R}\right|^2\sin^2\theta = \hat{R}\frac{S_o}{R^2}\sin^2\theta$$

The power density of the field radiated by the Hertzian dipole decays with distance as $1/R^2$. For a fixed radius the power density $\overline{S}_{\text{av, ff}}$ varies with θ but not with ϕ. In the x-y plane the pattern is isotropic and in the x-z plane we get the donut pattern, as shown in Fig. 8.5.

htb

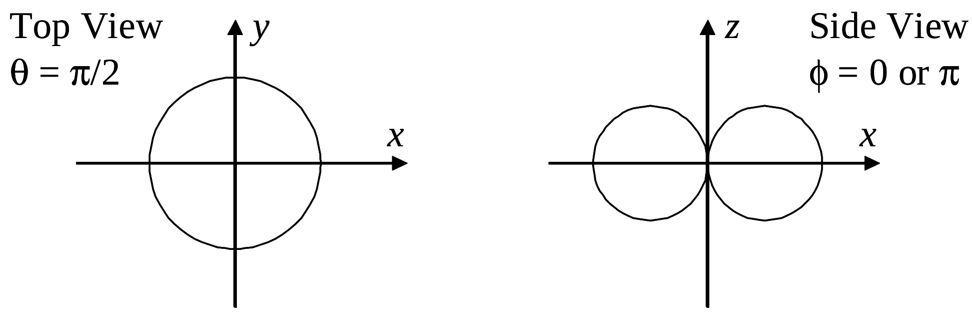


Figure 8.5: Dipole radiation pattern phi cut and theta cut.

8.4 Antenna Parameters

To describe and to quantify the properties of an antenna, we use directivity, gain, beamwidth, and radiation efficiency. Many of these antenna parameters deal with the antenna radiation pattern.

8.4.1 Radiation Pattern

The antenna pattern or radiation pattern is a functional representation of how the radiated power density in the far field varies with position. It is defined as:

$$F(\theta,\phi) = \frac{\overline{S}_{\text{av, ff}} \cdot \hat{R}}{(\overline{S}_{\text{av, ff}} \cdot \hat{R})_{\text{max}}} = \frac{S_R}{S_{\text{max}}} \tag{8.11}$$

where S_R is the radial component of the radiated time average power flux evaluated far from the antenna. The antenna pattern is often shown in dB and with respect to polar coordinates. The key parameters that are often extracted from the antenna pattern are the main beam width and the maximum sidelobe level.

Let's examine the antenna pattern a little more. To aid in this derivation, we define the solid angle. In spherical coordinates, the differential area element for a sphere of radius R is

$$dA = R^2 \sin\theta d\theta d\phi$$

We define the differential solid angle as

$$d\Omega = \frac{dA}{R^2} = \sin\theta d\theta d\phi$$

The integral of the differential solid angle over a sphere is

$$\int_0^{2\pi}\int_0^{\pi} d\Omega = 4\pi$$

So, the solid angle of a full sphere is 4π (units = steradians).

Now, the power radiated by the antenna is:

$$\begin{aligned} P_{\text{rad}} &= \int_0^{2\pi}\int_0^{\pi} \overline{S}_{\text{av}}(\theta,\phi)\cdot\hat{R}R^2\sin\theta d\theta d\phi \\ &= R^2\int_0^{2\pi}\int_0^{\pi} S_{\text{av},R}(\theta,\phi)d\Omega \\ &= R^2 S_{\max}\int_0^{2\pi}\int_0^{\pi} F(\theta,\phi)d\Omega \end{aligned}$$

Some properties of the radiation pattern are

1. **Nulls:** Directions where the radiation pattern is zero or very small.
2. **Main Lobe:** Angular region where most of the energy is transmitted.
3. **Sidelobes:** Smaller lobes of energy transmission.
4. **Elevation Pattern:** Radiation pattern in the plane for a constant value of ϕ or a theta cut.
5. **Azimuth Pattern:** Radiation pattern in the plane for $\theta = 90^\circ$ or a phi cut.
6. **E-plane and H-plane:** Pattern cuts in the planes containing the electric field and magnetic field directions.
7. **Beamwidth:** Angular extent of main lobe between two angles at which $|F(\theta,\phi)|$ is half its peak value (3 dB down from the peak). (Sometimes the null-to-null beamwidth is also used.)
8. **Omidirectional and isotropic patterns:** An omnidirectional antenna has a pattern that is constant with respect to azimuth angle. An isotropic pattern is constant over a full sphere. No real antenna is isotropic, but the isotropic antenna is used as an ideal reference to define directivity.

8.4.2 Directivity

The directivity characterizes the angular extent of the transmitted beam. A high directivity means that the power is confined to a small angular region. An antenna with a high directivity has good power confinement but requires more accurate pointing.

The directivity is defined as the maximum radiation pattern over its average as given by

$$D = \frac{F_{\max}}{F_{\text{av}}} = \frac{1}{\frac{1}{4\pi}\int_0^{2\pi}\int_0^{\pi} F(\theta,\phi)d\Omega} \tag{8.12}$$

By integrating (8.11) over a sphere we can see that

$$\int_0^{2\pi}\int_0^{\pi} F(\theta,\phi)R^2 d\Omega = \frac{P_{\text{rad}}}{S_{\max}}$$

Using this result the directivity can be written as

$$D = \frac{S_{\max}}{P_{\text{rad}}/4\pi R^2}$$

The directivity D represents the maximum power density divided by the power density that would occur if all the power were radiated equally in all directions:

$$D = \frac{S_{\text{max}}}{S_{\text{isotropic radiator}}} \tag{8.13}$$

We can also define the directivity pattern $D(\theta, \phi) = S_R(\theta, \phi)/S_{\text{isotropic radiator}}$ which takes on the maximum value given by Eq. (8.13). This is just another way of normalizing the radiation pattern, since $D(\theta, \phi) = DF(\theta, \phi)$.

8.4.3 Gain and Radiation Efficiency

Let P_t be the total power supplied to the antenna from the transmitter. The antenna radiation efficiency is

$$\eta_{\text{rad}} = \frac{P_{\text{rad}}}{P_t} \tag{8.14}$$

The antenna gain is

$$G = \frac{S_{\text{max}}}{P_t/4\pi R^2} = \frac{S_{\text{max}}}{P_{\text{rad}}/\eta_{\text{rad}} 4\pi R^2} = \eta_{\text{rad}} D \tag{8.15}$$

Antenna gain accounts for losses in the antenna, while directivity does not. There is another quantity that includes impedance mismatches between the connected transmission line and the antenna, called realized gain.

8.4.4 Radiation Resistance

Antennas can be thought of as matching devices between the transmission line and free space. To the transmission line feeding the antenna, the antenna is merely an impedance.

We define a radiation resistance R_{rad} which relates the radiated power to the antenna driving current according to

$$P_{\text{rad}} = \frac{1}{2}\text{Re}\{V_o I_o^*\} = \frac{1}{2} I_o^2 R_{\text{rad}} \tag{8.16}$$

for a lossless antenna. If the antenna is lossy, there is an additional resistance R_{loss} representing the ohmic and dielectric losses in the materials that make up the antenna. The power dissipated by the losses is

$$P_{\text{loss}} = \frac{1}{2}|I_o|^2 R_{\text{loss}}$$

The total power delivered to the antenna is

$$P_t = P_{\text{rad}} + P_{\text{loss}} = \frac{1}{2}|I_o|^2(R_{\text{rad}} + R_{\text{loss}}) = \frac{1}{2}|I_o|^2 \text{Re}[Z_{\text{in}}]$$

since $R_{\text{rad}} + R_{\text{loss}}$ is the real part of the antenna input impedance (8.1). The input power to the antenna at its transmission line port is either dissipated as losses or radiated to the far field.

We can express the radiation efficiency as

$$\eta_{\text{rad}} = \frac{P_{\text{rad}}}{P_t} = \frac{R_{\text{rad}}}{R_{\text{rad}} + R_{\text{loss}}}$$

For a highly efficient antenna such as a satellite dish receiver, R_{rad} is negligible and the efficiency η_{rad} is very close to one. For antennas used in wireless devices used in noisy environments, the radiation efficiency may be lower, around 70%.

Example: Communications satellites & estimating directivity from beamwidth

A satellite antenna is designed to illuminate the continental U.S. What is the antenna directivity?

We will approximate the antenna pattern as one inside of some cone and zero outside. The actual pattern will be smoothly varying and will have sidelobes, but this will give us a decent directivity estimate.

Assume that the distance between the satellite and the earth is $L = 40,000\,$km and that the diameter of the continental U.S. is $2a = 2,500\,$mi or $2a = 4 \times 10^6\,$m.

The antenna illuminates a cone with an angle of

$$\tan\theta = \frac{r}{L} = \frac{2 \times 10^6}{4 \times 10^7}$$
$$\theta = 2.86^o$$

The antenna pattern is

$$F = \begin{cases} 1 & 0 < \theta < 2.86^\circ \\ 0 & \text{else} \end{cases} \tag{8.17}$$

The directivity is $D = 1/F_{\text{av}}$, where

$$\begin{aligned} F_{\text{av}} &= \frac{1}{4\pi R^2}\int_{\phi=0}^{2\pi}\int_{\theta=0}^{2.86^o} R^2 \sin\theta d\theta d\phi \\ &= \frac{1}{4\pi}(2\pi)\cos\theta|_0^{2.86^o} \\ &= \frac{1}{2}\left[1 - \cos\left(2.86^o\right)\right] \\ &= 6.2 \times 10^{-4} \end{aligned}$$

So,

$$D = \frac{1}{6.2 \times 10^{-4}} = 1606$$

Another way to get the same result is to assume the total power radiated is P_{rad}. The power density on the continental U.S. is about $S_{\text{max}} = P_{\text{rad}}/(4\pi a^2)$. The directivity is

$$D = \frac{S_{\text{max}}}{S_{\text{isotropic}}} \simeq \frac{P_{\text{rad}}/(\pi a^2)}{P_{\text{rad}}/(4\pi R^2)} = \frac{4R^2}{a^2} = 1600 \rightarrow 32\text{dB}$$

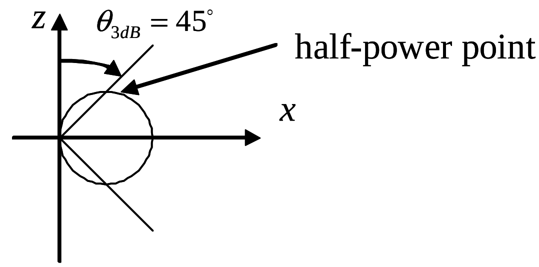
8.5 Hertzian Dipole - Antenna Parameters

Let's examine the antenna parameters for the Hertzian dipole. From before, the far field radiated power density and radiation pattern are

$$\overline{S}_{\text{av, ff}} = \hat{R}\frac{\eta_0}{2}\left|\frac{kI_o\ell}{4\pi R}\right|^2 \sin^2\theta$$
$$F(\theta,\phi) = \sin^2\theta$$

The 3 dB beamwidth is

$$\begin{aligned}\sin^2\theta_{3\,\text{dB}} &= \frac{1}{2}\\ \theta_{3\,\text{dB}} &= 45^\circ\\ \text{Beamwidth} &= 2(90^\circ - \theta_{3\,\text{dB}}) = 90^\circ\end{aligned}$$



The direction of maximum radiation occurs at $\theta = 90^\circ$ or perpendicular to the dipole.

The total radiated power can be found from

$$\begin{aligned}S_{\text{max}} &= \frac{\eta_0}{2}\left(\frac{kI_o\ell}{4\pi R}\right)^2\\ P_{\text{rad}} &= R^2 S_{\text{max}}\int_{\phi=0}^{2\pi}\int_{\theta=0}^{\pi} F(\theta,\phi)\sin\theta d\theta d\phi\\ &= \frac{\eta_0}{2}\left(\frac{kI_o\ell}{4\pi}\right)^2\int_{\phi=0}^{2\pi}\int_{\theta=0}^{\pi}\sin^3\theta d\theta d\phi\\ &= \frac{\eta_0}{2}\left(\frac{kI_o\ell}{4\pi}\right)^2 2\pi\int_{\theta=0}^{\pi}\sin^3\theta d\theta d\phi\end{aligned}$$

We need to integrate $\sin^3\theta$. We substitute $\sin^2\theta = 1-\cos^2\theta$ to get

$$\begin{aligned}\int\sin^3\theta d\theta &= \int\left(1-\cos^2\theta\right)\sin\theta d\theta\\ &= \int\sin\theta d\theta - \int\cos^2\theta\sin\theta d\theta\end{aligned}$$

We then do a u substitution for the $\cos$ term as given by

$$\begin{aligned}u &= \cos\theta\\ du &= -\sin\theta d\theta\end{aligned}$$

The indefinite integral becomes

$$\begin{aligned}\int\sin^3\theta d\theta &= \int\sin\theta d\theta - \int u^2\frac{\sin\theta}{-\sin\theta}du\\ &= -\cos\theta + \frac{1}{3}u^3\\ &= -\cos\theta + \frac{1}{3}\cos^3\theta\end{aligned}$$

The definite integral is then

$$\int_{\theta=0}^{\pi}\sin^3\theta d\theta = \frac{1}{3}\left[\cos(\pi)-\cos(0)\right] - \left[\cos(\pi)-\cos(0)\right] = -\frac{2}{3}+2 = \frac{4}{3}$$

The resulting total radiated power is

$$\begin{aligned}P_{\text{rad}} &= \frac{\eta_0}{2}\left(\frac{kI_o\ell}{4\pi}\right)^2(2\pi)\left(\frac{4}{3}\right) &(8.18)\\ &= \frac{\eta_0\left(kI_o\ell\right)^2}{12\pi} &(8.19)\end{aligned}$$

The directivity is

$$D = \frac{S_{\text{max}}}{P_{\text{rad}}/(4\pi R^2)} \tag{8.20}$$

$$= \frac{F_{\text{max}}}{F_{\text{av}}} \tag{8.21}$$

$$= \frac{1}{\frac{1}{4\pi}\int_{\phi=0}^{2\pi}\int_{\theta=0}^{\pi}\sin^3\theta d\theta d\phi} \tag{8.22}$$

$$= \frac{1}{\frac{1}{4\pi}\frac{8\pi}{3}} \tag{8.23}$$

$$= 1.5 \tag{8.24}$$

Since $R_{\text{rad}} = 2P_{\text{rad}}/I_o^2$, for a Hertzian dipole we have

$$R_{\text{rad}} = \eta_0 \frac{(kI_o^2\ell)^2}{12\pi}\frac{2}{I_o^2} = \eta_0 \frac{(k\ell)^2}{6\pi} \tag{8.25}$$

For very short dipoles ($\ell \to 0$), the radiation resistance is very small, which means it is hard to radiate real power with a short dipole, since I_o must be very large for P_{rad} to be significant and the conductor losses $P_{\text{loss}} = I_o^2 R_{\text{loss}}/2$ become unacceptable.

If the Hertzian dipole length is $\ell = \lambda/50$ then the radiation resistance is $R_{\text{rad}} \simeq 0.3\,\Omega$. A more common length is $\ell = \lambda/2$, in which case $R_{\text{rad}} \simeq 200\,\Omega$. Because the constant current approximation made in the Hertzian dipole model breaks down when ℓ approaches the wavelength, this value is inaccurate. As we will see in Sec. 8.6, the actual radiation resistance of a half-wave dipole is around 73 Ω (which is why antenna cable often has a 75 Ω characteristic impedance).

8.6 Dipole Antennas

The Hertzian dipole is great because it is easy to formulate the fields for this antenna. However, it is impractical because we cannot effectively radiate power with such an antenna (the radiation resistance is small). The analysis that we used assumes that the current along the dipole is constant. However, for dipoles of a practical length (say a half wavelength), the current is not constant along the dipole, and therefore our analysis is incorrect. We therefore desire to examine this more practical antenna structure.

Before we can do this analysis, however, we need to make some simplifications to our integral for $\overline{A}$. For realistic currents, we generally cannot perform the integration to compute $\overline{A}$. However, since we are typically interested in the far-fields, we can make a far-field approximation to the integral.

Let $\hat{R}$ be the unit vector in the direction of the observation vector $\overline{r}$. For a point $\overline{r}$ very far from the source point $\overline{r}'$, we can approximate the value

$$|\overline{r} - \overline{r}'| \approx R - \hat{R}\cdot\overline{r}' \tag{8.26}$$

So, for the phase term in our Green's function, we can write

$$e^{-jk|\overline{r}-\overline{r}'|} \approx e^{-jkR}e^{jk\hat{R}\cdot\overline{r}'} \tag{8.27}$$

For the magnitude, we can simplify this expression even further by neglecting the term $\hat{R}\cdot\overline{r}'$ to write

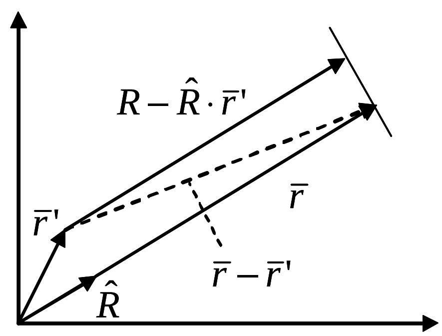
$$\frac{1}{|\overline{r}-\overline{r}'|}\approx\frac{1}{R} \tag{8.28}$$

We therefore have our far-field approximate form of the magnetic vector potential given as

$$\overline{A}_{\text{ff}}(\overline{r})=\frac{\mu}{4\pi}\frac{e^{-jkR}}{R}\int\overline{J}(\overline{r}')e^{jk\hat{R}\cdot\overline{r}'} \tag{8.29}$$

Furthermore, when we take the curl of $\overline{A}_{\text{ff}}$ to obtain the magnetic and electric fields, we neglect any terms that come from this expression that decay faster than $1/R$ (i.e. terms that behave as $1/R^2$, $1/R^3$, etc. This simplification leads to the forms

$$\begin{aligned}\overline{B}_{\text{ff}}&=\nabla\times\overline{A}_{\text{ff}}\approx-jk\hat{R}\times\overline{A}_{\text{ff}}\\ \overline{E}_{\text{ff}}&=\frac{1}{j\omega\epsilon}\nabla\times\overline{H}_{\text{ff}}\approx-\frac{jk}{j\omega\epsilon}\hat{R}\times\overline{H}_{\text{ff}}\approx j\omega\hat{R}\times(\hat{R}\times\overline{A}_{\text{ff}})\end{aligned}$$



We can now do the integration for a half-wavelength dipole. A reasonable approximation for the current on a dipole is a sinusoid that goes to zero at the ends of the dipole wires, or $\overline{J}(\overline{r}')=\hat{z}I_o\delta(x')\delta(y')\cos(kz')$, $-\lambda/4\le z'\le\lambda/4$. Then

$$\begin{aligned}\overline{A}_{\text{ff}}(\overline{r})&=\frac{\mu_0}{4\pi}\frac{e^{-jkR}}{R}\int_{-\lambda/4}^{\lambda/4}\hat{z}I_o\cos(kz')e^{jkz'\cos\theta}dz'\\ &=\hat{z}\frac{\mu_0}{8\pi}\frac{e^{-jkR}}{R}I_o\int_{-\lambda/4}^{\lambda/4}\left[e^{jkz'(\cos\theta+1)}+e^{jkz'(\cos\theta-1)}\right]dz'\\ &=(\hat{R}\cos\theta-\hat{\theta}\sin\theta)\frac{\mu_0}{2k\pi}\frac{e^{-jkR}}{R}I_o\frac{\cos\left[\pi/2\cos\theta\right]}{\sin^2\theta}\\ \overline{H}_{\text{ff}}(\overline{r})&=-\frac{jk}{\mu_0}\hat{R}\times\overline{A}_{\text{ff}}(\overline{r})=\hat{\phi}\frac{jI_o}{2\pi}\frac{e^{-jkR}}{R}\frac{\cos\left[\pi/2\cos\theta\right]}{\sin\theta}\\ \overline{E}_{\text{ff}}&=j\omega\hat{R}\times(\hat{R}\times\overline{A}_{\text{ff}})=\hat{\theta}\frac{j\eta_0I_o}{2\pi}\frac{e^{-jkR}}{R}\frac{\cos\left[\pi/2\cos\theta\right]}{\sin\theta}\end{aligned}$$

The time-average Poynting vector is:

$$S_{av,R}=\frac{|\overline{E}_{\text{ff}}|^2}{2\eta_0}=\frac{\eta_0|I_o|^2}{8(\pi R)^2}\left\{\frac{\cos\left[\pi/2\cos\theta\right]}{\sin\theta}\right\}^2$$

This Poynting vector is maximum at $\theta = \pi/2$ with the maximum being

$$S_{\text{max}} = \frac{\eta_0 |I_o|^2}{8(\pi R)^2}$$

Therefore, the radiation pattern is:

$$F(\theta) = \left\{ \frac{\cos\left[\pi/2 \cos\theta\right]}{\sin\theta} \right\}^2$$

With this radiation pattern, we can determine:

Radiated Power:	$P_{\text{rad}} = 36.6 |I_o|^2$
Directivity:	$D = 1.64$
Radiation Resistance:	$R_{\text{rad}} = 73\Omega$

8.7 Receiving Antennas

Antennas are also used for capturing energy from an incident wave and converting it into a wave on a transmission line. A receiving antenna has the Thévenin equivalent circuit shown in Fig. 8.6.

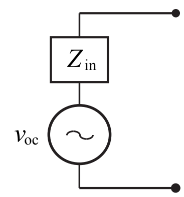


Figure 8.6: Thévenin equivalent circuit model for a receiving antenna.

The power collected by the receiving antenna depends on the power density of the incident wave and the effective collecting area of the antenna as given by

$$P_{\text{rec}} = S_i A_e, \tag{8.30}$$

where S_i is the power density of the incident wave, P_{rec} is the available power, or the power collected by the receiver when connected to a conjugate matched load ($Z_L = Z_{\text{in}}^*$), and A_e is the effective collecting area of the receiving antenna.

The basic derivation process we will use is:

1. Calculate the amount of power collected by a Hertzian dipole.
2. Relate this to an effective area for the dipole.
3. Generalize to an arbitrary antenna by relating the effective area to the directivity.

These following derivation assume that (1) the antenna is impedance matched to the transmission line and (2) the antenna loss is small ($R_{\text{loss}} \ll R_{\text{rad}}$).

The first step is to calculate the collected power for a given incident power density. The load is matched to the antenna using $Z_L = Z_{\text{in}}^*$. The load current is thus given by

$$I_L = \frac{V_{\text{oc}}}{Z_{\text{in}} + Z_L} = \frac{V_{\text{oc}}}{2R_{\text{rad}}}$$

The received power is

$$\begin{aligned} P_{\text{rec}} &= \frac{1}{2} |I_L|^2 R_{\text{rad}} \\ &= \frac{1}{2} \frac{|V_{\text{oc}}|}{(2R_{\text{rad}})^2} R_{\text{rad}} = \frac{|V_{\text{oc}}|^2}{8R_{\text{rad}}} \end{aligned} \tag{8.31}$$

The incident power density is related to the incident electric field by

$$S_i = \frac{|E_i|^2}{2\eta_o}$$

The effective area of the antenna is

$$A_e = \frac{P_{\text{rec}}}{S_i} = \frac{|V_{\text{oc}}|^2}{8R_{\text{rad}}} \frac{2\eta_0}{|E_i|^2} \tag{8.32}$$

For a Hertzian dipole the field is constant across the antenna, resulting in

$$V_{\text{oc}} = E_i \, \ell$$

Using (8.25), the effective area can then be calculated to be

$$A_e = \frac{3\lambda^2}{8\pi} \tag{8.33}$$

Relating this to the gain of a Hertzian dipole results in

$$A_e = \frac{\lambda^2 G}{4\pi} \tag{8.34}$$

Although we derived this for a Hertzian dipole, this same expression can be used to define the effective area of any antenna.

8.7.1 Friis Transmission Formula and Link Budgets

Now we want to couple the transmitting and receiveing antennas together to get a complete link. We start with calculating the power density at the location of the receiver produced by the transmitting antenna, which can be found from the definition of gain:

$$\begin{aligned} G_t &= \frac{\text{Power Density}}{\text{Power density of an isotropic radiator}} \\ &= \frac{S_i}{\frac{P_t}{4\pi R^2}} \\ &= S_i \left(\frac{4\pi R^2}{P_t} \right) \end{aligned}$$

From this, we find that the incident power density is

$$S_i = G_t \left(\frac{P_t}{4\pi R^2} \right) \tag{8.35}$$

Now we determine the power collected by the receiving antenna using

$$P_{\text{rec}} = S_i A_r \tag{8.36}$$

We relate the effective area to the antenna gain to get

$$P_{\text{rec}} = S_i G_r \frac{\lambda^2}{4\pi} \tag{8.37}$$

Finally, we plug in the expression for the incident power density to give

$$P_{\text{rec}} = \left(G_t \frac{P_t}{4\pi R^2} \right) G_r \left(\frac{\lambda^2}{4\pi} \right) \tag{8.38}$$

This results in the Friis transmission formula,

$$P_{\text{rec}} = P_t G_t G_r \left(\frac{\lambda}{4\pi R}\right)^2 \tag{8.39}$$

The last factor on the right is sometimes called the **free space path loss**. This formula is often expressed in terms of dB, so that the gains, powers, and free space path loss in dB units can simply be added. The transmitted and received powers are commonly given in dB relative to one Watt (dBW) for high power systems or one milliwatt (dBm) for low power applications.

Example: Ku band SatCom link budget

A Ku band (12 GHz) satellite downlink system transmits a signal power of $P_t = 100\,\text{W}$ and bandwidth $50\,\text{MHz}$. The diameter of the dish antenna on the satellite is $1\,\text{m}$. The distance from the satellite to the ground terminal receiver is $R = 40,000\,\text{km}$. The diameter of the receive antenna is $60\,\text{cm}$. The receiver equivalent system noise temperature is $100\,\text{Kelvin}$. We wish to find the signal to noise ratio (SNR) at the output of the receive antenna.

Dish antennas usually have an antenna efficiency of about 60%. This means that the gains are

$$G_t = \frac{4\pi A_e}{\lambda^2} = \frac{4\pi \cdot 0.6\pi(0.5)^2}{(0.025)^2} \Rightarrow 40\,\text{dB} \tag{8.40}$$

$$G_r \Rightarrow 35\,\text{dB} \tag{8.41}$$

If we add the transmit antenna gain in dB to the transmit power expressed in dBW (i.e., $10\log_{10} P_t$), we obtain the equivalent isotropic radiated power (EIRP) in dBW, which is often used in satellite communication system link budgets.

Using the Friis tranmission formula with each term expressed in dB units, the receive power is

$$P_{\text{rec}}\,(\text{dBm}) = \underbrace{40 + 20\,\text{dBW}}_{\text{EIRP}} + 35\,\text{dB} + \underbrace{20\log_{10}\left(\frac{\lambda}{4\pi R}\right)}_{\text{FPL (dB)}} + 30$$

where power in dBm is the power relative to 1 mW in dB, or

$$P_{\text{rec}}\,(\text{dBm}) = 10\log_{10}(1000 P_{\text{rec}}) = 30 + 10\log_{10} P_{\text{rec}}$$

The free space path loss is $-206\,\text{dB}$. Adding all of the terms, we arrive at $P_{\text{rec}}\,(\text{dBm}) = -81\,\text{dBm}$.

The noise floor can be found from the formula $P_n = k_B T_e B$, where k_B is Boltzmann's constant, T_e is the equivalent system noise temperature, and B is the bandwidth of the signal. If the noise temperature is the reference value, $T_0 = 290\,\text{Kelvin}$ (room temperature), then $10\log_{10}(10^3 k_B T_0)$ has the value $-174\,\text{dBm/Hz}$. This is the noise power spectral density in dB units produced by a noise source at room temperature. The noise power is

$$P_n\,(\text{dBm}) = -174\,\text{dBm/Hz} + 10\log_{10}(100/290) + 10\log_{10}(50\times 10^6) = -102\,\text{dBm}$$

where the second term on the right hand side converts the equivalent noise temperature from the room temperature reference value to the given equivalent system noise temperature value of $100\,\text{K}$.

The SNR in dB is

$$\text{SNR (dB)} = -81 - (-102) = 21\,\text{dB}$$

For most communication protocols, this is a fairly high SNR and the bit error rate should be low, meaning that the link quality of service will be good.

8.8 Antenna Arrays

An antenna array consists of a number of antenna elements the inputs or outputs of which are combined. There are a variety of benefits of antenna arrays. Here are a few of the most common advantages:

- Increasing the directivity of a simple antenna
- Concentrating the radiated power where you want it
- Eliminating the radiated power (or received power) where you don't want it
- Electronically steering the main beam of a highly directional antenna

Antenna arrays come in different types of increasing complexity:

- **Uniform linear array:** Elements evenly spaced along a line, with different driving currents for each element.
- **2D array:** A two-dimensional array of antenna elements, often with a rectangular or triangular configuration.
- **Random and irregular arrays:** Arrays with elements at uneven spacings to reduce cost, lower sidelobes, or achieve some other design goal.
- **Other types of arrays:** Array feeds that are combined with a dish-type reflector antenna, three-dimensional arrays, and many others.

In this class, we will study the uniform linear array. Most practical applications of arrays require 2D arrays, but we can learn most of the basic principles of antenna arrays by studying the mathematically simpler 1D case.

Beamforming network. In a transmitting phased array, a modulated signal such as a communications waveform or a radar pulse is split into multiple signals, one for each element in the array. The amplitude and the time delay or phase of the signals for each element are then adjusted using analog or digital components to shape the antenna array radiation pattern and steer the main beam to a desired direction. This part of the system is referred to as a feed network or beamforming network.

For a receiving array, the signals from each antenna element are adjusted in amplitude and time delay or phase and then combined by summation. This can be done using analog or digital hardware.

The beamforming network is often the most expensive part of an antenna array. Analog phase shifters are relatively expensive. For digitally beamformed arrays, the required hardware is also expensive, particularly for broadband antenna arrays with a high data throughput requirement. There are many types of arrays and beamforming systems with various tradeoffs between size, weight, power consumption, cost, and performance.

The feeding networks used to drive array elements with signals of different amplitudes and phases come in different types:

- **Fixed or corporate fed array:** Time delays built into the feeding network (e.g., transmission lines of different lengths. This type of array has a fixed beam pattern.
- **Passive phased array:** Phase delays for each element using passive electronic components.
- **Active phased array:** Amplification or other active electronic devices at each array element.
- **Electronically scanned array (ESA):** Electronically controlled time delays or phase for the driving signal at each element to change the direction of the main lobe of the array radiation pattern.
- **Analog beamforming:** The driving signals for array elements are generated and controlled using analog electronics such as phase shifters or variable gain amplifiers.
- **Digital beamforming:** For transmitting arrays, signals are generated using digital signal processing and analog to digital converters. For receiving arrays, signals are sampled and combined using digital signal processing.

Narrowband and wideband arrays. For wideband arrays, time delays between the signals that drive each antenna element are required to steer the main lobe of the radiation pattern to different directions. This is time delay beam steering or beam scanning. Analog devices that can provide variable time delays are expensive or complicated. For narrowband arrays, a time delay is equivalent to a phase shift for a time-harmonic signal. For this reason, phased arrays for narrowband applications often use phase shifters instead of time delays in the beamforming network. If a beamforming network using phase shifters is used over too wide a bandwidth, the beam pattern deviates from the desired directivity or scan angle. This is called beam squint.

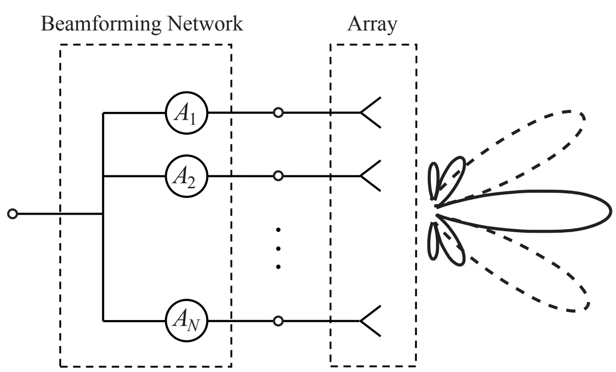


Figure 8.7: Block diagram of a narrowband transmitting antenna array. An input signal is split N ways, then scaled by phasor quantities A_n to adjust the phase and amplitude of the signals driving each element, allowing for control of the antenna array radiation pattern.

8.8.1 Array Analysis and Synthesis

For antenna arrays, we can divide problems into two classes:

Analysis: The individual antenna elements are known and we calculate the pattern that is produced.

Synthesis: The antenna pattern is known and we try and determine the individual antenna element locations and amplitudes that will produce this pattern.

Taking a given antenna array design and driving currents that excite each element and computing the radiation pattern is array analysis. The reverse problem, starting from a prescribed radiation pattern or gain requirement and designing the array element locations and driving currents that achieve the desired pattern, is array synthesis.

This division into analysis and synthesis applies to many types of engineering problems. Typically, analysis is easier, so we learn that first, and then use the understanding gained from analyzing simple example cases to design solutions for more challenging constraints.

8.8.2 The Array Factor

One of the most basic analysis techniques for array antenna analysis and synthesis is the array factor method. We will start with the field radiated by a Hertzian dipoles and then generalize it to an array of arbitrary elements.

For the Hertzian dipole antenna,

$$\overline{E}_{\text{ff}}(R,\theta,\phi) = \hat{\theta}\frac{\eta_0}{4\pi}I_o\ell jk\frac{e^{-jkR}}{R}\sin\theta = \hat{\theta}E_0\frac{e^{-jkR}}{R}\sin\theta$$

where we have lumped all the constants into E_0. Suppose we now have several such antennas, each having a different excitation current (magnitude and phase) arranged in a line. For an array of arbitrary element types, the nth element will have electric field that is similar in form to the Hertzian dipole, but with a more complicated dependence on angle and with a different distance to the element, reflecting its location in the array:

$$\overline{E}_n(R_n,\theta_n,\phi_n) = A_n\frac{e^{-jkR_n}}{R_n}\overline{f}_e(\theta_n,\phi_n) \tag{8.42}$$

The terms in this expression are

$A_n = a_n e^{j\psi_n} =$ current weight or driving amplitude of the nth array element

e^{-jkR_n}/R_n = spherical wave factor for the nth element

$\overline{f}_e(\theta_n,\phi_n) =$ vector function that depends only on the observation angles (θ_n,ϕ_n) and is unique to a particular element type

The distance R_n is the distance from the nth element to the observation point. The angles have a similar definition, as shown in Fig. 8.8. In the figure, the array elements are located along the z axis, but array elements can be oriented in any way relative to the coordinate system.

The total field for this array is the sum of the fields radiated by all of the elements:

$$\overline{E} = \sum_{n=0}^{N-1} A_n\frac{e^{-jkR_n}}{R_n}\overline{f}_e(\theta_n,\phi_n) \tag{8.43}$$

We now go back to our far-field assumption. From the figure, we can see that the far-field approximation is that $\theta_n \simeq \theta$ and $\phi_n \simeq \phi$ for all n. For the $n = 1$ element in the array, we use the approximations

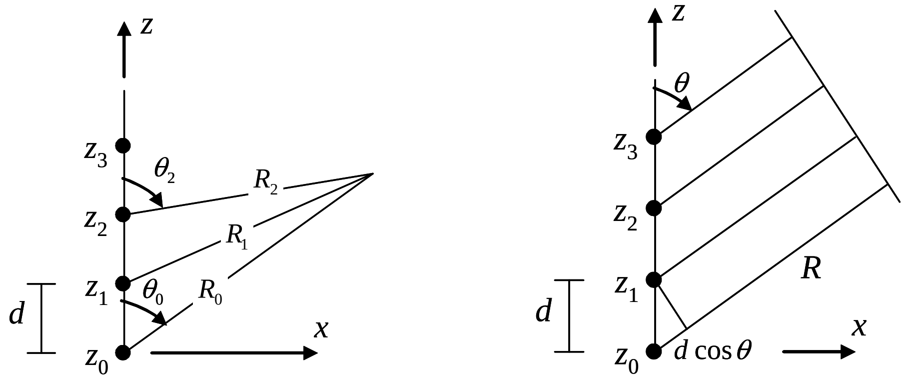


Figure 8.8: Antenna array with elements evenly spaced along the z axis.

$R_1 \simeq R - d\cos\theta$ in the phase of the spherical wave and $R_1 \simeq R$ for the magnitude. For the nth element, the approximations are

$$\begin{aligned} \text{Phase:} \quad R_n &\approx R - nd\cos\theta \\ \text{Magnitude:} \quad R_n &\approx R \end{aligned}$$

In the far field, the total field radiated by the array is

$$\overline{E} = \underbrace{\overline{f}_e(\theta,\phi)\frac{e^{-jkR}}{R}}_{\text{Single Element Radiation}} \underbrace{\sum_{n=0}^{N-1} A_n e^{jknd\cos\theta}}_{f_a(\theta)} \tag{8.44}$$

From this expression, we see that the total electric field can be written as the product of the single element radiation and an additional factor that takes into account the geometry of the array and the element excitations.

The power density radiated by the array is is

$$S_R(R,\theta,\phi) = S_{R,\mathrm{ff}}(R,\theta,\phi)F_a(\theta) \tag{8.45}$$

where

$$F_a(\theta) = \left|\sum_{n=0}^{N-1} A_n e^{jknd\cos\theta}\right|^2 \tag{8.46}$$

We call $F_a(\theta)$ the **array factor** for the pattern. It gives the shape of the radiation pattern due to the combination of the multiple elements independent of the shape of the individual element patterns. Often, the array factor dominates the behavior of the total radiation pattern. This is a very important result in array antenna theory: we can get the antenna pattern of the array by finding the array factor and then simply multiplying by the pattern of one individual element:

$$[\text{Array Pattern}\,] = [\text{Element Pattern}\,] \times [\text{Array Factor}] \tag{8.47}$$

This factorization allows us to design the element locations and excitations (which determine the array factor) separately from the antenna element (which determines the element pattern).

Example: Two Element Uniform Linear Array

Let's look at a simple ULA example consisting of two y-directed Hertzian dipoles that are:

- located on the z axis and separated by $d = \lambda/2$
- driven by current excitations with equal amplitudes ($a_o = a_1$)
- and with the currents out of phase by $\pi/2$ ($\phi_0 = 1, \phi_1 = \pi/2$)

For this example, the array factor is

$$\begin{aligned} F_a &= \left|1 + e^{-j\frac{\pi}{2}} e^{jkd\cos\theta}\right|^2 \\ &= \left|1 + e^{-j\frac{\pi}{2}} e^{j\frac{2\pi}{\lambda}\frac{\lambda}{2}\cos\theta}\right|^2 \\ &= \left|1 + e^{-j\frac{\pi}{2}} e^{j\pi\cos\theta}\right|^2 \end{aligned}$$

To simplify this we use

$$\left|1 + e^{jx}\right| = \left|2e^{j\frac{x}{2}}\left(e^{-j\frac{x}{2}} + e^{j\frac{x}{2}}\right)\right|^2 = 4\cos^2\left(\frac{x}{2}\right)$$

The array factor becomes

$$F_a = 4\cos^2\left(\frac{\pi}{2}\cos\theta - \frac{\pi}{4}\right)$$

The total array power density is then given by

$$S_R(\theta,\phi) = 4S_{R,\mathrm{HD}}(\theta,\phi)\cos^2\left(\frac{\pi}{2}\cos\theta - \frac{\pi}{4}\right)$$

If we confine our attention to the $x - z$ plane, the pattern of the individual dipoles is isotropic, so the array pattern reduces to just the array factor. The array pattern has a maximum when

$$\frac{\pi}{2}\cos\theta - \frac{\pi}{4} = 0 \quad \Rightarrow \quad \cos\theta = \frac{1}{2} \quad \Rightarrow \quad \theta = \pm 60^\circ$$

and minimums when

$$\frac{\pi}{2}\cos\theta - \frac{\pi}{4} = -\frac{\pi}{2} \quad \Rightarrow \quad \theta = \pm 120^\circ$$

Since the linear array is along the z axis, the $\theta = 90^\circ$ direction is normal to the axis of the array. This is called the broadside direction. The z direction is along the axis of the array, so that is called the endfire direction. Due to the relative phase shift between the currents that excite the elements, the main beam is steered to 30 degrees from the broadside direction of the array.

Because the dipoles are y-directed, the element pattern is isotropic in the z-z plane, which means that the radiation pattern is the same as the array factor (if the element pattern were not isotropic, we would have to multiply the array factor by the element pattern to get the radiation pattern of the overall antenna array). The pattern is shown in Fig. 8.9.

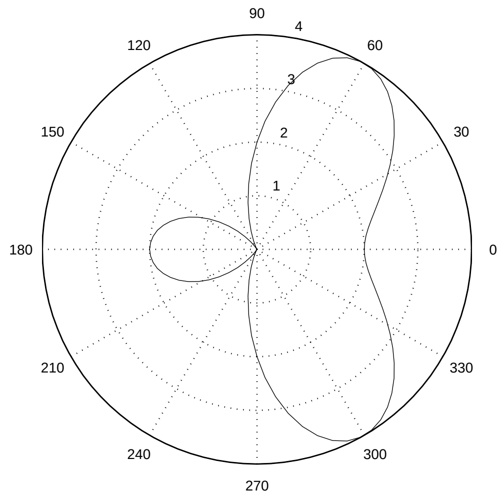


Figure 8.9: Polar plot of radiation pattern for a two element array example. The array elements are located along the horizontal axis of the plot.

Example: Two Element ULA - Pattern Synthesis

Next let's look at an example of pattern synthesis. In this example we want to use two antennas to produce no radiation in the north/south directions and maximum radiation in the east/west direction.

Again we start with the array factor as given by

$$F_a(\theta) = \left|1 + a_1 e^{j\psi_1} e^{j\frac{2\pi d}{\lambda}\cos\theta}\right|^2$$

We want $F_a = 0$ when $\theta = \pm 90^\circ$ (north/south direction). Since $\cos 90^\circ = 0$ the array factor becomes

$$F_a(\theta = 90^\circ) = \left|1 + a_1 e^{j\psi_1}\right|^2 = 0.$$

This requires $a_1 = a_o = 1$ and $e^{j\psi_1} = -1$ or $\psi_1 = \pi$. In order to have a maximum at $\theta = 0$ $(\cos(0) = 1)$ the array factor becomes

$$F_a(\theta = 90^o) = \left|1 + a_1 e^{j\psi_1} e^{j2\pi d\lambda}\right|^2 = 1$$

resulting in

$$e^{j\pi} e^{j2\pi d\lambda} = 1 \quad \Rightarrow \quad d = \frac{\lambda}{2}$$

The resulting array factor is

$$\begin{aligned} F_a &= \left|1 - e^{j\pi\cos\theta}\right|^2 \\ &= \left|\left(2j\ e^{j\frac{\pi}{2}}\right)\left(e^{-j\frac{\pi}{2}\cos\theta} - e^{j\frac{\pi}{2}\cos\theta}\right)\right|^2 \\ &= 4\sin^2\left(\frac{\pi}{2}\cos\theta\right) \end{aligned}$$

If the phase were set at $\psi_1 = 0$ then the array pattern becomes $F_a = 4\cos^2\left(\frac{\pi}{2}\cos\theta\right)$, and the picture rotates by 90°.

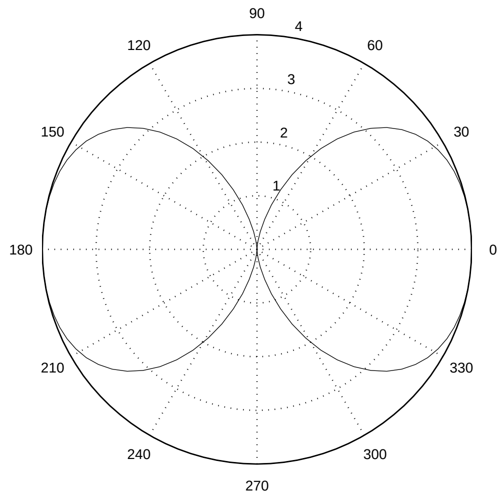


Figure 8.10: Two element array pattern synthesis example.

8.8.3 Visible Window Method for a ULA with Progressive Phase Shifts

Let's examine the special case of driving currents that are progressively out of phase by some angle ψ as we move from one element to another along the array. The driving currents in this case are

$$A_n = e^{jn\psi}$$

We call this a uniformly excited array with a linearly progressive phase shift. For narrowband signals, the differences in phase in the driving currents that excite the elements along the array represent time delays. The relative time delays in the driving currents cause the main lobe of the radiation pattern to move away from the broadside direction. By controlling the phase shift ψ, we can steer the beam. This is referred to as electronic beam scanning.

For this type of array, we can put the array factor into a particularly simple form. The array factor is

$$\begin{aligned} F_a(\theta) &= \left| \sum_{n=0}^{N-1} e^{jn(\psi + kd\cos\theta)} \right|^2 \\ &= \left| \sum_{n=0}^{N-1} \left[e^{j(\psi + kd\cos\theta)} \right]^n \right|^2 \end{aligned}$$

This is a geometric series, which can be summed in closed form:

$$\sum_{n=0}^{N-1} \gamma^n = \frac{1-\gamma^N}{1-\gamma} \qquad \text{for } |a| \leq 1$$

The array factor becomes

$$\begin{aligned}
F_a(\theta) &= \left| \frac{1 - e^{jN(\psi + kd\cos\theta)}}{1 - e^{j(\psi + kd\cos\theta)}} \right|^2 \\
&= \left| \frac{e^{jN(\psi + kd\cos\theta)/2}}{e^{j(\psi + kd\cos\theta)/2}} \frac{e^{-jN(\psi + kd\cos\theta)/2} - e^{jN(\psi + kd\cos\theta)/2}}{e^{-j(\psi + kd\cos\theta)/2} - e^{j(\psi + kd\cos\theta)/2}} \right|^2 \\
&= \left| \frac{-j2\sin\left[N(\psi + kd\cos\theta)/2\right]}{-j2\sin\left[(\psi + kd\cos\theta)/2\right]} \right|^2 \\
&= \frac{\sin^2\left[N(\psi + kd\cos\theta)/2\right]}{\sin^2\left[(\psi + kd\cos\theta)/2\right]}
\end{aligned}$$

To simplify the array factor, we define the variable

$$u = kd\cos(\theta) + \psi \tag{8.48}$$

The array factor is then

$$F_a(u) = \frac{\sin^2(Nu/2)}{\sin^2(u/2)} \tag{8.49}$$

This is called the Dirichlet function or the periodic sinc function. Let's look at the properties of this function:

- Locations of maxima: At $u = 2m\pi$,

  $$\lim_{u\to 2m\pi} \frac{\sin Nu/2}{\sin u/2} = \lim_{u\to 2m\pi} N/2 \frac{\cos Nu/2}{\cos u/2} = N\frac{\cos(Nm\pi)}{\cos(m\pi)} = N(-1)^{(N-1)m}$$

  So, the peak value of the periodic sinc function is

  $$F_a(2m\pi) = N^2$$

  at every positive and negative integer multiple of 2π. Between $u = 0$ to $u = 2\pi$ the periodic sinc function has N maxima, including the largest values at 0 and 2π. The function repeats its values in every interval of length 2π.

- Locations of zeros: The zeros of $F_a(u)$ occur at $u = 2m\pi/N$ except for $m = 0, \pm N, \pm 2N, \ldots$ since these points are equivalent to the maxima determined above.

- Range of u (the visible window): Because $\cos\theta = (u - \psi)/kd$ and $-1 \le \cos\theta \le 1$, we have:

  $$\begin{aligned}
  -1 \le \frac{u - \psi}{kd} &\le 1 \\
  -kd \le u - \psi &\le kd \\
  \psi - kd \le u &\le \psi + kd
  \end{aligned}$$

  This is the range over which the argument u of the periodic sinc function ranges as the angle θ is varied around a full circle. We call this range of u the **Visible Window**.

We can use this information to sketch the array factor for a ULA with progressive phase shifts. This is done with the following steps:

1. Sketch $F_a(u)$ for the given number of array elements N.
2. Draw a circle with radius kd centered belos $u = \psi$.
3. Use the relationship $u = kd\cos(\theta) + \psi$ to plot the array factor in polar format.

This is the visible window method for plotting the array factor of a ULA with progressive phase shifts.

Example: Visible Window Method for a 7 Element ULA With Progressive Phase Shifts

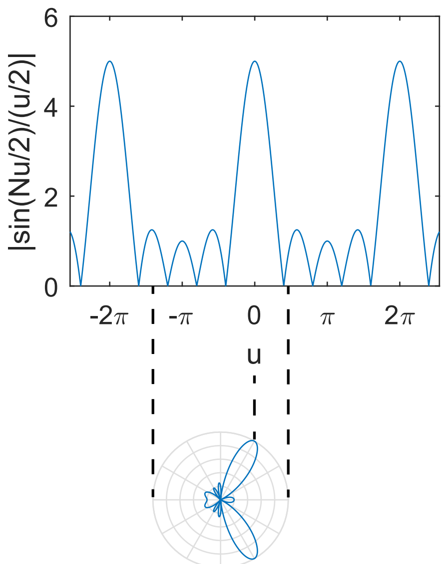
As an example, consider a ULA with $N = 5$, $\psi = -\pi/2$, and $d = \lambda/2$ ($kd = \pi$). The visible window is

$$-\pi/2 - \pi \leq u \leq -\pi/2 + \pi \rightarrow \frac{-3\pi}{2} \leq u \leq \frac{\pi}{2}$$

The definition

$$u = kd\cos\theta + \psi$$

implies that to map u to θ, we draw a circle centered at ψ with radius kd as shown. We call this technique the visible window technique for plotting array factors. The periodic sinc function and the constructed polar plot of the array factor are showin in Fig. 8.11.



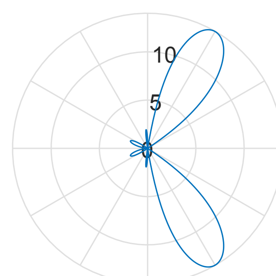


Figure 8.11: Visible window method example. The top plot is a graph of the periodic sinc function. To make the smaller peaks easily visible, we plot the square root of the array factor rather than the array factor itself. The lower plot is a circle of radius kd with the center shifted to $u = \psi$ relative to the horizontal axis of the upper plot. Using the definition of u, we transfer the linear plot of the periodic sinc function to a polar plot within the circle. The plot on the right is a polar plot of the array factor in dB (i.e., the right polar plot is $20\log_{10}$ of the values of the left polar plot). Using dB units allows the sidelobes of the array factor to remain visible despite very small values in relation to the main lobe.

8.8.4 Special Cases

1. Broadside array: $\psi = 0$. Main lobe is perpendicular to the array.
2. Endfire array: $\psi = \pi$. Main lobe is parallel to the array.
3. Grating lobes: For arrays with wide element spacing, additional main lobes can appear in the pattern. If $kd > 2\pi$, the pattern has more than one main lobe (in addition to the symmetry of the pattern with respect to the axis of the array). Since $k = 2\pi/\lambda$, this corresponds to $d > \lambda$, so arrays with elements that are widely spaced have multiple main lobes or grating lobes in the pattern.

8.8.5 Beam Scan Angle

If the progressive phase shift is zero, then the peak of the main lobe or the main beam direction is in the broadside direction ($\theta = 90°$). To move the beam away from the broad side direction, we use a nonzero progressive phase shift along the array. The main beam peak occurs where $F(u)$ is at a maximum, or $u = 0$. From the definition of u in (8.48), we can solve for the angle of the steered main beam peak,

$$\theta = \cos^{-1}\left[-\psi/(kd)\right] \tag{8.50}$$

The beam scan angle is the angle of the steered main beam away from the broadside direction:

$$\text{Scan angle} = \cos^{-1}\left[-\psi/(kd)\right] - \pi/2 \tag{8.51}$$

This represents one of the main applications of phased arrays, electronic steering of the main beam direction by controlling the phases of the array element excitations. For transmitting arrays, we can adjust the phase of the signals driving the elements using analog phase control or digital synthesis of driving signals with different phases. Energy is directed preferentially in the main beam direction.

For receiving arrays, we shift the phases of the received signals from the array elements before adding them. Waves arriving from the main beam direction add coherently in the beamforming network and are preferentially received (this is the receiving array pattern main lobe or main beam direction). Waves from other directions add out of phase and are suppressed (nulls or sidelobes of the array pattern).

8.8.6 Pattern Multiplication

All we have plotted so far is the array factor. The array factor along only gives us the array radiation pattern if the elements are isotropic radiators. Recall from (8.47) that the actual array pattern is equal to the array factor multiplied by the individual pattern of one of the elements. We can multiply patterns graphically by noting that if either pattern has a null, then the overall pattern has a null. An example is shown in Fig. 8.12.

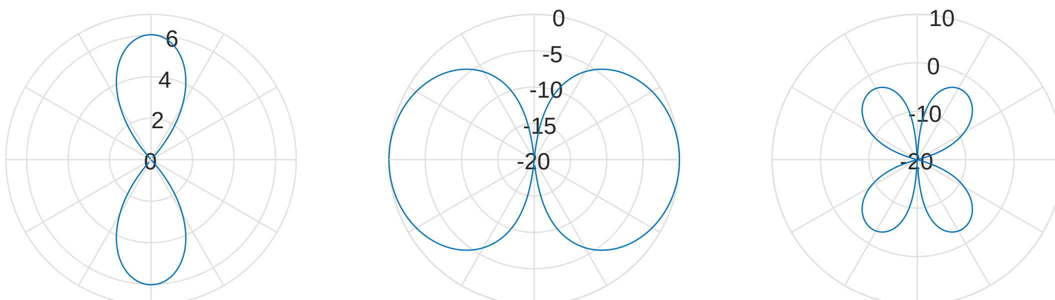


Figure 8.12: Pattern multiplication for a two element array of vertically oriented dipoles. Left: Array factor for a two element array, no progressive phase shift, and half-wave element spacing. Middle: Element pattern. Right: Product of the array factor and the element pattern. The array radiation pattern has nulls where either the array factor or the element pattern has nulls.

Fundamentals

Magnetic vector potential: $\overline{B} = \nabla \times \overline{A}$

Radiation integral: $\overline{A}(\overline{R}) = \frac{\mu}{4\pi} \int_V \overline{J}(\overline{R}') \frac{e^{-jk|\overline{R}-\overline{R}'|}}{|\overline{R}-\overline{R}'|} dV'$

Hertzian dipole model: $\overline{J}(\overline{r}') = \overline{J}(x', y', z') = I_o \hat{z} \delta(x')\delta(y'), \quad -\ell/2 \le z' \le \ell/2$

Far field of Hertzian dipole:

$$\overline{E}_{\text{ff}} = \hat{\theta}\eta_0 I_o \ell jk \frac{e^{-jkR}}{4\pi R} \sin\theta$$
$$\overline{H}_{\text{ff}} = \hat{\phi} I_o \ell jk \frac{e^{-jkR}}{4\pi R} \sin\theta$$
$$\overline{S}_{\text{av, ff}} = \hat{R} \frac{\eta_0}{2} \left| \frac{k I_o \ell}{4\pi R} \right|^2 \sin^2\theta$$

Antenna pattern: $F(\theta, \phi) = S_{\text{av, ff},R}(\theta, \phi)/S_{\text{max}}$

Directivity:

$$D = \frac{S_{\text{max}}}{P_{\text{rad}}/(4\pi R^2)}$$

Radiation efficiency: $\eta_{\text{rad}} = P_{\text{rad}}/P_t$

Gain:

$$G = \frac{S_{\text{max}}}{P_{\text{t}}/(4\pi R^2)}$$

Radiation resistance: $P_{\text{rad}} = \frac{1}{2}|I_0|^2 R_{\text{rad}}$

Effective receiving area: $A_{\text{e}} = \frac{\lambda^2}{4\pi} G$

Antenna arrays:

$$S_R(R, \theta, \phi) = \underbrace{S_{\text{R,el}}(\theta, \phi)}_{\text{Single Element Radiation}} \underbrace{\left| \sum_{n=0}^{N-1} A_n e^{jknd\cos\theta} \right|^2}_{\text{Array factor } F_a(\theta,\phi)}$$

Uniform linear array with progressive phase shift: $F_a(u) = \sin^2(Nu/2)/\sin^2(u/2)$, $u = kd\cos\theta + \psi$

Midterm 3 Topics

Reflection and transmission

- Normal incidence
- Transmission line analogy
- Oblique incidence, parallel and perpendicular polarization
- Snell's law
- Reflection and transmission coefficients
- Reflected and transmitted power
- Brewster's angle
- Total internal reflection

Waveguides

- Modes, mode numbers, cutoff frequency, single mode operating bandwidth

Radiation

- Current model for antennas
- Magnetic vector potential
- Radiation integral, source and observation coordinates, impulse response, Green's functions
- Far field approximation

Antenna parameters

- Radiation pattern
- Null, main lobe, sidelobes, beamwidth
- Directivity and gain, total radiated power, radiation efficiency
- Radiation resistance

Hertzian dipole example

Receiving antennas

- Effective area
- Friis transmission formula

Antenna arrays

- Single element pattern and array factor
- Uniform linear array with progressive phase shift
    Visible window graphical method
- Broadside and endfire patterns, grating lobes
- Pattern multiplication

Final Exam Topics

Transmission lines
- Modeling frameworks - circuit theory, transmission line theory, Maxwell's equations
- Phasors
- Transmission line voltage and current solutions, forward and reverse waves, reflection coefficient
- Matched, open circuit, and short circuit loads; quarter wave matching
- Input impedance, stub matching, Smith chart, power
- Transients on transmission lines, bounce diagram

Maxwell's Equations
- Vector calculus, gradient, curl, divergence, Stokes' theorem
- Point forms and integral forms of Maxwell's equations
- Statics, Gauss's law, capacitance, inductance
- Faraday's law, Lenz's law, transformer and generator EMF
- Ampere's law, displacement current

Plane waves
- Wave equation, separation of variables, dispersion relation, plane wave solutions
- Electric and magnetic fields, wave vector, characteristic impedance, orthogonality relations
- Polarization
- Lossy media, skin depth

Reflection and transmission
- Normal incidence, transmission line analogy
- Oblique incidence, parallel and perpendicular polarization, Snell's law
- Reflection and transmission coefficients, reflected and transmitted power
- Brewster's angle, total internal reflection

Waveguides
- Modes, mode numbers, cutoff frequency, single mode operating bandwidth

Radiation and antennas
- Current model for antennas
- Magnetic vector potential, radiation integral, source and observation points, Green's functions
- Far field approximation
- Antenna parameters, radiation pattern, nulls, main lobe, sidelobes, beamwidth
- Directivity and gain, total radiated power, radiation efficiency, radiation resistance
- Hertzian dipole example, aperture antennas, microstrip patch, electrically small antennas
- Receiving antennas, effective area, Friis transmission formula, link budget
- Antenna arrays, element pattern, array factor, pattern multiplication
- ULA, beam steering, visible window method