Fundamentals of Electromagnetic Waves

Thomas E. Roth

Elmore Family School of Electrical and Computer Engineering

Purdue University

Spring 2023

ii

Preface

This set of lecture notes was prepared for a junior level course on the fundamentals of electromagnetic waves that is offered each semester in the Elmore Family School of Electrical and Computer Engineering at Purdue University. The class was originally created by Prof. Kevin Webb at Purdue. The class serves as a follow-up to a first course on electromagnetic physics that primarily covers electrostatics and magnetostatics, with a little time left at the end of the semester to begin discussing the very basics of plane waves. The second course on electromagnetic waves covers a wide range of topics, beginning from time-varying Maxwell's equations and then developing in depth the fundamentals of plane waves. After that, we cover in detail transmission lines before discussing the field theory treatment of waveguides and resonators. We finally wrap up the course by discussing the basics of antenna theory. In addition to these lecture notes, we use David K. Cheng's *Field and Wave Electromagnetics (2nd edition)* as a required textbook for the course. In preparing these lecture notes, I drew from Cheng's textbook, David Pozar's *Microwave Engineering (4th edition)*, Constantine Balanis's *Antenna Theory: Analysis and Design (3rd edition)*, and Jian-Ming Jin's *Theory and Computation of Electromagnetic Fields (2nd edition)*. Due to the well-established nature of the material covered, I make no attempt to exhaustively include citations throughout these lecture notes.

Looking through the table of contents of these lecture notes, one will quickly see that our ordering of topics does not follow a typical progression found in most textbooks. Although it leads to an unusually long and broad first chapter, we follow this order in the class to balance the content presented to students with the goal of keeping the class more engaging. In particular, we intersperse fundamental electromagnetic topics (e.g., boundary conditions, Poynting's theorem, material models etc.) with discussions of plane waves or other more applied topics that lend themselves to introducing how electromagnetics plays a role in a wide range of practical technologies. This also allows us to cover topics in a sequence such that each weekly homework assignment can have more meaningful problems that build the desired mathematical and problem-solving skills we wish to impart to our students. We have found that most students really enjoy this sequence and stay more engaged throughout the semester as a result.

Contents

CONTENTS

CONTENTS

Chapter 1

Fundamentals of Electromagnetics and Plane Waves

1.1 Maxwell's Equations and Electromagnetic Quantities

Maxwell's equations (completed in 1865) describe the fundamental properties and interplay between electricity and magnetism. A number of the principles behind these equations were originally introduced by other researchers. Maxwell took these experimental results and theoretical concepts and developed a comprehensive and unifying summary of these disparate concepts.

Maxwell's equations represent one of the greatest triumphs ever achieved in physics. They are for almost all intents and purposes, perfect. In 2012, their theoretical accuracy had been experimentally validated to one part in a trillion. This level of accuracy is equivalent to measuring the distance from the Earth to the Moon and being correct to within the width of a *single human hair* [1]. Many other disciplines in physics can only dream of having foundational equations that are this accurate or well-behaved. For instance, in the field of fluid dynamics, the Navier-Stokes equations are of practical interest for many applications ranging from modeling weather to complex aerodynamics. Even though they have had great success in practical applications, the proof of existence and smoothness of solutions to the Navier-Stokes equations are still considered one of the most important open problems in mathematics. Such is the importance that the Clay Mathematics Institute has offered a $1M USD bounty for anyone who can solve it!

Further, many areas of physics can only develop equations with various approximations in place – i.e., they are only valid for certain situations and uses, making their application difficult for the many "boundary cases" that inevitably occur in practical engineering analysis and design. In the field of electromagnetics we rarely have to be concerned about this, we can always fall back on the validity of Maxwell's equations to know that we have firm footing to investigate further concepts, develop numerical solutions, etc. We are truly privileged and indebted to the great work of Maxwell and others who were so successful in developing the theory of electromagnetism!

With this in mind, we turn our attention to Maxwell's equations.

Maxwell's Equations

$$\text{Ampere's Law}: \quad \nabla \times \mathbf{H} = \partial_t \mathbf{D} + \mathbf{J} \tag{1.1}$$

$$\text{Faraday's Law}: \quad \nabla \times \mathbf{E} = -\partial_t \mathbf{B} \tag{1.2}$$

$$\text{Gauss' Law of Electricity}: \quad \nabla \cdot \mathbf{D} = \rho \tag{1.3}$$

$$\text{Gauss' Law of Magnetism}: \quad \nabla \cdot \mathbf{B} = 0 \tag{1.4}$$

These equations relate the following electromagnetic quantities to each other.

- $\mathbf{E}(\mathbf{r}, t)$: electric field intensity [V/m]

- $\mathbf{H}(\mathbf{r}, t)$: magnetic field intensity [A/m]

- $\mathbf{D}(\mathbf{r}, t)$: electric flux density [C/m^2]

- $\mathbf{B}(\mathbf{r}, t)$: magnetic flux density [T = Wb/m^2]

One of the fundamental contributions of Maxwell was generalizing these equations to suggest that a time-varying electric flux density can produce a magnetic field in the same way that a time-varying magnetic flux density was known to produce an electric field. As we will see, this relationship is what allows for electromagnetic waves to propagate, and thus allows for many kinds of technologies to exist (e.g., long-distance wireless communication, radar, etc.). In addition to electric and magnetic fields/fluxes being able to produce one another, there are also two additional sources in Maxwell's equations that can produce electric and magnetic fields.

- $\mathbf{J}$: electric current density [A/m^2]

- ρ: electric charge density [C/m^3]

As one would expect, these quantities are not independent of each other. This fact is codified in the current continuity equation. The current continuity equation is a statement of the fact that charges are conserved. In differential form, this is

$$\nabla \cdot \mathbf{J} = -\partial_t \rho. \tag{1.5}$$

Intuitively, we can think of the divergence operation as helping us to measure the current density that is entering/exiting any point in space (more on this later). This must always be balanced out by changes in the amount of charge density at a particular location to ensure that charge is conserved.

Continuity equations are usually related to *conservation laws*, which play a very important role in many areas of physics. They usually take a similar kind of form to the current continuity equation. For instance, conservation laws play a very central role in the theories of special and general relativity. In those settings, these concepts often come about when considering stress-energy (also called energy-momentum) tensors which are useful quantities to help us "track" how energy flows/transfers through a system.

Before moving on, it is good to take a closer look at the units of $\mathbf{E}$, $\mathbf{H}$, $\mathbf{D}$, and $\mathbf{B}$. From this, we see that the electric and magnetic *fields* each have a unit of 1/m, while

the electric and magnetic *fluxes* each have a unit of $1/\text{m}^2$. As we become more and more acquainted with these quantities, we can often have a tendency to treat the fields and fluxes as interchangeable. However, it is always good to remember that there are some fundamental (and important) differences, as suggested by the units. In particular, we should always think of **E** and **H** as quantities that are "built" to be integrated along a line/curve, while **D** and **B** are "built" to be integrated along surfaces.

Even very talented researchers have fallen into the trap of treating the fields and fluxes as interchangeable time and time again! There are many instances throughout the history of computational electromagnetics (a field devoted to numerically solving Maxwell's equations with computers) where the researchers forgot this core principle. These early numerical methods would produce "spurious" or erroneous solutions that were sometimes difficult to properly diagnose. However, in many cases, the issue was that **E** and **H** weren't being treated as quantities "built" to be integrated along a curve, while **D** and **B** weren't being treated as quantities "built" to be integrated over a surface. Once these quantities were treated appropriately, all these spurious numerical issues vanished!

Now, if we take a closer look at Maxwell's equations, we can see that at this point we have two vector equations and two scalar equations. This is equivalent to having 8 equations that we could try and solve (each vector equation can actually be considered to be "worth" 3 scalar equations). However, we need to compute 4 vector quantities (**E**, **H**, **D**, and **B**), which implies that we need 12 equations to fully specify Maxwell's equations. So what is missing? The *constitutive relations* are our answer. These specify additional interrelationships between **E**, **H**, **D**, and **B** in the presence of simple forms of matter.

For simple materials (we will be more specific on what this means later), **D** can be related to **E**, while **B** can be related to **H**. The particular relationships are

$$\mathbf{D} = \epsilon\mathbf{E}, \tag{1.6}$$

$$\mathbf{B} = \mu\mathbf{H}, \tag{1.7}$$

where ϵ is the *permittivity* of the material (with units of [F/m]) and μ is the *permeability* of the material (with units [H/m]). Recall our discussion on units and the distinction in character between the fields and fluxes (in terms of what they should be integrated over). We see that even though these constitutive relationships appear very simple, they are fundamentally changing the character of the fields into fluxes (and, in some sense, vice-versa). Looking at the units we also see that ϵ will augment/contribute to capacitive effects, while μ will do the same for inductive effects.

In *free space* (i.e., a vacuum), the permittivity and permeability are denoted as $\epsilon_0 \approx 8.854 \times 10^{-12}$ [F/m] and $\mu_0 = 4\pi \times 10^{-7}$ [H/m], respectively (Note: the subscript 0 is generally used in electromagnetics to denote that the quantity being considered corresponds to its free space value). Since these are rather inconvenient numbers that are difficult to remember, we typically normalize permittivities and permeabilities of different materials by the corresponding free space value. These normalized quantities are called the *relative permittivity* and *relative permeability*, and are denoted as ϵ_r and μ_r, respectively. This leads to much simpler values to remember, like that Teflon has a relative permittivity of about 2.1, as opposed to Teflon has a permittivity of $2.1 \times 8.854 \times 10^{-12}$ (whatever that ugly number is). Before continuing, it is worth mentioning that many different terminologies are used for

these quantities in practice. For instance, it is very common to refer to the relative permittivity (ϵ_r) as the *dielectric constant* of a material. We prefer to not use this terminology, since, as we will learn later, the relative permittivity is never truly a "constant".

The constitutive relationships written in (1.6) and (1.7) are only valid for special kinds of simple (albeit extremely prevalent) types of materials. In particular, this is for *isotropic materials*. This means that the permittivity/permeability are the same regardless of which direction they are measured in. There are special kinds of materials, called *anisotropic materials*, where this is not the case. Simple examples of where anisotropic materials are used include in polarized sunglasses and liquid crystal displays. However, there are more exotic cases, such as for *metamaterials*, which are synthetic materials designed to have advantageous electromagnetic properties not typically available in nature. For now, we won't worry about the consequences of anisotropic materials, but we will revisit this briefly at a later point in the semester.

It is good to take stock at this point to see if we have enough equations to fully specify our electromagnetic system. If we add everything up (including constitutive relations), we now appear to have 4 vector equations and 2 scalar equations, making for a total of 14 equations with 12 unknowns to solve. Did we go too far, have we now *over-specified* everything? It turns out that the issue is in us naively treating the two Gauss' laws as being independent from the other Maxwell's equations. We can actually show that the two Gauss' laws can be derived from Ampere's and Faraday's laws (see the practice problems). So why then do we write them as separate equations? The answer is that these equations are only derivable for *time-varying systems*. Since we want our Maxwell's equations to also describe electrostatic and magnetostatic systems, it is best to keep Gauss' laws in our definition of what we consider to be "Maxwell's equations". However, this does illustrate the important point that we do not always need to try and independently solve all of Maxwell's equations. Sometimes, we can get by with just considering the constitutive relations and a few other key parts of Maxwell's equations.

Another kind of constitutive relationship that exists is Ohm's law. You are no doubt very familiar with Ohm's law from your circuit theory classes. However, as we will learn throughout this course, circuit theory at a fundamental level is a (very useful) simplification of Maxwell's equations. In the full electromagnetic picture, Ohm's law is

$$\mathbf{J} = \sigma \mathbf{E} \tag{1.8}$$

In this equation, σ is the *conductivity* of the material. Typically, a unit of [S/m] is used for the conductivity. Again note how the units taking part in this constitutive relation is fundamentally changing the character of the electric field.

1.2 Vector Calculus Review

Vector calculus plays a very important role in the study of Maxwell's equations. Because of this, we will want to develop intuition for many of the operations we will use repeatedly throughout the semester. Many of these concepts can feel abstract while first learning them in a mathematics class, but studying them in the context of an applied theory like electromagnetism can help bring the material to life. With that in mind, let's review some
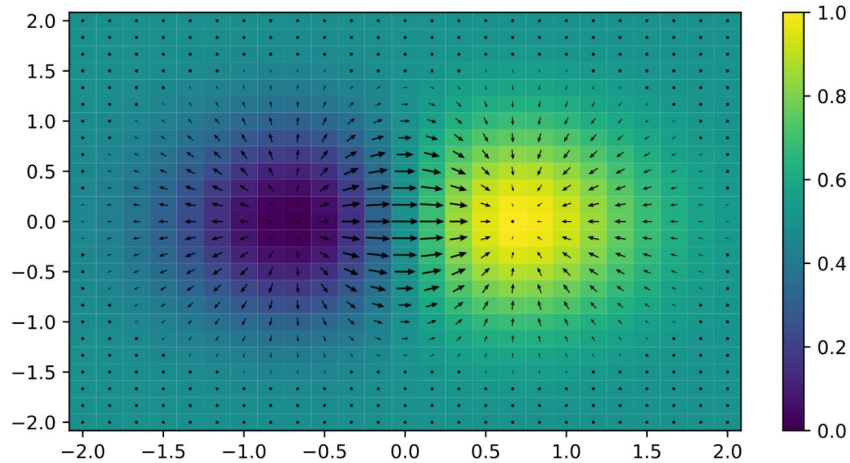
Figure 1.1: Illustration of the gradient. The colors denote the value of the scalar field and the arrows denote the vector field produced by taking the gradient of the underlying scalar field (image from Wikipedia [2]).

of the most important pieces of vector calculus. To help build intuition, we will only focus on the definitions of these quantities appropriate for Cartesian coordinates. It is essential to remember that these formulas cannot be naively applied to *other coordinate systems*; in general, that will not work!

We will start with the most basic derivative in vector calculus, the *gradient*. The gradient takes a differentiable scalar function and turns it into a vector field. For scalar function $f(\mathbf{r})$, the gradient operation is denoted as $\nabla f(\mathbf{r})$. The components of the vector field $\nabla f(\mathbf{r})$ (in Cartesian coordinates) are

$$\nabla f(\mathbf{r}) = \begin{bmatrix} \partial_x f(\mathbf{r}) \\ \partial_y f(\mathbf{r}) \\ \partial_z f(\mathbf{r}) \end{bmatrix} \tag{1.9}$$

The direction of the vector field ∇f is pointed toward where f is increasing or changing the fastest. Hence, it helps us understand the "flow" of a scalar function. By taking the dot product of ∇f with a unit vector, we can measure how quickly the scalar field is changing along the direction specified by the unit vector. This is illustrated in Fig. 1.1.

To go along with the gradient, we have the *gradient theorem*. This tells us what happens when we try and take the line integral of the gradient. You should recall from the definition of a derivative that it involves dividing by the infinitesimal distance that we are computing the change in a function over. Hence, we can think of it as carrying with it a change in units of a function by a factor of $1/\text{m}$. If you recall our discussion about the character of $\mathbf{E}$ and $\mathbf{H}$ (i.e., they should be integrated along lines), we see that the gradient is taking a general function and preparing it to be integrated along a line/curve. This is captured in a very special way in the gradient theorem, which says

$$\int_C \nabla f(\mathbf{r}) \cdot d\boldsymbol{\ell} = f(\mathbf{r}_2) - f(\mathbf{r}_1) \tag{1.10}$$
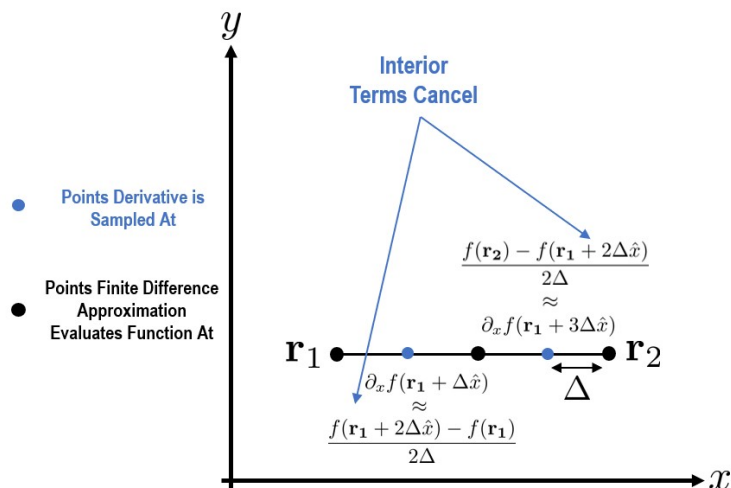
Figure 1.2: Illustration of the gradient theorem using a finite difference approximation to the derivative.

where $\mathbf{r}_1$ ($\mathbf{r}_2$) are the beginning (ending) points of the curve C. This tells us that if we integrate the gradient of a function, all we really need to consider is what happens at the *boundary of our integration range* (in this case, the endpoints of the line). See Fig. 1.2 for more details on how to interpret this result using finite difference approximations. This can be very important in electromagnetism, and is certainly prevalent in electrostatics.

As we will see shortly, for every derivative we will frequently use in vector calculus there is a matching integration theorem that tells us the only thing that we need to monitor is what happens at the *boundary* of the integration range. This has many important consequences throughout electromagnetism.

Next, we will consider the *curl* of a vector field. The curl of a vector field helps measure the "rotation/circulation" of the vector field. It takes a vector field $\mathbf{f}$ and defines a new vector field $\nabla \times \mathbf{f}$ with magnitude equaling the maximum circulation of $\mathbf{f}$ and direction denoting the axis of this maximum circulation. In Cartesian coordinates, the curl is

$$\nabla \times \mathbf{f}(\mathbf{r}) = \begin{vmatrix} \hat{x} & \hat{y} & \hat{z} \\ \partial_x & \partial_y & \partial_z \\ f_x(\mathbf{r}) & f_y(\mathbf{r}) & f_z(\mathbf{r}) \end{vmatrix} = \begin{bmatrix} \partial_y f_z(\mathbf{r}) - \partial_z f_y(\mathbf{r}) \\ \partial_z f_x(\mathbf{r}) - \partial_x f_z(\mathbf{r}) \\ \partial_x f_y(\mathbf{r}) - \partial_y f_x(\mathbf{r}) \end{bmatrix} \tag{1.11}$$

We see that each component (e.g., $\hat{x}$-component) involves derivatives "measuring" the rate of change around a closed loop in the plane orthogonal to the component (e.g., the (y, z)-plane). This is more clearly seen by expanding the derivatives in a finite difference approximation, as shown in Fig. 1.3.

To go along with the curl, we have the integration theorem known as *Stokes' theorem.* Stokes' theorem is used to integrate the curl of a vector field over a surface S. It tells us that this result only depends on what is happening on the *boundary* of S. Mathematically, this is given as

$$\iint_S \nabla \times \mathbf{f}(\mathbf{r}) \cdot d\mathbf{S} = \oint_C \mathbf{f} \cdot d\boldsymbol{\ell}, \tag{1.12}$$

6
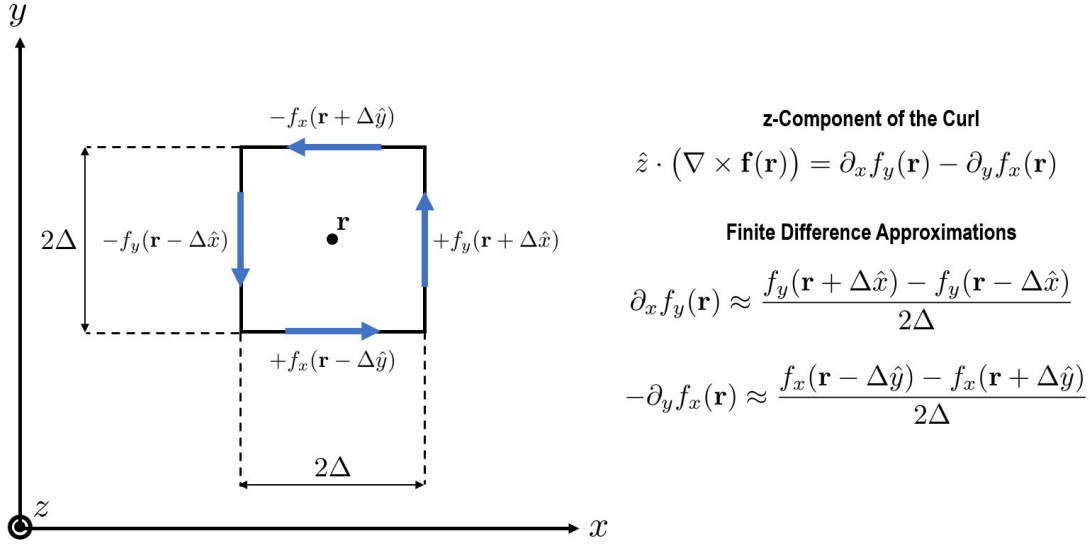
Figure 1.3: Illustration of one component of the curl using a finite difference approximation for the derivatives. The portion of the vector field that aligns with the blue arrows will accumulate into a positive value of the curl.
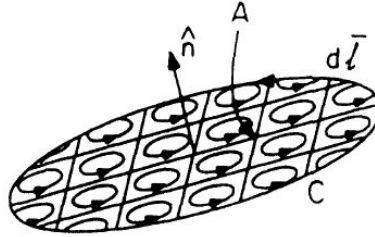
Figure 1.4: Illustration of Stokes' theorem. The circulations from nearby points cancel with one another, except over the boundary where there are not "matching" points to allow for the cancellation to occur (image from [3]).

where C forms the boundary of S and $\boldsymbol{\ell}$ defines the integration direction around C. The intuitive picture is that at all adjacent interior points the derivatives involve terms with equal magnitudes but opposite directions, making them cancel each other out. As a result, the only location where a result can actually accumulate to a non-zero value is on the boundary of the integration region where there are no adjacent points to cancel with. A rough picture of this effect can be seen in Fig. 1.4.

Next up, the *divergence* of a vector. The divergence of a vector field helps measure the "outward flux" of the vector field. It takes a vector field $\mathbf{f}$ and defines a new scalar field $\nabla \cdot \mathbf{f}$ with magnitude equaling the net flux exiting/entering the point. Positive magnitudes denote flux moving out of the point of measurement. In Cartesian coordinates, the divergence is

$$\nabla \cdot \mathbf{f}(\mathbf{r}) = \partial_x f_x + \partial_y f_y + \partial_z f_z. \tag{1.13}$$

2D Divergence

$$\nabla \cdot \mathbf{f}(\mathbf{r}) = \partial_x f_x(\mathbf{r}) + \partial_y f_y(\mathbf{r})$$

Finite Difference Approximations

$$\partial_x f_x(\mathbf{r}) \approx \frac{f_x(\mathbf{r} + \Delta\hat{x}) - f_x(\mathbf{r} - \Delta\hat{x})}{2\Delta}$$

$$\partial_y f_y(\mathbf{r}) \approx \frac{f_y(\mathbf{r} + \Delta\hat{y}) - f_y(\mathbf{r} - \Delta\hat{y})}{2\Delta}$$
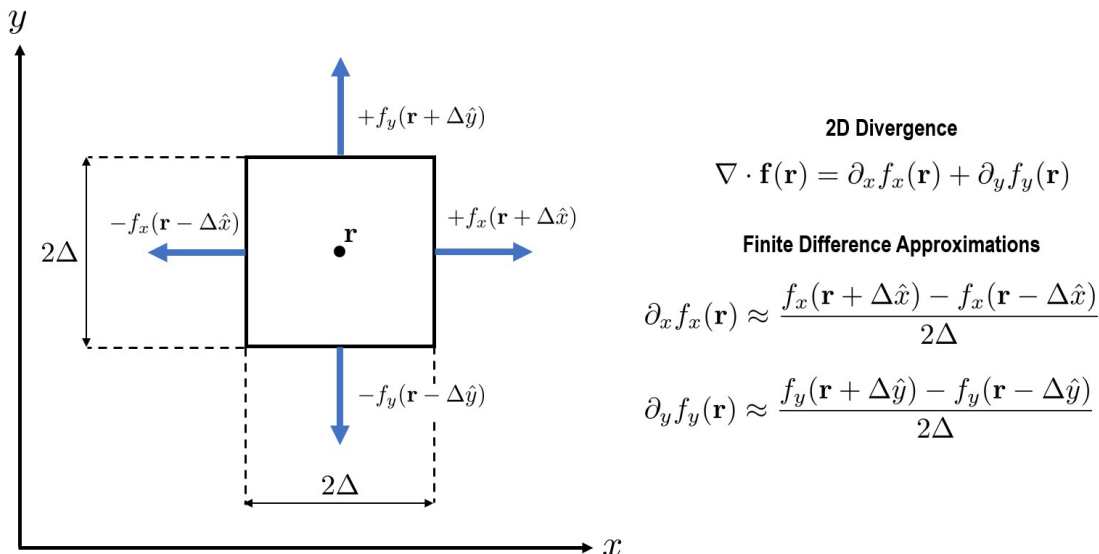
Figure 1.5: Illustration of the divergence in two-dimensions using a finite difference approximation. The portion of the vector field that aligns with the blue arrows will accumulate into a positive value of the divergence.

An illustration of the divergence shown in Fig. 1.5 using finite differences helps illustrate why this computes the "outward flux".

As with the last two derivatives, there is a matching integration theorem. This is typically referred to as the *divergence theorem*, but is also sometimes referred to as Gauss' theorem or Ostrogradsky's theorem. The divergence theorem is used to integrate the divergence of a vector field (typically over a volume). It tells us that this result only depends on what is happening on the *boundary* of the integration region, for similar reasons to what was seen with Stokes' theorem. For integrating over a volume, the divergence theorem is given as

$$\iiint_V \nabla \cdot \mathbf{f}(\mathbf{r})dV = \oiint_S \mathbf{f} \cdot \hat{n}dS, \tag{1.14}$$

where S forms the boundary of V and $\hat{n}$ is the outward pointing unit normal vector. An illustration of the derivatives calculated at interior points leading to cancellation in the integration is shown in Fig. 1.6.

The trend should hopefully be clear now. Each kind of derivative we take in vector calculus gives us a new quantity that is "built" to be integrated over some new kind of region; be it a curve, a surface, or a volume. For instance, when we originally discussed $\mathbf{E}$ and $\mathbf{H}$ we mentioned that they were built to be integrated along curves. However, they are related to $\mathbf{D}$ and $\mathbf{B}$ in Faraday's and Ampere's laws, which we said should always be thought of as being integrated over surfaces. How do we reconcile this? The answer is that the curl operations in both of these equations change the characters of $\mathbf{E}$ and $\mathbf{H}$ so that they are ready to be integrated over surfaces.

Before concluding our discussion on vector calculus, there are two last identities that are important to remember. These are that

$$\nabla \times (\nabla f) = 0, \tag{1.15}$$
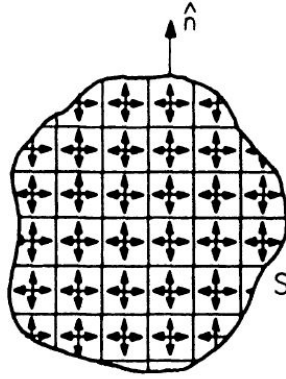
Figure 1.6: Illustration of the divergence theorem (image from [3]).

$$\nabla \cdot \left(\nabla \times \mathbf{f} \right) = 0. \tag{1.16}$$

In words, we see that the gradient vector field *always* has no "circulation" and that the circulation vector field *always* has no net "outward flux". We can think of these derivatives as measuring *"orthogonal" rates of change*, i.e., they are in some sense independent measurements/specifications of how a quantity is changing. This is a very fundamental concept that is typically encapsulated in what is known as the *Helmholtz decomposition theorem.* This theorem is central to the use of electromagnetic potential functions that will be discussed later in the course when we consider analyzing simple antennas.

1.3   Integral Form of Maxwell's Equations

Armed with these vector calculus concepts, we can now use them to determine the integral form of Maxwell's equations. This form of Maxwell's equations is useful for deriving certain results such as the boundary conditions needed to solve the differential form of Maxwell's equations. As we will see, the integral form of Maxwell's equations are also useful in deriving circuit theory from electromagnetic theory.

To begin finding the integral form of Maxwell's equations, let's start with Ampere's law. From our previous discussions, we see that all the quantities in this equation are "built" to be integrated over a surface. So let's do exactly this! Integrating Ampere's law over surface S gives us

$$\iint_S \left(\nabla \times \mathbf{H} \right) \cdot \hat{n} \, dS = \iint_S \left(\partial_t \mathbf{D} + \mathbf{J} \right) \cdot \hat{n} \, dS. \tag{1.17}$$

Looking at the left-hand side, we see that we can simplify this by using Stokes' theorem. The result is

$$\oint_C \mathbf{H} \cdot d\boldsymbol{\ell} = \iint_S \left(\partial_t \mathbf{D} + \mathbf{J} \right) \cdot \hat{n} \, dS. \tag{1.18}$$

We see that the magnetic field circulating a surface depends on the total current flowing through that surface (where $\partial_t \mathbf{D}$ is typically referred to as the *displacement current*).

9

Next, let's consider Faraday's law. The same treatment with Stokes' theorem can be applied here to give us

$$\oint_C \mathbf{E} \cdot d\boldsymbol{\ell} = -\iint_S \partial_t \mathbf{B} \cdot \hat{n}\, dS. \tag{1.19}$$

This tells us that the electric field around a loop will counteract changes in the magnetic flux intersecting that loop. This effect can be leveraged to make many different kinds of electromechanical machines.

Now let's turn our attention to Gauss' laws of electricity and magnetism. In both cases, we are taking the divergence of the fluxes. Hence, these equations are giving us quantities that should be integrated over a volume. So let's do just that! This gives us

$$\iiint \nabla \cdot \mathbf{D}\, dV = \iiint \rho\, dV \tag{1.20}$$

for Gauss' law of electricity. Looking at the left-hand side, we see that we can simplify this using the divergence theorem. This gives

$$\oiint_S \mathbf{D} \cdot \hat{n}\, dS = \iiint \rho\, dV. \tag{1.21}$$

In words, we see that the total flux exiting some volume depends on the total amount of charge contained within that volume. Following an identical process for Gauss' law of magnetism gives us

$$\oiint_S \mathbf{B} \cdot \hat{n}\, dS = 0. \tag{1.22}$$

This equation tells us two (related) things. First, because there are no magnetic charges the flux leaving a closed surface will always be exactly balanced out by an equal amount of flux entering the surface. Stated another way, this tells us that **B** *always forms closed loops.*

In summary, the integral form of Maxwell's equations are the following.

Maxwell's Equations (Integral Form)

$$\text{Ampere's Law}: \quad \oint_C \mathbf{H} \cdot d\boldsymbol{\ell} = \iint_S \left(\partial_t \mathbf{D} + \mathbf{J} \right) \cdot \hat{n}\, dS \tag{1.23}$$

$$\text{Faraday's Law}: \quad \oint_C \mathbf{E} \cdot d\boldsymbol{\ell} = -\iint_S \partial_t \mathbf{B} \cdot \hat{n}\, dS \tag{1.24}$$

$$\text{Gauss' Law of Electricity}: \quad \oiint_S \mathbf{D} \cdot \hat{n}\, dS = \iiint \rho\, dV \tag{1.25}$$

$$\text{Gauss' Law of Magnetism}: \quad \oiint_S \mathbf{B} \cdot \hat{n}\, dS = 0 \tag{1.26}$$

We can also follow similar steps for the current continuity equation. By integrating this over a volume and using the divergence theorem, we get

$$\oiint_S \mathbf{J} \cdot \hat{n}\, dS = -\iiint \partial_t \rho\, dV. \tag{1.27}$$

$$\oiint_S \mathbf{J} \cdot \hat{n} dS \rightarrow \sum_j i_j = 0$$
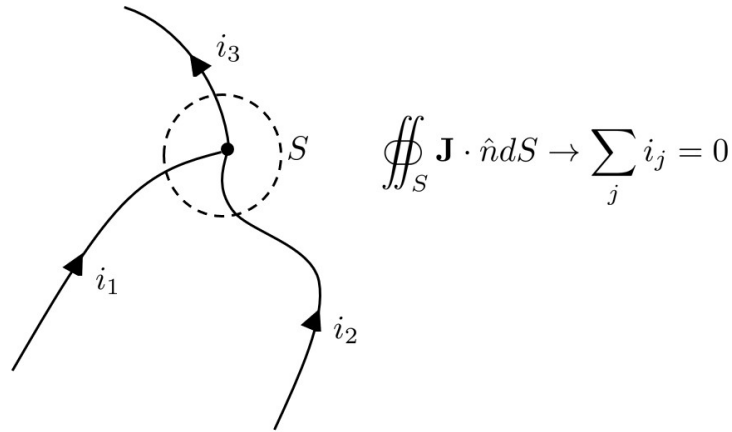
Figure 1.7: Illustration of Kirchoff's current law. For simple circuits, the integration of the current densities reduces to a summation of the total current flowing on the wires, which makes the current continuity equation reduce to Kirchoff's current law.

We can relate this to circuit theory by integrating around a "circuit node" where there are only perfect wires connected to the node. In this case, no charge can build up at the node, so we have that the sum of all the currents entering/exiting the node will be equal to 0. This is *Kirchoff's current law*, which is illustrated in Fig. 1.7.

Similarly, *Kirchoff's voltage law* can be inferred from Maxwell's equations. This is done by applying Faraday's law around the loop of a circuit that doesn't have a time-varying magnetic flux density intersecting it. This is illustrated in Fig. 1.8. Although this leads to a simple result, it should be recognized that it is an idealization of Faraday's law by assuming there is no time-varying magnetic flux intersecting the loop. In reality, many electrical components can produce time-varying magnetic fluxes (intentionally or unintentionally) that will intersect "loops" in practical circuits. This induces unexpected voltages that can cause a circuit to function incorrectly. Determining the sources of this kind of *interference* and mitigating the issues is typically referred to as the field of *electromagnetic compatibility (EMC)* or *electromagnetic interference (EMI)*.

Next, we can understand the *I-V relationships for inductors and capacitors* through Maxwell's equations. This is done by applying Faraday's law around a closed loop that contains an inductor. The total flux going through the inductor is related to the inductance by definition. In particular, we have that $L = \Phi_{tot}/i$. We can compute the total flux passing through the inductor by integrating the $\mathbf{B}$ that passes through the surface of the inductor as

$$\Phi_{tot} = \iint_S \mathbf{B} \cdot \hat{n} dS. \tag{1.28}$$

This result can then be substituted into Faraday's law to see the expected I-V relationship. See Fig. 1.9 for more details.

Finally, we can understand the I-V relationship for a capacitor in a similar way. By applying Ampere's law at various points of the circuit, we can see that the displacement

$$v_j = \int_{+}^{-} \mathbf{E} \cdot d\boldsymbol{\ell}$$

$$\oint_C \mathbf{E} \cdot d\boldsymbol{\ell} \rightarrow \sum_j v_j = 0$$
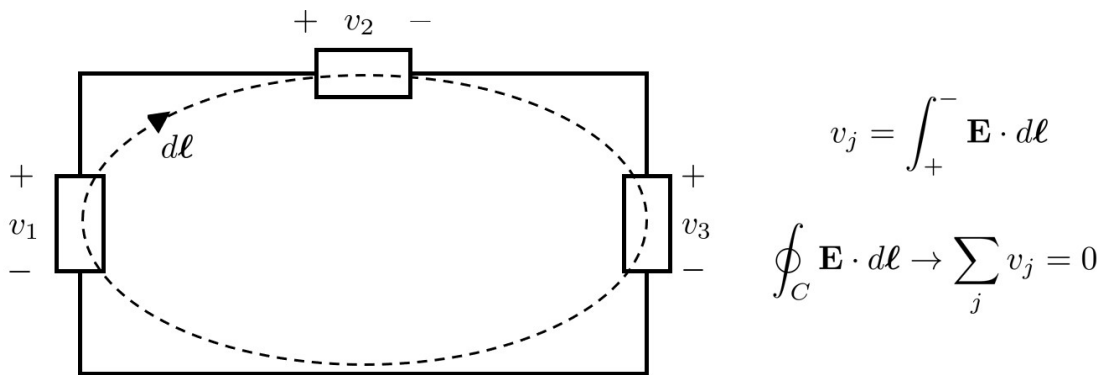
Figure 1.8: Illustration of Kirchoff's voltage law. For simple circuits with ideal wires, the only voltage changes occur through circuit components. In this simplified setting, Faraday's law becomes Kirchoff's voltage law when it is assumed there is no time-varying magnetic flux intersecting the circuit loop.

$$\oint_C \mathbf{E} \cdot d\boldsymbol{\ell} = -\partial_t \iint_S \mathbf{B} \cdot \hat{n}\, dS$$

$$-v = -\partial_t \Phi_{\text{tot}}$$

$$v = \partial_t (Li)$$

$$v = L\partial_t i$$
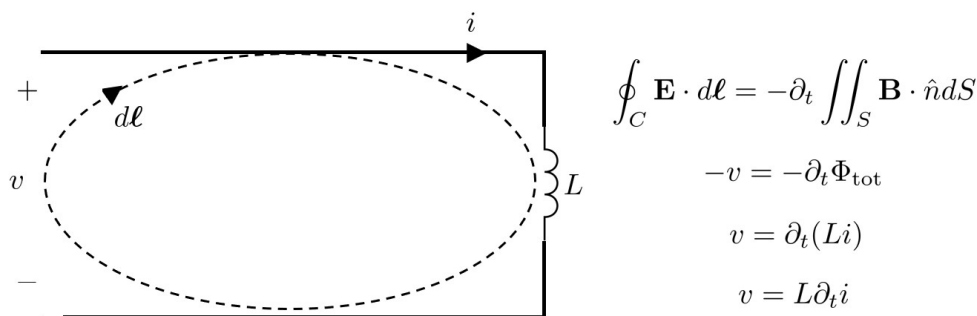
Figure 1.9: Illustration of the I-V relationship for an inductor.

current flowing through the capacitor must equal the current flowing through the circuit. From this we can infer the expected I-V relationship for a capacitor. This is most easily done for the simple case of a parallel plate capacitor. In this case, we know that the capacitance will approximately be $C = \epsilon A/d$, where A is the area of the capacitor plates and d is the distance between the two plates. If we ignore fringing fields, we know from electrostatic theory that the magnitude of the electric field in between the two plates should be approximately constant and equal to $|\mathbf{E}| = v_c/d$, where v_c is the voltage across the capacitor. Using the constitutive relationship of $\mathbf{D} = \epsilon \mathbf{E}$ and that the current flowing through the middle of the capacitor must be equal to

$$i = \iint_S \partial_t \mathbf{D} \cdot \hat{n}\, dS \tag{1.29}$$

due to Ampere's law, we can determine that if S covers the entire area of the parallel plates (and $\mathbf{D}$ is constant over the area) that

$$i = \partial_t \left(A\frac{\epsilon v_c}{d} \right) = C\partial_t v_c. \tag{1.30}$$

See Fig. 1.10 for more details.
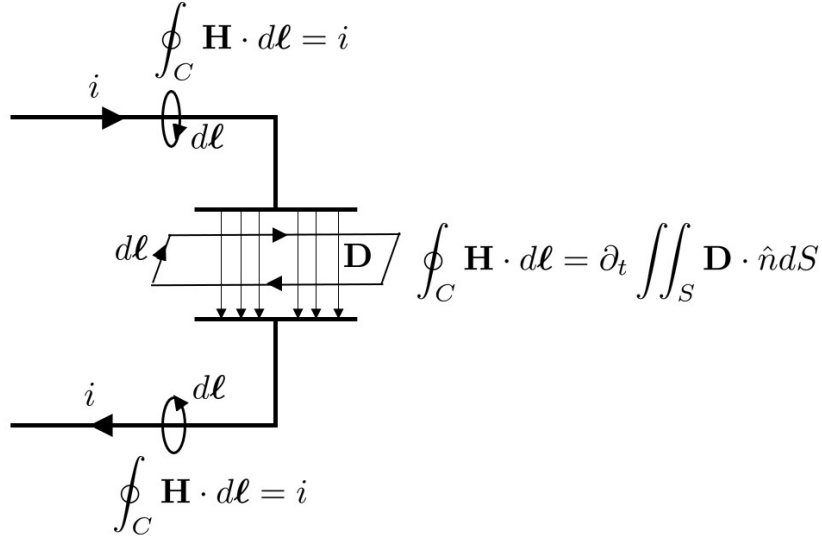
Figure 1.10: Illustration of the I-V relationship for a capacitor.

1.4 Free-Space Wave Equation

We can now go about deriving the wave equation for the electric and magnetic fields in free space. Originally, Maxwell was able to determine the wave equation because he had added the displacement current term to Ampere's law. Without this modification, propagating electromagnetic waves would not be able to exist.

To begin deriving the wave equation, we will rewrite Ampere's and Faraday's laws using the constitutive relations. This will allow us to combine the two coupled 1st-order partial differential equations (PDEs) to arrive at 2nd-order PDE. It is this 2nd-order PDE that clearly shows evidence of wave propagation. With this in mind, Ampere's law becomes

$$\begin{aligned} \nabla \times \mathbf{H} &= \partial_t \mathbf{D} \\ &= \partial_t(\epsilon_0 \mathbf{E}) \end{aligned} \tag{1.31}$$

and Faraday's law becomes

$$\begin{aligned} \nabla \times \mathbf{E} &= -\partial_t \mathbf{B} \\ &= -\partial_t(\mu_0 \mathbf{H}). \end{aligned} \tag{1.32}$$

We want to combine these two equations into a single equation. To do this, we need to make it so that we can substitute one of the equations into the other one. This will be possible by taking the curl of (1.32), which gives

$$\nabla \times \nabla \times \mathbf{E} = -\mu_0 \nabla \times \partial_t \mathbf{H}. \tag{1.33}$$

Note that because μ_0 does not vary as a function of position, we were able to bring it "outside" of the curl operation. Next, we will swap the order of the temporal and spatial derivatives on $\mathbf{H}$. This is safe to do in this case because our "time" and "space" variables are independent. However, this "swapping" of the order of second-order (or higher) derivatives

is generally acceptable under fairly easy requirements of the function being differentiated. Doing this, we have

$$\nabla \times \nabla \times \mathbf{E} = -\mu_0 \partial_t (\nabla \times \mathbf{H}). \tag{1.34}$$

We can now substitute the righ-hand side of (1.31) into this equation for $\nabla \times \mathbf{H}$. This gives

$$\nabla \times \nabla \times \mathbf{E} = -\mu_0 \epsilon_0 \partial_t^2 \mathbf{E}. \tag{1.35}$$

Moving everything to one side of the equation gives us the *vector wave equation* for the electric field. This is

$$\boxed{\nabla \times \nabla \times \mathbf{E} + \mu_0 \epsilon_0 \partial_t^2 \mathbf{E} = 0.} \tag{1.36}$$

We can follow a very similar derivation to get the vector wave equation for the magnetic field as well. This is

$$\nabla \times \nabla \times \mathbf{H} + \mu_0 \epsilon_0 \partial_t^2 \mathbf{H} = 0. \tag{1.37}$$

Since we are in a source-free region, we can use some more vector calculus identities to rewrite these equations into another form. In particular, we note that

$$\nabla \times \nabla \times \mathbf{f} = \nabla (\nabla \cdot \mathbf{f}) - \nabla^2 \mathbf{f}, \tag{1.38}$$

where $\nabla^2 = \nabla \cdot \nabla$ (i.e., the divergence of the gradient of the vector) is the vector Laplacian operator. Using this in (1.36), we have

$$\nabla (\nabla \cdot \mathbf{E}) - \nabla^2 \mathbf{E} + \mu_0 \epsilon_0 \partial_t^2 \mathbf{E} = 0. \tag{1.39}$$

Has this helped us out at all? Well, not yet. However, if we note that Gauss' law of electricity in a source-free region of free space is

$$\epsilon_0 \nabla \cdot \mathbf{E} = 0, \tag{1.40}$$

we can further simplify our wave equation. Using (1.40) in (1.39), we have

$$\boxed{\nabla^2 \mathbf{E} - \mu_0 \epsilon_0 \partial_t^2 \mathbf{E} = 0.} \tag{1.41}$$

This is the *vector Helmholtz wave equation*. It is traditional for wave equations to define the square of the propagation speed as the inverse of the constants multiplying the time derivative term. Here, this gives us the speed of light, denoted as

$$c_0 = \frac{1}{\sqrt{\mu_0 \epsilon_0}} \approx 3 \times 10^8 \, [\text{m/s}]. \tag{1.42}$$

In our current digital age, there are very few numbers that I will recommend for you to memorize. However, this is certainly one of them. Another good one to note is that $c_0 \approx 1 \, [\text{ft/ns}]$. We will examine this equation in greater detail shortly, and will spend a significant amount of

time understanding the properties of solutions to this equation for different electromagnetic systems.

So how do we go about interpreting the wave equation (1.41)? The first item to tackle is understanding what the *Laplacian operator* does when it is applied to a vector quantity. However, before we jump to that, let's first recall how it works with a *scalar function*. In Cartesian coordinates, the Laplacian operator applied to a scalar function takes the following form:

$$\nabla^2 f(\mathbf{r}) = \nabla \cdot \big[\nabla f(\mathbf{r})\big]. \tag{1.43}$$

That is, we first take the gradient of f and then take the divergence of the result. This evaluates to

$$\nabla^2 f(\mathbf{r}) = \partial_x^2 f(\mathbf{r}) + \partial_y^2 f(\mathbf{r}) + \partial_z^2 f(\mathbf{r}). \tag{1.44}$$

If we recall our interpretations of the gradient and divergence operations, we see that in words the Laplacian operator is computing the net outward flux of the gradient flow of f. It is also a higher order derivative. These occur very commonly in physics to describe things like diffusion. Further, if you recall the Taylor series of a function, we can associate with the second derivative as being useful in establishing a "higher order" understanding of how the function f is changing around the point we are evaluating our derivatives at. Understanding quantities through their "curvature" is a hugely fundamental concept that occurs in many areas of modern physics.

Now that we have a better feel for what the Laplacian means when applied to a scalar function, we can turn our attention to using it on a vector function. However, we are immediately hit with needing to compute the gradient of a vector function. Is this allowed? If so, what does it even mean and what does the result look like? It turns out that this is allowed, and we go about computing it by taking the gradient of each of the *components* of the vector function. Considering this, the output of this operation is 3 vector quantities that are all related to each other. We need a good way to handle the bookkeeping of this quantity to not make mistakes or make our notation too cumbersome. Traditionally, this has been done using *dyads*. In the same way that we sometimes represent vector functions using a column vector notation, dyads can be represented as a matrix of values. Hence, the gradient of a vector can be represented as

$$\nabla \mathbf{f}(\mathbf{r}) = \begin{bmatrix} \partial_x f_x(\mathbf{r}) & \partial_x f_y(\mathbf{r}) & \partial_x f_z(\mathbf{r}) \\ \partial_y f_x(\mathbf{r}) & \partial_y f_y(\mathbf{r}) & \partial_y f_z(\mathbf{r}) \\ \partial_z f_x(\mathbf{r}) & \partial_z f_y(\mathbf{r}) & \partial_z f_z(\mathbf{r}) \end{bmatrix}. \tag{1.45}$$

Successfully manipulating dyads and performing calculus on them requires learning a number of new rules. This approach has fallen out of favor and is becoming a bit of a dying art in modern times for a number of reasons. One main reason is that better methods were needed to handle generalizations of dyads (called tensors) that are important in studying various areas of modern physics, such as general relativity. You will occasionally come across dyads still being used in some limited research fields, and occasionally they will still be used in some graduate level electromagnetics courses. However, we favor leaving them as a dying art, so we will not dwell on them in detail.

Instead, we will just note the following simple form of the vector Laplacian that occurs in Cartesian coordinates (and *only Cartesian coordinates!*). In this case, the Laplacian operator applied to a vector function takes the following simple form:

$$\nabla^2 \mathbf{f}(\mathbf{r}) = \begin{bmatrix} \nabla^2 f_x(\mathbf{r}) \\ \nabla^2 f_y(\mathbf{r}) \\ \nabla^2 f_z(\mathbf{r}) \end{bmatrix}. \tag{1.46}$$

From this, we see that the vector Laplacian operator simply is the scalar Laplacian operator applied to each component of the vector. However, it cannot be emphasized enough that this simple form *only applies in Cartesian coordinates!*

Now that we understand the vector Laplacian operator better, let's revisit the vector Helmholtz wave equation. This was

$$\nabla^2 \mathbf{E} - \frac{1}{c_0^2} \partial_t^2 \mathbf{E} = 0. \tag{1.47}$$

Considering the complexity of using the vector Laplacian operator, this form of the wave equation is typically only useful for understanding simple wave solutions. So let's consider one of these now. In particular, let's assume that we can represent $\mathbf{E}(\mathbf{r}, t)$ as

$$\mathbf{E}(\mathbf{r}, t) = \hat{x} E_x(z, t). \tag{1.48}$$

This electric field only varies as a function of z and t, and has only a single vector direction (we say that is it *polarized* along the x-direction). In this case, the vector Helmholtz equation reduces down to the much simpler *scalar Helmholtz equation*. That is,

$$\nabla^2 E_x(z, t) - \frac{1}{c_0^2} \partial_t^2 E_x(z, t) = 0. \tag{1.49}$$

Considering that E_x only varies with z in space, we can further simplify this to be

$$\partial_z^2 E_x(z, t) - \frac{1}{c_0^2} \partial_t^2 E_x(z, t) = 0. \tag{1.50}$$

What will a solution to this need to look like? Well, we can first of all obviously say that it will need to be twice-differentiable for the derivatives to be able to be correctly applied. Next, we see that whatever the spatial derivative does to E_x is going to need to be very carefully balanced out by what the temporal derivative does to E_x. More formally, we can likely expect that the z and t variables will enter into E_x in the same functional form.

Considering these points, let's assume for now that $E_x(z, t)$ actually takes the form of $E_x(a_1 t \pm a_2 z)$, where a_1 and a_2 are some constants that we want to determine. Let's substitute this proposed solution form into (1.50) and see what we get. At first, we have

$$\partial_z^2 E_x(a_1 t \pm a_2 z) = \frac{1}{c_0^2} \partial_t^2 E_x(a_1 t \pm a_2 z), \tag{1.51}$$

which isn't immediately useful. However, we can use a simple trick with the chain rule to recognize that

$$\partial_z E_x(a_1 t \pm a_2 z) = E_x'(a_1 t \pm a_2 z) \frac{d}{dz}[a_1 t \pm a_2 z] = \pm a_2 E_x'(a_1 t \pm a_2 z), \tag{1.52}$$

where the $'$ symbol is being used to denote a derivative with respect to the entire argument of a function, which is $a_1 t \pm a_2 z$ in this case. Similarly, we see that

$$\partial_z^2 E_x(a_1 t \pm a_2 z) = a_2^2 E_x''(a_1 t \pm a_2 z). \tag{1.53}$$

Likewise, we see that

$$\partial_t^2 E_x(a_1 t \pm a_2 z) = a_1^2 E_x''(a_1 t \pm a_2 z). \tag{1.54}$$

If we substitute (1.53) and (1.54) into (1.51), we now have

$$a_2^2 E_x''(a_1 t \pm a_2 z) = \frac{a_1^2}{c_0^2} E_x''(a_1 t \pm a_2 z). \tag{1.55}$$

We can cancel out the identical terms to see that this functional form of a solution will work so long as

$$c_0^2 = \frac{a_1^2}{a_2^2}. \tag{1.56}$$

It is customary to choose $a_1 = 1$ and $a_2 = 1/c_0$, so that our functional form of a solution to the wave equation becomes

$$\boxed{\mathbf{E}(\mathbf{r}, t) = E_x\left(t \pm \frac{z}{c_0}\right).} \tag{1.57}$$

This solution exhibits a simple form of *wave propagation*. If we advance t to $t + \Delta t$, we see that the same point of the function E_x can be reached by correspondingly adjusting our z location to $z \mp \Delta z$ if $\Delta z = c_0 \Delta t$. We say that in time Δt the wave has moved/advanced by a distance of $c_0 \Delta t$. This is why c_0 is referred to as the *speed of light*. Further, we see that $E_x(t - z/c_0)$ can be associated with waves traveling in the $+z$-direction and that $E_x(t + z/c_0)$ can be associated with waves traveling in the $-z$-direction.

1.5 Maxwell's Equations in the Frequency Domain

In general, solving Maxwell's equations in the time domain is a very difficult undertaking. It is a good rule of thumb that each time we increase the number of dimensions in a partial differential equation we are trying to solve analytically the problem becomes exponentially harder. Due to this, it is very common that we will try and find useful ways to simplify the problems we are studying to make them more tractable. An excellent example of this is provided by *time-harmonic analysis*. You should already be familiar with using the Fourier transform to simplify the analysis of problems in your AC circuit theory and linear systems courses. In your AC circuits courses it was typically used to convert a differential equation into an algebraic one that could be solved much more readily. Similarly, your linear systems courses should have taught you how we can build significant intuition into how linear, time-invariant systems behave by "probing" them with sinusoidal functions. In electromagnetics,

we can use a combination of these concepts to greatly simplify the study of many practical problems.

To begin, let's recall some basic details about the Fourier transform. The Fourier transform provides us with a way to express a time domain signal in terms of an "infinite summation" of sinusoidal functions (and vice-versa). If you recall your linear algebra theory, the Fourier transform can be predominantly viewed as a kind of basis transformation. The particular form of the Fourier transform that we will use in this course is

$$f(\omega) = \mathcal{F}\{f(t)\} = \int_{-\infty}^{\infty} f(t)e^{-j\omega t}dt. \tag{1.58}$$

We can interpret this integration as an inner product of our function $f(t)$ with a particular *Fourier harmonic*, namely, $e^{j\omega t}$. Hence, we are measuring how much of $f(t)$ "overlaps" with the particular Fourier harmonic of interest.

We can invert our Fourier transform using a similar transform. In particular, this will be

$$f(t) = \mathcal{F}^{-1}\{f(\omega)\} = \frac{1}{2\pi} \int_{-\infty}^{\infty} f(\omega)e^{j\omega t}d\omega, \tag{1.59}$$

which can be interpreted as an "infinite summation" of all the different Fourier harmonics weighted by $f(\omega)$ to properly reproduce $f(t)$.

One of the most important properties of the Fourier transform for simplifying differential equations is how a derivative of a Fourier transform variable changes. To see this, let's take the time derivative of $f(t)$ as expressed through the inverse Fourier transform written above. This is

$$\begin{aligned}
\frac{d}{dt}f(t) &= \frac{1}{2\pi}\frac{d}{dt}\int_{-\infty}^{\infty} f(\omega)e^{j\omega t}d\omega \\
&= \frac{1}{2\pi}\int_{-\infty}^{\infty} j\omega f(\omega)e^{j\omega t}d\omega.
\end{aligned} \tag{1.60}$$

Hence, we see that

$$\boxed{\frac{d}{dt} \iff j\omega} \tag{1.61}$$

where the $\iff$ denotes how these quantities transform under the action of the Fourier and inverse Fourier transforms.

Let's see how we can use this to simplify Maxwell's equations. We will start by assuming that we are dealing with a linear, time-invariant system so that our Fourier representations of functions are appropriate. Now, let's rewrite all of the quantities in Ampere's law using the representation provided by the inverse Fourier transform. This gives us

$$\nabla \times \mathbf{H}(\mathbf{r}, t) = \partial_t \mathbf{D}(\mathbf{r}, t) + \mathbf{J}(\mathbf{r}, t)$$

$$\nabla \times \left[\frac{1}{2\pi}\int_{-\infty}^{\infty} \mathbf{H}(\mathbf{r}, \omega)e^{j\omega t}d\omega\right] = \partial_t\left[\frac{1}{2\pi}\int_{-\infty}^{\infty} \mathbf{D}(\mathbf{r}, \omega)e^{j\omega t}d\omega\right] + \frac{1}{2\pi}\int_{-\infty}^{\infty} \mathbf{J}(\mathbf{r}, \omega)e^{j\omega t}d\omega \tag{1.62}$$

$$\frac{1}{2\pi}\int_{-\infty}^{\infty} \nabla \times \mathbf{H}(\mathbf{r}, \omega)e^{j\omega t}d\omega = \frac{1}{2\pi}\int_{-\infty}^{\infty} j\omega\mathbf{D}(\mathbf{r}, \omega)e^{j\omega t}d\omega + \frac{1}{2\pi}\int_{-\infty}^{\infty} \mathbf{J}(\mathbf{r}, \omega)e^{j\omega t}d\omega$$

Since our system is linear and time-invariant, we can analyze this system at each frequency component individually and then add up all the results later to recover the full time domain response. Due to this, we don't need to "worry" about the integrations and common terms in this equation. A more exact justification for this is possible by "projecting" our equation onto a particular Fourier harmonic of interest. We can do this by taking the inner product of (1.62) with a Fourier harmonic, e.g., $\exp[j\omega't]$. For notational simplicity, we will only focus on what happens to the left-hand side of this equation (the same process will work for all of the terms in the overall equation). Now, taking the inner product gives us

$$\int_{-\infty}^{\infty} e^{-j\omega't} \left[\frac{1}{2\pi} \int_{-\infty}^{\infty} \nabla \times \mathbf{H}(\mathbf{r}, \omega) e^{j\omega t} d\omega \right] dt. \tag{1.63}$$

Swapping the order of integrations then gives us

$$\frac{1}{2\pi} \int_{-\infty}^{\infty} \nabla \times \mathbf{H}(\mathbf{r}, \omega) \left[\int_{-\infty}^{\infty} e^{j(\omega - \omega')t} dt \right] d\omega. \tag{1.64}$$

The inner integration will only be non-zero if $\omega = \omega'$. More generally, we can recognize this inner integration as being proportional to the Dirac delta function $\delta(\omega - \omega')$. Hence, we have that (1.64) simplifies to

$$\int_{-\infty}^{\infty} \nabla \times \mathbf{H}(\mathbf{r}, \omega) \delta(\omega - \omega') d\omega. \tag{1.65}$$

This final integration can be evaluated easily using the sifting property of the delta function. This finally gives us the simplified result that

$$\int_{-\infty}^{\infty} e^{-j\omega't} \left[\frac{1}{2\pi} \int_{-\infty}^{\infty} \nabla \times \mathbf{H}(\mathbf{r}, \omega) e^{j\omega t} d\omega \right] dt = \nabla \times \mathbf{H}(\mathbf{r}, \omega'). \tag{1.66}$$

This same process can be applied to the rest of the terms in Ampere's law given in (1.62). By finally swapping the arbitrary notations of ω' to ω at the end to keep a consistent notation, we have

$$\nabla \times \mathbf{H}(\mathbf{r}, \omega) = j\omega \mathbf{D}(\mathbf{r}, \omega) + \mathbf{J}(\mathbf{r}, \omega). \tag{1.67}$$

We can follow suit with the rest of Maxwell's equations. The end result is summarized below.

Maxwell's Equations: Frequency Domain

$$\begin{aligned}
\text{Ampere}'\text{s Law}: && \nabla \times \mathbf{H} &= j\omega \mathbf{D} + \mathbf{J} && (1.68) \\
\text{Faraday}'\text{s Law}: && \nabla \times \mathbf{E} &= -j\omega \mathbf{B} && (1.69) \\
\text{Gauss}'\text{ Law of Electricity}: && \nabla \cdot \mathbf{D} &= \rho && (1.70) \\
\text{Gauss}'\text{ Law of Magnetism}: && \nabla \cdot \mathbf{B} &= 0 && (1.71)
\end{aligned}$$

Similarly, the current continuity equation becomes

$$\nabla \cdot \mathbf{J} = -j\omega\rho. \tag{1.72}$$

In electrical engineering, we typically use *phasors* to represent our time-harmonic quantities. These are completely specified if we establish their magnitude, frequency, and initial phase. We do this so that we can represent our phasors using complex-valued exponential functions (for which algebra is very easy) as opposed to the more cumbersome trigonometric functions. For a cosine reference at a fixed position in space, an example of this is

$$\begin{aligned} E_x(z = 0, t) &= E_0 \cos\left(\omega t + \varphi\right) \\ &= \mathrm{Re}\left\{ \underline{E_0 e^{j\varphi}} \, e^{j\omega t} \right\}. \end{aligned} \tag{1.73}$$

We typically call the underlined portion the phasor. It has a magnitude of E_0, a phase of φ, and a frequency of ω. We can see that the two lines in (1.73) are equivalent by using Euler's formula,

$$e^{jx} = \cos(x) + j\sin(x), \tag{1.74}$$

to write the exponential functions as trigonometric functions. We will use the phasor representation shown in (1.73) in most cases throughout this course. As a result of this, it is good to remember that we can recover the *instantaneous expression* of a phasor solution by multiplying it by $\exp(j\omega t)$ and taking the real part. You will need to do this frequently in this course, and can also be useful in practice; e.g., for generating time-harmonic visualizations of simulations produced by various computational electromagnetics software.

Before moving on, it will be good to review a few additional mathematical preliminaries. Two identities that we will find particularly useful throughout this course are

$$2\cos(x) = e^{jx} + e^{-jx} \tag{1.75}$$

and

$$2j\sin(x) = e^{jx} - e^{-jx}. \tag{1.76}$$

We will often use these identities to simplify solutions to problems involving reflected fields and to compute the power flow in an electromagnetic system. We will also often need to take the complex conjugate of expressions involving many imaginary numbers. It is important to remember that to successfully perform this operation we simply flip the sign anywhere that we see the imaginary number. For example, we have that

$$\left(je^{5-jx} \right)^* = -je^{5+jx}. \tag{1.77}$$

Now, let's put some of these frequency domain and phasor concepts together to look at a particularly useful form of the solution to the wave equation. This is

$$E_x(z, t) = E_0 \cos(\omega t - k_0 z), \tag{1.78}$$
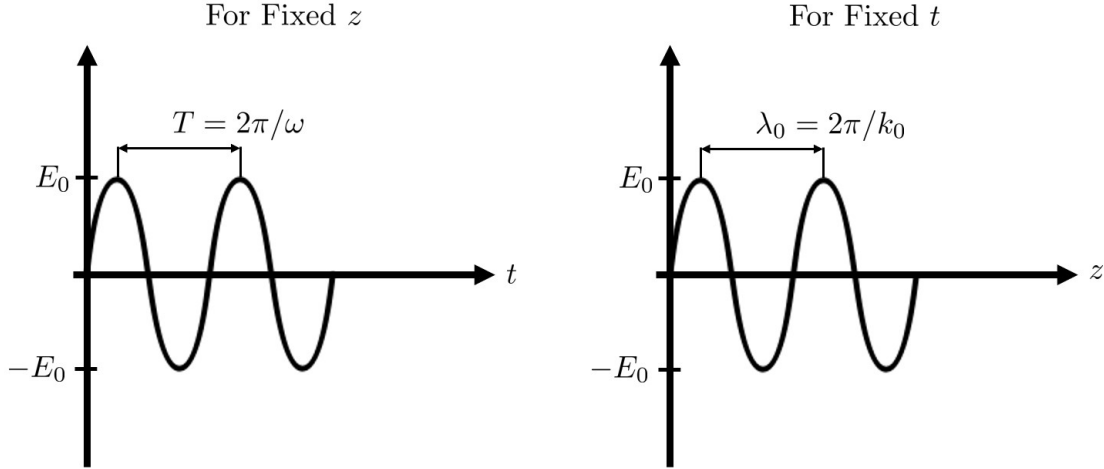
Figure 1.11: Illustration of the period and wavelength of a sinusoidal wave.

where $k_0 = 2\pi/\lambda_0 = \omega/c_0$ is the *free-space wavenumber* of the wave, and $\lambda_0 = c_0/f$ is the wavelength in free space *(Note: this formula for the wavelength is only valid in a limited number of circumstances, and should be used with caution throughout this course!)*. You can see that this is in fact a solution to the wave equation by substituting it back into (1.50). Considering the units of the wavenumber, we see that it helps us understand how much the phase will change by moving backward/forward some physical distance along the wave. Similarly, if we hold at a fixed spatial location, $\omega = 2\pi f = 2\pi/T$ helps us understand how much the phase will change as time elapses. Both of these quantities are illustrated in Fig. 1.11.

We can also use this expression to introduce the concept of *phase velocity*. This is the velocity that a particular position of the wave propagates at. We can find this by setting the argument of the cosine function equal to a constant and seeing what result we get. For simplicity, we will choose the constant to be 0. For this situation, we have

$$\omega t - k_0 z = 0. \tag{1.79}$$

We can now rearrange this equation so that we will have units that look like a velocity. This gives

$$\frac{\omega}{k_0} = \frac{z}{t}. \tag{1.80}$$

We will call the left-hand side of this equation the *phase velocity* of the wave, which will be denoted as v_p. If we expand out the definitions of ω and k_0 we can find that

$$\frac{\omega}{k_0} = \frac{\omega}{\omega/c_0} = c_0. \tag{1.81}$$

That is, the phase velocity matches the speed of light. So what was the point in defining the phase velocity as a concept if it is just equal to the speed of light? It turns out that the phase velocity will only match the speed of light for a few special cases of propagating

21

waves. Somewhat surprisingly, we will even find that for some kinds of propagating waves the phase velocity can be greater than the speed of light! The reason for this lies in the fact that the wavenumber does not always have such a simple form as ω/c_0. We will discuss this in much more detail when we reach other portions of the course, such as the sections on waveguides.

1.5.1 Frequency Domain Wave Equation

Now that we know how to convert Maxwell's equations to the frequency domain, let's look at how our source-free wave equations change. As we will see, the derivation changes in only a minor way compared to the time domain one. Considering this, let's make things a bit more interesting by deriving the wave equation for a *homogeneous medium* instead of free space. This means that all of space is completely filled with a material characterized by a general permittivity and permeability, denoted as ϵ and μ, respectively.

With this in mind, we again start with Ampere's law. In the frequency domain and after using the constitutive relations, this is

$$\nabla \times \mathbf{H} = j\omega\epsilon\mathbf{E}. \tag{1.82}$$

Similarly, Faraday's law is

$$\nabla \times \mathbf{E} = -j\omega\mu\mathbf{H}. \tag{1.83}$$

To combine these two equations into a single one, we will take the curl of (1.83). This gives

$$\nabla \times \nabla \times \mathbf{E} = -j\omega\mu\nabla \times \mathbf{H}. \tag{1.84}$$

We can now substitute the right-hand side of (1.82) into this equation for $\nabla \times \mathbf{H}$. This gives

$$\nabla \times \nabla \times \mathbf{E} = \mu\epsilon\omega^2\mathbf{E}. \tag{1.85}$$

Moving everything to one side of the equation gives us the *vector wave equation* for the electric field in the frequency domain. This is

$$\nabla \times \nabla \times \mathbf{E} - \mu\epsilon\omega^2\mathbf{E} = 0. \tag{1.86}$$

If we define the *speed of light (in this material)* to be $c = 1/\sqrt{\mu\epsilon}$, then we can rewrite this to be

$$\nabla \times \nabla \times \mathbf{E} - \left(\frac{\omega}{c}\right)^2 \mathbf{E} = 0. \tag{1.87}$$

Again, we can define the *wavenumber (in this material)* to be $k = \omega/c$ so that we arrive at

$$\boxed{\nabla \times \nabla \times \mathbf{E} - k^2\mathbf{E} = 0.} \tag{1.88}$$

Before moving on, it is useful to note that for simple materials with $\mu, \epsilon > 0$ the speed of light will be reduced in the material. As a result, the wavenumber becomes larger. That

is, the phase varies more quickly as a function of position in the material than it would in free space. Hence, the wavelength is also smaller. Although most materials have $\mu, \epsilon > 0$, it is possible for either or both of μ and ϵ to be less than 0. An example of a natural material that can have $\epsilon < 0$ is a *plasma*. More recently, there has been significant research interest in designing synthetic materials that behave as if they have negative μ and/or ϵ. These man-made materials are known as *metamaterials*, and they can be engineered to have many exciting properties that are not possible with some of the simpler materials that are commonly encountered in the natural world.

Now, returning to the derivation of the wave equation, since we are still in a source-free homogeneous region we can use the same vector calculus identities to rewrite these equations into the Helmholtz wave equation form we are familiar with. In particular, we note that

$$\nabla \times \nabla \times \mathbf{f} = \nabla(\nabla \cdot \mathbf{f}) - \nabla^2 \mathbf{f}. \tag{1.89}$$

Using this in (1.88), we have

$$\nabla(\nabla \cdot \mathbf{E}) - \nabla^2 \mathbf{E} - k^2 \mathbf{E} = 0. \tag{1.90}$$

Noting that Gauss' law of electricity in a source-free region of a homogeneous medium is

$$\epsilon \nabla \cdot \mathbf{E} = 0, \tag{1.91}$$

we can further simplify the wave equation. Using (1.91) in (1.90), we have

$$\boxed{\nabla^2 \mathbf{E} + k^2 \mathbf{E} = 0.} \tag{1.92}$$

This is the *vector Helmholtz wave equation* in the frequency domain.

We can now turn our attention to coming up with solutions to this equation in our homogeneous medium. As with the time domain, this form of the wave equation is mostly useful when we consider a relatively simple case so we can avoid the difficulty of dealing with dyadic functions. So, let's go ahead and again assume that we have an electric field polarized along the x-direction that is only varying in the z-direction. That is, $\mathbf{E} = \hat{x} E_x(z)$. Note that in this expression we have dropped the explicit reference to the fact that this electric field should be defined at a particular frequency. This is done to save space in our equations, and is done very often when it is understood by context that a particular time convention is being used to suppress the frequency variables.

So what should the functional form of $E_x(z)$ be to solve (1.92)? To start out, let's substitute our expression for $\mathbf{E}$ into (1.92) to get our simpler equation. Recall that because we are working in Cartesian coordinates and our electric field is polarized along a single axis and only varies along a single dimension, the vector Laplacian simplifies greatly. We end up with our equation being

$$\frac{d^2}{dz^2} E_x(z) + k^2 E_x(z) = 0. \tag{1.93}$$

The first thing that we should notice is that whatever functional form we choose, it is going to need to "return" back to itself in some manner after we have taken two derivatives.

The options that should jump to mind for this kind of behavior should be $\cos(kz)$, $\sin(kz)$, and $\exp(\pm jkz)$. As we will see throughout the semester, we can certainly use any of these functions as our "baseline" solution to the wave equation. However, we can often simplify the algebra in our overall analysis of a problem if we make a good choice as to which general solution to use. For "open" problems that don't have any boundaries that we need to worry about the behavior of our solution at, it will generally be the case that the exponential functions are good choices of general solutions.

Considering this, let's write our solution as

$$E_x(z) = E^+ e^{-jkz} + E^- e^{jkz} \tag{1.94}$$

for the time being, where E^+ and E^- are complex-valued constants denoting the amplitude and phase of the two solutions. The superscripts on these constants are suggestively written to point out that the first function on the right-hand side of (1.94) corresponds to a wave propagating in the $+z$-direction, while the second function on the right-hand side of (1.94) corresponds to a wave propagating in the $-z$-direction. How can we tell that this is the case? The easiest way will be to transform our phasor solutions back to the time domain. Recall that we go about doing this by multiplying these solutions by $\exp(j\omega t)$ (our time convention) and taking the real part of the result. From this, we get

$$
\begin{aligned}
\mathrm{Re}\left\{ E_x(z) e^{j\omega t} \right\} &= \mathrm{Re}\left\{ E^+ e^{-jkz} e^{j\omega t} \right\} + \mathrm{Re}\left\{ E^- e^{jkz} e^{j\omega t} \right\} \\
&= |E^+| \cos(\omega t - kz + \angle E^+) + |E^-| \cos(\omega t + kz + \angle E^-).
\end{aligned}
\tag{1.95}
$$

In this equation, we have used the symbol $\angle E^{\pm}$ to represent the phase of the complex-valued constant $E^{\pm}$. Hence, we see that the first function on the right-hand side of (1.95) does represent a wave propagating in the $+z$-direction, and likewise for the negative propagating wave.

1.6 Introduction to Plane Waves

The wave solution that we studied in the previous section is an example of a very special kind of wave that is often used in studying electromagnetic systems. In particular, this kind of wave is called a *plane wave*. This name is in reference to the fact that if we were to freeze time we would see that the *isophase surfaces* (the surface formed by all points having the same phase) of this wave form *planes* in 3D space (see Fig. 1.12). Further, we can call this wave a *uniform plane wave* because all of the points on our isophase surface have the same magnitude. This kind of wave, although very simple, is extremely important in many practical applications. The reason for this is that if we move a *sufficiently far* distance away from the source of a propagating wave we can almost always consider the wave reaching us to effectively be a uniform plane wave (see Fig. 1.13). This greatly simplifies the analysis we need to perform when designing various systems, such as antennas, as we will see later in this course.

Before moving on, a *quick note on terminology*. If someone refers to something simply as a plane wave, it is a good rule of thumb to assume they actually mean a uniform plane wave.
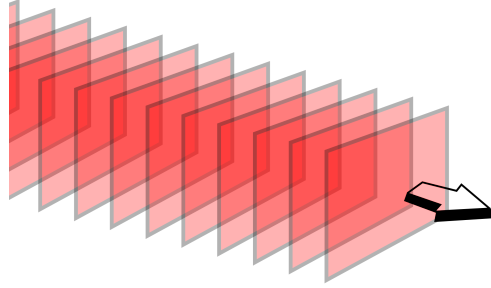
Figure 1.12: Example of the wavefronts of a plane wave. Note that the wavefronts are orthogonal to the direction of propagation (image from Wikipedia [4]).
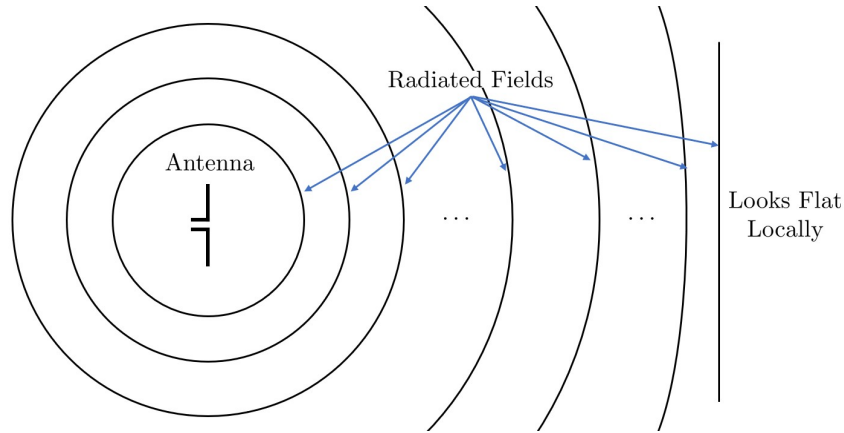
Figure 1.13: Radiated fields far away from their source quickly begin to look locally flat; i.e., like plane waves.

The "uniform" part is very commonly omitted since it is generally safe to assume that what we are discussing are uniform plane waves.

Let's now take some time to look at the properties of plane waves in a little more detail. To do this, we will first go back to using our vector wave equation in the form of

$$\nabla \times \nabla \times \mathbf{E}(\mathbf{r}) - k^2 \mathbf{E}(\mathbf{r}) = 0. \tag{1.96}$$

Let's see what happens when we apply the curl operation to the simple plane wave solution we have been considering, i.e., $\mathbf{E}(\mathbf{r}) = \hat{x}\exp(\pm jkz)$. This is

$$\nabla \times \left(\hat{x}e^{\pm jkz} \right) = \begin{vmatrix} \hat{x} & \hat{y} & \hat{z} \\ \partial_x & \partial_y & \partial_z \\ e^{\pm jkz} & 0 & 0 \end{vmatrix} = \pm jk\hat{y}e^{\pm jkz}. \tag{1.97}$$

If we take the curl again, we arrive at

$$\nabla \times \left(\pm jk\hat{y}e^{\pm jkz} \right) = \begin{vmatrix} \hat{x} & \hat{y} & \hat{z} \\ \partial_x & \partial_y & \partial_z \\ 0 & \pm jke^{\pm jkz} & 0 \end{vmatrix} = k^2\hat{x}e^{\pm jkz}, \tag{1.98}$$
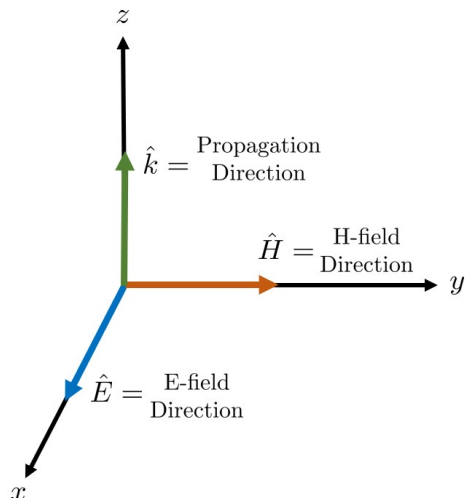
Figure 1.14: Illustration of how the propagation direction and polarization directions of **E** and **H** form a right-handed coordinate system for a plane wave. We can take cross-products between the unit vectors of the different components of this coordinate system to find the direction of the remaining component (e.g., $\hat{E} \times \hat{H} = \hat{k}$ or $\hat{k} \times \hat{E} = \hat{H}$).

which was the expected result from our previous study. So why did we do this exercise? The reason is that when we evaluate the derivatives in this way, we can notice an important "shortcut" when dealing with plane waves. To see this, we note that the result of the first curl operation can be viewed as

$$\nabla \times \left(\hat{x} e^{\pm jkz} \right) = -jk(\mp \hat{z}) \times \hat{x} e^{\pm jkz}, \tag{1.99}$$

while the second curl operation can be viewed as

$$\nabla \times \left(\pm jk\hat{y} e^{\pm jkz} \right) = (-jk)^2 (\mp \hat{z}) \times \left((\mp \hat{z}) \times \hat{x} \right) e^{\pm jkz}. \tag{1.100}$$

The trend is that the final direction of the curl operation applied to a plane wave simplifies down to taking the cross product between the direction of propagation and the polarization direction of the plane wave. This is multiplied by $-jk$ and the original plane wave function $\exp(\pm jkz)$ to get the complete result of the curl operation. We will revisit this concept in a more general setting when we learn more about plane waves later in this course. For now, the important takeaway is that due to this relationship, *a plane wave can only be polarized in a direction orthogonal to its propagation direction.* This is why we never considered solutions of the form $\hat{z} \exp(\pm jkz)$. Simply put, this is not a valid solution to the free-space wave equation.

This same kind of "shortcut" can also be used to help us quickly determine what the magnetic field associated with our plane wave will look like if we have already determined the behavior of the electric field. In particular, we can look at Faraday's law

$$\nabla \times \mathbf{E} = -j\omega\mu\mathbf{H}, \tag{1.101}$$

26

to see that

$$\mathbf{H} = -\frac{1}{j\omega\mu}\nabla \times \mathbf{E} = \frac{jk}{j\omega\mu}(\mp\hat{z}) \times \hat{x}e^{\pm jkz} = \mp\hat{y}\frac{1}{\eta}e^{\pm jkz}, \qquad (1.102)$$

where $\eta = \sqrt{\mu/\epsilon}$ is the *intrinsic impedance of the medium.* In free space, the intrinsic impedance is $\eta_0 = 120\pi \approx 377\Omega$. We see that the magnetic field is also pointing in a direction orthogonal to the propagation direction. Further, it is orthogonal to the direction the electric field is pointing in. Overall, we get a "right-handed" coordinate system relationship where we see that the cross product between the polarization direction of the electric and magnetic fields (in that order) gives us the propagation direction of the plane wave. This is illustrated in Fig. 1.14.

A wave that has both it's electric and magnetic fields orthogonal to the propagation direction is called a *transverse electromagnetic wave.* A plane wave is a typical example of this, but there are other special kinds of transmission lines that also can support the existence/propagation of transverse electromagnetic waves. We will learn about these more in the transmission lines portion of the course.

1.7   Boundary Conditions for Maxwell's Equations

In many ways, the integral and differential forms of Maxwell's equations can be viewed as being equivalent. However, in reality, the integral forms are a more "complete" or "fundamental" representation of Maxwell's equations. The reason for this is that if we only have the differential forms of Maxwell's equations we find some significant difficulty in determining how the fields and fluxes should behave at *boundaries* between regions of space with different electromagnetic properties (e.g., free space meeting a dielectric material, free space meeting a perfect conductor, etc.). The key to overcoming these difficulties is by specifying *boundary conditions* for the different quantities involved in Maxwell's equations. These boundary conditions tell us how the fields and fluxes should behave at any kind of boundary/interface between materials, allowing us to move forward with solving Maxwell's equations. As we will see shortly, these boundary conditions are actually "embedded" in the integral form of Maxwell's equations. That is, we can derive the boundary conditions from the integral form of Maxwell's equations. The fact that this information is already contained in the integral form of Maxwell's equations is why we can think of them as being a more "complete" representation of Maxwell's equations than the differential form.

1.7.1   Curl Equations

To begin deriving boundary conditions, let's consider the situation illustrated in Fig. 1.15. This setup consists of an interface between two materials that have distinct electromagnetic properties, denoted by $\epsilon_1 \neq \epsilon_2$ and $\mu_1 \neq \mu_2$. We will use Faraday's law with this picture to determine how the electric field should behave at this kind of interface. To do this, we will integrate over a small rectangular surface centered at the interface (as shown in Fig. 1.15) while the height of the rectangle becomes infinitesimally small. For this situation, Faraday's
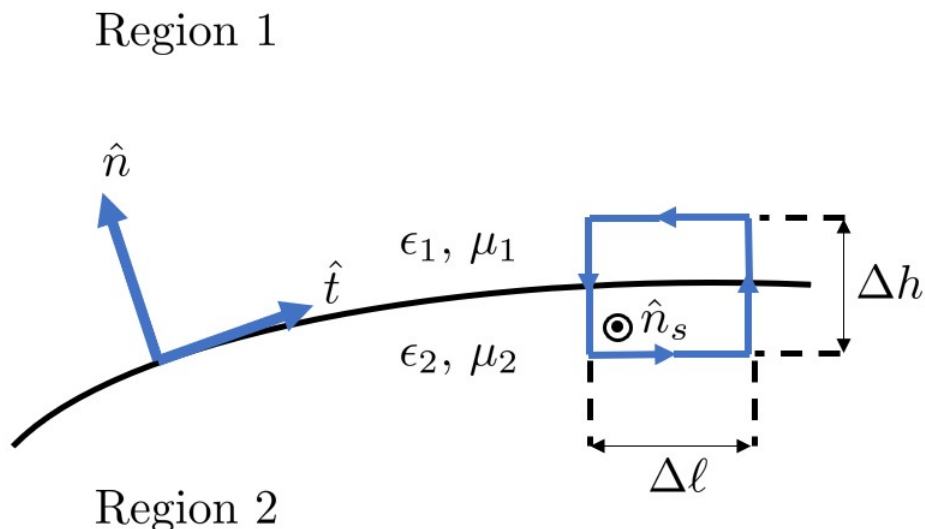
Figure 1.15: Illustration of regions to derive boundary conditions from Maxwell's curl equations.

law becomes

$$\lim_{\Delta h \to 0} \oint_C \mathbf{E} \cdot d\boldsymbol{\ell} = -\partial_t \lim_{\Delta h \to 0} \iint_S \mathbf{B} \cdot \hat{n}_s dS. \tag{1.103}$$

Due to the limit, the terms aligned with $\hat{n}$ will go to 0, and as a result we will not need to consider them. Likewise, the surface integral will evaluate to 0 in the limit because the surface area goes to 0. Hence, we simplify Faraday's law to be

$$\Delta \ell \hat{t} \cdot \left(\mathbf{E}_2 - \mathbf{E}_1 \right) = 0. \tag{1.104}$$

This is starting to look like a usable boundary condition. However, both $\Delta \ell$ and $\hat{t}$ (a unit vector tangential to the material interface) are not uniquely determined quantities. We would prefer to phrase our boundary condition using only uniquely specified quantities, so it would be good for us to work to rewrite (1.104).

Since $\Delta \ell$ is applied in equal proportion to both $\mathbf{E}_1$ and $\mathbf{E}_2$, we can simply "divide" it out of this equation. We can't simply remove $\hat{t}$ in a similar way because then we would lose the essential information that this boundary condition only applies to the components of $\mathbf{E}_1$ and $\mathbf{E}_2$ that are *tangential* to our interface.

There are various ways to mathematically specify that the boundary condition should only apply to the tangential components of $\mathbf{E}_1$ and $\mathbf{E}_2$. We will follow a somewhat roundabout derivation for a particular specification that is commonly used in electromagnetics. Although this is roundabout for deriving a boundary condition from Faraday's law, it is more useful in considering Ampere's law (which we will do shortly).

In particular, if we are clever we can see that $\hat{t} = \hat{n} \times (\hat{t} \times \hat{n})$. This allows us to use a few

vector identities to rewrite the left-hand side of (1.104) as

$$
\begin{aligned}
\hat{t} \cdot \left(\mathbf{E}_2 - \mathbf{E}_1\right) &= \left(\mathbf{E}_2 - \mathbf{E}_1\right) \cdot \hat{t} \\
&= \left(\mathbf{E}_2 - \mathbf{E}_1\right) \cdot \left[\hat{n} \times (\hat{t} \times \hat{n})\right] \\
&= -(\hat{t} \times \hat{n}) \cdot \left[\hat{n} \times \left(\mathbf{E}_2 - \mathbf{E}_1\right)\right].
\end{aligned}
\tag{1.105}
$$

In moving from line two to line three, we used the vector identity that

$$
\mathbf{a} \cdot (\mathbf{b} \times \mathbf{c}) = \mathbf{b} \cdot (\mathbf{c} \times \mathbf{a}) = \mathbf{c} \cdot (\mathbf{a} \times \mathbf{b}).
\tag{1.106}
$$

This is a very useful identity to memorize. It is easy to memorize it by recognizing that the result stays the same under "cyclic permutations" of the vectors in the identity.

Returning to (1.105), we see that we now have a useful result. We can omit the $(\hat{t} \times \hat{n})$ part safely from our boundary condition because the $\hat{n} \times$ that is being applied to $\mathbf{E}_1$ and $\mathbf{E}_2$ has the desired result of *projecting* the field vectors into the plane of the interface between the two regions. Further, the unit normal vector *is* uniquely determined by the geometry of the interface. Hence, our final boundary condition from Faraday's law becomes

$$
\boxed{\hat{n} \times \left(\mathbf{E}_1 - \mathbf{E}_2\right) = 0.}
\tag{1.107}
$$

In words, this boundary condition tells us that *the tangential components of the electric field are continuous across a boundary.*

We can of course use our constitutive relations to relate this boundary condition to the electric fluxes as well. This tells us that

$$
\hat{n} \times \epsilon_1^{-1}\mathbf{D}_1 = \hat{n} \times \epsilon_2^{-1}\mathbf{D}_2.
\tag{1.108}
$$

That is, the tangential components of the electric flux is *discontinuous* across a boundary if the two materials have different permittivities. This is again another helpful reminder to us that the electric fields and fluxes are fundamentally different quantities.

We can now turn our attention to Ampere's law. We will consider the situation shown in Fig. 1.16, and will explain some of the changes from Fig. 1.15 shortly. For this situation, Ampere's law becomes

$$
\lim_{\Delta h \to 0} \oint_C \mathbf{H} \cdot d\boldsymbol{\ell} = \lim_{\Delta h \to 0} \iint_S \left(\partial_t \mathbf{D} + \mathbf{J}\right) \cdot \hat{n}_s dS.
\tag{1.109}
$$

Using similar arguments to what we did for Faraday's law, we can reason that the normal components of $\mathbf{H}$ and the surface integral of $\partial_t \mathbf{D}$ aren't going to be important in this limit.

It is tempting to also use this logic to suggest that the surface integral of $\mathbf{J}$ is also equal to 0 in this limit. However, there are many situations where it is convenient to consider the concept of a *surface current density.* This is a form of current that can flow in two dimensions, but is confined to an infinitesimally thin sheet. Although surface currents are always an approximation to an actual current density that can physically exist, they can be very useful in simplifying the mathematical analysis of common problems. For instance, this
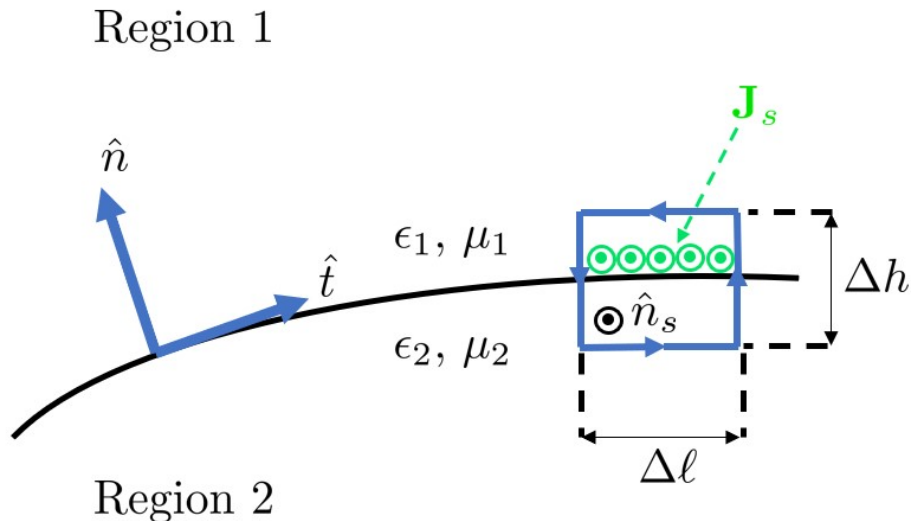
Figure 1.16: Illustration of regions to derive boundary conditions from Ampere's law when a surface current density is present.

kind of surface current can be used when analyzing conductors with very high conductivity (which is the case in many practical scenarios), and can also be a useful mathematical tool in modeling certain kinds of antennas. In general, a surface current has units of [A/m]. We will often denote a surface current as $\mathbf{J}_s$ in this set of notes. However, it is very common to simply use the symbol $\mathbf{J}$ for any kind of current density when the type of current density (e.g., volume or surface currents) can be inferred from the context.

Now, we can evaluate (1.109) assuming that there is a surface current density present. The integral evaluates to

$$\Delta\ell\hat{t} \cdot \left(\mathbf{H}_2 - \mathbf{H}_1\right) = \Delta\ell(\hat{t} \times \hat{n}) \cdot \mathbf{J}_s. \tag{1.110}$$

Note that we have used $\hat{n}_s = (\hat{t} \times \hat{n})$ on the right-hand side to act as the unit normal vector to the surface we integrated over in Fig. 1.16. Dropping common terms and recalling the vector identities we used to simplify the boundary condition from Faraday's law, we arrive at

$$(\hat{t} \times \hat{n}) \cdot \left[\hat{n} \times \left(\mathbf{H}_1 - \mathbf{H}_2\right)\right] = (\hat{t} \times \hat{n}) \cdot \mathbf{J}_s. \tag{1.111}$$

We can again omit the common vector terms $(\hat{t} \times \hat{n})$ since they no longer contribute any meaningful information to the boundary condition. As a result, the final boundary condition becomes

$$\boxed{\hat{n} \times \left(\mathbf{H}_1 - \mathbf{H}_2\right) = \mathbf{J}_s.} \tag{1.112}$$

This tells us that the tangential components of the magnetic field can be discontinuous at an interface between two regions *only if* there is a surface current density flowing at the interface. As alluded to previously, this will occur at the interface with a *perfect electric*
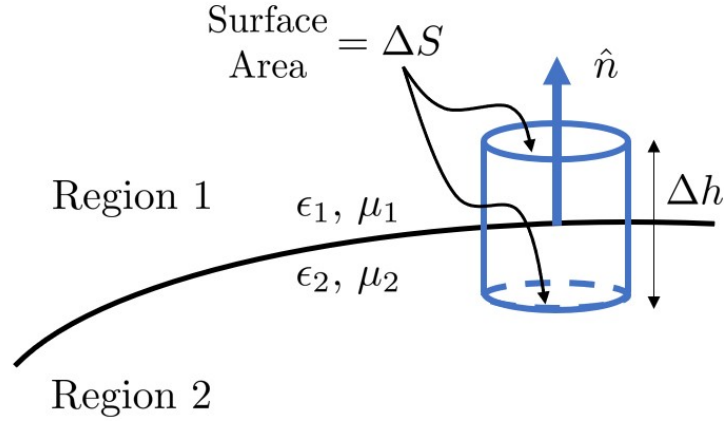
Figure 1.17: Illustration of regions to derive boundary conditions from Maxwell's divergence equations.

conductor (an approximation that assumes the conductivity $\sigma \approx \infty$). However, for many cases there will not be a $\mathbf{J}_s$ and we can usually conclude that in similarity to the electric field, the tangential components of the magnetic field will be continuous across an interface. We can also use the constitutive relations to write this boundary condition in terms of the magnetic flux. The interested reader can do this themselves and see that a very similar conclusion to that from Faraday's law can be found here as well.

1.7.2 Divergence Equations

We have so far determined what happens to the tangential components of the fields and fluxes at interfaces between two distinct regions. Although this is good progress, we are not finished yet. In particular, we need to also determine what happens to the normal components of the electric and magnetic fields/fluxes at an interface. We can do this by considering the other two Maxwell's equations. The illustration of our integration regions that will be appropriate for these equations are shown in Fig. 1.17.

Let's start with Gauss' law of magnetism. For this situation, we will perform our integrations with the limit of Δh going to 0. This gives us

$$\lim_{\Delta h \to 0} \oiint_S \mathbf{B} \cdot d\boldsymbol{S} = 0. \tag{1.113}$$

We can see that we have 3 distinct surfaces to consider in Fig. 1.17. The top and bottom disks of our cylinder will be able to retain non-zero values, but the limit of Δh going to 0 will cause the integration around the rest of the cylinder to go to zero due to the shrinking surface area. Hence, our result becomes

$$\Delta S\, \hat{n} \cdot \left(\mathbf{B}_1 - \mathbf{B}_2\right) = 0. \tag{1.114}$$

We can remove the ΔS for our unique description of the boundary condition. As a result, we end up with

$$\boxed{\hat{n} \cdot \left(\mathbf{B}_1 - \mathbf{B}_2\right) = 0.} \tag{1.115}$$

This tells us that the normal component of the magnetic flux density is continuous across a boundary. We can use the constitutive relations to easily see that this means that the normal component of the magnetic *field* will not be continuous across the material interface (if the materials have different permeabilities).

We can follow a similar process now for Gauss' law of electricity. This becomes

$$\lim_{\Delta h \to 0} \oiint_S \mathbf{D} \cdot d\mathbf{S} = \lim_{\Delta t \to 0} \iiint \rho dV. \tag{1.116}$$

The left-hand side will evaluate in the same way as for Gauss' law of magnetism. We may be tempted to suggest that the right-hand side is equal to 0. However, in the same way that we needed to consider surface current densities for Ampere's law, we also need to consider *surface charge densities.* This is essentially the same concept as a surface current density, but is now applied to charge densities. In particular, a surface charge density, denoted as ρ_s with units $[\mathrm{C/m^2}]$, is a set of charges that are confined to an infinitesimally thin region along the boundary between the materials.

Combining these concepts, we have that

$$\boxed{\hat{n} \cdot \left(\mathbf{D}_1 - \mathbf{D}_2\right) = \rho_s.} \tag{1.117}$$

This tells us that the normal component of the electric flux density can be discontinuous across a material interface if there is a surface charge density there. Similarly, we can see that if there is no surface charge density then the normal component of the electric flux density will be continuous across the interface. Finally, if the regions have different permittivities, we can conclude that the normal component of the electric *field* will be discontinuous across the interface even if there is no surface charge density there.

1.7.3 Summary

The boundary conditions are summarized below.

<div style="border:1px solid">

Boundary Conditions

$$\hat{n} \times \left(\mathbf{E}_1 - \mathbf{E}_2\right) = 0 \tag{1.118}$$
$$\hat{n} \times \left(\mathbf{H}_1 - \mathbf{H}_2\right) = \mathbf{J}_s \tag{1.119}$$
$$\hat{n} \cdot \left(\mathbf{D}_1 - \mathbf{D}_2\right) = \rho_s \tag{1.120}$$
$$\hat{n} \cdot \left(\mathbf{B}_1 - \mathbf{B}_2\right) = 0 \tag{1.121}$$

</div>

It should be emphasized that in this class, $\mathbf{J}_s$ and ρ_s will only be non-zero at the interface between a homogeneous medium and a perfect conductor or will be explicitly specified as part of the problem statement (e.g., in analyzing a simple "antenna" in a few lectures).

1.8 Material Properties

Up to this point, we have been fairly unconcerned about whether we are talking about free space or a homogeneous medium. We saw a few lectures ago that the only real difference was

that we adjusted our permittivity and permeability to account for the fact that we were in a material as opposed to being in free space. However, from our other physics courses we know that real materials are made up of a huge collection of atoms/molecules arranged in very special fashions. This raises the question of whether or not our simple picture of accounting for the effects of a material with a modified permittivity and permeability is appropriate. As we will see, this is reasonable for many situations that we are interested in for practical applications. However, there are situations where a more detailed description of the features of a material are needed.

One of the driving concepts behind this simplification can be captured in the name *macroscopic electromagnetics*. This is a framework for studying electromagnetic systems where we can make simplifications such as treating materials through a shift in the permittivity and permeability. The reason this is acceptable is that the length scale that the atomic/molecular physics operates over is generally incredibly small compared to the wavelengths of electromagnetic fields. As a result, we can take a "macroscopic" viewpoint where we don't need to consider what is happening to each of the individual atoms/molecules. If we were to try and consider how each individual atom/molecule is behaving, we would be taking a "microscopic" viewpoint of the physics. For certain situations, we need to attempt to use some kind of in-between picture of the physics. This is called the "mesoscopic" regime. Each of these different viewpoints have important applications in practical engineering. However, in this course, we will mainly focus on a macroscopic viewpoint.

With this in mind, we will now consider how materials respond to electric and magnetic fields to better rationalize the macroscopic electromagnetism viewpoint that we have been using somewhat freely up to this point. We will start by considering the *polarization* of a dielectric material followed by the *magnetization* of a magnetic material.

1.8.1   Polarization Intensity

We will first consider what happens to the electric charges in a medium where an electric field is applied, and correspondingly, how these changes to the electric charges affect the electromagnetic fields and fluxes. To simplify this, we will consider the material to be made up of generic atoms/molecules that are composed of a positively charged nucleus surrounded by negatively charged electrons. Further, we assume that the number of positive and negative charges over the range of the entire material are perfectly balanced; i.e., the material is *neutrally charged*. We will also assume that the strength of electric fields we are considering are not powerful enough to "rip" the electrons away from the nucleus (i.e., we *do not ionize the material*).

In the absence of an electric field, we can assume that all of the atoms/molecules are essentially *randomly oriented* so that even if they are polar molecules there is no net polarity to the material (see Fig. 1.18). However, this changes if the material is placed in a *uniform electric field*. In this situation, the electric field can exert a force on the charges that will "stretch" the nuclei and electrons away from their equilibrium positions and cause them to be oriented in a regular arrangement (see Fig. 1.18). As a result, we can view each of these atoms/molecules as being a *small electric dipole*. We can consider these tiny dipoles through

In Absence of
Electric Field

In Presence of
(Uniform) Electric Field

E
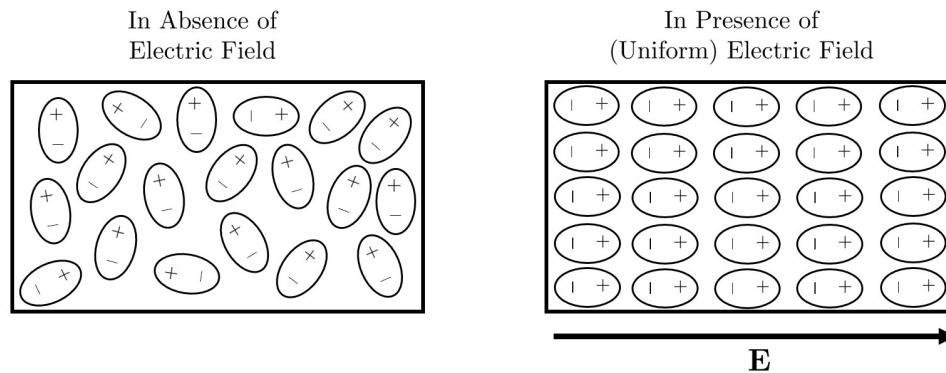
Figure 1.18: (Left) In the absence of an applied electric field, the "dipoles" that make up the material are oriented randomly. (Right) When an external electric field is applied, the "dipoles" are stretched and aligned into a regular pattern.

their *dipole moments*, defined as

$$\mathbf{p} = q\boldsymbol{\ell}, \tag{1.122}$$

where q is the charge of the dipole and $\boldsymbol{\ell}$ is a vector pointing from the center of mass of the negative charges to the center of mass of the positive charges.

We can describe the overall effect of these many tiny dipoles through the *polarization vector/intensity*, defined as

$$\mathbf{P} = \lim_{\Delta v \to 0} \sum_i \frac{\mathbf{p}_i}{\Delta v}. \tag{1.123}$$

How can we add this into our description of Maxwell's equations? The key is recognizing that the divergence of $\mathbf{P}$ is going to be able to be considered to be a new kind of charge density (look at the units). In particular, we have a *bound charge density* (because it is bound to the material we are considering) that is given as

$$\rho_b = -\nabla \cdot \mathbf{P}. \tag{1.124}$$

The negative sign is needed because the divergence computes an *outward* flux, while we are actually more concerned with what is happening "inside" the region we are considering.

Now that we have associated the polarization intensity with a charge density, we can rewrite Gauss' law of electricity within a material as

$$\begin{aligned} \nabla \cdot (\epsilon_0 \mathbf{E}) &= \rho + \rho_b \\ &= \rho - \nabla \cdot \mathbf{P}. \end{aligned} \tag{1.125}$$

We can move the polarization intensity term to the left-hand side and consolidate terms under the divergence symbol to be

$$\nabla \cdot \left(\epsilon_0 \mathbf{E} + \mathbf{P} \right) = \rho. \tag{1.126}$$

Comparing this to what we originally stated Gauss' law to be, we can see that the effect of the material can be elegantly incorporated into Maxwell's equations if we say that

$$\boxed{\mathbf{D} = \epsilon_0 \mathbf{E} + \mathbf{P}.} \tag{1.127}$$

In many cases, working directly with the polarization intensity is inconvenient. Fortunately, it is often the case that the polarization intensity that occurs in a material can be easily related back to the applied electric field. In particular, we typically have in the frequency domain that

$$\mathbf{P} = \epsilon_0 \chi_e \mathbf{E}, \tag{1.128}$$

where χ_e is the *electric susceptibility*. Each material will have a different χ_e that characterizes the unique properties of that particular material. We will consider how to derive χ_e from a simple model appropriate for many types of materials a few lectures from now. As mentioned previously, this type of relationship is commonly adequate. Important cases where it is not acceptable include for nonlinear materials, which require a more sophisticated description.

If we are performing our analysis in the frequency domain, the electric susceptibility will typically be a complex number. In this situation, it is represented as

$$\chi_e = \chi_e' - j\chi_e'', \tag{1.129}$$

where χ_e' is the real part and χ_e'' is the imaginary part of the susceptibility. The imaginary part is necessary to describe *lossy materials*, i.e., materials which absorb some of the energy from the electric field. In all physical materials, there will be some loss. However, there are plenty of materials for which the loss is small enough that we can ignore it for quick calculations. We can typically associate this loss of energy with a kind of frictional force that occurs due to collisions between nearby atoms/molecules inside the material as they are being stretched and "flipped" (for time-varying fields). It turns out that water molecules have very high loss at microwave frequencies. This is why your microwave oven works so well for heating up food, since many foods that we eat have a high amount of water content.

Moving on, we can substitute our new representation for the polarization intensity into (1.127) to get

$$\begin{aligned} \mathbf{D} &= \epsilon_0 \mathbf{E} + \epsilon_0 \chi_e \mathbf{E} \\ &= \epsilon_0 (1 + \chi_e) \mathbf{E}. \end{aligned} \tag{1.130}$$

By defining the *relative permittivity* as

$$\epsilon_r = \epsilon_r' - j\epsilon_r'' = (1 + \chi_e), \tag{1.131}$$

we arrive back at our simple constitutive relationship of

$$\mathbf{D} = \epsilon_0 \epsilon_r \mathbf{E} = \epsilon \mathbf{E}. \tag{1.132}$$

Our interpretation of the imaginary part of the material property corresponding to loss also applies to the relative permittivity. Note that within the context of a lossy dielectric, it is common to refer to ϵ_r' as the *dielectric constant* of the material.
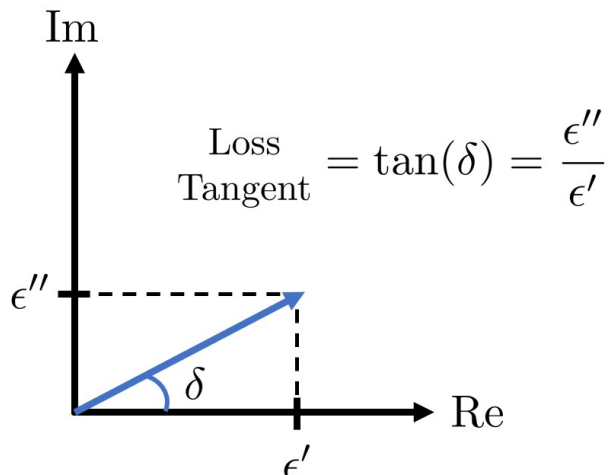
Figure 1.19: Illustration of the origin of the name "loss tangent".

Now, one common way to specify the loss associated with a material is through its *loss tangent*. This is typically written as $\tan(\delta) = \epsilon''/\epsilon' = \epsilon''_r/\epsilon'_r$, and is illustrated in Fig. 1.19. This quantity is particularly useful for determining whether a material corresponds to a low-loss dielectric or not. Generally, we want $\tan(\delta) \ll 1$ to classify a material as being low-loss. When this is the case, certain approximate formulas can be used to analyze the physics of the situation, as we will see shortly.

As promised, we see from (1.132) that so long as this macroscopic electromagnetics perspective works we can indeed account for a material medium by simply changing the permittivity we use in our equations. This type of perspective can break down in a few different ways. One manner is when the frequencies of our electromagnetic waves become very high. In these situations, the small wavelengths of the waves can more directly interact with the atoms/molecules (including their internal structure). At this point, we typically need to resort to some kind of quantum mechanical description of the atoms/molecules to begin to understand phenomena like spontaneous and stimulated emission. Another situation where our macroscopic viewpoint can break down is if the material region we are considering becomes extremely thin or extremely small. Again, we start to run into quantum mechanical effects becoming more pronounced in these extremely small volumes. We will not study these cases in this course, however, we will mention that understanding and engineering these kinds of fundamental quantum interactions is a very active and vibrant research area at the heart of achieving things like quantum computation. In most physical implementations being pursued to date, electromagnetic devices/physics play a crucial role in controlling and interacting with the quantum systems being used as quantum bits (qubits).

1.8.2 Magnetization Intensity

The effect that a magnetic material has on the applied magnetic field can be captured in a similar manner to the polarization intensity. In this situation, we consider the material to have many tiny magnetic dipole moments that can be viewed as being due to the orbiting of electrons, as well as their spin orientation in certain cases. These details are beyond the
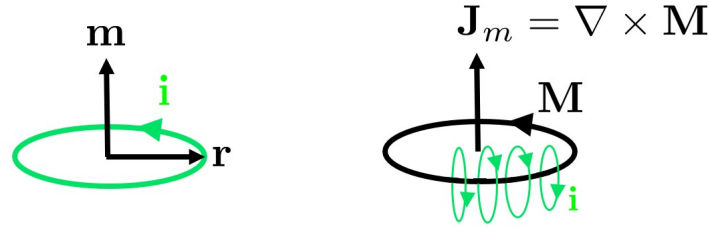
Figure 1.20: (Left) Illustration of a single magnetic dipole moment arising from a small current loop. (Right) Computing the circulation of the magnetization intensity is related to how much the individual current loops of the magnetic dipoles align in a particular direction.

scope of this course, so we will simply say that the material has many tiny dipole moments that are due to small current loops flowing inside the material. The magnetic dipole moment for these currents can be given by $\mathbf{m} = \mathbf{r} \times \mathbf{i}$, where $\mathbf{r}$ points from the origin of the current loop to the loop itself and $\mathbf{i}$ is the vector characterizing the direction of current flow, as shown in Fig. 1.20.

We will define the *magnetization intensity/vector* as

$$\mathbf{M} = \lim_{\Delta v \to 0} \sum_i \frac{\mathbf{m}_i}{\Delta v}, \tag{1.133}$$

where $\mathbf{m}_i$ is the magnetic dipole moment of the ith current loop. We will associate with this magnetization a new volume current density given by

$$\mathbf{J}_m = \nabla \times \mathbf{M}. \tag{1.134}$$

We can see that this leads to a sensible volume current density by considering how the circulation of the magnetization intensity relates to the small current loops that make up the magnetic dipoles, as shown in Fig. 1.20.

Now, this volume current density can be incorporated into Ampere's law as

$$\begin{aligned} \nabla \times \mu_0^{-1}\mathbf{B} &= j\omega\mathbf{D} + \mathbf{J} + \mathbf{J}_m \\ &= j\omega\mathbf{D} + \mathbf{J} + \nabla \times \mathbf{M}. \end{aligned} \tag{1.135}$$

We can move the magnetization to the left-hand side and consolidate terms underneath the curl operation to have

$$\nabla \times \left(\mu_0^{-1}\mathbf{B} - \mathbf{M}\right) = j\omega\mathbf{D} + \mathbf{J}. \tag{1.136}$$

Comparing this to our original form of Maxwell's equations we can recognize that we should adjust the definition of $\mathbf{H}$ to be

$$\mathbf{H} = \mu_0^{-1}\mathbf{B} - \mathbf{M}. \tag{1.137}$$

Alternatively, we can write this as

$$\boxed{\mathbf{B} = \mu_0\left(\mathbf{H} + \mathbf{M}\right).} \tag{1.138}$$

Similar to the case of electrical response, we can define a magnetic susceptibility to relate the magnetization more directly to the magnetic field. In particular, we have in the frequency domain that

$$\mathbf{M} = \chi_m \mathbf{H}, \tag{1.139}$$

where

$$\chi_m = \chi'_m - j\chi''_m. \tag{1.140}$$

Similarly, we can define a relative permeability as

$$\mu_r = \mu'_r - j\mu''_r = 1 + \chi_m. \tag{1.141}$$

This allows us to write

$$\mathbf{B} = \mu_0 \mu_r \mathbf{H} = \mu \mathbf{H}, \tag{1.142}$$

as we had been previously doing in this course. A loss tangent can also be written for μ_r similar to what was done with ϵ_r. This is not as commonly done for magnetic materials, since most magnetic materials are significantly lossy at microwave frequencies. As a result, tracking the dependence on other physical quantities in terms of a magnetic loss tangent isn't as useful as it is with an electric loss tangent.

Now, why were we careful to point out at various times in this derivation (and for the polarization) that we are working in the *frequency domain*? The reason is that for any real material the permittivity and permeability will be functions of frequency. This can be viewed as a consequence of *inertia*, i.e., the various atoms/molecules/currents that make up the dipole moments can't respond immediately to changes in the applied fields. This delay in motion will cause the response of the materials to change as a function of frequency. From your linear systems courses you should recall that if we have two quantities multiplied in the frequency domain that corresponds to a *convolution in the time domain*. Hence, all of our constitutive relations in the presence of a material become convolutions in the time domain if the material properties vary as a function of frequency. There are certainly times where we can approximate the frequency variation of the material as being unimportant so that the convolutions can be safely ignored. However, in general, this is not always going to be the case. This is another reason why performing analysis in the frequency domain can be significantly simpler than working directly in the time domain.

1.9   Plane Waves in Lossy Media

Previously, we discussed how materials could be incorporated into Maxwell's equations and briefly commented on how they were mathematically described when they were lossy. We will now see how a lossy material impacts our solutions to the wave equation in a homogeneous medium. To be general, we will assume that the material has *dielectric, magnetic, and conductive losses*. Since we are looking for plane wave solutions, we will also consider a source-free region of space. Despite this, there can still be *induced current* flowing inside of the material due to the conductive losses. We will denote this conduction current as $\mathbf{J}_c$.

Now, for this situation, Ampere's law becomes

$$\nabla \times \mathbf{H} = j\omega \mathbf{D} + \mathbf{J}_c$$
$$= j\omega\epsilon\mathbf{E} + \sigma\mathbf{E}, \tag{1.143}$$

where we have used the constitutive relation and Ohm's law to make all the terms on the right-hand side proportional to the electric field. Following our standard approach for deriving the wave equation, we next look at Faraday's law, which for this situation is

$$\nabla \times \mathbf{E} = -j\omega\mu\mathbf{H}. \tag{1.144}$$

Remember that both ϵ and μ are complex-valued quantities to account for the dielectric and magnetic losses, respectively.

Now, we take the curl of Faraday's law and substitute in from Ampere's law to arrive at

$$\nabla \times \nabla \times \mathbf{E} + j\omega\mu(\sigma + j\omega\epsilon)\mathbf{E} = 0. \tag{1.145}$$

If we define

$$\gamma = \sqrt{j\omega\mu(\sigma + j\omega\epsilon)}, \tag{1.146}$$

then we can write this wave equation as

$$\boxed{\nabla \times \nabla \times \mathbf{E} + \gamma^2\mathbf{E} = 0.} \tag{1.147}$$

We typically call γ the *propagation constant*, since it characterizes many details about how the wave will propagate in a particular medium.

It should be noted that in many cases, it is common to "lump" the dielectric and conductive losses together into a single complex-valued ϵ. We can see how to do this by returning to Ampere's law and writing it as

$$\nabla \times \mathbf{H} = j\omega\epsilon\mathbf{E} + \mathbf{J}_c$$
$$= j\omega(\epsilon' - j\epsilon'')\mathbf{E} + \sigma\mathbf{E}$$
$$= j\omega\left[\epsilon' - j\left(\epsilon'' + \frac{\sigma}{\omega}\right)\right]\mathbf{E}. \tag{1.148}$$

If we consider $\epsilon'' + \sigma/\omega$ to be the imaginary part of the permittivity, we find that our expressions for γ will reduce to

$$\gamma = j\omega\sqrt{\mu\epsilon}, \tag{1.149}$$

where both μ and ϵ can be complex-valued. *If a conductivity is not explicitly given or referred to in a problem, you should typically use (1.149) in your calculations rather than (1.146).*

Now, returning to (1.147), we can use our vector identity in a source-free region of space to rewrite this wave equation as a vector Helmholtz wave equation. This is

$$\boxed{\nabla^2\mathbf{E} - \gamma^2\mathbf{E} = 0.} \tag{1.150}$$

From similarity to our previous equations, we see that an identical form of plane wave solution is going to be appropriate here. That is, if $\mathbf{E} = \hat{x}E_x(z)$, we have that

$$E_x(z) = E^+ e^{-\gamma z} + E^- e^{\gamma z}. \tag{1.151}$$

It is generally useful to break γ into its real and imaginary parts as

$$\gamma = \alpha + j\beta, \tag{1.152}$$

where α is the called the *attenuation constant* and β is called the *phase constant*. The reasons for this should hopefully be clear. Since α is real, it will correspond to exponential decay, while the $j\beta$ term will correspond to a varying phase as a function of position. It is important to note that because α could accidentally portray exponential growth, we must be careful with our signs and specify for which regions of a problem our mathematical representations are valid.

The SI units for α are neper per meter (Np/m) and the SI units for β are radians per meter. A neper is a dimensionless quantity that is defined so that if $\alpha = 1\,[\mathrm{Np/m}]$ a unit wave amplitude will have decreased to a magnitude of $1/e$ after traveling a distance of 1 meter. We often work with decibel units in electromagnetic engineering. In these cases, the α is often reported in terms of dB/m. We can convert between units by noting that $1\,[\mathrm{Np/m}] = 20\log_{10} e = 8.686\,[\mathrm{dB/m}]$. *You should never plug an attenuation constant expressed in* dB/m *into an exponential function when computing aspects of a solution, this will lead to the wrong answer!*

It should also be noted that in a lossy medium we will need to use β to compute quantities like the phase velocity or wavelength of our plane waves. This is because the speed of light naively defined as $c = 1/\sqrt{\mu\epsilon}$ will become a complex number and lose its useful meaning. Instead, we will need to compute the phase velocity as

$$v_p = \frac{\omega}{\beta} \tag{1.153}$$

and the wavelength as

$$\lambda = \frac{2\pi}{\beta}. \tag{1.154}$$

These expressions will generally work in most situations, and so are much safer to use than something like $\lambda = c/f$, which only holds in lossless and unbounded homogeneous regions of space (which is a very specific situation!).

Along with the modified propagation constant, we also need to modify our intrinsic impedance. For this kind of lossy material, this will become

$$\eta = \sqrt{\frac{j\omega\mu}{\sigma + j\omega\epsilon}}, \tag{1.155}$$

or simply

$$\eta = \sqrt{\frac{\mu}{\epsilon}} \tag{1.156}$$

if any conductive losses have been "lumped" into the definition of ϵ. *Again, if a conductivity is not explicitly given or referred to in a problem, you should typically use (1.156) in your calculations rather than (1.155).*

Before moving on, let's look at how to find the instantaneous expression for a plane wave in a lossy medium. To do this, we will remember that we can simply multiply our phasors by $\exp[j\omega t]$ and then take the real part. Doing this for the wave that propagates in the $+z$-direction from (1.151), we get

$$
\begin{aligned}
E^+ e^{-\gamma z} \implies & \mathrm{Re}\left\{ E^+ e^{-\gamma z} e^{j\omega t} \right\} \\
= & \mathrm{Re}\left\{ E^+ e^{-\alpha z} e^{-j\beta z} e^{j\omega t} \right\} \\
= & e^{-\alpha z}\mathrm{Re}\left\{ E^+ e^{-j\beta z} e^{j\omega t} \right\} \\
= & |E^+| e^{-\alpha z} \cos\left(\omega t - \beta z + \angle E^+ \right).
\end{aligned}
\tag{1.157}
$$

It is also important to remember that when computing the electric or magnetic fields from one another in a lossy media that these operations should be done with phasors and then converted into instantaneous expressions from there. The issue is that you cannot simply divide by something like η in the instantaneous expressions, since in the lossy media η is a complex number which cannot be used directly in the time domain.

1.9.1 Useful Approximations

We can now specialize our new formulas to a few cases of particular interest. It is important to note that these expressions are only approximations valid for limited ranges of material parameters. *It is always acceptable to simply use the full formulas, rather than approximate ones in working with lossy dielectrics.*

The first special case of interest will be for a *good dielectric*, which is characterized by $\omega\epsilon \gg \sigma$. That is, the conductive losses are very small. Approximate formulas appropriate to this regime can be found in Section 8.3-1 of your textbook. One approximate formula that is of particular interest is that for γ. This is

$$
\gamma = \alpha + j\beta \approx j\omega\sqrt{\mu\epsilon'}\left[1 - j\frac{1}{2}\frac{\epsilon''}{\epsilon'} + \frac{1}{8}\left(\frac{\epsilon''}{\epsilon'}\right)^2 \right].
\tag{1.158}
$$

This formula highlights how the loss tangent (recall this is given as ϵ''/ϵ') becomes a meaningful parameter in an approximate expansion of the propagation characteristics for good dielectrics.

Another useful case to consider is that of a *good conductor*. Approximate formulas appropriate to this regime can be found in Section 8.3-2 of your textbook. For this material, we will have that $\sigma \gg \omega\epsilon$. We can then simplify our propagation constant to be

$$
\gamma \approx \sqrt{j\omega\mu\sigma}.
\tag{1.159}
$$

Similarly, the intrinsic impedance becomes

$$\eta \approx \sqrt{j\omega\mu/\sigma}. \tag{1.160}$$

For a good conductor, it is useful to consider the concept of *skin depth*. The skin depth is defined as the distance into a material where the field decays to $1/e$ of its original value. For a good conductor, we have that

$$\alpha = \sqrt{\frac{\omega\mu\sigma}{2}}. \tag{1.161}$$

Hence, our skin depth (denoted as δ), will be

$$\delta = \frac{1}{\alpha} = \sqrt{\frac{2}{\omega\mu\sigma}}. \tag{1.162}$$

Since we are dealing with exponential decay, it usually only takes a handful of skin depths for a field to effectively decay to 0. At that point, the field will not be able to "push" charges around anymore to cause currents to flow. Hence, currents will typically only flow within a small number of skin depths of the surface of a good conductor.

To get a better feel for this, let's see how thick the skin depth is for a commonly used material in circuit board and antenna design. At the frequency of WiFi signals (2.4 GHz), the skin depth is approximately 1.33 μm for copper cladding. Many circuit board materials use very thin thicknesses of copper to define their traces (which can also be used to make antennas, e.g., in your phone). However, they still are typically on the order of $30 - 40$ μm thick, so we can conclude that for the most part all of the current is effectively flowing on the *surfaces of the conductors*. This is part of the reason why surface current densities are such a useful concept. It is much easier to solve problems analytically (and numerically) if we consider only an infinitesimal layer of current as opposed to rigorously accounting for the decay profile of the current in the conductor. However, there are situations where this is no longer adequate. For instance, when the sizes of the conductors start becoming comparable to the skin depth. This can become important for the design of modern computer chips because of the huge amount of miniaturization that has happened over the years to continue improving chip performance.

1.10 Poynting's Theorem

The flow of energy and power throughout a system is important in many areas of electrical engineering, including in electromagnetics. You are no doubt familiar with the concept of the *conservation of energy*. In the area of electromagnetics, the conservation of energy is captured in *Poynting's theorem*. We will now take some time to consider what this looks like in the time and frequency domains.

1.10.1 Time Domain

To begin, we will write out Faraday's and Ampere's laws for a medium characterized by μ, ϵ, and σ. In addition to this, we will also assume that there is an *impressed current* present,
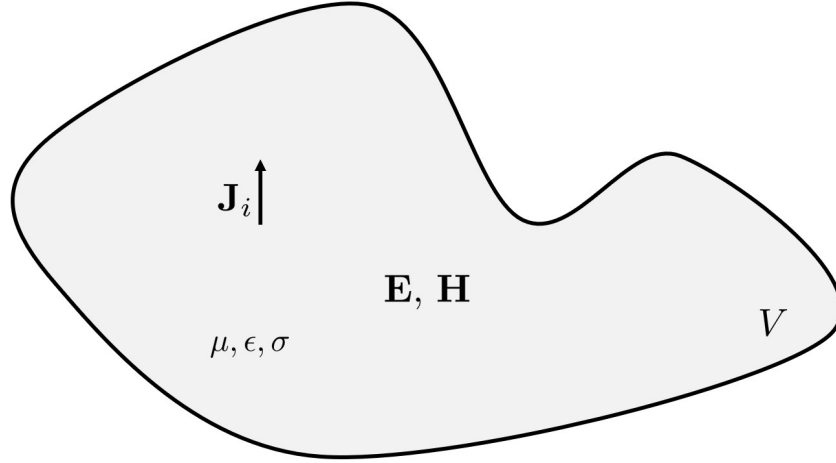
Figure 1.21: Illustration of the region of space considered for deriving Poynting's theorem.

as depicted in Fig. 1.21. An impressed current can be viewed as a source of electromagnetic fields that is "forced" onto the system. The important aspect of this is that we don't allow the electromagnetic fields to have any meaningful "back action" on the current density itself. That is, the current density does not change due to the presence of the produced electromagnetic fields. This is, in general, not a completely accurate view of many physical phenomena in their entirety. However, it does provide us with a very useful mathematical tool to go about solving for the accurate description of the physics. This is much like the ideal voltage or current sources that you used in circuit analysis. For instance, no real current source can drive an exact current value through a circuit regardless of the composition of the circuit. However, this is still a very useful model that allows us to neglect the complicated composition of the source for the many situations where the practical source is operated in a way where it behaves close to ideal.

Now, our equations are

$$\nabla \times \mathbf{E} = -\mu \partial_t \mathbf{H} \tag{1.163}$$

$$\nabla \times \mathbf{H} = \epsilon \partial_t \mathbf{E} + \sigma \mathbf{E} + \mathbf{J}_i. \tag{1.164}$$

We are interested in looking at energy and power. To start getting to quantities with the correct kinds of units for this to be the case, we can take the dot product of (1.163) with $\mathbf{H}$. We can see this is a reasonable idea by remembering that the magnetic energy stored in an inductor is given by $\frac{1}{2}LI_0^2$. Performing the dot product with (1.163) gives

$$\mathbf{H} \cdot \nabla \times \mathbf{E} = -\mu \mathbf{H} \cdot \partial_t \mathbf{H}, \tag{1.165}$$

from which we can see that the right-hand side is starting to have units related to an energy density. To make this more explicit, we can rewrite the right-hand side by noting that

$$\mu \mathbf{H} \cdot \partial_t \mathbf{H} = \frac{1}{2}\mu \partial_t (\mathbf{H} \cdot \mathbf{H}) = \partial_t \left(\frac{1}{2}\mu \mathbf{H} \cdot \mathbf{H} \right). \tag{1.166}$$

We define

$$\boxed{w_m = \frac{1}{2}\mu\mathbf{H}\cdot\mathbf{H}} \tag{1.167}$$

as the *magnetic energy density*. This has units of $\text{H/m}\cdot(\text{A/m})^2 = \text{J/m}^3$, so it is in fact a volumetric energy density. Using this energy density, we can write (1.165) as

$$\mathbf{H}\cdot\nabla\times\mathbf{E} = -\partial_t w_m. \tag{1.168}$$

We can follow a similar process for (1.164) by taking the dot product of it with $\mathbf{E}$. This gives us

$$\mathbf{E}\cdot\nabla\times\mathbf{H} = \epsilon\mathbf{E}\cdot\partial_t\mathbf{E} + \sigma\mathbf{E}\cdot\mathbf{E} + \mathbf{E}\cdot\mathbf{J}_i. \tag{1.169}$$

We can use a similar trick with the first term on the right-hand side to define an *electric energy density* as

$$\boxed{w_e = \frac{1}{2}\epsilon\mathbf{E}\cdot\mathbf{E}.} \tag{1.170}$$

Using this, we can rewrite (1.169) as

$$\mathbf{E}\cdot\nabla\times\mathbf{H} = \partial_t w_e + \sigma\mathbf{E}\cdot\mathbf{E} + \mathbf{E}\cdot\mathbf{J}_i. \tag{1.171}$$

We would like to combine (1.168) and (1.171) into a single equation to describe energy and power. To do this, we will subtract (1.171) from (1.168) to get

$$\mathbf{H}\cdot\nabla\times\mathbf{E} - \mathbf{E}\cdot\nabla\times\mathbf{H} = -\partial_t w_m - \partial_t w_e - \sigma\mathbf{E}\cdot\mathbf{E} - \mathbf{E}\cdot\mathbf{J}_i. \tag{1.172}$$

We can simplify the left-hand side by noting the vector calculus identity

$$\nabla\cdot(\mathbf{a}\times\mathbf{b}) = \mathbf{b}\cdot\nabla\times\mathbf{a} - \mathbf{a}\cdot\nabla\times\mathbf{b}. \tag{1.173}$$

Using this gives us,

$$\nabla\cdot(\mathbf{E}\times\mathbf{H}) = -\partial_t w_m - \partial_t w_e - \sigma\mathbf{E}\cdot\mathbf{E} - \mathbf{E}\cdot\mathbf{J}_i. \tag{1.174}$$

We can integrate this equation over a volume to get a description of the total power and energy contained within the volume. If we use the divergence theorem on the left-hand side of this volume-integrated equation we get

$$\oiint_S \hat{n}\cdot(\mathbf{E}\times\mathbf{H})dS = -\iiint_V \left[\partial_t w_m + \partial_t w_e + \sigma\mathbf{E}\cdot\mathbf{E} + \mathbf{E}\cdot\mathbf{J}_i\right]dV, \tag{1.175}$$

where V is some finite volume we are integrating over and S is the boundary surface of V. By noting that the left-hand side has units of $\text{V/m}\cdot\text{A/m} = \text{W/m}^2$, we can see that the left-hand side can be viewed as measuring the total amount of power flowing into/out of the

surface S that encloses V. We see that this is equal to the change in energy density in V plus a few other terms.

Let's take a moment to consider the other terms. If we look at the third set of terms on the right-hand side, we see that this has units of $\mathrm{S/m} \cdot (\mathrm{V/m})^2 = \mathrm{W/m}^3$. Hence, we can associate this with the amount of power that is being dissipated in our conductive medium. The units on our last term on the right-hand side also gives us $\mathrm{W/m}^3$. However, from our original discussion of this as an impressed current we can conclude that this is power being supplied to our system.

We can now rearrange terms to arrive at Poynting's theorem, which characterizes the balance of electromagnetic energy and power. In particular, we have that

$$\oiint_S \hat{n} \cdot \mathbf{S}\, dS + \iiint_V \left[\partial_t w_m + \partial_t w_e + \sigma \mathbf{E} \cdot \mathbf{E} \right] dV = - \iiint_V \mathbf{E} \cdot \mathbf{J}_i dV, \qquad (1.176)$$

where $\mathbf{S}(\mathbf{r}, t) = \mathbf{E}(\mathbf{r}, t) \times \mathbf{H}(\mathbf{r}, t)$ is the *instantaneous Poynting vector*. The remaining terms on the left-hand side account for the magnetic and electric energy densities and the power dissipated in the medium. The right hand side is the power supplied to the medium by the current $\mathbf{J}_i$ (recall that negative powers are by convention supplied powers). Hence, Poynting's theorem is a statement of the conservation of energy in a simple electromagnetic medium.

1.10.2 Frequency Domain

We can now construct Poynting's theorem in the frequency domain. It should be no surprise by now that much of this derivation is going to follow closely from the time domain derivation. Hence, we will begin with Faraday's and Ampere's laws as

$$\nabla \times \mathbf{E} = -\mu j \omega \mathbf{H} \qquad (1.177)$$

$$\nabla \times \mathbf{H} = \epsilon j \omega \mathbf{E} + \sigma \mathbf{E} + \mathbf{J}_i. \qquad (1.178)$$

To go about combining these equations, we will still need to take the dot products with the electric and magnetic fields. However, because we are now dealing with complex-valued quantities in the frequency domain we will need to be a little more careful.

The decision on which way to take complex conjugates comes down to convention. By convention, we will define the *average Poynting vector* as

$$\boxed{\mathbf{S}_{\mathrm{avg}}(\mathbf{r}, \omega) = \frac{1}{2} \mathbf{Re}\big\{ \mathbf{E}(\mathbf{r}, \omega) \times \mathbf{H}^*(\mathbf{r}, \omega) \big\}.} \qquad (1.179)$$

To get the left-hand side of our combined equations to give us the desired average Poynting vector, we will need to take the dot product of (1.177) with $\mathbf{H}^*$. Similarly, we will take the dot product of $\mathbf{E}$ with the complex conjugate of (1.178). This gives us

$$\mathbf{H}^* \cdot \nabla \times \mathbf{E} = -\mu j \omega |\mathbf{H}|^2 \qquad (1.180)$$

$$\mathbf{E} \cdot \nabla \times \mathbf{H}^* = -\epsilon^* j \omega |\mathbf{E}|^2 + \sigma |\mathbf{E}|^2 + \mathbf{E} \cdot \mathbf{J}_i^*. \qquad (1.181)$$

We can subtract (1.181) from (1.180) to get

$$\mathbf{H}^* \cdot \nabla \times \mathbf{E} - \mathbf{E} \cdot \nabla \times \mathbf{H}^* = -\mu j\omega |\mathbf{H}|^2 + \epsilon^* j\omega |\mathbf{E}|^2 - \sigma |\mathbf{E}|^2 - \mathbf{E} \cdot \mathbf{J}_i^* \tag{1.182}$$

We can now use the same vector calculus identity and the divergence theorem to get

$$\oiint_S \hat{n} \cdot (\mathbf{E} \times \mathbf{H}^*) dS + \iiint_V \left[j\omega \left(\mu |\mathbf{H}|^2 - \epsilon^* |\mathbf{E}|^2 \right) + \sigma |\mathbf{E}|^2 \right] dV = - \iiint_V \mathbf{E} \cdot \mathbf{J}_i^* dV. \tag{1.183}$$

This can be viewed as being Poynting's theorem in the frequency domain. We can arrive at the time average power relationships by taking $1/2 \operatorname{Re}\{(1.183)\}$. The main point of interest here is to note that the only way that ϵ and μ will impact the real power (i.e., dissipation) is if ϵ and μ have imaginary parts. If you recall from our early discussions on these quantities, we mentioned that the imaginary parts were associated with loss. We now have another way to see this through Poynting's theorem.

1.10.3 Power Flow of a Plane Wave

We will now take a closer look at how the Poynting vector relates to power flow by considering the case of a plane wave in free space. To do this, we will return back to our standard example of a plane wave. This is

$$\mathbf{E} = \hat{x} E_0 \cos(\omega t - k_0 z) \tag{1.184}$$

in the time domain and

$$\mathbf{E} = \hat{x} E_0 e^{-jk_0 z} \tag{1.185}$$

in the frequency domain. To evaluate the Poynting vector for our plane wave, we will also need the expression for the magnetic field. We can get this from Faraday's law, which says that

$$\nabla \times \mathbf{E} = -j\omega \mu_0 \mathbf{H}. \tag{1.186}$$

From this, we can determine that

$$\mathbf{H} = \hat{y} \frac{E_0}{\eta_0} e^{-jk_0 z} \tag{1.187}$$

in the frequency domain and

$$\mathbf{H} = \hat{y} \frac{E_0}{\eta_0} \cos(\omega t - k_0 z) \tag{1.188}$$

in the time domain.

Let's start by considering the time-average Poynting vector in the frequency domain. This will be

$$\mathbf{S}_{\text{avg}} = \frac{1}{2} \operatorname{Re}\{\mathbf{E} \times \mathbf{H}^*\} = \hat{x} \times \hat{y} \frac{1}{2} \operatorname{Re}\left\{ E_0 \left(\frac{E_0}{\eta_0} \right)^* e^{-jk_0 z} e^{jk_0 z} \right\}, \tag{1.189}$$
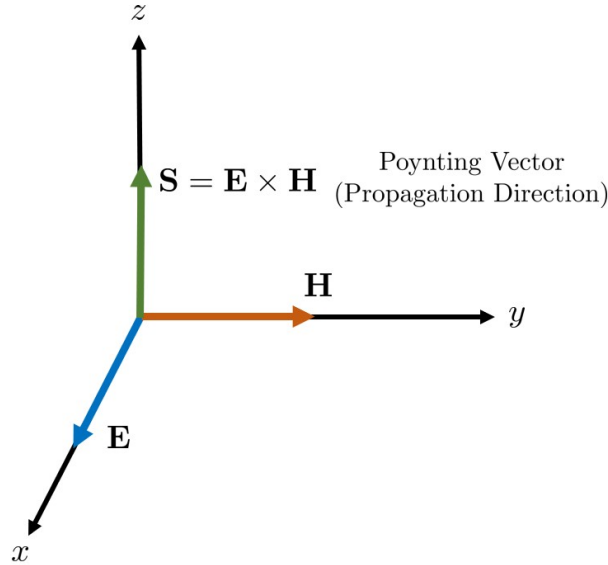
Figure 1.22: Illustration of the Poynting vector following the right-hand rule.

which simplifies down to

$$\mathbf{S}_{\text{avg}} = \hat{z}\frac{|E_0|^2}{2\eta_0}. \tag{1.190}$$

This basic structure is a good one to *memorize*. The power flow density for a plane wave has a direction given by the cross product of the directions that the electric and magnetic fields point in (in that order to get the correct propagation direction). The magnitude is given by the remainder of the equation. This is easy to remember by recalling using Ohm's law to express powers in your circuits class. The expression in (1.190) is very similar to $P = V^2/R$ that you can recall from circuits classes. If instead of using E_0 we had used the amplitude of the magnetic field we would then have

$$\mathbf{S}_{\text{avg}} = \hat{z}\frac{\eta_0|H_0|^2}{2}, \tag{1.191}$$

which should remind you of $P = I^2R$.

Although we showed this with a plane wave, it is a general property of the Poynting vector that it points in the *direction of power flow*. As a result, once we know the electric and magnetic fields we can also determine the power flow throughout a system. From the definition of the Poynting vector, we can also see that the direction of power flow follows the right-hand rule, as illustrated in Fig. 1.22.

Let's now turn our attention to looking at how the Poynting vector evaluates in the time domain. Using (1.184) and (1.188), we get

$$\mathbf{S}(\mathbf{r}, t) = \mathbf{E}(\mathbf{r}, t) \times \mathbf{H}(\mathbf{r}, t) = \hat{z}\frac{E_0^2}{\eta_0}\cos^2(\omega t - k_0 z). \tag{1.192}$$
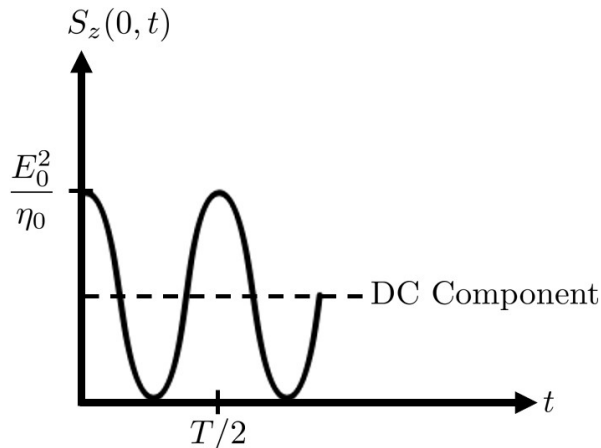
Figure 1.23: Illustration of the instantaneous Poynting vector at $z = 0$.

For simplicity, we will focus on the instantaneous Poynting vector at $z = 0$. This gives us

$$\mathbf{S}(\mathbf{r}, t) = \hat{z} \frac{E_0^2}{\eta_0} \cos^2(\omega t). \tag{1.193}$$

We can now use some trigonometric identities to see how this relates to the time-average Poynting vector we had in (1.190). In particular, we can recall that

$$\cos^2(x) = \frac{1}{2} + \frac{1}{2} \cos(2x). \tag{1.194}$$

We can use this to write our instantaneous Poynting vector as

$$\mathbf{S}(\mathbf{r}, t) = \hat{z} \frac{E_0^2}{2\eta_0} \big(1 + \cos(2\omega t)\big). \tag{1.195}$$

We see that we get a *DC component* (the first term in the parentheses) and another component that oscillates at *twice the frequency* of our plane wave (the second term in the parentheses), which is illustrated in Fig. 1.23.

We can now compute the time average of the instantaneous Poynting vector. This should equal the time-average Poynting vector result that we arrived at from our frequency domain analysis. To see this, we note that the time average is computed as

$$\langle \mathbf{S}(0, t) \rangle = \frac{1}{T} \int_0^T \mathbf{S}(0, t) dt, \tag{1.196}$$

where $T = 2\pi/\omega$ is a single period of the waveform. Evaluating this gives us

$$\boxed{\langle \mathbf{S}(0, t) \rangle = \mathbf{S}_{\text{avg}} = \hat{z} \frac{E_0^2}{2\eta_0}} \tag{1.197}$$

because the integral of the cosine function is 0. As expected, our frequency and time domain results do match.
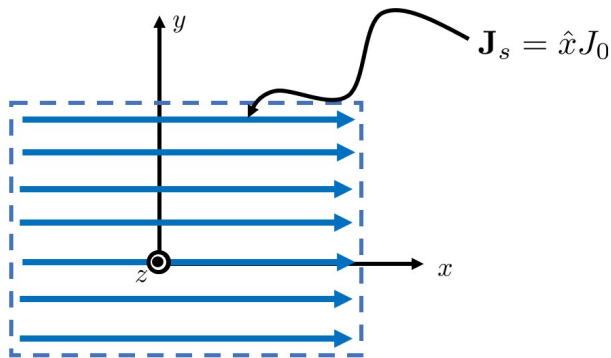
Figure 1.24: Illustration of a current sheet located in the (x, y)-plane.

1.11 Radiation from Current Sheets

We will now get our first taste of how to go about solving Maxwell's equations for a situation where we can't consider all of space to be a simple, source-free region of space. We will consider a relatively straightforward example to begin learning the necessary strategies that will be used repeatedly throughout the rest of this course. In particular, we will consider the radiation produced by an infinitely large planar current sheet.

The basic setup of this problem is shown in Fig. 1.24. In it, we have a current sheet that is covering the entire (x, y)-plane. The mathematical definition of the current is given by

$$\mathbf{J}_s(\mathbf{r}, t) = \begin{cases} \hat{x} J_0 \cos(\omega t), & z = 0, \\ 0, & \text{elsewhere.} \end{cases} \tag{1.198}$$

As with many things, we could also write this current density in another equivalent form. For instance, we could use the Dirac delta function to write this current density as

$$\mathbf{J}_s(\mathbf{r}, t) = \hat{x} J_0 \cos(\omega t) \delta(z). \tag{1.199}$$

Regardless of how we write it, the concept should hopefully be clear. We have a current density that covers the (x, y)-plane and points in the $+x$-direction with a fixed amplitude (J_0) and frequency (ω).

Now, we will go about determining what the electromagnetic fields should be equal to in all regions of space. The basic strategy that we will reuse frequently in this course is the following:

1. Identify the distinct regions of the problem and the boundaries that exist between the regions.

2. Propose a suitable general solution to Maxwell's equations in the distinct regions of the problem. These general solutions will involve a number of unknown constants that we will need to solve for. For instance, the amplitude, phase, or frequency of the solution.

3. Determine the specific values of the unknown constants by making use of the boundary conditions we know must be satisfied at various boundaries between the different regions of the problem.

Let's now run through these steps for the current problem.

For *Step 1*, we can see that we have two distinct regions – free space for $z < 0$ and free space for $z > 0$. The boundary between these regions is at the (x, y)-plane and is filled by the current sheet. We will identify the region where $z < 0$ as Region 1 and the region where $z > 0$ as Region 2. For our derivation of boundary conditions, we assumed that $\hat{n}$ points into Region 1. Hence, for this problem we will have $\hat{n} = -\hat{z}$ for any boundary conditions we use.

Throughout this course we will learn a number of different general wave solutions to Maxwell's equations that we can consider to be part of our "catalog" for answering the questions posed in *Step 2*. Each solution will have different characteristics and be well-suited for satisfying different boundary conditions. At this point, we only have one wave solution that we are familiar with – *a plane wave*. Fortunately, this wave solution is exactly what we will need for this problem (funny how that works out...).

We can rationalize that a plane wave is a good solution because the current sheet fills the entire (x, y)-plane and has a constant amplitude and phase over this entire region of space. Considering the boundary condition for the magnetic field, i.e.,

$$\hat{n} \times (\mathbf{H}_1 - \mathbf{H}_2) = \mathbf{J}_s, \tag{1.200}$$

we see that the magnetic field is going to have a spatial and temporal distribution similar to the current sheet. Hence, we have a good feeling that using a plane wave as a general solution will be a good choice.

Before writing out our proposed form of solution, the last item we will address is the expected propagation direction for the two regions. Since the current sheet is a *source* of the electromagnetic waves, it is reasonable to expect that the waves are going to propagate *away* from the source. If this weren't the case, antennas wouldn't work very well for communications or radar systems, so this seems like a reasonable thing to expect of our solution.

With this in mind, our proposed solutions in the two regions will be

$$\mathbf{H}(\mathbf{r}, t) = \begin{cases} \mathbf{H}_1(\mathbf{r}, t) = \hat{x} H_{x1} \cos(\omega_1 t + k_1 z + \varphi_{x1}) + \hat{y} H_{y1} \cos(\omega_1 t + k_1 z + \varphi_{y1}), & z < 0, \\ \mathbf{H}_2(\mathbf{r}, t) = \hat{x} H_{x2} \cos(\omega_2 t - k_2 z + \varphi_{x2}) + \hat{y} H_{y2} \cos(\omega_2 t - k_2 z + \varphi_{y2}), & z > 0. \end{cases} \tag{1.201}$$

We now need to go about the process of solving for the unknown constants to complete our solution. For reference, we need to solve for H_{x1}, H_{y1}, H_{x2}, H_{y2}, φ_{x1}, φ_{y1}, φ_{x2}, φ_{y2}, k_1, k_2, ω_1 and ω_2. This may look like a lot, but many of these constants will be quite easy to solve to the point we won't even really need to worry about them in future problems, we will just know what they should be.

This brings us to *Step 3* of the process. We are going to solve for these constants by ensuring they give us the correct behavior at the *boundary* between the two regions. Let's start with the boundary condition for the magnetic field that we already wrote out earlier. This gives us

$$-\hat{z} \times (\mathbf{H}_1 - \mathbf{H}_2) = \mathbf{J}_s. \tag{1.202}$$

We can expand these terms at $z = 0$ to get

$$- \hat{y}H_{x1}\cos(\omega_1 t + \varphi_{x1}) + \hat{x}H_{y1}\cos(\omega_1 t + \varphi_{y1})$$
$$- (-\hat{y}H_{x2}\cos(\omega_2 t + \varphi_{x2}) + \hat{x}H_{y2}\cos(\omega_2 t + \varphi_{y2})) = \hat{x}J_0\cos(\omega t). \quad (1.203)$$

This looks like a bit of a mess, so let's simplify it down. First, we can immediately see that because the current source is only directed in the x-direction we can safely say that $H_{x1} = H_{x2} = 0$. Since this removes these terms from our general solution, we also no longer need to worry about solving for φ_{x1} or φ_{x2}. Next, we can see that for this equation to be satisfied we are going to need to have the arguments of all the remaining cosine functions equal one another for all t. From this, we can conclude that the frequencies of our two waves match that of the source. That is, $\omega_1 = \omega_2 = \omega$. Further, we must also have that $\varphi_{y1} = \varphi_{y2} = 0$. Recalling that our wavenumbers are linked to the material properties in the regions (which are identical for this case) and the frequency of the wave, we can also conclude that $k_1 = k_2 = k_0 = \omega/c_0$. Now that we have that out of the way, the equation can be simplified down to be

$$\boxed{H_{y1} - H_{y2} = J_0.} \quad (1.204)$$

At this point, you can probably guess that this is going to work out to be $H_{y1} = -H_{y2} = J_0/2$. However, let's be more careful about this and make sure that we can compute this without "guessing".

To do this, we will make use of another boundary condition. In particular, we will use the boundary condition for the electric field. Hence, we will need to compute what the electric field is considering our proposed form of solution for the magnetic field. We can do this either by using Ampere's law or the tricks we learned previously about manipulating plane waves. Following the first approach, we have that computing the curl of our plane wave gives

$$\nabla \times \left(\hat{y}\cos(\omega t \pm k_0 z) \right) = k_0(\mp\hat{z}) \times \hat{y}\cos'(\omega t \pm k_0 z), \quad (1.205)$$

where the $'$ denotes a derivative with respect to the entire argument of the function. We can then use Ampere's law to see that

$$\mathbf{E} = \begin{cases} \mathbf{E}_1(\mathbf{r}, t) = -\hat{x}\dfrac{k_0}{\omega\epsilon_0}H_{y1}\cos(\omega t + k_0 z), & z < 0, \\[3mm] \mathbf{E}_2(\mathbf{r}, t) = +\hat{x}\dfrac{k_0}{\omega\epsilon_0}H_{y2}\cos(\omega t - k_0 z), & z > 0. \end{cases} \quad (1.206)$$

We can simplify these expressions by recalling that $k_0/(\omega\epsilon_0) = \eta_0$. This gives us

$$\mathbf{E} = \begin{cases} \mathbf{E}_1(\mathbf{r}, t) = -\hat{x}\eta_0 H_{y1}\cos(\omega t + k_0 z), & z < 0, \\[2mm] \mathbf{E}_2(\mathbf{r}, t) = +\hat{x}\eta_0 H_{y2}\cos(\omega t - k_0 z), & z > 0. \end{cases} \quad (1.207)$$

If we had instead used our "tricks" to manipulate plane waves we would first have determined the appropriate polarization directions. For $\mathbf{E}_1 \times \mathbf{H}_1$ to give us the correct propagation

direction of $-\hat{z}$, we would see that we would need $\mathbf{E}_1$ to be polarized in the $-\hat{x}$-direction. To adjust the amplitude of the wave, we would recall that we can use the intrinsic impedance in a manner similar to Ohm's law to convert amplitudes. In this case, we would recognize that $V = ZI$, and so we need to *multiply* our magnetic field amplitude by η_0 to get the amplitude for $\mathbf{E}_1$. Combining these steps, we see that we arrive at the same result as explicitly using Ampere's law. We can of course also follow this process to find $\mathbf{E}_2$, but we will leave that as a practice exercise for the interested reader.

Now that we have our electric fields, we can expand the boundary condition on the electric field to get another equation. In particular, the boundary condition

$$-\hat{z} \times (\mathbf{E}_1 - \mathbf{E}_2) = 0 \tag{1.208}$$

tells us that

$$\boxed{H_{y1} + H_{y2} = 0.} \tag{1.209}$$

At this point, we can solve (1.204) and (1.209) together to give us the result that we guessed earlier; i.e., $H_{y1} = -H_{y2} = J_0/2$.

We now summarize our final results. We have that the magnetic field is

$$\mathbf{H}(\mathbf{r}, t) = \begin{cases} \mathbf{H}_1(\mathbf{r}, t) = \hat{y}\dfrac{J_0}{2}\cos(\omega t + k_0 z), & z < 0, \\[2mm] \mathbf{H}_2(\mathbf{r}, t) = -\hat{y}\dfrac{J_0}{2}\cos(\omega t - k_0 z), & z > 0, \end{cases} \tag{1.210}$$

and the electric field is

$$\mathbf{E}(\mathbf{r}, t) = \begin{cases} \mathbf{E}_1(\mathbf{r}, t) = -\hat{x}\dfrac{J_0}{2}\cos(\omega t + k_0 z), & z < 0, \\[2mm] \mathbf{E}_2(\mathbf{r}, t) = -\hat{x}\dfrac{J_0}{2}\cos(\omega t - k_0 z), & z > 0. \end{cases} \tag{1.211}$$

These are also shown in Fig. 1.25.

We can briefly look over these results to see if they make sense intuitively. The first thing we see is that the waves have equal amplitudes and opposite propagation directions. This matches our expectations, since we don't have any reason to believe that the current sheet should produce a stronger wave in one direction compared to the other. Next, we see that the polarization of the electric field is the same on either side of the current sheet. Again, this matches our expectations since the boundary condition told us that the tangential component of the electric field should be continuous. To achieve the different propagation directions, we see that the tangential component of the magnetic field is discontinuous. Again, this is expected since this is exactly what the boundary condition tells us to expect.

This exercise of "checking your solution" is a good habit to get in. It is likely to serve you well in your academic and professional career, especially if you ever find yourself running numerical analysis programs. It is easy for these tools to accidentally be set up incorrectly. Despite that, they will still typically run and produce some kind of solution. Doing a "sanity check" that the solution makes sense is always a good idea to make sure that things are working properly.
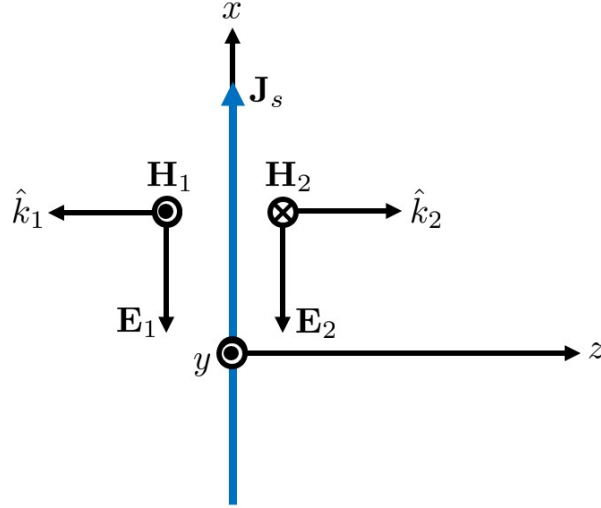
Figure 1.25: Illustration of the solution to the current sheet problem.

1.12 Plane Wave Scattering from a PEC Region (Normal Incidence)

We will now use the strategies we learned in the previous section to analyze the scattering of a plane wave from a region containing a perfect electric conductor (PEC). To simplify the analysis, we will only consider the *"normal incidence"* case. This case is when the isophase surfaces of the plane wave are in the same plane as the outer surface of the PEC region. Equivalently, we can say that there is no angle formed between the propagation direction of the plane wave and the unit normal vector of the PEC region ("no angle" here can mean 0° or 180°). We will consider the more general case, referred to as *oblique incidence*, later in the course.

The problem we will solve in this section is illustrated in Fig. 1.26. In it, we have a PEC region that begins in the $z = 0$ plane and extends into $z > 0$. We also have a plane wave *incident* on the PEC region from the $z < 0$ portion of space. When we say that it is "incident", we mean that it is part of the specification of the problem. That is, we are not concerned with how it was produced, all we care about is what must happen "next" to ensure that Maxwell's equations are satisfied in all regions of space given that this particular specified field already exists. From an intuitive perspective, we have a wave approaching an obstacle so we expect that some portion of the wave is likely to be *reflected* from the boundary. In this section, we will show how we can use Maxwell's equations to prove that this is indeed the case (and also provide a more detailed description of how it *specifically reflects* from the boundary).

For our current situation, the electric field of our incident plane wave is given by

$$\mathbf{E}^{\text{inc}}(z, t) = \hat{x} E^{+} \cos(\omega t - k_0 z). \tag{1.212}$$

We will analyze this problem in the frequency domain to show how our process from the previous section looks in the frequency domain. Considering this, our incident plane wave
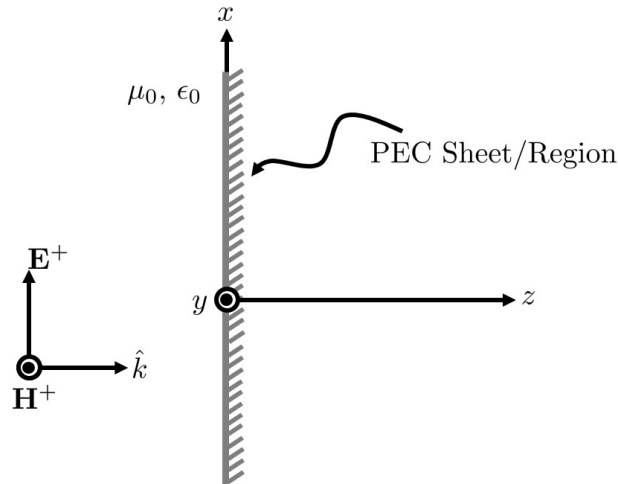
Figure 1.26: Illustration of a PEC region that begins in the $z = 0$ plane and continues into the $z > 0$ region of space. There is an incident field approaching the PEC region from the $z < 0$ region.

becomes

$$\mathbf{E}^{\text{inc}}(z) = \hat{x}E^+ e^{-jk_0 z}. \tag{1.213}$$

Let's now apply the three-step process from the previous section. In *Step 1*, we need to identify the distinct regions of the problem, the boundaries between them, and any sources or existing fields within each region. In this problem, we have Region 1 with $z < 0$ and Region 2 with $z > 0$. Region 1 is made up of free space and has the incident field $\mathbf{E}^{\text{inc}}$, while Region 2 is filled with PEC. We have a boundary made of PEC at the $z = 0$ plane.

Step 2 can be approached in a couple ways. We will use the simplest way here, which is to note that because we already have a plane wave that is incident on our boundary in Region 1 we will go ahead and assume that we should also consider plane waves in the other regions. Included in this, is noting that since we already have a plane wave propagating in the $+z$-direction in Region 1 we will need to introduce another plane wave propagating in the $-z$-direction. We will also assume that we have a plane wave propagating in the $+z$-direction in Region 2. This would correspond to the *transmitted field*. Considering this, we have the total electric field being given as

$$\mathbf{E} = \begin{cases} \mathbf{E}_1(\mathbf{r}, t) = \hat{x}E^+ e^{-jk_0 z} + (\hat{x}E_x^- + \hat{y}E_y^-)e^{jk_0 z}, & z < 0, \\ \mathbf{E}_2(\mathbf{r}, t) = (\hat{x}E_{x2}^+ + \hat{y}E_{y2}^+)e^{-jkz}, & z > 0. \end{cases} \tag{1.214}$$

Let's dissect this for a moment. First of all, we have included two polarizations for the plane waves with unknown constants because we don't necessarily know at this point whether only having x-polarized plane waves will be enough to solve the problem. We also see that we have k in the second line instead of k_0. This is because this field is in a different medium than free space. Although we haven't specified an ϵ and μ for a PEC region, the important thing to note here is that this k would still be calculated at the same frequency that we are performing our analysis at since we have a linear, time-invariant system.

We now turn our attention to *Step 3*, which covers solving for the different constants of the problem using the information we have at the various boundaries of the problem. We start out by writing the boundary condition down at $z = 0$, which is

$$\hat{n} \times (\mathbf{E}_1 - \mathbf{E}_2) = 0. \tag{1.215}$$

We can simplify this by noting that a PEC material is characterized by having "$\sigma = \infty$". Since $\mathbf{J}$ can't take on an infinite value, Ohm's law tells us that $\mathbf{E}$ *must* be equal to 0 in a PEC region. Hence, we have that $\mathbf{E}_2 = 0$. Due to this, even if the PEC region only extends a short distance into the $z > 0$ the physics of the scenario being considered will be unchanged. As a result, it is common to just illustrate PEC "sheets" in drawings, like in Fig. 1.26. Now, the boundary condition given in (1.215) then tells us that the *total tangential electric field in Region 1 must be equal to 0 at $z = 0$*. This gives us

$$-\hat{z} \times \left(\hat{x}E^+ + (\hat{x}E_x^- + \hat{y}E_y^-) \right) = 0, \tag{1.216}$$

which simplifies to

$$-\hat{y}E^+ - \hat{y}E_x^- + \hat{x}E_y^- = 0. \tag{1.217}$$

The first thing we note from (1.217) is that E_y^- must be 0 because there is no other term with the same polarization left in the equation. The remaining part of the equation tells us that $E_x^- = -E^+$.

When dealing with systems where reflections can occur, we generally will define a *reflection coefficient*. The exact definition can depend somewhat on the situation of interest, but in general a reflection coefficient is defined as the ratio between the complex-valued amplitudes of the reflected and incident fields referenced to a particular location in space. Where this reference point makes sense to be taken at depends on the specifics of a problem. For this problem, we will reference our reflection coefficient to $z = 0$. The reflection coefficient is typically denoted with the symbol Γ. For this problem, we have

$$\Gamma(z = 0) = \frac{E^-(z = 0)}{E^+(z = 0)}, \tag{1.218}$$

which in general will be a complex-valued number. For *passive systems* (i.e., systems that don't provide some kind of amplification) the magnitude of the reflection coefficient will be less than or equal to 1. That is, $|\Gamma| \leq 1$. *By convention, the reflection coefficient is **always** specified in terms of the ratio of electric fields for this kind of problem. If you calculate the reflection coefficient in terms of the magnetic field, you will get the wrong result!* The reflection coefficient is often a very important figure of merit of many different microwave systems. We generally want to minimize the reflection coefficient in the design of transmission line and antenna systems. We will learn much more about this during those portions of the course.

Before finishing our discussion of this problem, it is useful to take one last look at the final answer that we came up with. In particular, we have that

$$\mathbf{E} = \begin{cases} \mathbf{E}_1(\mathbf{r}, t) = \hat{x}E^+\left(e^{-jk_0z} - e^{jk_0z}\right), & z < 0, \\ \mathbf{E}_2(\mathbf{r}, t) = 0, & z > 0. \end{cases} \tag{1.219}$$
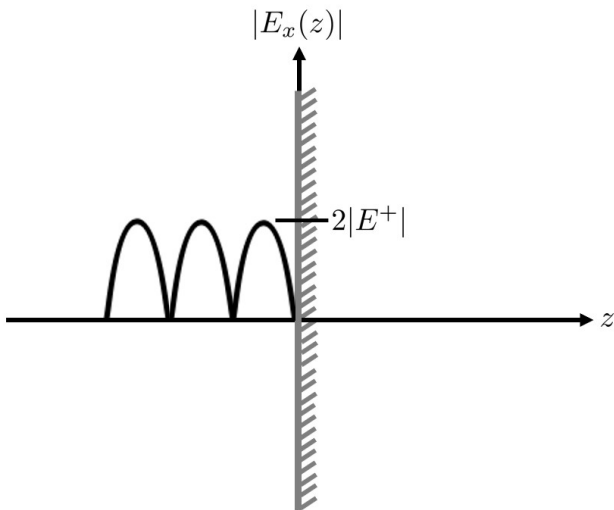
Figure 1.27: Illustration of a standing wave produced by the reflected field.

The portion inside the parentheses in the first line should look familiar to you. In particular, we can use Euler's identities to rewrite this as a sine function. We end up with

$$\mathbf{E} = \begin{cases} \mathbf{E}_1(\mathbf{r}, t) = -\hat{x}j2E^+\sin(k_0 z), & z < 0, \\ \mathbf{E}_2(\mathbf{r}, t) = 0, & z > 0. \end{cases} \tag{1.220}$$

The form of the solution in the top line is often referred to as a *standing wave*. There are many reasons for this name, and we will look into this more later in the course. The intuitive reason is that this solution no longer shows us the same kind of wave propagation effect we had seen previously. It has fixed points in space where the amplitude of the wave is *always* 0. These "nulls" do not move in space as a function of time, and neither does any other part of the wave. Hence, it looks like it is "standing" in space. An illustration of the standing wave pattern is shown in Fig 1.27. This is an example of *wave interference*, which is a broadly useful resource. As you will see, much of engineering with waves involves carefully interfering them to achieve a desired outcome.

Before concluding, it is worth looking at why we used the boundary condition for the tangential components of $\mathbf{E}$ rather than the tangential components of $\mathbf{H}$. The reason lies in that at a PEC boundary the tangential magnetic field is *discontinuous*. In particular, it goes from having a non-zero finite value on one side of the boundary to immediately having a value of zero in the PEC region. As we discussed previously, this kind of discontinuity can only occur if there is a surface current flowing along the PEC boundary. Since we don't know what value the surface current has going into the problem, we can't use

$$\hat{n} \times (\mathbf{H}_1 - \mathbf{H}_2) = \mathbf{J}_s \tag{1.221}$$

to solve for $\mathbf{H}_1$ (noting that $\mathbf{H}_2 = 0$ in the PEC region). However, now that we have solved the problem, we can easily use

$$\hat{n} \times \mathbf{H}_1 = \mathbf{J}_s \tag{1.222}$$

to compute the surface current from the solved for fields.

1.13 Lorentz Force and Material Response

We have already discussed informally in this class that the electric field can apply a force to a charged particle, and consequently, put it into motion. In addition to this, the magnetic flux may also influence the motion of a charged particle. These effects are captured in the *Lorentz force*, which describes how the electromagnetic fields influence the motion of a point charge. In this course, we are not often concerned with looking at the motion of individual electrons (this is something physicists tend to be more interested in, and can also be important in the design of certain kinds of traveling wave tube amplifiers that are needed in high-power communication systems). However, the Lorentz force provides a useful way to determine the material response of certain systems, such as plasmas. Surprisingly, we will see shortly that even using relatively simple equations can often provide a reasonable description of the physics. These simple models have been leveraged to great effect to analyze many different physical systems.

Overall, the complete Lorentz force is given by

$$\boxed{\mathbf{F} = q\mathbf{E} + q\mathbf{v} \times \mathbf{B},} \tag{1.223}$$

where $\mathbf{F}$ is the total force applied to the point charge that has velocity $\mathbf{v}$ and charge q. We have previously discussed how charges and currents can produce electromagnetic fields, now we see how these fields can influence charges.

1.13.1 Plasma Material Response

We now turn our attention to using the Lorentz force to determine the material response of a plasma medium. A plasma can often be viewed as an ionized gas with a number of "species" of ions present. Typically, the electrons are substantially lighter than the positively-charged ions that also make up the plasma. As a result, the electrons are generally the predominant component of the plasma that the electromagnetic field will interact with. Hence, we will develop an equation of motion for the electrons in the plasma medium and see how this can be used to derive an effective material response/description of the plasma.

Since we are dealing with a classical, non-relativistic system, our equation of motion for each electron will simply be Newton's law. We typically write this for a plasma as

$$m\ddot{\mathbf{r}} + m\gamma\dot{\mathbf{r}} + \kappa\mathbf{r} = q\mathbf{E}. \tag{1.224}$$

The first term in (1.224) is the acceleration of the particle, the second term is a friction/collision term that accounts for loss in the system, and the third term is a restoring force similar to Hooke's law. For simplicity, we have omitted the magnetic flux density. This can often be acceptable when the velocity of the charge is small and the magnetic field is relatively weak (it is often weaker than the electric field since $|\mathbf{B}| = \mu|\mathbf{H}| = \mu|\mathbf{E}|/\eta$ for a plane wave). These assumptions are generally good approximations for a propagating plane wave. However, in more general applications where magnetic fields can be applied from other sources than the wave itself the more complete description of the full Lorentz force can be needed.

Depending on which terms in (1.224) are kept, we can arrive at a number of different models for material response that are appropriate for different situations. We will start

our analysis by considering all of the terms and then looking at a particular special case of interest from there. The easiest way to analyze this system is to consider a time harmonic case. Then we can apply a Fourier transform to arrive at

$$-\omega^2 \mathbf{r} + j\omega\gamma \mathbf{r} + \omega_0^2 \mathbf{r} = \frac{q}{m}\mathbf{E}, \qquad (1.225)$$

where $\omega_0 = \sqrt{\kappa/m}$. We can solve for $\mathbf{r}$ to get

$$\mathbf{r} = \frac{1}{\omega_0^2 - \omega^2 + j\omega\gamma}\frac{q}{m}\mathbf{E}. \qquad (1.226)$$

Now, recall from our discussions about the polarization intensity of a material that we can treat the displacements of the charged particles as small dipole moments given by $q\,\mathbf{r}$. If we sum the effects of all of these little dipoles up we get the overall polarization intensity, from which we can identify the electric susceptibility of the material. In particular, if we have N electrons per unit volume we will have that

$$\mathbf{P} = Nq\mathbf{r} = \frac{1}{\omega_0^2 - \omega^2 + j\omega\gamma}\frac{Nq^2}{m}\mathbf{E}. \qquad (1.227)$$

Recalling that we also have

$$\mathbf{P} = \epsilon_0 \chi_e \mathbf{E}, \qquad (1.228)$$

we see that the electric susceptibility is going to be given by

$$\boxed{\chi_e = \frac{Nq^2}{\epsilon_0 m}\frac{1}{\omega_0^2 - \omega^2 + j\omega\gamma}.} \qquad (1.229)$$

We will now consider the particular case where the collision frequency γ and restoring force can be neglected. This can occur, for example, when the density of the plasma is very low (this case is sometimes referred to as a cold collisionless plasma). Under this approximation (since it is always an approximation), our electric susceptibility can be simplified greatly to be

$$\chi_e = -\frac{Nq^2}{\epsilon_0 m \omega^2} = -\frac{\omega_p^2}{\omega^2}, \qquad (1.230)$$

where

$$\boxed{\omega_p = \sqrt{\frac{Nq^2}{m\epsilon_0}}} \qquad (1.231)$$

is called the *plasma frequency*. The plasma frequency corresponds to the natural oscillation frequency of the plasma medium.
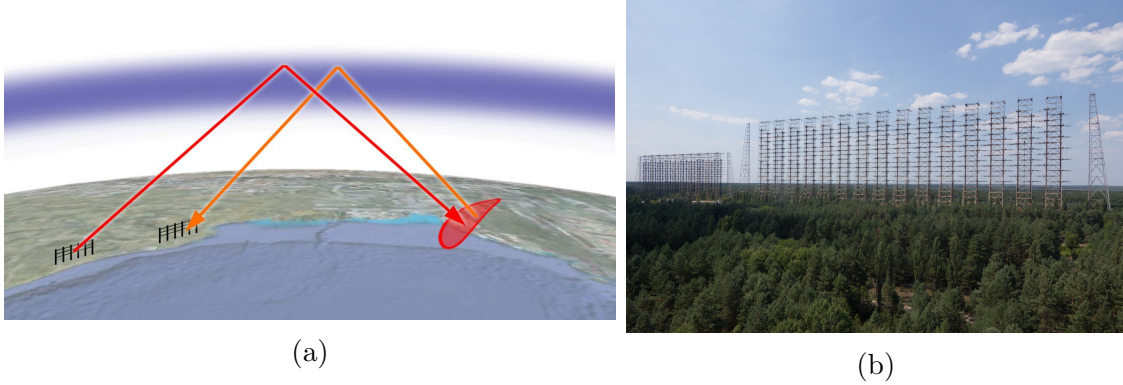
(a)

(b)

Figure 1.28: (a) Illustration of an over-the-horizon radar system interrogating a location that is not visible to a line-of-sight radar wave propagation. (b) The low operating frequencies of over-the-horizon radars require massive antennas to operate effectively (images from Wikipedia [5]).

We get different kinds of material responses depending on the value of the electromagnetic field's frequency relative to the plasma frequency. To see this easily, we recall that the relative permittivity of the medium can be described by

$$\epsilon_r = 1 + \chi_e = 1 - \frac{\omega_p^2}{\omega^2}, \tag{1.232}$$

where ω is the frequency of the electromagnetic wave. Correspondingly, our propagation constant for a plane wave in this plasma medium would be

$$\gamma = j\omega\sqrt{\mu_0\epsilon_0}\sqrt{1 - \frac{\omega_p^2}{\omega^2}}. \tag{1.233}$$

If $\omega < \omega_p$, the propagation constant becomes a purely real number, and as a result the plane wave will be quickly attenuated as it travels into the plasma medium. The name for this kind of behavior is an *evanescent wave*. Although we won't look at this in detail right now, this case also often results in a large reflection of the plane wave off of the "boundary" of the plasma medium since it isn't able to propagate into the medium. This can occur off of the ionosphere for particularly low frequency waves (generally less than 40 MHz), and is leveraged to build *over-the-horizon radar systems*. These systems bounce radar waves off of the ionosphere so that they can search for targets of interest that would otherwise not be visible for a standard line-of-sight radar operation (see Fig. 1.28).

On the other hand, if we have that $\omega \gg \omega_p$, the relative permittivity is approximately 1 and the plasma has almost no effect on the propagating wave. Since satellite communication signals are operated well above the plasma frequency, they can typically propagate through the ionosphere without any issue. However, the ionosphere is a complicated system with the plasma density always slowly varying depending on the time of day, solar wind effects, etc. As a result, the plasma density can still impact the propagation of electromagnetic waves in important ways. For instance, global navigation systems (e.g., GPS) are only able to achieve the level of accuracy that we have grown accustomed to on a day-to-day basis

by constantly calculating updates to internal error models to approximately correct for the ionospheric propagation effects. The correction terms for these error models are calculated at well-surveyed locations (often cell phone network towers), which then can distribute these corrections to nearby users connected to the network. Without these corrections, the GPS positioning accuracy for standard users can diminish by multiple meters (sometimes much more in extreme cases).

1.13.2   Other Material Response Models

As mentioned previously, depending on which terms of (1.224) we keep in our description of a material we can arrive at various other material responses. One common model is the *Drude model* for metals. This case assumes that the collision force dominates, and leads to a description of a conduction current and conductivity of a material. Other models include that of a *Lorentz oscillator*, which corresponds to keeping the acceleration and restoring force terms; as well as the *Debye relaxation model*, which keeps the friction/collision and restoring force terms. These models are useful for describing the particular dispersion profiles (more on this shortly) of various kinds of materials.

1.14   Polarization of Electromagnetic Waves

Up to this point in the course, we have occasionally mentioned in passing how a particular plane wave is *polarized*. Since we were only dealing with rather simple descriptions of plane waves, we did not need to be very detailed in our descriptions of what this means exactly. However, the polarization of electromagnetic waves brings with it incredibly rich and useful physics that can be leveraged to great effect in many application areas. We will now turn our attention to providing a more complete description of electromagnetic wave polarization, followed by a brief look at how this important phenomenon can be utilized. Following that, we will finish introducing the essential topics related to the polarization of electromagnetic waves.

1.14.1   Introduction to Polarization

The polarization of a plane wave is defined by *the locus of points that the tip of the electric field intensity vector traces out as a function of time with space fixed.* In particular, the locus of points is traced out in a plane orthogonal to the propagation direction. Let's take a closer look at our standard plane wave example to see how this definition works out.

In particular, we have

$$\mathbf{E}(\mathbf{r}) = \hat{x}E_0 e^{-jkz}. \tag{1.234}$$

We want to understand the behavior in time, so we will convert our phasor back into the time domain by multiplying by $\exp j\omega t$ and taking the real part. This gives us

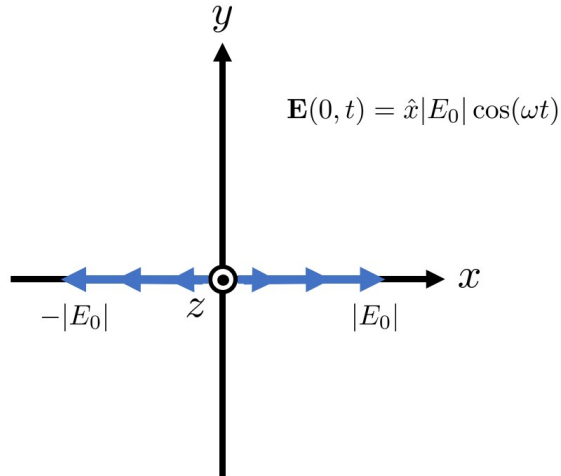$$\mathbf{E}(\mathbf{r}, t) = \hat{x}|E_0| \cos(\omega t - kz + \varphi), \tag{1.235}$$

Figure 1.29: Illustration of linear polarization for a plane wave propagating in the $+z$-direction.

where φ is any phase factor that may have been embedded in the complex-valued E_0. For now, we will assume that $\varphi = 0$ to keep things simple. Further, we will look at the polarization in the plane $z = 0$. *It is conventional to look at how the polarization varies from the perspective of an observer that the plane wave is traveling toward.* Since our current plane wave is propagating in the $+z$-direction, we will be looking at the $z = 0$ plane from above when we plot the electric field vector as a function of time.

An illustration of how the tip of the electric field vector will move around in the $z = 0$ plane as a function of time is shown in Fig. 1.29. As expected from the cosine function in (1.235), the tip of the electric field vector simply oscillates back and forth between $x = -|E_0|$ and $x = |E_0|$. Since the shape in the $z = 0$ plane is a line, we refer to this kind of polarization as *linear polarization*. We would refer to this plane wave as being *linearly polarized*.

We can of course also have a plane wave polarized along the y-axis as well. This function would look like

$$\mathbf{E}(\mathbf{r}, t) = \hat{y}|E_0| \cos(\omega t - kz + \varphi). \tag{1.236}$$

The only change to our illustration in Fig. 1.29 would be that the line would now be varying along the y-axis instead of the x-axis.

What happens if we combine two linearly polarized plane waves? Is the resulting polarization still linear? The answer depends on the *relative phase* between the two plane waves. For instance, if we have a relative phase of 0, our plane wave can be represented as

$$\mathbf{E}(z = 0, t) = \hat{x}|E_0| \cos(\omega t) + \hat{y}|E_0| \cos(\omega t), \tag{1.237}$$

in the $z = 0$ plane. If we draw this out in the $z = 0$ plane we get the result shown in Fig. 1.30. We see that the result is still a line, so this combination of plane waves is still linearly polarized. If we were to change the magnitude of the amplitudes between the two plane waves, the angle the line would make with the x-axis would change. *In the end, we find that if there is no relative phase difference between the two plane waves with orthogonal spatial*
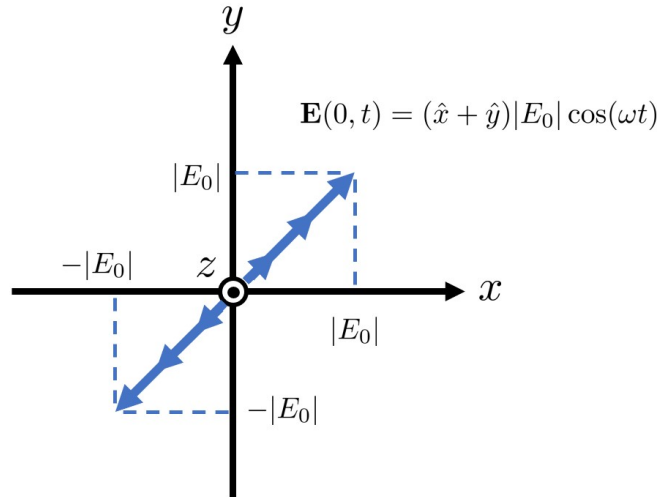
Figure 1.30: Illustration of a combination of two plane waves with orthogonal linear polarizations and no relative phase difference propagating in the $+z$-direction.

polarizations the resulting total plane wave (i.e., the summation of the two components) will be linearly polarized.

However, if we have a *non-zero relative phase difference* between our plane waves the story can be different. For instance, if we have that our y-polarized plane wave has a leading phase of 90 degrees our expression for the electric field becomes

$$\mathbf{E}(z = 0, t) = \hat{x}|E_0| \cos(\omega t) + \hat{y}|E_0| \cos(\omega t + \pi/2) \\ = \hat{x}|E_0| \cos(\omega t) - \hat{y}|E_0| \sin(\omega t).$$
(1.238)

If we trace out the tip of the electric field vector for this case, we get the result shown in Fig. 1.31. We see that the angle the field vector makes with respect to the x-axis is now changing as a function of time. Further, because the magnitudes of the two plane waves are equal to each other we end up with the tip of the electric field vector tracing out a circular shape. Hence, we refer to this as being a *circular polarization*. However, this is not a full specification of our polarization. We also must relate the *sense* of the rotation. For this particular case, we see that it is following a *clockwise rotation*. Alternatively, we can also call it a *left-hand circular polarization* because when the fingers of the left hand curl to follow the sense of the rotation the thumb points in the direction of propagation. *It is important to remember that this sense of polarization **must** be specified from the perspective of an observer that the wave is traveling toward.*

If we instead have a lagging phase of 90 degrees, the sense of our rotation will change. To see this, we note that the expression for our electric field now becomes

$$\mathbf{E}(z = 0, t) = \hat{x}|E_0| \cos(\omega t) + \hat{y}|E_0| \cos(\omega t - \pi/2) \\ = \hat{x}|E_0| \cos(\omega t) + \hat{y}|E_0| \sin(\omega t).$$
(1.239)

If we trace out the tip of the electric field vector for this case, we get the result shown in Fig. 1.32. However, this is not a full specification of our polarization. We see that the *sense*
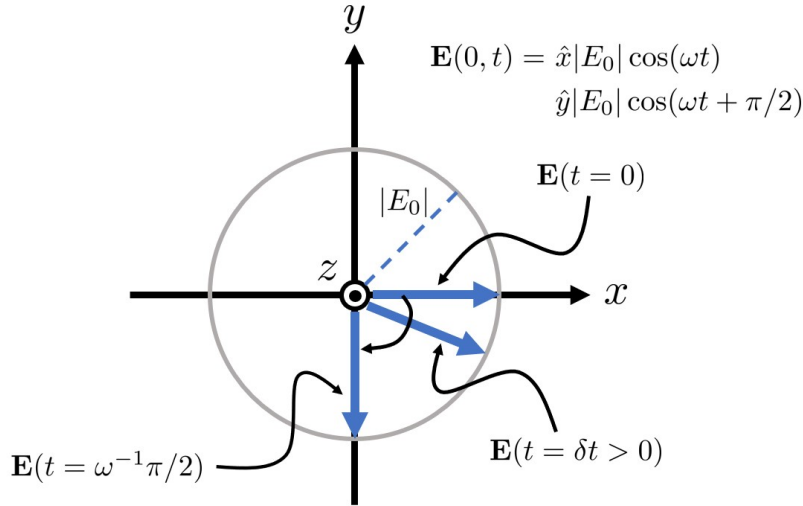
Figure 1.31: Illustration of a combination of two plane waves with orthogonal linear polarizations and a 90° relative phase difference propagating in the $+z$-direction. The y-polarized wave has a leading phase difference so we get a clockwise/left-hand sense to our rotation.

of the rotation is following a *counterclockwise rotation*. Alternatively, we can also call it a *right-hand circular polarization* because when the fingers of the right hand curl to follow the sense of the rotation the thumb points in the direction of propagation. An illustration of how right- and left-hand circularly polarized waves propagating through space look is shown in Fig. 1.33. As a final note, we emphasize that for a *circularly polarized wave to exist, we **must have** two components that have equal magnitude and are orthogonal in time and space.*

In general, it takes a very special kind of source to produce only a single linear polarization or to produce a pure form of circular polarization. More generally, the non-ideal nature of realistic systems means we will have some perturbation from these desired results. This can come in the form of cross polarization (unintended production of an orthogonal polarization), or by corrupting the polarization that was produced. This leads us to a very general type of polarization, known as *elliptical polarization*. Elliptical polarization will occur if we have two spatially orthogonal components that *do not meet all of the conditions* for being either linearly or circularly polarized. An illustration of an elliptically polarized wave is shown in Fig. 1.34.

There are a few important parameters used to characterize an elliptically polarized wave. First, the amplitudes of the semi-minor and semi-major axes. These will often be reported together in the *axial ratio* of the wave, which is equal to the ratio between the minor and major axis of the polarization ellipse. Due to the difficulty in generating pure circularly polarized waves, practical applications will generally specify an *axial ratio bandwidth* as being the frequencies for which the axial ratio is less than 3 dB. Second, the *angle of inclination* (ψ in Fig. 1.34) is defined, which gives the angle between the semi-major axis and the x-axis. Finally, the sense of rotation also needs to be specified in the same way that it was for a circularly polarized wave.

It is often a useful skill to be able to quickly determine the polarization of a wave by

$$\mathbf{E}(0,t) = \hat{x}|E_0|\cos(\omega t)$$
$$\hat{y}|E_0|\cos(\omega t - \pi/2)$$
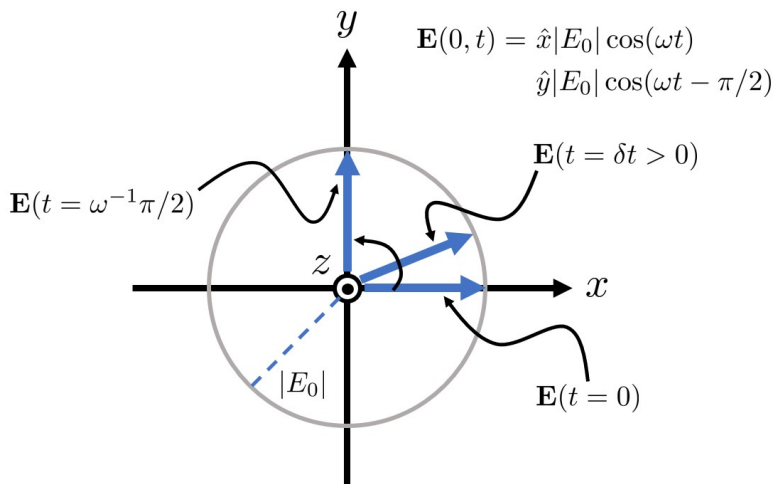
Figure 1.32: Illustration of a combination of two plane waves with orthogonal linear polarizations and a $90°$ relative phase difference propagating in the $+z$-direction. The y-polarized wave has a lagging phase difference so we get a counterclockwise/right-hand sense to our rotation.

Figure 1.33: Illustration of propagating right- and left-hand circularly polarized waves (images from [6]).

looking at its phasor representation. We will look at a few cases here to get the general idea.

- $\mathbf{E}(z) = \left(\hat{x}|E_1| + \hat{y}|E_2|\right)e^{-jkz}, \quad |E_1| \neq |E_2| \implies$ Linear Polarization

Figure 1.34: Illustration of a general elliptical polarization (image from Wikipedia [7]).

- $\mathbf{E}(z) = \left(\hat{x}|E_1| + \hat{y}|E_2|\right)e^{-jkz}, \ |E_1| = |E_2| \implies$ Linear Polarization

- $\mathbf{E}(z) = \left(\hat{x}|E_1| + \hat{y}j|E_2|\right)e^{-jkz}, \ |E_1| = |E_2| \implies$ Left-Hand Circular Polarization

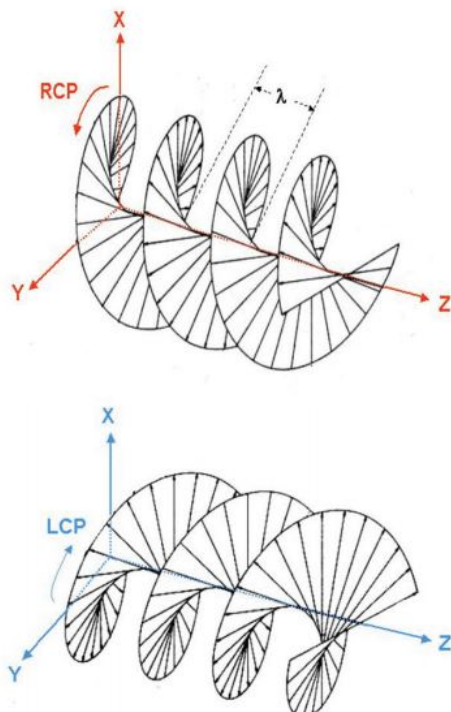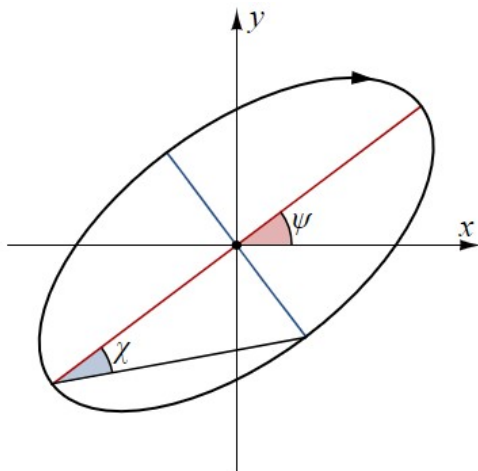- $\mathbf{E}(z) = \left(\hat{x}|E_1| - \hat{y}j|E_2|\right)e^{-jkz}, \ |E_1| = |E_2| \implies$ Right-Hand Circular Polarization

- $\mathbf{E}(z) = \left(\hat{x}|E_1| + \hat{y}j|E_2|\right)e^{-jkz}, \ |E_1| \neq |E_2| \implies$ Left-Hand Elliptical Polarization

- $\mathbf{E}(z) = \left(\hat{x}|E_1| - \hat{y}j|E_2|\right)e^{-jkz}, \ |E_1| \neq |E_2| \implies$ Right-Hand Elliptical Polarization

Although it is good to be able to quickly identify these, you can always convert them back to a time domain representation and plot a few points to identify the polarization of the wave you are given.

Before completing our discussion of polarization, it is useful to note that we can always decompose a generally polarized wave into some set of orthogonal components. In general, we can think of two orthogonal polarizations being able to be used as a *basis* to describe other polarizations. For instance, in the above exercise we very clearly see that we can always decompose these waves into two orthogonal linear polarization components. It is also common to decompose waves into orthogonal circular polarizations, where right-hand and left-hand circularly polarized waves form our basis. Other options can be used as well, but these are the most often used in practice.

1.14.2 Applications of Polarized Waves

We are often most interested in polarization in the context of antennas, since these are typically the devices that we will use to generate and receive plane waves. A quick way to

Figure 1.35: A crossed dipole antenna (image from Wikipedia [8]).

see how polarization impacts antenna design is by recalling from our discussion on Poynting's theorem that the power going into/out of the source function $\mathbf{J}$ was modeled as $\mathbf{E} \cdot \mathbf{J}^*$. Hence, the electric field needs to have a polarization that matches the direction of the current flow to be able to extract any power from the currents.

This has a few practical consequences. First, if we can make an antenna that predominantly has all of its current only able to flow along a single direction it will only be able to transmit or receive the components of plane waves polarized along the direction of the current flow. A good example of this kind of antenna is a wire antenna – like a *dipole antenna*. If we take two dipoles and orient them orthogonally to each other we get a *crossed dipole antenna*, as seen in Fig. 1.35. This allows us to break the signals we are receiving into their two orthogonal constituents.

Why might we be interested in doing this? One example is in the area of *wireless communications*. Our ability to transmit information simultaneously is dependent on the number of *sufficiently independent channels* we have access to. A classic example would be a frequency division multiplexing scheme where channels can be made up of various dedicated portions of the electromagnetic spectrum. However, because our two orthogonal antenna channels are independent of each other, we can effectively *double our number of channels* by making use of each polarization. This is sometimes referred to as *polarization diversity*, and is often an important component of making practical *multiple-input multiple-output (MIMO) communication systems* that are integral parts of things like 5G networks.

In communication systems with particularly challenging isolation requirements, polarization diversity can be a vital tool in lowering the coupling between communication channels. An excellent example is in the area of *satellite communications*. The huge distance between a satellite and the earth results in *massive propagation losses*. For instance, in S-band (say 2 GHz), the loss to a typical *low earth orbit is approximately 155 dB* while the loss to a *geostationary orbit is approximately 185 dB*. As a result, the receivers on a satellite must be *exceptionally sensitive*. Likewise, the *transmitter power* on the satellite must also be large so that users on the Earth (who typically have much cheaper receivers, e.g., in their phone) can receive the signals. Locating a sensitive receiver next to a high power transmitter is

Figure 1.36: Example of a traditional "spotlight mode" SAR image of some buildings, parking lots, and New Mexico terrain. Courtesy of Sandia National Laboratories, Radar ISR.

a challenging engineering problem that can typically require multiple levels of isolation to meet system requirements (e.g., 100 dB of isolation). In these systems, polarization diversity between the transmit and receive channels can be an important component of the overall isolation strategy.

Polarization diversity can also be a useful tool in areas like remote sensing. A great example is in *synthetic aperture radar (SAR)* systems. These systems coherently process many different radar pulses together to form images of an area being investigated (this requires a combination of careful timing and processing of phase shifts related to the Doppler effect). A typical processed SAR image can require careful training to be able to properly interpret since they aren't as intuitive as optical images that we are used to looking at. An example of a traditional SAR image is shown in Fig. 1.36. However, if you instead use a *polarimetric SAR sensor* that transmits and receives the SAR signals from two orthogonal polarizations you can make much richer images. This is possible because the manner in which the polarization of a wave changes after interacting with an object can be indicative of certain properties about the target being observed. This kind of information can be added into a SAR image as "colors" to help the viewer better interpret what different parts of the image may correspond to. For instance, a colored polarimetric SAR image is shown in Fig. 1.37. We can see that this image has much richer content and more clearly differentiates between different kind of terrain (e.g., trees vs. buildings, grass vs. road, etc.)

Figure 1.37: Example of a polarimetric SAR image of some buildings, parking lots, and New Mexico terrain. The coloring is done through processing of the polarization information received by the radar. Courtesy of Sandia National Laboratories, Radar ISR.

1.15   Doppler Effect

The Doppler effect is a general wave phenomena that occurs when there is relative motion in the direction of wave propagation between a transmitter and receiver of a wave. You are familiar with this effect in everyday life when an emergency response vehicle with its sirens on drives past you. As the vehicle comes toward you, the sirens have a higher pitch and as the car moves past you the siren begins to have a lower pitch as it drives away. The general idea is that because the transmitter or receiver is moving relative to the wave propagation, the wave becomes stretched or compressed, effectively changing the frequency of the overall wave. This is illustrated in Fig. 1.38.

With this in mind, we will first look at the derivation of the Doppler effect that is applicable for most applications in Section 1.15.1. Following this, we will look at a special case of how it is used in Doppler radars in Section 1.15.2.
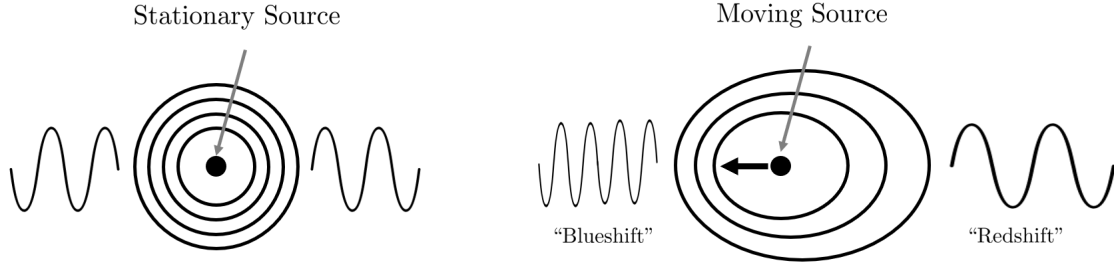
Figure 1.38: A simplified illustration of the Doppler effect. Note that the blueshift and redshift terminology is common in the physics community, and is in reference to the extremes of the visible light spectrum. Relativistic Doppler effects have important consequences in general relativity and our ability to interpret details about the universe from measurements of received electromagnetic energy on Earth.
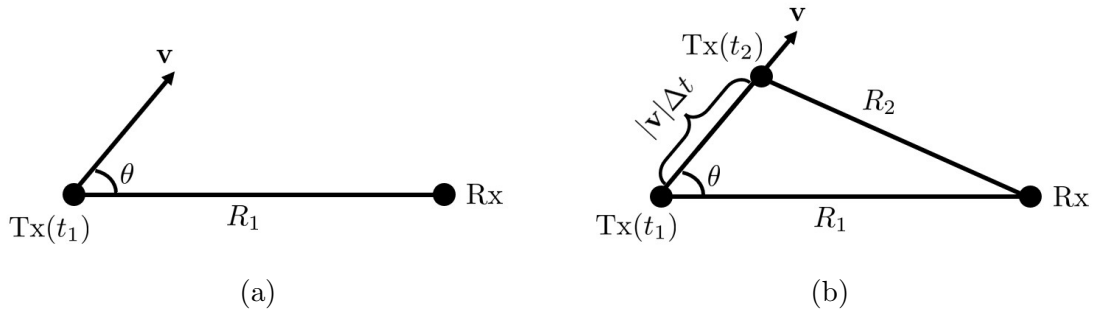
Figure 1.39: Geometry for the derivation of the Doppler effect: (a) initial time and (b) after some additional time has passed.

1.15.1 Derivation

To derive the Doppler effect, we will assume that we have a moving transmitter and a stationary receiver. The transmitter is initially located a distance $\mathbf{R}_1$ away from the receiver. The transmitter is moving with speed v at a relative angle of θ to the line connecting the initial locations of the transmitter and receiver. This is illustrated in Fig. 1.39(a). If we consider the portion of the wave that is emitted at the initial time $t = 0$, we can see that it will arrive at the receiver at time

$$t_1 = \frac{R_1}{c}, \tag{1.240}$$

where c is the speed of the light.

If we now consider the case at a later time, $t = \Delta t$, we see that the transmitter will be at a new position (see Fig. 1.39(b)). The wave emitted from this position will reach the receiver at time

$$t_2 = \Delta t + \frac{R_2}{c}, \tag{1.241}$$

where R_2 is the distance between the transmitter and receiver at $t = \Delta t$. Of course, the distance R_2 can be related back to R_1 and the parameters describing the motion of the

transmitter. This requires a bit of trigonometry, but the overall result is that

$$R_2 = \sqrt{R_1^2 - 2R_1(v\Delta t)\cos(\theta) + (v\Delta t)^2}. \tag{1.242}$$

If we substitute this into (1.241), we get a bit of an ugly expression that is

$$t_2 = \Delta t + \frac{1}{c}\sqrt{R_1^2 - 2R_1(v\Delta t)\cos(\theta) + (v\Delta t)^2}. \tag{1.243}$$

However, we can simplify the final result into a nicer form if we make a useful approximation at this point. In particular, for most normal operating conditions we will have that $v \ll c$, since it is very difficult to move at a speed close to the speed of light. As a result, if we only keep terms that are "first order" in the parameter v/c we can simplify this to be

$$t_2 = \Delta t + \frac{R_1}{c}\sqrt{1 - \frac{2(v\Delta t)}{R_1}\cos(\theta)} \tag{1.244}$$

Since the second term in the square root is still a small number, we can use a small argument approximation to $\sqrt{1+x}$ to get

$$\begin{aligned} t_2 &= \Delta t + \frac{R_1}{c}\left(1 - \frac{(v\Delta t)}{R_1}\cos(\theta)\right) \\ &= \Delta t + \frac{R_1}{c} - \frac{(v\Delta t)}{c}\cos(\theta). \end{aligned} \tag{1.245}$$

At this point, we can determine what the difference in receive times look like at the receiver. This gives

$$\begin{aligned} \Delta t' &= t_2 - t_1 \\ &= \Delta t\left(1 - \frac{v}{c}\cos(\theta)\right). \end{aligned} \tag{1.246}$$

If we assume that Δt was a full period of the transmitters harmonic waveform we can arrive at the perceived receive frequency to get

$$f' = \frac{1}{\Delta t'} = \frac{f}{1 - \frac{v}{c}\cos(\theta)}. \tag{1.247}$$

For standard cases where $v/c \ll 1$, this can be further simplified to give us that

$$f' \approx f\left(1 + \frac{v}{c}\cos(\theta)\right). \tag{1.248}$$

This gives us the expected result that if the transmitter is coming toward the receiver ($\theta = 0$) the perceived frequency is higher, while if the transmitter is moving away from the receiver ($\theta = 180$) the perceived frequency is lower.

1.15.2   Doppler Radar

In this section, we will look at a simplified schematic of a Doppler radar to get a better understanding of how a real device can use the effects derived in the previous section. The goal of a Doppler radar is to use the Doppler shift of a moving target to determine how fast it is moving. Hence, we are hoping to "isolate" the amount of frequency shift that has occurred. We will denote the Doppler shift as f_d in this section.

A simplified schematic of how we can isolate the Doppler shift is shown in Fig. 1.40. In interpreting this schematic it is important to note that we are dealing with *waves*. As a result, we can use various components to "direct" the waves through our system in a desirable way so that is has a particular "flow" through the schematic. Hence, it is quite different from a traditional circuit schematic where you would consider voltages and currents instantaneously interacting between every component of the circuit.
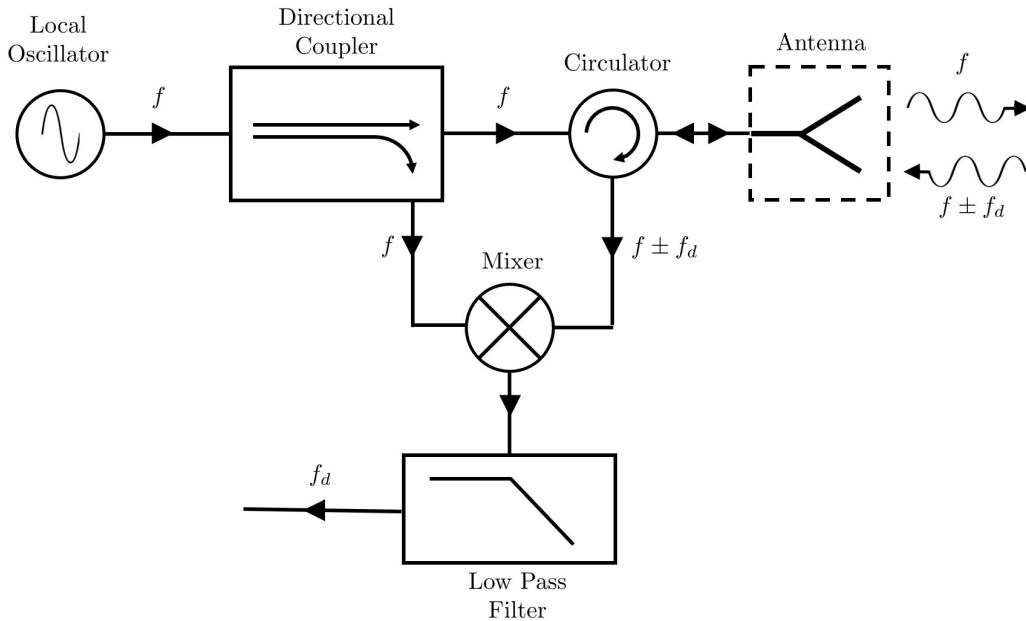
Figure 1.40: A simplified schematic of a continuous wave Doppler radar.

With this in mind, we begin to work our way through the schematic. The first component is the *local oscillator (LO)*, which is our source for the radar. It produces a sinusoidal waveform with frequency f. The LO is connected to a *directional coupler*, which is a special kind of device that allows us to split our wave into two components. One component has a much higher power than the other component by design. Our higher power signal will be what we transmit at our target, while the lower power component will be "saved" in our system to use as a reference signal later on.

We will now follow the path of our higher power signal that we will transmit at the target. It next goes to a device called a *circulator*. This is a kind of non-reciprocal device that provides different behavior depending on which direction a wave approaches it from. As the symbol suggests, it takes a signal entering one port and "directs" it to exit from the next port in a clockwise fashion. Nothing goes out of the port on the counter-clockwise port.

The utility of a circulator is that it allows us to separate the signals that the antenna will transmit and receive.

After the transmit signal flows out of the circulator, it reaches the antenna and is transmitted toward the target. The transmitted wave will reflect off of the target and pick up a Doppler shift in its frequency based off of the speed of the target. This reflected signal propagates back to the antenna and is received. The received signal propagates back toward the circulator. It is directed through the circulator in a clockwise fashion and exits the device.

This received signal is then fed into a *mixer*. At the other port of the mixer we add in our reference signal that we had coming out of the directional coupler. The mixer performs a multiplication operation between the two input signals. The output leaves the mixer from a third port. If we expand the multiplication of our two input signals we get

$$
\begin{aligned}
\cos(f)\cos(f \pm f_d) &= \frac{1}{2}\left[\cos\left(f - (f \pm f_d)\right) + \cos\left(f + (f \pm f_d)\right)\right] \\
&= \frac{1}{2}\left[\cos\left(\mp f_d\right) + \cos\left(2f \pm f_d\right)\right] \\
&= \frac{1}{2}\left[\cos\left(f_d\right) + \cos\left(2f \pm f_d\right)\right].
\end{aligned}
\tag{1.249}
$$

We see that we get one sinusoid oscillating at the Doppler shift frequency and another sinusoid oscillating at a much higher frequency. It is generally much easier to process lower frequency signals (e.g, digitizing them), so we will want to isolate this part only. This can be readily accomplished by using a *low pass filter*. We can then feed this signal into a variety of different processing electronics to make our final detection of the Doppler shift.

1.15.3   Introduction to Synthetic Aperture Radar

As we discussed previously, a synthetic aperture radar (SAR) is a type of imaging radar. It is able to form images due to a combination of clever processing techniques involving Doppler shifts and pulsed radar waveforms. The general idea is shown in Fig. 1.41.

A SAR system generally involves a moving platform that is attempting to image a stationary scene. The relative motion between the transmitter and the various points in the scene being imaged leads to a Doppler shift on the reflected signals. Since the transmitter antenna illuminates a relatively large imaging scene, a single transmit pulse will interact with many points in the scene. The reflected signal at each point of the scene will have some amount of Doppler shift when it is received due to the relative motion of the transmitter (which also acts as the receiver). However, this Doppler shift is not unique, since each point in the scene at a fixed angle from the transmitter will pick up the same amount of relative motion. Hence, we cannot use a Doppler shift alone to try and discriminate between the energy reflected from various points in the illuminated scene.

To provide more unique information, we transmit our waveform in pulses where the waveform is periodically turned off. This allows us to now also time when signals are received. This alone does not provide us with enough unique information to locate where a pulse came from because each point that is the same distance away from the transmitter will produce the same time delay. However, by combining the information from the time delay and the
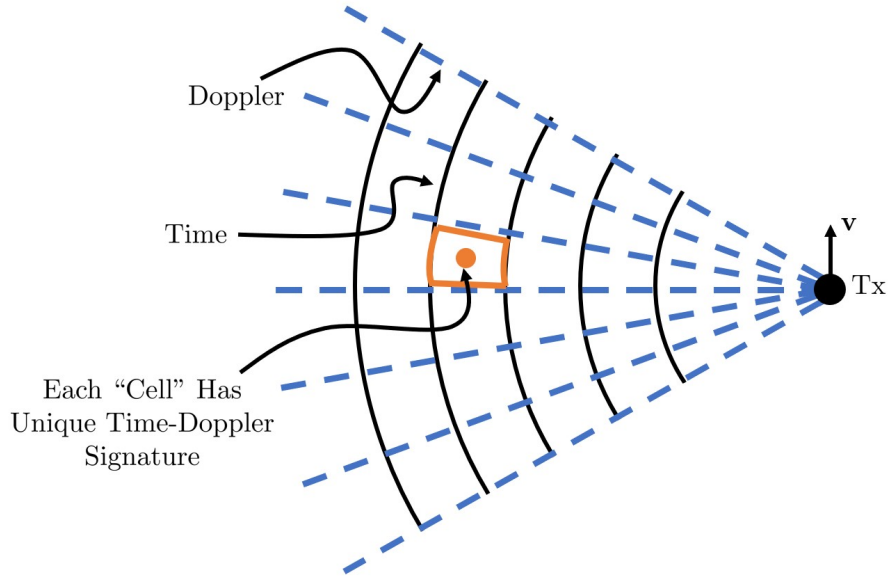
Figure 1.41: Illustration of a single SAR pulse geometry.

Doppler shift we can begin to have unique enough information to determine where a signal came from.

Although this provides useful information, it often is not able to provide a sufficiently high enough resolution in the image to be particularly useful. The key to SAR is by observing the same scene over multiple transmit pulses as the transmitter moves around the extent of the scene. By *coherently processing* the results from all of these different transmitter pulse locations, we can provide a much higher resolution image. This is what leads to the exquisite kinds of images that we looked at previously in the course.

This kind of "multi-look" imaging approach to improve resolution is a very common technique. For instance, a computed tomography (CT) scan does a similar approach with X-rays. The sensor is rotated around the person to take cross-sectional images at various points. The results of all the scans can be combined using various reconstruction algorithms to form a complete image of the object being analyzed.

1.16 Dispersion and Group Velocity

When we originally discussed plane waves propagating in lossless media, we introduced the concept of *phase velocity*. This was the velocity at which a constant phase point on our wave moved. We saw that for an unbounded and lossless medium, the phase velocity could be expressed as

$$v_p = \frac{\omega}{\beta}, \tag{1.250}$$

where β was the phase constant of the wave. In reality, this formula for the phase velocity holds in more general cases so long as we can find the phase constant. However, for the
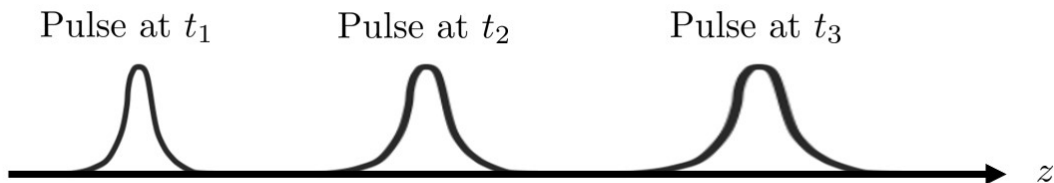
Figure 1.42: Example of a pulse distorting as it propagates through a dispersive medium.

simple case of an unbounded and lossless medium, we had that $\beta = \omega\sqrt{\mu\epsilon}$. Since this β is a linear function of ω, we saw that the phase velocity became a constant number as a function of frequency in this situation. Further, the phase velocity was equal to the speed of light.

Previously, we have seen that many materials will necessarily have some kind of frequency dependence in their permittivity (or permeability). As a result, the phase constant will no longer have a pure linear dependence on frequency in most materials. Later in the course, we will also discuss the case where a geometry causes the phase constant to not linearly depend on frequency. Regardless of the source of the augmented frequency dependence, the end result is that waves with different frequencies will have constant phase points propagate with different phase velocities. This effect is known as *dispersion*, because it causes the shape of a signal composed of multiple frequency components (which is essentially every practical transmitted signal) to *distort/disperse* as it propagates (see Fig. 1.42). As a result, a medium that has a frequency-dependence in its material properties is known as a *dispersive medium*.

Let's now take a closer look at what happens to a signal as it propagates through a dispersive medium. To start, we will assume that we have a relatively *narrowband signal*. This is common in wireless communication systems because the amount of data that needs to be transmitted typically occupies a small bandwidth compared to the *carrier frequency* that the information is transmitted on. Let's consider the simplest case of a *wave packet* composed of two slightly different frequencies equally offset from a particular center frequency (denoted by ω_0).

For this situation, our electric field can be expressed as

$$E(z,t) = E_0 \cos\left([\omega_0 + \Delta\omega]t - [\beta_0 + \Delta\beta]z\right) + E_0 \cos\left([\omega_0 - \Delta\omega]t - [\beta_0 - \Delta\beta]z\right), \quad (1.251)$$

where $\Delta\omega \ll \omega_0$ and $\Delta\beta$ is the corresponding change in β as $\beta : \beta(\omega_0) \rightarrow \beta(\omega_0 \pm \Delta\omega)$. Note that we are implicitly assuming that we can treat β as being a linear function locally to ω_0 when we assume that $\Delta\beta$ is appropriate for both the positive and negative frequency shift. Hence, we can see that we are using a local *Taylor series* representation of β in our discussion here. We restricted our discussion to *narrowband signals* so that this approximation is reasonable.

Now, we can simplify our expression for $E(z,t)$ by using standard trigonometric identities. These give us

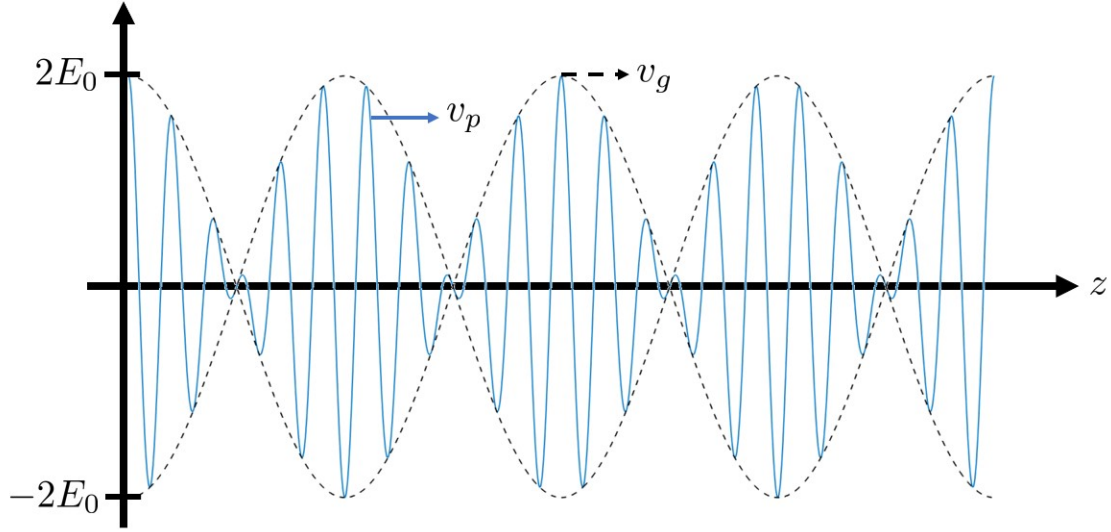$$E(z,t) = 2E_0 \cos(\Delta\omega t - \Delta\beta z)\cos(\omega_0 t - \beta_0 z). \quad (1.252)$$

Figure 1.43: Illustration of phase and group velocities for a simple wave packet. The envelope is shown with dashed lines, while the compete signal is shown with a solid line.

Since by assumption $\Delta\omega \ll \omega_0$, we see that our overall wave packet can be decomposed into two parts. The second cosine function has a fast oscillation at the center frequency of the wave. This response is *modulated* by a slowly-varying function with frequency dependent on $\Delta\omega$. This is usually referred to as an *envelope function*. A simple illustration of this situation is shown in Fig. 1.43.

We now see that the wave *inside the envelope* propagates with a phase velocity of ω_0/β_0. We can also look at the velocity *of the envelope* by setting its argument equal to a constant function. Assuming this is 0 for simplicity, we have that

$$\Delta\omega t - \Delta\beta z = 0, \tag{1.253}$$

from which we can define the *group velocity* as

$$v_g \approx \frac{\Delta\omega}{\Delta\beta}. \tag{1.254}$$

We can take a limit to convert these Δ's into a derivative. We typically write this as

$$\boxed{v_g = \left(\frac{d\beta}{d\omega}\right)^{-1}.} \tag{1.255}$$

We write this in this somewhat odd form because we generally are able to compute β as a function of ω, rather than the other way around. We see that this result matches the intuition that we had about a Taylor series representation from earlier. In particular, we are using a first order (linear) Taylor series result to approximate the local behavior of β to estimate the velocity that the overall wave packet moves at. Although this is usually how v_g is written, it will also sometimes be written as $d\omega/d\beta$ as well.

Let's now consider the particular case of the simple plasma medium we considered previously (i.e., no restoring force or loss). For this situation, we have that

$$\beta = \frac{\omega}{c_0}\sqrt{1 - \left(\frac{\omega_p}{\omega}\right)^2}.$$ (1.256)

We can easily find that the phase velocity is given by

$$v_p = \frac{\omega}{\beta} = \frac{c_0}{\sqrt{1 - \left(\frac{\omega_p}{\omega}\right)^2}}.$$ (1.257)

As discussed previously, we will have wave propagation when $\omega > \omega_p$. We see that this corresponds to a phase velocity that is larger than the speed of light. Is this an issue? It turns out that it isn't, mostly because the phase velocity is not the velocity that information travels at. When a phase point approaches the "edge" of the wave packet envelope it isn't able to move *outside of* or *past* the envelope. Hence, that phase point has reached the end of the packet and no longer is of interest to be tracked.

For a narrowband signal, "information" will propagate at the group velocity. For this situation, we get that

$$v_g = c_0\sqrt{1 - \left(\frac{\omega_p}{\omega}\right)^2},$$ (1.258)

which is less than c_0 for our case of $\omega > \omega_p$. Hence, we have no "issues" with one of the central principles of special relativity (that information can't propagate faster than the speed of light).

There are many "tools" that we can use to help visualize and understand the behavior of dispersive media. One of the common ones is a *dispersion diagram*. These can come in different forms, but generally involve plotting either ω as a function of β or vice-versa. An example of a dispersion diagram (also called an ω-β diagram for this case) for our plasma case is shown in Fig. 1.44.

Dispersion diagrams are particularly useful when we consider more complicated propagation environments. In these situations, it is common that only a few very specially structured "modes" will be valid solutions to Maxwell's equations. We can plot a trace on a single dispersion diagram for each mode we are considering to understand how energy will be able to travel through our medium. This kind of tool is very prevalent in the analysis of different kinds of *periodic structures*, such as metamaterials, photonic crystals, and electromagnetic bandgap structures.

It has become such an important topic that researchers have begun discussing *dispersion engineering*. This is a field aimed at designing different materials to have desirable dispersion characteristics. For instance, one may wish to design a material that compensates for the dispersion of another natural material so that when a signal propagates through both of them the end result is a wave packet that has significantly reduced distortion (see Fig. 1.45). Other areas of interest are making *bandgaps* where no signal is able to propagate through the
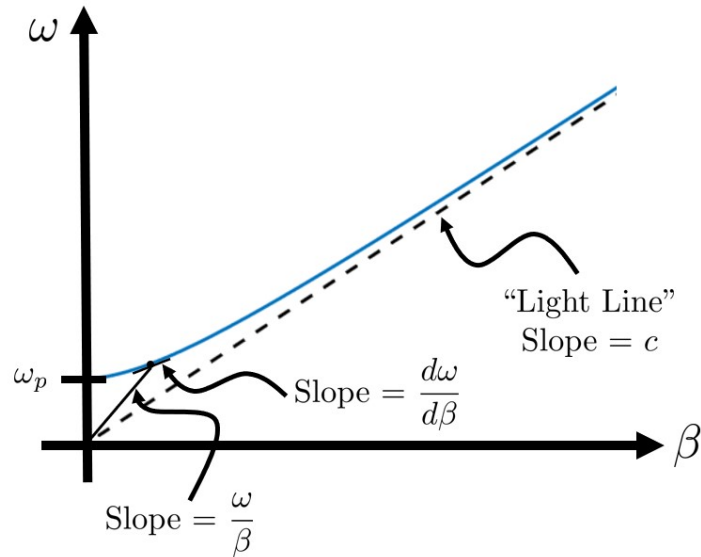
Figure 1.44: Example of a dispersion diagram for a simple plasma medium. The slope of the line running from the origin to the current ω point gives the phase velocity. Similarly, the slope of the tangent line to the curve gives the local group velocity.
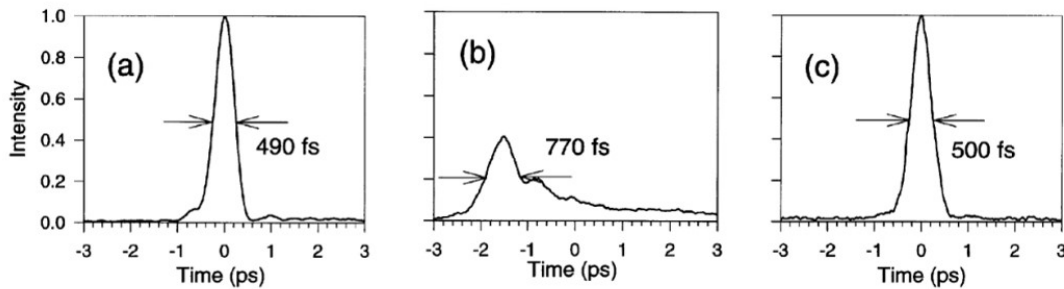
Figure 1.45: Example of dispersion compensation to correct the properties of a pulse. (a) Input pulse, (b) pulse after propagating through a long dispersive medium, and (c) pulse after applying dispersion compensation (images from [9]).

medium. This can be useful for lowering the interactions between nearby electromagnetic devices, such as antennas. This is a technique that has garnered reasonable research interest for designing MIMO (multiple-input, multiple-output) antenna systems, for example (see Fig. 1.46).

1.17 Plane Waves Propagating in Arbitrary Directions

Up to this point, we have typically used relatively simple descriptions of plane waves. In particular, we always considered them to be propagating along a particular axis of our coordinate system so that we could treat our wave equation as a one-dimensional problem. However, there are many practical cases where it is easiest to approach them by considering

Figure 1.46: Example of an electromagnetic bandgap (EBG) structure to lower the electromagnetic interference between nearby antennas. (a) Geometry of the antenna array with EBG, (b) surface currents without and (c) with the EBG (images from [10]).

a plane wave that is propagating in an arbitrary direction, potentially having a component propagating along *all* of our coordinate axes. A good example of when this is useful is in considering the oblique scattering of a plane wave off of a planar interface (we will consider this problem in upcoming lectures).
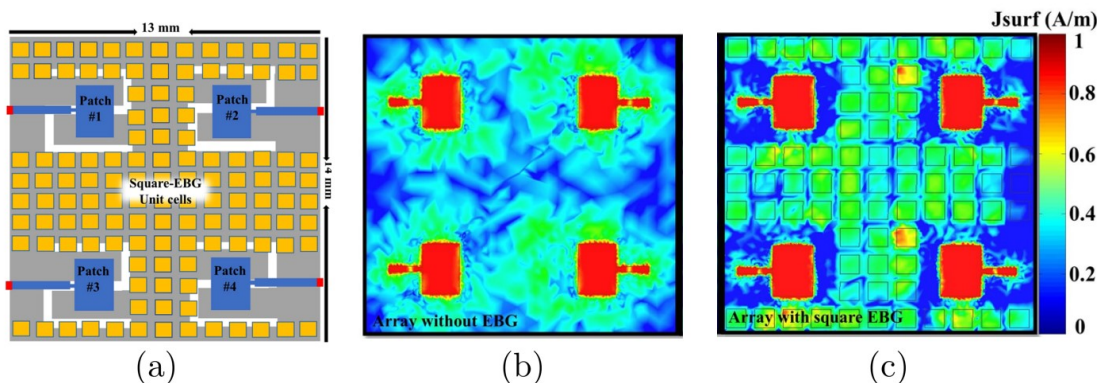
We will now consider how to approach this more general problem. This will also give us our first look at one of the very common approaches at solving complicated partial differential equations – the *separation of variables strategy.* To begin, let's assume that we are trying to solve the vector Helmholtz wave equation in free space. We then have that

$$\nabla^2 \mathbf{E}(x,y,z) + k_0^2 \mathbf{E}(x,y,z) = 0. \tag{1.259}$$

As mentioned earlier, we are going to assume that $\mathbf{E}$ can have a spatial dependence along all 3 of our coordinate axes. In order to make the solution of the problem more tractable, we are going to see if we can find a solution that has the form

$$\Psi(x,y,z) = X(x)Y(y)Z(z), \tag{1.260}$$

where Ψ could be any of E_x, E_y, or E_z. That is, we are looking for a solution that "cleanly" factorizes into a function that only depends on x, another that only depends on y, and yet another that only depends on z. As we will see shortly, searching for a solution of this form leads to a greatly simplified differential equation. This type of solution approach is what is referred to as using the *separation of variables.*

Let's now plug (1.260) into our wave equation. Recalling that we are in Cartesian coordinates and that Ψ is a scalar function, we see that we get

$$\partial_{xx}(XYZ) + \partial_{yy}(XYZ) + \partial_{zz}(XYZ) + k_0^2 XYZ = 0. \tag{1.261}$$

We can simplify our derivatives by noting that they each will only apply to one of the functions in each of the parentheses. Hence, we get

$$YZ\partial_{xx}X + XZ\partial_{yy}Y + XY\partial_{zz}Z + k_0^2 XYZ = 0. \tag{1.262}$$

Next, we divide our equation by XYZ to arrive at the desired form of equation for the separation of variables strategy to be useful. This is

$$\frac{1}{X}\frac{d^2}{dx^2}X + \frac{1}{Y}\frac{d^2}{dy^2}Y + \frac{1}{Z}\frac{d^2}{dz^2}Z + k_0^2 = 0. \tag{1.263}$$

Now, the key result of the separation of variables strategy is that we may now argue that each of the terms on the left-hand side of our equation *must be equal to a constant*. The reason for this is that k_0^2 is a constant and each of the remaining terms all vary independently as functions along different coordinate axes. Hence, there is no way for them to "cancel each other out" in order for the overall summation of terms to equal a constant at every single point in space.

Considering this argument, we can set each of the terms on the left-hand side (with the exception of k_0) of our equation to be equal to different *separation constants*. For instance, we have that

$$\frac{1}{X}\frac{d^2}{dx^2}X = -k_x^2, \tag{1.264}$$

$$\frac{1}{Y}\frac{d^2}{dy^2}Y = -k_y^2, \tag{1.265}$$

$$\frac{1}{Z}\frac{d^2}{dz^2}Z = -k_z^2. \tag{1.266}$$

Let's look at the first equation for simplicity and then note that each of the other equations will behave in the same manner due to the identical form. Now, we can rearrange our terms to get

$$\frac{d^2}{dx^2}X + k_x^2 X = 0. \tag{1.267}$$

This is exactly the kind of equation that we encountered when we were considering our "simple" plane waves! As a result, we see that a valid solution for X will be a plane wave with "wavenumber" being equal to k_x. More explicitly, we have that

$$X(x) = X_0 e^{\pm jk_x x}, \tag{1.268}$$

where X_0 is some constant-valued amplitude factor. Since all of the separated equations will behave in the same way, we get as an overall solution that

$$\Psi(x, y, z) = E_0 e^{\pm jk_x x} e^{\pm jk_y y} e^{\pm jk_z z}, \tag{1.269}$$

where E_0 is again some constant-valued amplitude factor that has absorbed all the constants from the individual solutions. A common shorthand way to write this function out is as

$$\Psi(x, y, z) = E_0 e^{j\mathbf{k}\cdot\mathbf{r}}, \tag{1.270}$$

where $\mathbf{k} = \hat{x}k_x + \hat{y}k_y + \hat{z}k_z$ and $\mathbf{r} = \hat{x}x + \hat{y}y + \hat{z}z$. We often refer to $\mathbf{k}$ as the *wave vector*. It points in the direction of propagation.

To determine the magnitude of the wave vector, it is important to note that k_x, k_y, and k_z are *not independent* of each other. In particular, to be a valid solution we will need to go back to our original equation in (1.263) before we separated them all out. Plugging in our separation constants, we see that

$$k_x^2 + k_y^2 + k_z^2 = k_0^2, \qquad (1.271)$$

where it is important to remember that k_0 still has the same definition as we have used previously, i.e., $k_0 = \omega/c_0$. From this, we can readily see that k_0 will be the correct *magnitude for our wave vector*. We often refer to this kind of equation as a *dispersion relation*. It tells us what are valid combinations of separation constants to still solve Maxwell's equations. Depending on the situation, we can plot aspects of this kind of equation on a dispersion diagram. Hence, the dispersion diagram is plotting out values that correspond to valid solutions to Maxwell's equations.

1.18   Dispersion Relations

We will see repeatedly throughout this course that determining dispersion relations is a very common and fundamental part of solving Maxwell's equations for a variety of situations. To see how it is useful, we will first consider a simple form of Maxwell's equations when dealing with plane wave solutions in Section 1.18.1. Following this, we will see how this approach provides us with a powerful set of tools to determine how waves will propagate in more complicated types of material media in Section 1.18.2.

1.18.1   Basics

In Practice Problem 8 from Section 1.23, you will prove that if $\mathbf{E}$ and $\mathbf{H}$ are of the form

$$\mathbf{E}(\mathbf{r}) = \mathbf{E}_0 e^{j\mathbf{k}\cdot\mathbf{r}}, \quad \mathbf{H}(\mathbf{r}) = \mathbf{H}_0 e^{j\mathbf{k}\cdot\mathbf{r}}, \qquad (1.272)$$

then Maxwell's equations in a source-free region reduce to

$$\mathbf{k} \times \mathbf{E} = \omega\mathbf{B}, \qquad (1.273)$$

$$\mathbf{k} \times \mathbf{H} = -\omega\mathbf{D}, \qquad (1.274)$$

$$\mathbf{k} \cdot \mathbf{D} = 0, \qquad (1.275)$$

$$\mathbf{k} \cdot \mathbf{B} = 0. \qquad (1.276)$$

We can use this form of Maxwell's equations to derive a "wave equation" for plane waves that only involves vector algebra as opposed to vector calculus. As we will see, this will allow us to quickly derive dispersion relations to determine our solutions to Maxwell's equations. The real value of this approach is that it will allow us to analyze fairly complicated situations,

e.g., those involving permittivity and permeability tensors characteristic of complex kinds of media, with a relatively straightforward process.

Let's go about doing this process now. Following a similar process to deriving an actual wave equation, we will start by taking the cross product of "Faraday's law" with the wave vector. This gives us

$$\mathbf{k} \times \mathbf{k} \times \mathbf{E} = \omega\mu\mathbf{k} \times \mathbf{H}. \tag{1.277}$$

We can now use "Ampere's law" to substitute for $\mathbf{k} \times \mathbf{H}$ on the right-hand side. This gives us

$$\mathbf{k} \times \mathbf{k} \times \mathbf{E} = -\omega^2\mu\epsilon\mathbf{E} = -k^2\mathbf{E}. \tag{1.278}$$

We can move everything to one side to get a "wave equation". This is

$$\mathbf{k} \times \mathbf{k} \times \mathbf{E} + k^2\mathbf{E} = 0. \tag{1.279}$$

We can simplify this by using the vector identity that

$$\mathbf{a} \times (\mathbf{b} \times \mathbf{c}) = (\mathbf{a} \cdot \mathbf{c})\mathbf{b} - (\mathbf{a} \cdot \mathbf{b})\mathbf{c}. \tag{1.280}$$

This allows us to write

$$\mathbf{k} \times \mathbf{k} \times \mathbf{E} = (\mathbf{k} \cdot \mathbf{E})\mathbf{k} - (\mathbf{k} \cdot \mathbf{k})\mathbf{E}. \tag{1.281}$$

From our simplified form of "Maxwell's equations" used earlier we conclude that $\mathbf{k} \cdot \mathbf{E} = 0$ from "Gauss' law". Hence, we can write our "wave equation" as

$$(\mathbf{k} \cdot \mathbf{k} - k^2)\mathbf{E} = 0. \tag{1.282}$$

This equation will only be satisfied if we have

$$\boxed{\mathbf{k} \cdot \mathbf{k} = k^2,} \tag{1.283}$$

which matches the dispersion relation we had found earlier since

$$\mathbf{k} \cdot \mathbf{k} = k_x^2 + k_y^2 + k_z^2. \tag{1.284}$$

1.18.2   Anisotropic Media

We will now use the techniques from the previous section to determine the dispersion relation appropriate to a more complicated case. In particular, we will consider the types of plane wave solutions that can exist in a *uniaxial medium*. This is a special kind of *anisotropic medium* that has a different permittivity only along a single axis. Mathematically, we represent this type of medium as a dyad/tensor – which we can write down as a matrix of values in this case. For a uniaxial dielectric medium, we have

$$\bar{\epsilon} = \begin{bmatrix} \epsilon_1 & 0 & 0 \\ 0 & \epsilon_1 & 0 \\ 0 & 0 & \epsilon_z \end{bmatrix}, \quad \mu = \mu_0. \tag{1.285}$$

We can use the same steps that we followed in the homework problem to see that Maxwell's equations can be simplified to

$$\mathbf{k} \times \mathbf{E} = \omega\mu\mathbf{H}, \tag{1.286}$$

$$\mathbf{k} \times \mathbf{H} = -\omega\bar{\epsilon} \cdot \mathbf{E}, \tag{1.287}$$

$$\mathbf{k} \cdot \mathbf{D} = \mathbf{k} \cdot [\bar{\epsilon} \cdot \mathbf{E}] = 0, \tag{1.288}$$

$$\mathbf{k} \cdot \mathbf{B} = \mu\mathbf{k} \cdot \mathbf{H} = 0, \tag{1.289}$$

for this particular situation. Following the process of the previous section, we want to now derive a "wave equation" that we will be able to find a dispersion relation from. However, we must be careful in the steps that we follow to not forget that $\bar{\epsilon}$ is a matrix. As a result, $\bar{\epsilon}$ can't be "factored" out of expressions in simple ways; we must respect the rules of matrix algebra.

To begin, we will start by taking the cross product of "Faraday's law" with the wave vector. This gives us

$$\mathbf{k} \times \mathbf{k} \times \mathbf{E} = \omega\mu\mathbf{k} \times \mathbf{H}. \tag{1.290}$$

We may now use "Ampere's law" to substitute for $\mathbf{k} \times \mathbf{H}$. This gives us

$$\mathbf{k} \times \mathbf{k} \times \mathbf{E} = -\omega^2\mu\bar{\epsilon} \cdot \mathbf{E}. \tag{1.291}$$

In the previous section, we used

$$\mathbf{k} \times \mathbf{k} \times \mathbf{E} = (\mathbf{k} \cdot \mathbf{E})\mathbf{k} - (\mathbf{k} \cdot \mathbf{k})\mathbf{E} \tag{1.292}$$

to simplify our "wave equation". Can we do that this time? The answer is no, because our "Gauss' law" no longer tells us that $\mathbf{k} \cdot \mathbf{E} = 0$, instead it says that $\mathbf{k} \cdot [\bar{\epsilon} \cdot \mathbf{E}] = 0$. Hence, this vector identity won't be able to help us at this point. Instead, we are going to need to just evaluate the vectors in (1.291) and use some concepts from linear algebra to help us find our dispersion relation.

With this in mind, let's look at a special case to simplify our algebra a bit. In particular, we will assume that wave propagates in the x-direction, which correspondingly means that $\mathbf{E}$ is confined to the (y, z)-plane. In this case, we have that

$$\mathbf{k} \times \mathbf{k} \times \mathbf{E} = -\hat{y}k_x^2 E_y - \hat{z}k_x^2 E_z. \tag{1.293}$$

We can then write out (1.291) in a full matrix notation to get some ideas for how to find the dispersion relation. This makes (1.291) become

$$\begin{bmatrix} 0 & 0 & 0 \\ 0 & -k_x^2 & 0 \\ 0 & 0 & -k_x^2 \end{bmatrix} \begin{bmatrix} 0 \\ E_y \\ E_z \end{bmatrix} = -\omega^2\mu \begin{bmatrix} \epsilon_1 & 0 & 0 \\ 0 & \epsilon_1 & 0 \\ 0 & 0 & \epsilon_z \end{bmatrix} \begin{bmatrix} 0 \\ E_y \\ E_z \end{bmatrix}, \tag{1.294}$$

where we have explicitly written that we are considering the case where $E_x = 0$. Now, we have the same vector on each side of our equation, so we can simplify things by combining everything into a single matrix. This gives

$$
\begin{bmatrix} \omega^2\mu\epsilon_1 & 0 & 0 \\ 0 & \omega^2\mu\epsilon_1 - k_x^2 & 0 \\ 0 & 0 & \omega^2\mu\epsilon_z - k_x^2 \end{bmatrix} \begin{bmatrix} 0 \\ E_y \\ E_z \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}. \tag{1.295}
$$

Now, if you recall from linear algebra, the only way for this matrix equation to have a non-trivial solution is if the determinant of the matrix is equal to 0. Since we also have that $E_x = 0$ due to the assumed propagation direction, we are really interested in making sure that the lower 2×2 block matrix has a determinant of 0. This gives us

$$
(\omega^2\mu\epsilon_1 - k_x^2)(\omega^2\mu\epsilon_z - k_x^2) = 0, \tag{1.296}
$$

which can of course be satisfied if one of the expressions in the parentheses are equal to 0. This means that we have two solutions, or in this case, two dispersion relations. These can be summarized as

1. The ordinary wave, with dispersion relation

$$
k_o^2 = \omega^2\mu\epsilon_1 \tag{1.297}
$$

and $E_y \neq 0, E_z = 0$. *(Note: the subscript here is an "o" for ordinary, not a "0".)*

2. The extraordinary wave, with dispersion relation

$$
k_e^2 = \omega^2\mu\epsilon_z \tag{1.298}
$$

and $E_z \neq 0, E_y = 0$.

We see that the propagation characteristics (phase velocity, attenuation constant, etc.) depend on the polarization of the plane wave. This polarization-dependent propagation effect can be utilized to make a number of useful devices.

For example, if we have a material for which ϵ_z is significantly lossy and ϵ_1 is approximately lossless, we can preferentially attenuate one polarization. This effect is used in making *polarized sunglasses* to reduce the glare produced by light reflecting off of surfaces. Later, we will learn more about how the reflection coefficient changes for an electromagnetic wave depending on its polarization with respect to the orientation of the reflecting surface. For now, we will just note that at large incidence angles the horizontally polarized light reflects strongly while the vertically polarized light reflects to a much weaker degree. As a result, polarized sunglasses are designed to block the horizontally polarized light that can be particularly "glaring/bright".

Another type of device that can be made with a uniaxial medium is a *quarter-wave plate*. These are typically made with materials that have both a relatively lossless ϵ_1 and ϵ_z. However, the two permittivities are different enough that the waves associated with the two polarizations will accumulate a different phase shift relative to each other. This effect can
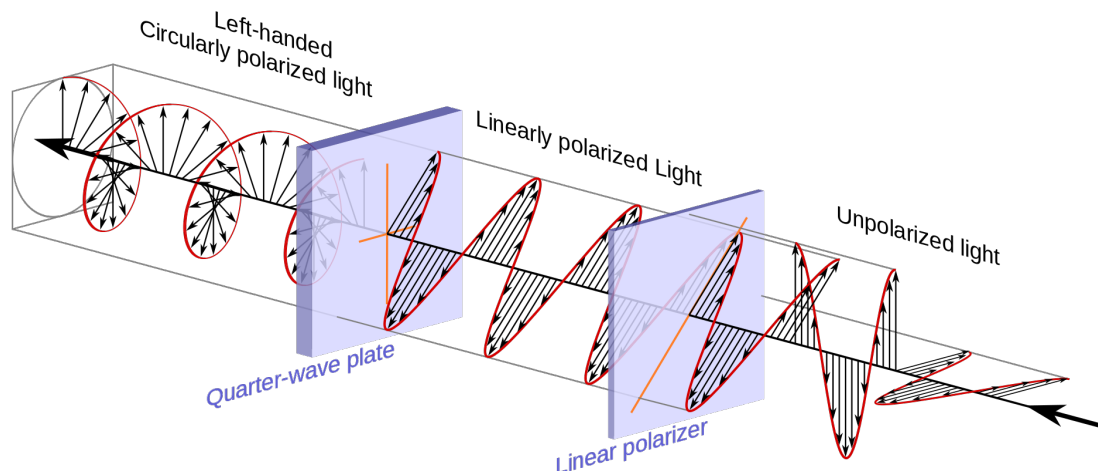
Figure 1.47: Illustration of a sequence of uniaxial materials being used to generate circularly polarized light. First, unpolarized light is passed through a linear polarizer so that the angle of the polarization is known. By keeping this angle at 45° with respect to the axis of the quarter-wave plate circular polarization is generated at the output of the quarter-wave plate (image from Wikipedia [11]).

be used to take a linearly polarized wave and convert it into a circularly polarized wave (see Fig. 1.47).

One final effect we will discuss is *Faraday rotation*. This can occur in *gyrotropic media*, such as magnetized plasmas. Depending on the frequency of a wave, the ionosphere can act as a magnetized plasma due to the influence of the Earth's (approximately) static magnetic field. If we were to account for the presence of the Earth's magnetic field in our analysis of a plasma medium previously (i.e., taking into account the $\mathbf{v} \times \mathbf{B}$ term in the Lorentz force) we would end up with a permittivity tensor that looks like

$$\bar{\epsilon} = \begin{bmatrix} \epsilon_x & jg & 0 \\ -jg & \epsilon_y & 0 \\ 0 & 0 & \epsilon_z \end{bmatrix}. \tag{1.299}$$

Fully analyzing the propagation of waves for this kind of permittivity is beyond the scope of this course, however, the main effect of interest is that the polarization of the wave "rotates" as it propagates through the medium (see Fig. 1.48). For instance, if a linearly polarized wave enters the gyrotropic medium with an electric field polarized along the x-axis it will begin to rotate so that further into the medium the polarization will now be at some angle with respect to the x-axis. This can be an important effect for certain components of satellite communication systems, and is one reason why these systems tend to use at least one circularly polarized antenna in the overall communication link (since the unpredictable rotation of the waves could make the communication systems cross-polarized and no longer able to communicate with one another). More broadly, it is common to include at least one circularly polarized antenna in a communication link regardless of if it will propagate through the ionosphere to ensure that the signal reception is always possible without inconvenient "drop outs" due to the two antennas being oriented incorrectly with respect to each other.
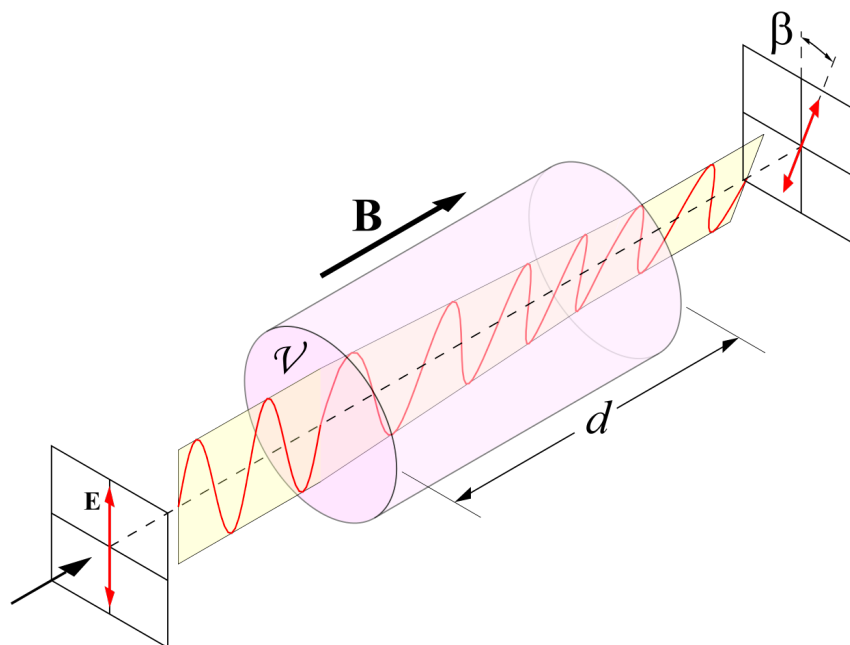
Figure 1.48: Illustration of Faraday rotation. The polarization angle of the electric field progressively changes as the wave propagates through the gyrotropic medium (image from Wikipedia [12]).

The effect of Faraday rotation can also be used in designing various non-reciprocal microwave components. For instance, one common way to design *circulators* is by using ferrite materials to cause Faraday rotation on the waves propagating through the device. Waves that propagate through the device in different directions (e.g., clockwise and counterclockwise) can be carefully controlled so that they destructively interfere at one port of the device and constructively add at the desired output port. Another common non-reciprocal component built using ferrite materials that leverage the Faraday rotation effect are *isolators*. These allow wave propagation through the device in one direction only, and are useful to protect sensitive components from any reflections that may be produced at other locations in a system. This can be important for protecting sensitive nonlinear components, such as oscillators or amplifiers, and to help prevent positive feedback loops from happening unintentionally in a system.

1.19 Plane Wave Scattering from a PEC Sheet (Oblique Incidence)

In a previous lecture, we considered how a plane wave scattered from a planar PEC region when the plane wave was normally incident on the planar boundary. We now turn our attention to how to characterize the scattering that occurs when the plane wave has an *oblique incidence* angle. This means that the unit normal vector of the planar boundary is no longer perfectly aligned with the direction of propagation of the incident plane wave.
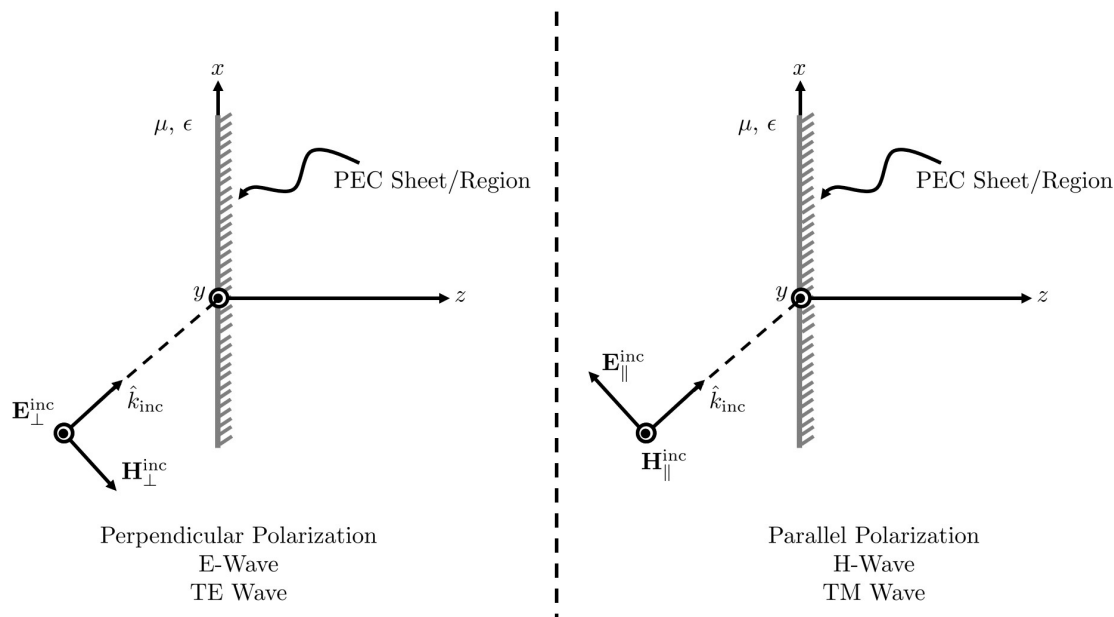
Figure 1.49: Illustration of the perpendicular (left) and parallel (right) polarizations. Other commonly used names are also written next to the illustrations.

1.19.1   Geometry Basics

For this oblique incidence case, we will have two general kinds of situations depending on the orientation of the plane wave's polarization with respect to the planar boundary. There are a number of different terminologies that are commonly used to differentiate between these two cases. We will refer to these as *perpendicular* and *parallel polarizations* in this class, but we will also mention the other names as we discuss the two cases in more detail.

To see why these are the relevant names, we must first understand the concept of the *plane of incidence* for a planar scattering problem. This is defined as the plane that contains the planar surface's unit vector and the unit vector describing the direction of propagation of the plane wave. The naming convention for our two polarizations mentioned above are in reference to the orientation of the electric field and the plane of incidence. Hence, our *perpendicular polarization* case occurs when the direction of the electric field is perpendicular to the plane of incidence. Similarly, our *parallel polarization* case occurs when the direction of the electric field is in the plane of incidence. Incident wave definitions for these two cases are illustrated in Fig. 1.49. We will often denote a perpendicular polarization with the subscript $\perp$ and a parallel polarization with the subscript $\parallel$. As we will see later, the different polarizations will result in very different scattering behavior for the two polarizations.

Let's now turn our attention to writing out the general expressions for the electric and magnetic fields of perpendicular and parallel polarized plane waves. Having a quick reference of these will be quite useful when we go about enforcing boundary conditions to solve for the reflected and transmitted fields for specific scattering problems.
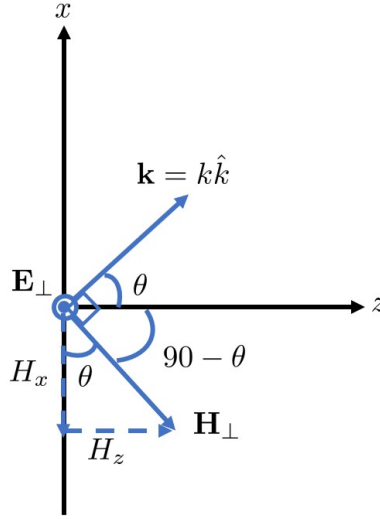
Figure 1.50: Illustration of the perpendicular polarization that we will derive expressions of **E** and **H** from.

Perpendicular Polarization

We will start with the perpendicular polarization case. This is also referred to as the *transverse electric (TE) wave or E-wave case* as well. The name transverse electric originates from the fact that the electric field in this case is *transverse* to the plane of incidence (i.e., it is not contained in it). Let's assume that we have the fields given in Fig. 1.50 to make things concrete and determine the electric and magnetic fields.

For the perpendicular polarization case, we see that we can give a very simple definition for the electric field. In particular, we have that

$$\mathbf{E}_\perp = \hat{y} E_\perp e^{-j\mathbf{k}\cdot\mathbf{r}}. \tag{1.300}$$

We will now go about writing the magnetic field in terms of the electric field using what we know about plane waves. First, we know that the magnetic field must be orthogonal to the electric field and the direction of propagation. Further, we know that the cross product of the electric and magnetic fields (in that order) must give us the direction of propagation. Hence, we can draw out $\mathbf{H}_\perp$ as shown in Fig. 1.50. We also know that the the magnitude of $\mathbf{H}_\perp$ will need to be $E_\perp/\eta$ for a plane wave. We now want to write out the mathematical expression for $\mathbf{H}_\perp$. With a little geometry (illustrated in Fig. 1.50) we get that

$$\mathbf{H}_\perp = \left(-\hat{x}\frac{E_\perp}{\eta}\cos(\theta) + \hat{z}\frac{E_\perp}{\eta}\sin(\theta) \right) e^{-j\mathbf{k}\cdot\mathbf{r}}. \tag{1.301}$$

It can be interesting to compare the *wave impedances* for the different polarizations, which can be roughly thought of as the effective impedance for the plane wave components propagating along the different coordinate axes. Since we have two components of the magnetic field, we can define a wave impedance for each component. In particular, we define

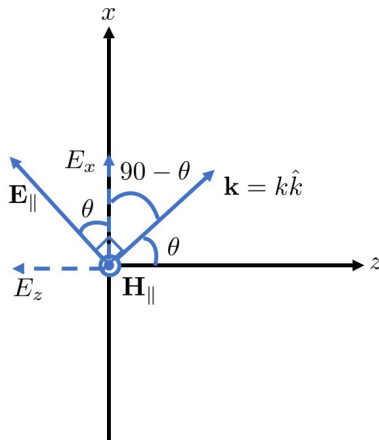$$Z_z = \frac{E_y}{-H_x} = \frac{\eta}{\cos(\theta)}, \tag{1.302}$$

Figure 1.51: Illustration of the parallel polarization that we will derive expressions of **E** and **H** from.

$$Z_x = \frac{E_y}{H_z} = \frac{\eta}{\sin(\theta)}, \tag{1.303}$$

where the sign conventions have been selected to give us positive impedances. The signs can be determined by how the signs on the components of **E** and **H** need to be to lead to a plane wave propagating in the correct direction – i.e., positive z or positive x for this example.

Parallel Polarization

Let's now take a look at the parallel polarization case. This is also referred to as the *transverse magnetic (TM) wave or H-wave case* as well. For this situation we will assume that we have the fields given in Fig. 1.51 to make things concrete and determine the electric and magnetic fields.

For the parallel polarization case, we see that we can give a very simple definition for the magnetic field. In particular, we have that

$$\mathbf{H}_\parallel = \hat{y} H_\parallel e^{-j\mathbf{k}\cdot\mathbf{r}}. \tag{1.304}$$

We will now go about writing the electric field in terms of the magnetic field using what we know about plane waves. First, we know that the electric field must be orthogonal to the magnetic field and the direction of propagation. Further, we know that the cross product of the electric and magnetic fields (in that order) must give us the direction of propagation. Hence, we can draw out $\mathbf{E}_\parallel$ as shown in Fig. 1.51. We also know that the the magnitude of $\mathbf{E}_\parallel$ will need to be $\eta H_\parallel$ for a plane wave. We now want to write out the mathematical expression for $\mathbf{E}_\parallel$. With a little geometry (illustrated in Fig. 1.51) we get that

$$\mathbf{E}_\parallel = \left(\hat{x} \eta H_\parallel \cos(\theta) - \hat{z} \eta H_\parallel \sin(\theta) \right) e^{-j\mathbf{k}\cdot\mathbf{r}}. \tag{1.305}$$

The *wave impedances* for this polarization will be

$$Z_z = \frac{E_x}{H_y} = \eta \cos(\theta), \tag{1.306}$$

88
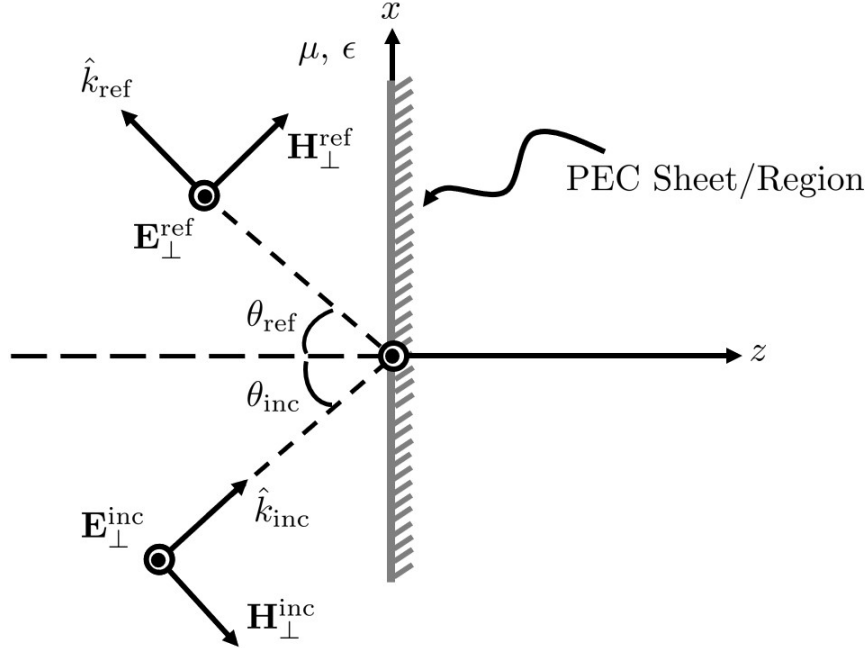
Figure 1.52: Problem setup for the perpendicular polarization oblique incidence scattering from a PEC sheet.

$$Z_x = \frac{-E_z}{H_y} = \eta \sin(\theta), \tag{1.307}$$

where the sign conventions have been selected to give us positive impedances. Comparing these to the wave impedances given in (1.302) and (1.303), we see that the dependence on the angle has changed. As we will see when analyzing the scattering for these two kinds of plane waves, this difference leads to fundamentally different scattering characteristics for the two polarizations.

With the necessary geometric basics determined, we can now turn our attention to solving for the reflected fields produced when a plane wave is obliquely incident on a planar PEC sheet. We will start out by analyzing this for the perpendicular polarization case before then considering the parallel polarization.

1.19.2 Perpendicular Polarization Case

The problem setup for the perpendicular polarization case is shown in Fig. 1.52. We will now go through our three step process for solving these kinds of problems to determine the reflected fields.

In *Step 1*, we identify the different regions of the problem, the boundaries between them, and any "excitations" for the problem (incident fields in this case). For the scenario shown in Fig. 1.52, we have

- Region 1: $z < 0$, homogeneous medium with incident field $\mathbf{E}_\perp^{\text{inc}}$,

- Region 2: $z > 0$, PEC,

- Boundary 1: PEC and the homogeneous medium interface at $z = 0$. The unit normal for boundary conditions is $-\hat{z}$.

Before moving on to Step 2, we first give our explicit expressions for the incident electric and magnetic fields. These are

$$\mathbf{E}_\perp^{\text{inc}} = \hat{y} E_\perp^{\text{inc}} e^{-j\mathbf{k}_{\text{inc}} \cdot \mathbf{r}} = \hat{y} E_\perp^{\text{inc}} e^{-jk(x \sin\theta_{\text{inc}} + z \cos\theta_{\text{inc}})} \tag{1.308}$$

$$\mathbf{H}_\perp^{\text{inc}} = \frac{E_\perp^{\text{inc}}}{\eta} \left(-\hat{x} \cos\theta_{\text{inc}} + \hat{z} \sin\theta_{\text{inc}} \right) e^{-jk(x \sin\theta_{\text{inc}} + z \cos\theta_{\text{inc}})} \tag{1.309}$$

Now, for *Step 2* we need to propose an appropriate general solution in the different regions of the problem. From our experience solving for the normal incidence case of scattering from a PEC sheet, we already know that the fields will need to be 0 in Region 2 due to the infinite conductivity. In Region 1, we expect that we will have a reflected field that will propagate away from the interface. We also have no reason to expect that the PEC sheet will produce new components of polarization directions, so we will assume that it will have a similar mathematical expression and see if that gets us a workable solution. Our particular trial solution will be based on the illustration in Fig. 1.52. For this field, we have

$$\mathbf{E}_\perp^{\text{ref}} = \hat{y} \Gamma_\perp E_\perp^{\text{inc}} e^{-j\mathbf{k}_{\text{ref}} \cdot \mathbf{r}} = \hat{y} \Gamma_\perp E_\perp^{\text{inc}} e^{-jk(x \sin\theta_{\text{ref}} - z \cos\theta_{\text{ref}})}, \tag{1.310}$$

$$\mathbf{H}_\perp^{\text{ref}} = \Gamma_\perp \frac{E_\perp^{\text{inc}}}{\eta} \left(\hat{x} \cos\theta_{\text{ref}} + \hat{z} \sin\theta_{\text{ref}} \right) e^{-jk(x \sin\theta_{\text{ref}} - z \cos\theta_{\text{ref}})}. \tag{1.311}$$

We now move on to *Step 3*, where we use the boundary conditions of our problem to solve for the unknown constants in our trial solutions. For this case, we need to solve for $\Gamma_\perp$ and θ_r. From our previous experience, we know that the boundary condition on the electric field is probably a good one to use for this situation. This tells us that the total tangential electric field at the PEC boundary ($z = 0$) needs to be 0. Mathematically, we express this condition as

$$-\hat{z} \times \left(\mathbf{E}_\perp^{\text{inc}}(z = 0) + \mathbf{E}_\perp^{\text{ref}}(z = 0) \right) = 0. \tag{1.312}$$

From this, we get

$$E_\perp^{\text{inc}} e^{-jkx \sin\theta_{\text{inc}}} + \Gamma_\perp E_\perp^{\text{inc}} e^{-jkx \sin\theta_{\text{ref}}} = 0. \tag{1.313}$$

For this equation to have a possible solution for all x, we see that we will need the remaining parts of the exponential functions to be equal to one another. This is known as a *phase matching condition*, which gives us the additional equation

$$\sin\theta_{\text{inc}} = \sin\theta_{\text{ref}}. \tag{1.314}$$

From this, we immediately see that $\theta_{\text{inc}} = \theta_{\text{ref}}$, that is the angle of reflection matches the angle of incidence. This is also often referred to as *Snell's law of reflection*.

Using Snell's law, we can simplify (1.313) to now be

$$E_\perp^{\text{inc}} + \Gamma_\perp E_\perp^{\text{inc}} = 0. \tag{1.315}$$

From this equation, we see that $\Gamma_\perp = -1$.

We can use this solution to now write out the total field in Region 1. This gives us

$$\mathbf{E}_1 = \mathbf{E}_\perp^{\text{inc}} + \mathbf{E}_\perp^{\text{ref}} = \hat{y} E_\perp^{\text{inc}} (e^{-jkz\cos\theta_{\text{inc}}} - e^{jkz\cos\theta_{\text{inc}}}) e^{-jkx\sin\theta_{\text{inc}}}, \tag{1.316}$$

which we can simplify to be

$$\mathbf{E}_1 = -\hat{y} 2j E_\perp^{\text{inc}} \sin(kz\cos\theta_{\text{inc}}) e^{-jkx\sin\theta_{\text{inc}}}. \tag{1.317}$$

We can dissect this solution to see that it looks like we have a standing wave that is formed in the z-direction and a kind of plane wave propagating in the x-direction with a phase velocity of

$$v_p = \frac{\omega}{k\sin\theta_{\text{inc}}} = \frac{c}{\sin\theta_{\text{inc}}}. \tag{1.318}$$

However, this is no longer a uniform plane wave propagating in the x-direction. This is because the amplitude of the overall wave changes as a function of z.

1.19.3 Parallel Polarization Case

Let's now look at the parallel polarization case. The problem setup for this scenario is shown in Fig. 1.53. We will now go through our three step process for solving these kinds of problems to determine the reflected fields.

As usual, in *Step 1*, we identify the different regions of the problem, the boundaries between them, and any "excitations" for the problem (incident fields in this case). For the scenario shown in Fig. 1.53, we have

- Region 1: $z < 0$, homogeneous medium with incident field $\mathbf{H}_\parallel^{\text{inc}}$,

- Region 2: $z > 0$, PEC,

- Boundary 1: PEC and the homogeneous medium interface at $z = 0$. The unit normal for boundary conditions is $-\hat{z}$.

Before moving on to Step 2, we first give our explicit expressions for the incident electric and magnetic fields. These are

$$\mathbf{H}_\parallel^{\text{inc}} = \hat{y} H_\parallel^{\text{inc}} e^{-j\mathbf{k}_{\text{inc}}\cdot\mathbf{r}} = \hat{y} H_\parallel^{\text{inc}} e^{-jk(x\sin\theta_{\text{inc}}+z\cos\theta_{\text{inc}})} \tag{1.319}$$

$$\mathbf{E}_\parallel^{\text{inc}} = \eta H_\parallel^{\text{inc}} (\hat{x}\cos\theta_{\text{inc}} - \hat{z}\sin\theta_{\text{inc}}) e^{-jk(x\sin\theta_{\text{inc}}+z\cos\theta_{\text{inc}})} \tag{1.320}$$

Now, for *Step 2* we need to propose an appropriate general solution in the different regions of the problem. From our experience solving for the normal incidence case of scattering from
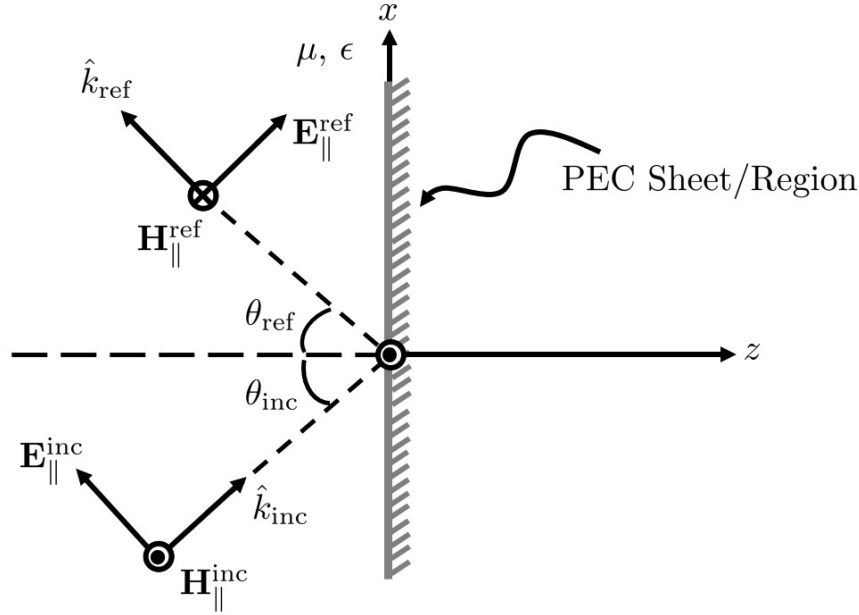
Figure 1.53: Problem setup for the parallel polarization oblique incidence scattering from a PEC sheet.

a PEC sheet, we already know that the fields will need to be 0 in Region 2 due to the infinite conductivity. In Region 1, we expect that we will have a reflected field that will propagate away from the interface. We also have no reason to expect that the PEC sheet will produce new components of polarization directions, so we will assume that it will have a similar mathematical expression and see if that gets us a workable solution. Our particular trial solution will be based on the illustration in Fig. 1.53. For this field, we have

$$\mathbf{H}_\parallel^{\text{ref}} = -\hat{y}\Gamma_\parallel H_\parallel^{\text{inc}} e^{-j\mathbf{k}_{\text{ref}} \cdot \mathbf{r}} = -\hat{y}\Gamma_\parallel H_\parallel^{\text{inc}} e^{-jk(x\sin\theta_{\text{ref}} - z\cos\theta_{\text{ref}})} \tag{1.321}$$

$$\mathbf{E}_\parallel^{\text{ref}} = \eta\Gamma_\parallel H_\parallel^{\text{inc}}\left(\hat{x}\cos\theta_{\text{ref}} + \hat{z}\sin\theta_{\text{ref}}\right) e^{-jk(x\sin\theta_{\text{ref}} - z\cos\theta_{\text{ref}})} \tag{1.322}$$

We now move on to *Step 3*, where we use the boundary conditions of our problem to solve for the unknown constants in our trial solutions. For this case, we need to solve for $\Gamma_\parallel$ and θ_{ref}. From our previous experience, we know that the boundary condition on the electric field is probably a good one to use for this situation. This tells us that the total tangential electric field at the PEC boundary $(z = 0)$ needs to be 0. Mathematically, we express this condition as

$$-\hat{z} \times (\mathbf{E}_\parallel^{\text{inc}}(z=0) + \mathbf{E}_\parallel^{\text{ref}}(z=0)) = 0. \tag{1.323}$$

From this, we get

$$\eta H_\parallel^{\text{inc}}\cos\theta_{\text{inc}} e^{-jkx\sin\theta_{\text{inc}}} + \eta\Gamma_\parallel H_\parallel^{\text{inc}}\cos\theta_{\text{ref}} e^{-jkx\sin\theta_{\text{ref}}} = 0. \tag{1.324}$$

We see that this equation can only be solved if we have the same *phase matching condition* as we did for the perpendicular polarization case. In particular, this gives us

$$\sin\theta_{\text{inc}} = \sin\theta_{\text{ref}}, \tag{1.325}$$

from which we see that $\theta_{\text{inc}} = \theta_{\text{ref}}$. Hence, we still get the same *Snell's law of reflection* for the parallel and perpendicular cases.

Using Snell's law, we can simplify (1.324) to now be

$$1 + \Gamma_{\|} = 0, \tag{1.326}$$

so that we again have $\Gamma_{\|} = -1$.

We can use this solution to now write out the total field in Region 1. This is easiest to interpret by looking at the magnetic field, which gives us

$$\mathbf{H}_1 = \mathbf{H}_{\|}^{\text{inc}} + \mathbf{H}_{\|}^{\text{ref}} = \hat{y} H_{\|}^{\text{inc}}(e^{-jkz\cos\theta_{\text{inc}}} + e^{jkz\cos\theta_{\text{inc}}})e^{-jkx\sin\theta_{\text{inc}}}, \tag{1.327}$$

which we can simplify to be

$$\mathbf{H}_1 = \hat{y} 2 H_{\|}^{\text{inc}} \cos(kz\cos\theta_{\text{inc}})e^{-jkx\sin\theta_{\text{inc}}}. \tag{1.328}$$

We can dissect this solution to see that it looks like we have a standing wave that is formed in the z-direction and a kind of plane wave propagating in the x-direction with a phase velocity of

$$v_p = \frac{\omega}{k\sin\theta_{\text{inc}}} = \frac{c}{\sin\theta_{\text{inc}}}. \tag{1.329}$$

However, this is no longer a uniform plane wave propagating in the x-direction. This is because the amplitude of the overall wave changes as a function of z.

1.20 Normal Incidence at a Homogeneous Half Space

We will now consider what happens when a plane wave is normally incident on a half space made of a homogeneous material. What this means is that at some location, say $z = 0$, we abruptly transition from one material (e.g., free space) to a homogeneous medium characterized by a constant ϵ and μ for all locations with $z > 0$. We are interested in determining what the electric and magnetic fields will need to be at all locations in space for Maxwell's equations to be satisfied. To do this, let's go ahead and apply our three step process for solving these kinds of problems. We will consider the specific case shown in Fig. 1.54.

In *Step 1*, we identify the different regions of the problem, the boundaries between regions, and the specified fields (if any). For this problem, we have the following setup.

- Region 1: $z < 0$, free space (μ_0, ϵ_0), $\mathbf{E}^{\text{inc}}(z) = \hat{x} E^{\text{inc}} e^{-jk_0 z}$

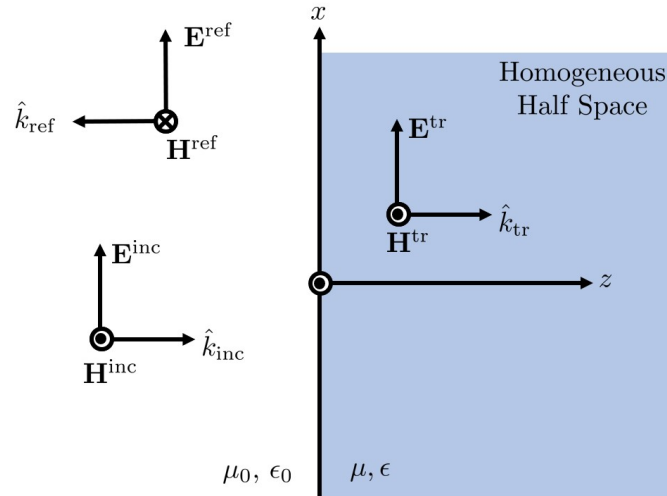- Region 2: $z > 0$, homogeneous medium with μ, ϵ

Figure 1.54: Illustration of the problem for scattering from a homogeneous half space. Incident field is part of the specification of the problem, the reflected and transmitted fields are postulated solutions that we need to solve for to satisfy Maxwell's equations.

- Boundary 1: $z = 0$, between free space and the homogeneous medium, the unit normal for boundary conditions is $-\hat{z}$

We now move on to *Step 2* where we propose reasonable trial solutions for the different regions of the problem. From our previous experience, we expect that we may need to have a reflected field in Region 1. Since Region 2 is not some kind of perfect conductor, we have a feeling that we will also now need to consider the possibility of some amount of the plane wave being able to *transmit* into Region 2 and continue propagating. Considering this, we will write out our reflected and transmitted field trial solutions as

$$\mathbf{E}^{\text{ref}}(z) = \hat{x} E^{\text{ref}} e^{jk_0 z}, \tag{1.330}$$

$$\mathbf{E}^{\text{tr}}(z) = \hat{x} E^{\text{tr}} e^{-jkz}. \tag{1.331}$$

Let's now move to *Step 3* where we enforce our boundary conditions to determine the different constants we need to finish solving our problem. To begin, we will use the boundary condition on the electric field. This was

$$\hat{n} \times (\mathbf{E}_1 - \mathbf{E}_2) = 0. \tag{1.332}$$

From this equation, we are able to deduce that at $z = 0$ we will need

$$E^{\text{inc}} + E^{\text{ref}} = E^{\text{tr}}. \tag{1.333}$$

We only know E^{inc} in this equation, so we are going to need to determine another equation to solve this problem.

Which equation should we use? The best approach is to use the boundary condition on the tangential components of the magnetic field. For this situation, there is no surface current density so we will have that

$$\hat{n} \times (\mathbf{H}_1 - \mathbf{H}_2) = 0. \tag{1.334}$$

Obviously, we will need to evaluate our magnetic fields for each of the different plane waves we are working with in order to use this equation. We can quickly find that

$$\mathbf{H}^{\text{inc}}(z) = \hat{y}\frac{E^{\text{inc}}}{\eta_0}e^{-jk_0 z}, \tag{1.335}$$

$$\mathbf{H}^{\text{ref}}(z) = -\hat{y}\frac{E^{\text{ref}}}{\eta_0}e^{jk_0 z}, \tag{1.336}$$

$$\mathbf{H}^{\text{tr}}(z) = \hat{y}\frac{E^{\text{tr}}}{\eta}e^{-jkz}. \tag{1.337}$$

We can plug these into our boundary condition to get

$$\frac{E^{\text{inc}}}{\eta_0} - \frac{E^{\text{ref}}}{\eta_0} = \frac{E^{\text{tr}}}{\eta}. \tag{1.338}$$

We can now go about solving (1.333) and (1.338) to find what E^{ref} and E^{tr} should be. We end up finding that

$$E^{\text{ref}} = \frac{\eta - \eta_0}{\eta + \eta_0}E^{\text{inc}}, \tag{1.339}$$

$$E^{\text{tr}} = \frac{2\eta}{\eta + \eta_0}E^{\text{inc}}. \tag{1.340}$$

We can now define a *reflection coefficient* as

$$\Gamma = \frac{E^{\text{ref}}}{E^{\text{inc}}} = \frac{\eta - \eta_0}{\eta + \eta_0} \tag{1.341}$$

and a *transmission coefficient* as

$$T = \frac{E^{\text{tr}}}{E^{\text{inc}}} = \frac{2\eta}{\eta + \eta_0} \tag{1.342}$$

(note that your textbook uses the symbol τ for this).

This reflection coefficient is the same as what we have encountered before. It is a dimensionless quantity that ranges from -1 to 1, and in general, can be a complex-valued number. It is defined as the ratio of the reflected and incident fields, and as a result is a function of position. If someone doesn't specify where the reflection coefficient is referenced to for a

problem like this they typically will have it referenced to the interface between the regions, which is what we did here.

We also now have a new concept of a *transmission coefficient*. This plays a similar role to the reflection coefficient, but helps us understand how much of the wave is transmitted into the new medium. To understand what values the transmission coefficient can range over, we can go back to (1.333) to see that at $z = 0$ we have

$$1 + \Gamma = T. \tag{1.343}$$

Hence, we see that $0 < T < 2$.

This raises an interesting question. If, for instance, we have a Γ of 0.9, we get a T of 1.9. How is this possible? How are we getting a "stronger" plane wave in medium 2 than what we had in our incident field? The answer can be put somewhat simply as *there is no law of conservation of voltages*. We only need to have power conserved, but the electric field (which is like a voltage) is not a conserved quantity. To see what happens with the power of the waves, let's take a look at the time-average Poynting vectors at either side of $z = 0$.

In Region 1, we have that

$$\begin{aligned}
\mathbf{E}_1(z) &= \mathbf{E}^{\text{inc}}(z) + \mathbf{E}^{\text{ref}}(z) = \hat{x} E^{\text{inc}}(e^{-jk_0 z} + \Gamma e^{jk_0 z}) \\
&= \hat{x} E^{\text{inc}} e^{-jk_0 z} \left[1 + \Gamma e^{j2k_0 z} \right]
\end{aligned} \tag{1.344}$$

$$\begin{aligned}
\mathbf{H}_1(z) &= \mathbf{H}^{\text{inc}}(z) + \mathbf{H}^{\text{ref}}(z) = \hat{y} \frac{E^{\text{inc}}}{\eta_0} (e^{-jk_0 z} - \Gamma e^{jk_0 z}) \\
&= \hat{y} \frac{E^{\text{inc}}}{\eta_0} e^{-jk_0 z} \left[1 - \Gamma e^{j2k_0 z} \right].
\end{aligned} \tag{1.345}$$

The time-average Poynting vector is then

$$\begin{aligned}
\mathbf{S}_{1,\text{avg}} &= \frac{1}{2} \text{Re}\{\mathbf{E}_1 \times \mathbf{H}_1^*\} \\
&= \hat{z} \frac{|E^{\text{inc}}|^2}{2\eta_0} \text{Re}\{(1 + \Gamma e^{j2k_0 z})(1 - \Gamma^* e^{-j2k_0 z})\} \\
&= \hat{z} \frac{|E^{\text{inc}}|^2}{2\eta_0} \text{Re}\{(1 - |\Gamma|^2) + |\Gamma|(e^{j(2k_0 z + \angle\Gamma)} - e^{-j(2k_0 z + \angle\Gamma)})\} \\
&= \hat{z} \frac{|E^{\text{inc}}|^2}{2\eta_0} \text{Re}\{(1 - |\Gamma|^2) + j|\Gamma|2\sin(2kz + \angle\Gamma)\} \\
&= \hat{z} \frac{|E^{\text{inc}}|^2}{2\eta_0} (1 - |\Gamma|^2).
\end{aligned} \tag{1.346}$$

In Region 2, we have only one plane wave so the math is much simpler. This gives us that

$$\mathbf{S}_{2,\text{avg}} = \hat{z} \frac{|E^{\text{inc}}|^2}{2\eta} |T|^2. \tag{1.347}$$

We can equate our time-average Poynting vectors at the boundary to see that power will only be conserved if

$$1 - |\Gamma|^2 = \frac{\eta_0}{\eta}|T|^2. \tag{1.348}$$

We can plug in our expressions for Γ and T to see that this equation does hold with our definitions as is. Hence, there is no issue with the fact that $1 + \Gamma = T$. Power is still conserved, it is simply a situation that voltages and electric fields are not!

We can also take a closer look at our total electric field in Region 1 to get a better understanding of what is occurring. In particular, we can be clever and add and subtract $\Gamma \exp(-jk_0 z)$ to write it as

$$
\begin{aligned}
\mathbf{E}_1(z) &= \hat{x}E^{\text{inc}}\left[(1+\Gamma)e^{-jk_0 z} + \Gamma\big(e^{jk_0 z} - e^{-jk_0 z}\big)\right] \\
&= \hat{x}E^{\text{inc}}\left[Te^{-jk_0 z} + \Gamma j2\sin(k_0 z)\right].
\end{aligned}
\tag{1.349}
$$

From this, we see that we can interpret the overall field as having a component that is a propagating plane wave and another part that is a standing wave. The overall field will then show interference between these two different kinds of waves.

1.21 Normal Incidence at Multiple Planar Interfaces

We will now consider the case where we have a plane wave incident on a "slab" of finite thickness. As we will see, this problem has many similarities with what we just considered. However, because we now have multiple interfaces between regions of differing constitutive parameters we will have a bit more algebra to do in order to solve the overall problem. However, the overall solution process will not need to change much at all from the previous scenario. The specific illustration of the problem we will consider is shown in Fig. 1.55.

Let's now begin our three step process to solving this kind of problem. We begin with *Step 1*, where we note the different regions of the problem, interfaces, and any excitations.

- Region 1: $z < 0$, homogeneous medium with ϵ_1, μ_1, incident field of $\mathbf{E}^{\text{inc}}$

- Region 2: $0 < z < d$, homogeneous medium with ϵ_2, μ_2

- Region 3: $z > d$, homogeneous medium with ϵ_3, μ_3

- Boundary 1: $z = 0$, between Region 1 and Region 2, unit normal is $-\hat{z}$

- Boundary 2: $z = d$, between Region 2 and Region 3, unit normal is $-\hat{z}$

We now move to *Step 2* where we specify appropriate general solutions in each region of the problem. As with the previous case, we expect we will need a reflected plane wave in Region 1. In Region 2, we expect to have a transmitted field as with the previous problem. However, because there is another interface at $z = d$ there is also the possibility of a reflected field in Region 2. Hence, we will need two possible solutions in this region, namely, forward- and backward-propagating plane waves. Now, in Region 3 we should only have a transmitted field that is propagating in the $+z$ direction. Summarizing, we have
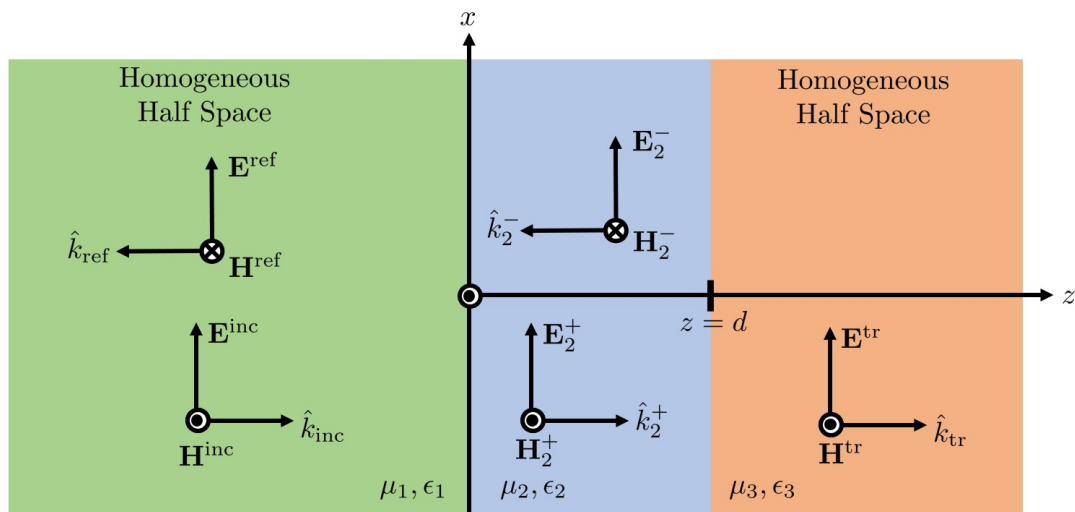
Figure 1.55: Illustration of the problem for scattering from multiple planar interfaces. The incident field is part of the specification of the problem, the reflected and transmitted fields are postulated solutions that we need to solve for to satisfy Maxwell's equations.

- Region 1

    - $\mathbf{E}^{\text{inc}} = \hat{x} E^{\text{inc}} e^{-jk_1 z}$

    - $\mathbf{H}^{\text{inc}} = \hat{y} \dfrac{E^{\text{inc}}}{\eta_1} e^{-jk_1 z}$

    - $\mathbf{E}^{\text{ref}} = \hat{x} \Gamma E^{\text{inc}} e^{jk_1 z}$

    - $\mathbf{H}^{\text{ref}} = -\hat{y} \Gamma \dfrac{E^{\text{inc}}}{\eta_1} e^{jk_1 z}$

- Region 2

    - $\mathbf{E}_2^+ = \hat{x} E_2^+ e^{-jk_2 z}$

    - $\mathbf{H}_2^+ = \hat{y} \dfrac{E_2^+}{\eta_2} e^{-jk_2 z}$

    - $\mathbf{E}_2^- = \hat{x} E_2^- e^{jk_2 z}$

    - $\mathbf{H}_2^- = -\hat{y} \dfrac{E_2^-}{\eta_2} e^{jk_2 z}$

- Region 3

    - $\mathbf{E}^{\text{tr}} = \hat{x} T E^{\text{inc}} e^{-jk_3 (z-d)}$

    - $\mathbf{H}^{\text{tr}} = \hat{y} \dfrac{T E^{\text{inc}}}{\eta_3} e^{-jk_3 (z-d)}$

We can now move to *Step 3* to solve for the unknown quantities by applying our boundary conditions at the interfaces. We see that we have 4 quantities to solve for $(\Gamma, T, E_2^+, E_2^-)$,

so we will need 4 equations. We can get these by ensuring the continuity of the tangential electric and magnetic fields at the two boundaries of the problem. Let's start at Boundary 2 and work our way back. At Boundary 2, we have that

$$E_2^+ e^{-jk_2 d} + E_2^- e^{jk_2 d} = TE^{\text{inc}}, \tag{1.350}$$

$$\frac{E_2^+}{\eta_2} e^{-jk_2 d} - \frac{E_2^-}{\eta_2} e^{jk_2 d} = \frac{TE^{\text{inc}}}{\eta_3}. \tag{1.351}$$

Similarly, at Boundary 1 we have that

$$E^{\text{inc}}(1 + \Gamma) = E_2^+ + E_2^-, \tag{1.352}$$

$$\frac{E^{\text{inc}}}{\eta_1}(1 - \Gamma) = \frac{E_2^+}{\eta_2} - \frac{E_2^-}{\eta_2}. \tag{1.353}$$

We can solve (1.350) to (1.353) for all of the unknown constants. However, the algebra becomes rather tedious if one attempts to do it directly. In some situations, certain strategies can be used to quickly determine certain results without requiring solving the full problem. For instance, if one is only concerned with the overall reflection produced by the slab geometry, one can use some tricks from analyzing transmission lines to determine that

$$\Gamma = \frac{\Gamma_{12} + \Gamma_{23} e^{-j2k_2 d}}{1 + \Gamma_{12}\Gamma_{23} e^{-j2k_2 d}}, \tag{1.354}$$

where

$$\Gamma_{ij} = \frac{\eta_j - \eta_i}{\eta_j + \eta_i} \tag{1.355}$$

is an *intrinsic reflection coefficient.* This is the reflection coefficient that would exist between these two media if they were both semi-infinite in extent. The expression in (1.354) appears quite frequently when analyzing systems with multiple interfaces, and is a good one to remember.

1.22   Oblique Incidence at a Homogeneous Half Space

We will now consider what happens when a plane wave approaches a homogeneous half space with oblique incidence. For this situation, we will have different behavior depending on the polarization of the incident field. Hence, we will need to consider the perpendicular and parallel polarization cases separately.

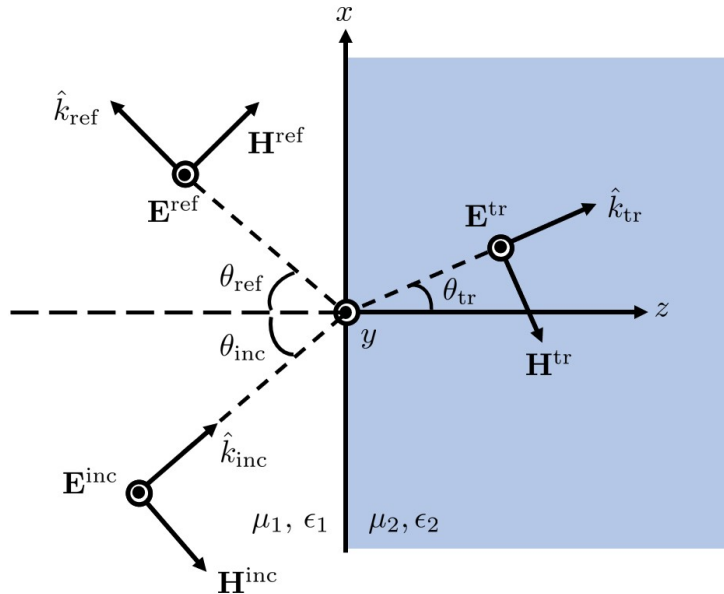Figure 1.56: Illustration of the problem for perpendicular polarization oblique incidence scattering from a homogeneous half space. The incident field is part of the specification of the problem, the reflected and transmitted fields are postulated solutions that we need to solve for to satisfy Maxwell's equations.

1.22.1 Perpendicular Polarization

We will start with the perpendicular polarization case, which is illustrated in Fig. 1.56. We will now go through the three step process of solving this problem. In *Step 1*, we note the different regions of the problem, interfaces, and excitations. For this problem, we have the following setup.

- Region 1: $z < 0$, homogeneous medium with ϵ_1, μ_1, and

$$\mathbf{E}^{\text{inc}}(\mathbf{r}) = \hat{y} E_i e^{-jk_1(x\sin\theta_{\text{inc}} + z\cos\theta_{\text{inc}})} \tag{1.356}$$

$$\mathbf{H}^{\text{inc}}(\mathbf{r}) = \frac{E_i}{\eta_1}(-\hat{x}\cos\theta_{\text{inc}} + \hat{z}\sin\theta_{\text{inc}})e^{-jk_1(x\sin\theta_{\text{inc}} + z\cos\theta_{\text{inc}})} \tag{1.357}$$

- Region 2: $z > 0$, homogeneous medium with ϵ_2, μ_2

- Boundary 1: $z = 0$, between free space and the homogeneous medium, the unit normal for boundary conditions is $-\hat{z}$

We now move on to *Step 2* where we propose reasonable trial solutions for the different regions of the problem. From our previous experience, we expect that we may need to have a reflected field in Region 1. Similarly, we now know we need to also consider the possibility of some amount of the plane wave being able to *transmit* into Region 2 and

100

continue propagating. Considering this, we will write out our reflected and transmitted field trial solutions as

$$\mathbf{E}^{\mathrm{ref}}(\mathbf{r}) = \hat{y} E_r e^{jk_1(z\cos\theta_{\mathrm{ref}} - x\sin\theta_{\mathrm{ref}})}, \tag{1.358}$$

$$\mathbf{H}^{\mathrm{ref}}(\mathbf{r}) = \frac{E_r}{\eta_1}(\hat{x}\cos\theta_{\mathrm{ref}} + \hat{z}\sin\theta_{\mathrm{ref}})e^{jk_1(z\cos\theta_{\mathrm{ref}} - x\sin\theta_{\mathrm{ref}})}, \tag{1.359}$$

$$\mathbf{E}^{\mathrm{tr}}(\mathbf{r}) = \hat{y} E_t e^{-jk_2(x\sin\theta_{\mathrm{tr}} + z\cos\theta_{\mathrm{tr}})}, \tag{1.360}$$

$$\mathbf{H}^{\mathrm{tr}}(\mathbf{r}) = \frac{E_t}{\eta_2}(-\hat{x}\cos\theta_{\mathrm{tr}} + \hat{z}\sin\theta_{\mathrm{tr}})e^{-jk_2(x\sin\theta_{\mathrm{tr}} + z\cos\theta_{\mathrm{tr}})}. \tag{1.361}$$

We can plug (1.356), (1.358), and (1.360) into our boundary condition for the tangential components of the electric field at $z = 0$ to get

$$E_i e^{-jk_1 x\sin\theta_{\mathrm{inc}}} + E_r e^{-jk_1 x\sin\theta_{\mathrm{ref}}} = E_t e^{-jk_2 x\sin\theta_{\mathrm{tr}}}. \tag{1.362}$$

The only way for this equation to have a possible solution is for the arguments of all of the exponential functions to match, i.e., our *phase matching* condition. Hence, we have that

$$\boxed{k_1\sin\theta_{\mathrm{inc}} = k_1\sin\theta_{\mathrm{ref}} = k_2\sin\theta_{\mathrm{tr}}.} \tag{1.363}$$

We have already seen the first equality before. This was *Snell's law of reflection*. The next equality is referred to as *Snell's law of refraction*. It tells us that the angle of transmission will depend on the angle of incidence and the ratio of material properties between the two media. More explicitly, we see that

$$\theta_{\mathrm{tr}} = \arcsin\left(\sqrt{\frac{\mu_1\epsilon_1}{\mu_2\epsilon_2}}\sin\theta_{\mathrm{ref}}\right). \tag{1.364}$$

After applying our phase matching condition, we can simplify (1.362) to be

$$E_i + E_r = E_t. \tag{1.365}$$

We need another equation to solve for the two unknowns. We can get this from the continuity of the tangential magnetic field. This will give us

$$\frac{1}{\eta_1}(E_i - E_r)\cos\theta_{\mathrm{inc}} = \frac{E_t}{\eta_2}\cos\theta_{\mathrm{tr}}. \tag{1.366}$$

We can solve our two equations to determine reflection and transmission coefficients of

$$\Gamma_\perp|_{z=0} = \frac{E_r}{E_i} = \frac{(\eta_2/\cos\theta_{\mathrm{tr}}) - (\eta_1/\cos\theta_{\mathrm{inc}})}{(\eta_2/\cos\theta_{\mathrm{tr}}) + (\eta_1/\cos\theta_{\mathrm{inc}})}, \tag{1.367}$$
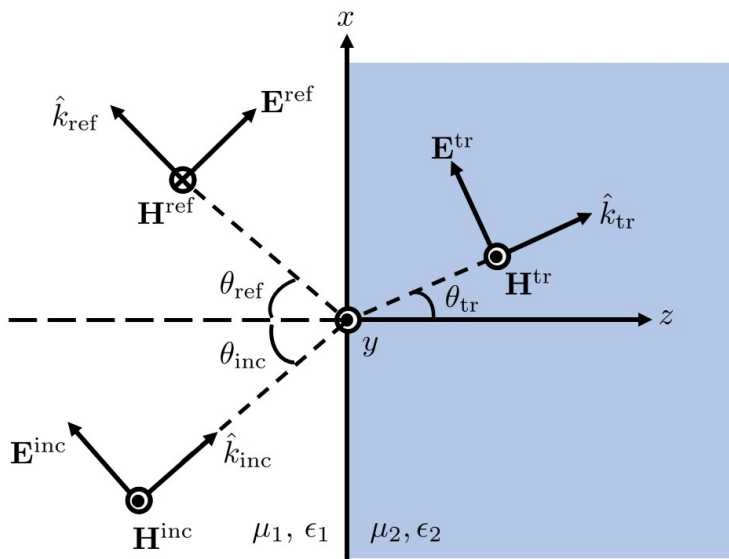
Figure 1.57: Illustration of the problem for parallel polarization oblique incidence scattering from a homogeneous half space. The incident field is part of the specification of the problem, the reflected and transmitted fields are postulated solutions that we need to solve for to satisfy Maxwell's equations.

$$T_\perp|_{z=0} = \frac{E_t}{E_i} = \frac{2(\eta_2/\cos\theta_{\text{tr}})}{(\eta_2/\cos\theta_{\text{tr}}) + (\eta_1/\cos\theta_{\text{inc}})}. \tag{1.368}$$

We also still have that

$$1 + \Gamma_\perp = T_\perp. \tag{1.369}$$

Hence, we see that our expressions look very close to the normal incidence case. We just need to scale the impedance of the different regions by $1/\cos\theta$ where θ is the angle of the wave's propagation in that medium. You should recall that we saw this same behavior when defining the wave impedance for perpendicular polarized plane waves in a previous lecture.

We will now turn our attention to solving the parallel polarization case. Afterward, we will compare the two situations and also look at a couple special cases of particular interest.

1.22.2 Parallel Polarization

The parallel polarization case is illustrated in Fig. 1.57.

We will now go through the three step process of solving this problem. In *Step 1*, we note the different regions of the problem, interfaces, and excitations. For this problem, we have the following setup.

- Region 1: $z < 0$, homogeneous medium with ϵ_1, μ_1, and

$$\mathbf{E}^{\text{inc}}(\mathbf{r}) = E_i(\hat{x}\cos\theta_{\text{inc}} - \hat{z}\sin\theta_{\text{inc}})e^{-jk_1(x\sin\theta_{\text{inc}}+z\cos\theta_{\text{inc}})} \tag{1.370}$$

$$\mathbf{H}^{\text{inc}}(\mathbf{r}) = \frac{E_i}{\eta_1}\hat{y}e^{-jk_1(x\sin\theta_{\text{inc}}+z\cos\theta_{\text{inc}})} \tag{1.371}$$

102

- Region 2: $z > 0$, homogeneous medium with ϵ_2, μ_2

- Boundary 1: $z = 0$, between free space and the homogeneous medium, the unit normal for boundary conditions is $-\hat{z}$

We now move on to *Step 2* where we propose reasonable trial solutions for the different regions of the problem. From our previous experience, we expect that we may need to have a reflected field in Region 1. Similarly, we now know we need to also consider the possibility of some amount of the plane wave being able to *transmit* into Region 2 and continue propagating. Considering this, we will write out our reflected and transmitted field trial solutions as

$$\mathbf{E}^{\text{ref}}(\mathbf{r}) = E_r(\hat{x}\cos\theta_{\text{ref}} + \hat{z}\sin\theta_{\text{ref}})e^{-jk_1(x\sin\theta_{\text{ref}} - z\cos\theta_{\text{ref}})}, \tag{1.372}$$

$$\mathbf{H}^{\text{ref}}(\mathbf{r}) = -\frac{E_r}{\eta_1}\hat{y}e^{-jk_1(x\sin\theta_{\text{ref}} - z\cos\theta_{\text{ref}})}, \tag{1.373}$$

$$\mathbf{E}^{\text{tr}}(\mathbf{r}) = E_t(\hat{x}\cos\theta_{\text{tr}} - \hat{z}\sin\theta_{\text{tr}})e^{-jk_2(x\sin\theta_{\text{tr}} + z\cos\theta_{\text{tr}})}, \tag{1.374}$$

$$\mathbf{H}^{\text{tr}}(\mathbf{r}) = \frac{E_t}{\eta_2}\hat{y}e^{-jk_2(x\sin\theta_{\text{tr}} + z\cos\theta_{\text{tr}})}. \tag{1.375}$$

We can plug (1.370), (1.372), and (1.374) into our boundary condition for the tangential component of the electric field at $z = 0$ to get

$$E_i\cos\theta_{\text{inc}}e^{-jk_1 x\sin\theta_{\text{inc}}} + E_r\cos\theta_{\text{ref}}e^{-jk_1 x\sin\theta_{\text{ref}}} = E_t\cos\theta_{\text{tr}}e^{-jk_2 x\sin\theta_{\text{tr}}}. \tag{1.376}$$

The only way for this equation to have a possible solution is for the arguments of all of the exponential functions to match, i.e., our *phase matching* condition. Hence, we have that

$$\boxed{k_1\sin\theta_{\text{inc}} = k_1\sin\theta_{\text{ref}} = k_2\sin\theta_{\text{tr}},} \tag{1.377}$$

which is exactly what we saw for the perpendicular polarization case.

After applying our phase matching condition, we can simplify (1.376) to be

$$(E_i + E_r)\cos\theta_{\text{inc}} = E_t\cos\theta_{\text{tr}}. \tag{1.378}$$

We need another equation to solve for the two unknowns. We can get this from the continuity of the tangential magnetic field. This will give us

$$\frac{1}{\eta_1}(E_i - E_r) = \frac{E_t}{\eta_2}. \tag{1.379}$$

We can solve our two equations to determine reflection and transmission coefficients of

$$\Gamma_{\|}|_{z=0} = \frac{E_r}{E_i} = \frac{\eta_2\cos\theta_{\text{tr}} - \eta_1\cos\theta_{\text{inc}}}{\eta_2\cos\theta_{\text{tr}} + \eta_1\cos\theta_{\text{inc}}}, \tag{1.380}$$

$$T_{\parallel}|_{z=0} = \frac{E_t}{E_i} = \frac{2\eta_2 \cos\theta_{\text{inc}}}{\eta_2 \cos\theta_{\text{tr}} + \eta_1 \cos\theta_{\text{inc}}}. \tag{1.381}$$

We also have that

$$1 + \Gamma_{\parallel} = T_{\parallel}(\cos\theta_{\text{tr}}/\cos\theta_{\text{inc}}). \tag{1.382}$$

We see that our expressions are still somewhat similar to the normal incidence case, but to a lesser degree than the perpendicular polarization expressions were. As a result, it is vital that we remember that for oblique incidence cases on a homogeneous half space that our parallel and perpendicular polarizations exhibit different behavior, and so, have different formulas that must be used in evaluating certain properties.

1.22.3   Comparison of Parallel and Perpendicular Polarizations

We have just seen that the reflection and transmission coefficients for the perpendicular and parallel polarization cases are different from each other. It is natural to then ask, are there any practical consequences of this?

There are of course many practical consequences of this, but one that you have probably experienced is in the use of polarized sunglasses. These are specifically designed to filter out "horizontally" polarized light. That is, light which is in parallel to the surface of the earth. The reason for this is that much of the sunlight that is getting reflected off of surfaces outside will be from "horizontal" surfaces (like the ground). If we assume that the sunlight itself isn't preferentially polarized (which turns out to be the case), we can assume that we will have approximately an equal amount of light approaching the ground in parallel and perpendicular polarizations. Thinking about the geometry of the problem, we see that the "horizontal" polarization will match the perpendicular polarization case, while a "vertical" polarization would correspond to the parallel polarization case.

We can write down the two reflection coefficients again to compare them. That is,

$$\Gamma_{\perp} = \frac{(\eta_2/\cos\theta_{\text{tr}}) - (\eta_1/\cos\theta_{\text{inc}})}{(\eta_2/\cos\theta_{\text{tr}}) + (\eta_1/\cos\theta_{\text{inc}})}, \tag{1.383}$$

$$\Gamma_{\parallel} = \frac{\eta_2 \cos\theta_{\text{tr}} - \eta_1 \cos\theta_{\text{inc}}}{\eta_2 \cos\theta_{\text{tr}} + \eta_1 \cos\theta_{\text{inc}}}. \tag{1.384}$$

With the exception of $\theta_{\text{inc}} = 0$, we can quickly see that as θ_{inc} gets larger $\Gamma_{\perp}$ will tend to get larger while $\Gamma_{\parallel}$ will get smaller. To see this more clearly, we have plotted the magnitude of the reflection coefficients for the case where $\eta_1 = \eta_0$ and $\eta_2 = \eta_0/\sqrt{10}$ in Fig. 1.58. From this, we see that at relatively large incidence angles the perpendicular reflection coefficient is much larger than the parallel reflection coefficient. Hence, by blocking the horizontal polarization (that corresponds to the perpendicular polarization) a pair of polarized sunglasses can block much of the glare that would come about due to reflected sunlight.
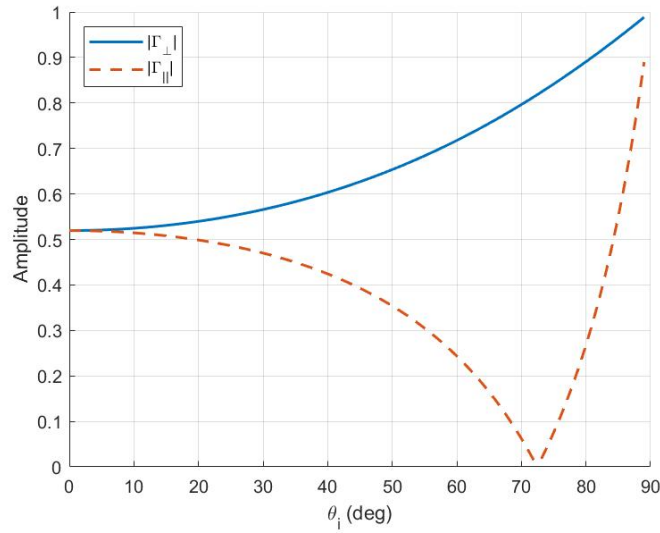
Figure 1.58: Comparison of parallel and perpendicular polarization reflection coefficients.

1.22.4   Special Angles

Moving on, there are two kinds of angles that are typically considered to be "special". These are the *critical angle* and the *Brewster angle*. The critical angle is the angle where the reflection coefficient is equal to 1 (i.e., perfect reflection) and the Brewster angle is the exact opposite – i.e., the angle where no reflection occurs. Let's take a look at what situations can allow for this to happen.

Critical Angle

We will begin with the *critical angle*. This typically occurs when we have a wave in a "denser" medium attempting to propagate into a "less dense" medium (density is a term sometimes used to refer to the material properties, a denser medium has a higher constitutive parameter). For many practical cases of interest, we won't have magnetic materials present so that we can assume that $\epsilon_1 > \epsilon_2$. The quickest way to find the critical angle is from Snell's law of refraction. We will have total reflection when $\theta_{\mathrm{tr}} = \pi/2$, since in this case no wave actually propagates into the less dense medium. Hence, we want to find

$$k_1 \sin \theta_c = k_2 \sin(\pi/2) = k_2. \tag{1.385}$$

We can write this as

$$\theta_c = \arcsin \left(\sqrt{\frac{\epsilon_2}{\epsilon_1}} \right). \tag{1.386}$$

We will continue to have a complete reflection for any situation where the incident angle is larger than the critical angle. Although there is complete reflection, we do have some fields "leak" into medium 2. These are in the form of *evanescent waves* that do not propagate, but instead exponentially attenuate. It should be noted that because there are still fields in
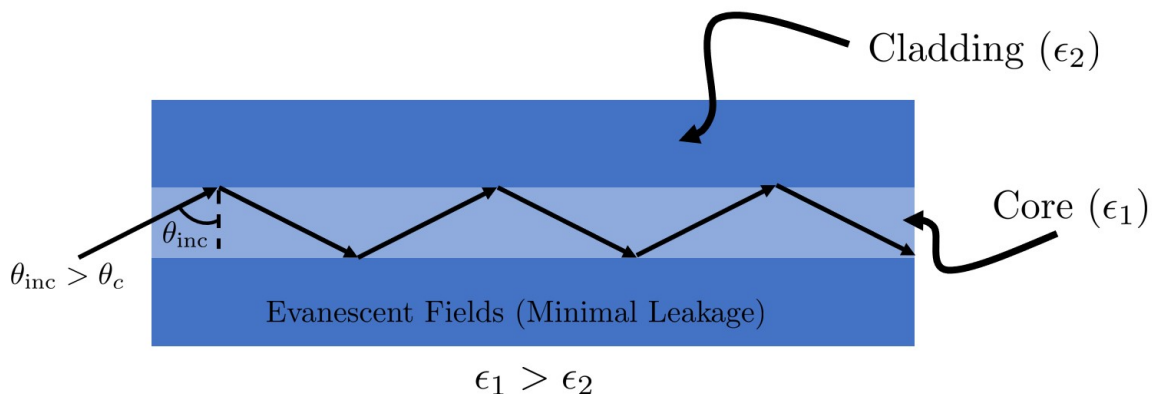
105

Figure 1.59: Illustration of total internal reflection within an optical fiber (illustration shows a cross section of a cylindrical fiber).

this "evanescent region" they can interact with objects near to the boundary between the propagating and evanescent regions. This is a general kind of wave phenomena, and is at the core of why *tunneling* of waves is possible. For instance, a number of semiconductor electronics are able to operate due to electrons tunneling through potential barriers. In the region of the tunneling barrier the electron wave is evanescent, but as long as the potential barrier is sufficiently thin it is possible for the electron to exit out the other side and begin propagating again.

A practical electromagnetic device that uses the *total internal reflection* effect provided by the critical angle is an optical fiber. The *core* of the fiber is made of a denser material than the *cladding* that surrounds the core. Light can then be input into the fiber. Any light that is propagating with an angle larger than the critical angle of the overall fiber geometry will be able to propagate down the fiber. You can picture a ray bouncing back and forth from the different walls of the fiber core, as shown in Fig. 1.59.

Brewster Angle

Next, we will discuss the *Brewster angle* for perpendicular and parallel polarizations. This is the angle at which no reflection occurs. *(Note: it is important to recognize that just because there is no reflection, it does not necessarily mean that the transmission coefficient is equal to 1, e.g., see (1.382)!)* Looking at the numerators of both of the reflection coefficients, we see that this seems like it should potentially be possible if we can find an appropriate combination of η_1, η_2, and θ_{inc} for the two terms to cancel each other. Denoting the Brewster angle for perpendicular polarization as $\theta_{B\perp}$, we want to find when

$$\frac{\eta_2}{\cos\theta_{\mathrm{tr}}} = \frac{\eta_1}{\cos\theta_{B\perp}}. \tag{1.387}$$

We can use Snell's law of refraction to write $\cos\theta_{\mathrm{tr}}$ as

$$\cos\theta_{\mathrm{tr}} = \sqrt{1 - \sin^2\theta_{\mathrm{tr}}} = \sqrt{1 - \frac{\mu_1\epsilon_1}{\mu_2\epsilon_2}\sin^2\theta_{\mathrm{inc}}}. \tag{1.388}$$

We can plug this into (1.387) to eventually find that

$$\sin^2 \theta_{B\perp} = \frac{1 - \mu_1\epsilon_2/(\mu_2\epsilon_1)}{1 - (\mu_1/\mu_2)^2}. \tag{1.389}$$

The first thing we can notice is that if $\mu_1 = \mu_2$ then the denominator will become infinite and we won't have a Brewster angle. Many practical materials are *dielectrics*, whose magnetic properties match that of free space (i.e., $\mu_r = 1$). Hence, for any interface between two dielectric regions there will be no Brewster angle for the perpendicular polarization. There are certain combinations of material properties where the Brewster angle can exist for a perpendicular polarization, however, these rarely are found in natural materials. As a result, it is *extremely rare* to encounter a Brewster angle for a perpendicular polarization in practical situations.

Although this is the case for perpendicular polarization, the story is rather different for a parallel polarization. The same set of steps can be used to find that the Brewster angle for a parallel polarization is

$$\sin^2 \theta_{B\parallel} = \frac{1 - \mu_2\epsilon_1/(\mu_1\epsilon_2)}{1 - (\epsilon_1/\epsilon_2)^2}. \tag{1.390}$$

It is extremely common to be able to have an interface between two regions with different permittivities, so the denominator is rarely going to have a situation where it would become infinite. As a result, it is generally possible to find a Brewster angle for the parallel polarization between two dielectric materials. For instance, this occurred in Fig. 1.58 at $72.5°$.

1.23 Practice Problems

1. (a) Show that Gauss' laws of electricity and magnetism can be derived from the other equations in Maxwell's equations.
   Hint: You will need to use some of the vector calculus identities discussed in the lecture notes to complete this question.

   (b) Explain why your derivation only works for time-varying systems.

2. (a) Prove that

   $$\nabla \cdot (f\mathbf{A}) = f\nabla \cdot \mathbf{A} + \mathbf{A} \cdot \nabla f \tag{1.391}$$

   for differentiable scalar f and vector $\mathbf{A}$. You may use Cartesian coordinates for simplicity.

   (b) Provide an intuitive explanation for what each term on the right-hand side of this identity means.

3. (a) Prove that

   $$\nabla \times (f\mathbf{A}) = f\nabla \times \mathbf{A} + (\nabla f) \times \mathbf{A} \tag{1.392}$$

   for differentiable scalar f and vector $\mathbf{A}$. You may use Cartesian coordinates for simplicity.

(b) Provide an intuitive explanation for what each term on the right-hand side of this identity means.

4. Verify in Cartesian coordinates that

$$\nabla \times (\nabla f) = 0 \tag{1.393}$$

and

$$\nabla \cdot (\nabla \times \mathbf{A}) = 0, \tag{1.394}$$

5. A medium where the permittivity and permeability may vary as a function of position is called an *inhomogeneous medium.*

(a) Derive a time domain wave equation for the electric field in an inhomogeneous medium where both ϵ and μ are functions of position.

(b) Derive a time domain wave equation for the magnetic field in an inhomogeneous medium where both ϵ and μ are functions of position.

Hint: start from Maxwell's equation written in terms of $\mathbf{E}$, $\mathbf{H}$, $\mathbf{D}$, and $\mathbf{B}$ and be careful to not factor ϵ and μ out of any spatial derivatives. Instead, think about if there are simple ways you can rearrange the equations prior to taking the spatial derivatives to simplify the rest of the derivation. Alternatively, follow the procedure given in class and use appropriate vector calculus identities (you can find the appropriate one in HW1) to find a form where Faraday's and Ampere's laws can be substituted into one another. These two approaches will yield different looking wave equations, but both are equivalent and valid.

6. The electric field of a uniform plane wave in free space is given by

$$\mathbf{E}(x,t) = \hat{z}\, 34 \cos(4.8\pi \times 10^9 t - kx) \ [\mathrm{mV/m}]. \tag{1.395}$$

(a) What direction is this plane wave traveling in?

(b) Find the frequency f, the wavenumber k, and the wavelength λ of this wave.
Note: do not forget to label the units of your answer. These reminders will not be made again.

(c) Sketch $E_z(x,t)$ as a function of t at $x = 0$ and $x = \lambda/4$.

(d) Sketch $E_z(x,t)$ as a function of x at $t = 0$ and $t = \pi/\omega$.

7. The magnetic field of a Ku-band satellite broadcast signal propagating in free space is given by

$$\mathbf{H}(z,t) = \hat{x}\, H_0 \cos(\omega t - 261.8z) \ [\mathrm{mA/m}]. \tag{1.396}$$

(a) Find f and λ for this wave.

(b) Find $\mathbf{E}(z,t)$.

Point of Interest: Ku-band is the name of a portion of the electromagnetic spectrum that contains the frequency of this broadcast signal. The exact range of frequencies contained in Ku-band varies depending on whose definition you use (e.g., commercial, government, military, NATO, etc.), but the rough size of the band corresponds to the frequencies that can be correctly transmitted down a standard-sized rectangular waveguide (in particular, a WR-62 waveguide). We will learn more about why a rectangular waveguide can only correctly transmit a portion of the electromagnetic spectrum in the waveguides section of the course.

8. For a time harmonic plane wave propagating in a homogeneous medium, both $\mathbf{E}$ and $\mathbf{H}$ vary with the exponential "phase factor" $\exp[-j\mathbf{k} \cdot \mathbf{R}]$, where $\mathbf{k} = \hat{x}k_x + \hat{y}k_y + \hat{z}k_z$ and $\mathbf{R} = \hat{x}x + \hat{y}y + \hat{z}z$. Assuming that $\mathbf{E} = \hat{E}_0 \exp[-j\mathbf{k} \cdot \mathbf{R}]$ and $\mathbf{H} = \hat{H}_0 \exp[-j\mathbf{k} \cdot \mathbf{R}]$, show by direct substitution that Maxwell's equations in a source-free region reduce to

$$\mathbf{k} \times \mathbf{E} = \omega\mu\mathbf{H}, \tag{1.397}$$

$$\mathbf{k} \times \mathbf{H} = -\omega\epsilon\mathbf{E}, \tag{1.398}$$

$$\mathbf{k} \cdot \mathbf{E} = 0, \tag{1.399}$$

$$\mathbf{k} \cdot \mathbf{H} = 0. \tag{1.400}$$

9. Although there is no physical evidence that magnetic current densities exist, the effects of certain real source distributions can be conveniently modeled by treating them as "effective" magnetic current densities. This is the case, for example, in the analysis of aperture antennas or various kinds of coupled waveguides. If we denote the effective magnetic current density as $\mathbf{M}$, Maxwell's equations become

$$\nabla \times \mathbf{H} = \partial_t\mathbf{D} + \mathbf{J}, \tag{1.401}$$

$$\nabla \times \mathbf{E} = -\partial_t\mathbf{B} - \mathbf{M}, \tag{1.402}$$

$$\nabla \cdot \mathbf{D} = \rho, \tag{1.403}$$

$$\nabla \cdot \mathbf{B} = 0. \tag{1.404}$$

Note: do not confuse $\mathbf{M}$ with the magnetization intensity of a material that we discussed previously; instead, think of it like a counterpart to $\mathbf{J}$.

(a) Derive (1.404) from (1.402). This will require you to specify a particular condition on a derivative of $\mathbf{M}$ as a "postulate" of our modified theory. What does this postulate say about the spatial variation of $\mathbf{M}$ (recall our discussions on the spatial variation of $\mathbf{B}$)?

(b) Derive the integral forms of the modified Maxwell's equations given in (1.401) to (1.404).

(c) Using your answer from 1(b), derive the boundary condition that is applicable to (1.402) assuming that there is a surface magnetic current density present at the boundary between the two regions.

10. A 5 GHz y-polarized electric field is propagating in the $+z$-direction through a low-loss dielectric medium which has $\epsilon'_r = 3.5$. Answer the following questions given that at the same instant in time at $z = 0$ [m] the phasor of the electric field is $10\angle 40°$ [V/m] and that at $z = 0.1$ [m] the phasor of the electric field is $8.22\angle-3.34°$ [V/m].

(a) What is the attenuation constant in dB/m?

(b) What is the phase constant in rad/m? Compare this to the wavenumber that would occur in a lossless dielectric medium with $\epsilon_r = 3.5$.
Hint: If your two answers are significantly different you have done something wrong.

(c) Write down an expression for the instantaneous magnetic field that is valid for all $z \geq 0$.

11. A 5 GHz z-polarized plane wave is propagating in the $-y$-direction in a lossy dielectric medium with $\epsilon_r = 3 - j2.5$. If the electric field phasor at $y = 0$ is $4\angle 30°$ V/m, write down the instantaneous expression for the magnetic field.

12. A 10 GHz electric field for a plane wave in a lossy dielectric medium is given by

$$\mathbf{E} = \left(\hat{x}E_x + j\hat{z}E_z\right)e^{\alpha y + j\beta y}.$$

If the relative permittivity of the medium is $10 - j6$, find an expression for the instantaneous magnetic field. Make sure to compute explicit numeric values for all quantities except the arbitrary complex-valued amplitudes E_x and E_z.

13. In this problem, we will again consider a form of Maxwell's equations that includes an effective magnetic current density, $\mathbf{M}$. Considering this, Maxwell's equations are given by (1.401) to (1.404). Note: you **do not** need to consider conduction currents for this problem.

(a) Derive Poynting's theorem in the time domain for the modified Maxwell's equations given in (1.401) to (1.404).

(b) Repeat this derivation in the frequency domain.

14. Assume that we have a surface magnetic current density in free space located at $x = 0$ that covers the entire (y, z)-plane. Our expression for this current sheet is

$$\mathbf{M}_s(\mathbf{r}) = \begin{cases} \hat{z}M_0 \cos(\omega t - \varphi_0), & x = 0, \\ 0, & \text{elsewhere.} \end{cases} \tag{1.405}$$

Using the boundary conditions appropriate to the modified Maxwell's equations given in (1.401) to (1.404), determine what $\mathbf{E}$ and $\mathbf{H}$ are equal to in all regions of space.

15. Previously, we discussed the concept of a perfect electric conductor (PEC). We saw that the boundary condition between free space and a PEC region was that the tangential component of the electric field was equal to 0. In similarity to this, we can also define the concept of a *perfect magnetic conductor (PMC)*. The boundary condition that is appropriate between free space and a PMC region is that the tangential component of the magnetic field must be equal to 0. Further, the perfect conductivity means that all fields will equal 0 inside the PMC region.

    (a) Now, assume that we have a PMC filling the entire region of space for $y \geq 0$. If we have a plane wave incident on this PMC region described by

    $$\mathbf{H}^{\text{inc}}(y) = \hat{z}H^+ e^{-jk_0 y}, \tag{1.406}$$

    determine the expressions for the reflected $\mathbf{E}$ and $\mathbf{H}$ that must be produced to ensure that Maxwell's equations are satisfied in all regions of space.

    (b) Determine the reflection coefficient for this scenario at $y = 0$.

    Note: the reflection coefficient by convention is *always* specified in terms of the ratio of the electric field.

    (c) Compare this reflection coefficient to the one we determined previously for the reflected fields produced at the interface with a PEC region.

    Point of Interest: Neither PECs or PMCs exist in the real world. However, a PEC is a reasonable approximation for many good conductors at microwave frequencies, and so it is used frequently in practical engineering design. In contrast to this, there are no "good magnetic conductors" that exist in the real world, so PMCs have traditionally just been useful as a mathematical concept for simplifying the analysis of certain advanced problems. Recently, clever antenna engineers have been able to design a type of metasurface known as an *artificial magnetic conductor (AMC)*. These are surfaces engineered to approximately replicate some of the properties of a PMC over a particular band of frequencies. These have found use in designing low-profile antennas and in improving the bandwidth and radiation patterns of certain types of antennas.

16. Determine the polarization for each of the following plane waves. At a minimum, please show your work by specifying the logic behind your answer (i.e., don't simply write linear, right-hand circular, etc.; provide some kind of justification for the answer). Don't forget to specify the sense of the polarization if it is circular or elliptical, and pay attention to the propagation direction of the wave.

    (a) $\mathbf{E}(z) = \left(j\hat{x}|E_1| + \hat{y}|E_2|\right)e^{-jkz}, \quad |E_1| = |E_2|$

    (b) $\mathbf{E}(z) = \left(\hat{x}|E_1| - j\hat{y}|E_2|\right)e^{jkz}, \quad |E_1| \neq |E_2|$

    (c) $\mathbf{E}(z) = \left(\hat{x}|E_1| - \hat{y}|E_2|\right)e^{jkz}, \quad |E_1| \neq |E_2|$

    (d) $\mathbf{E}(z) = \left(\hat{x}|E_1| + \hat{y}|E_2|e^{j\pi/3}\right)e^{-jkz}, \quad |E_1| = |E_2|$

17. Prove that an arbitrarily-polarized plane wave can be decomposed into a superposition of right-hand and left-hand circularly polarized waves. Stated another way, show that right-hand and left-hand circularly polarized waves form a *basis* for plane waves.

18. Prove that the instantaneous Poynting vector of a circularly polarized plane wave in a lossless homogeneous medium is constant as a function of time and space.

19. A plane wave in medium 1 (ϵ_1, μ_1) is incident normally onto a lossless dielectric slab (ϵ_2, μ_2) that is backed by a PEC plane. If the dielectric slab is of thickness d, determine the reflected electric field in medium 1, the total electric field in the dielectric slab, and the time-average Poynting vector everywhere.

    Hint: If you denote E_{i0} as the amplitude of the incident field and E_r as the complex-valued amplitude of the reflected field you should find that

    $$\frac{E_r}{E_{i0}} = \frac{j\eta_2 \tan(\beta_2 d) - \eta_1}{j\eta_2 \tan(\beta_2 d) + \eta_1}.$$

    To make it easiest to arrive at this expression, the phase factors in the exponentials of your proposed solutions inside medium 2 should be $\pm j\beta_2(z - d)$; i.e., their phase reference is set to the surface of the PEC located at $z = d$. Along these lines, if E_2^+ is the complex-valued amplitude of the field propagating in the $+z$-direction inside medium 2, you should find that

    $$\frac{E_2^+}{E_{i0}} = \frac{\eta_2}{\eta_1 \cos(\beta_2 d) + j\eta_2 \sin(\beta_2 d)}.$$

    Also note that you may not use these relationships as given in your solution, you must derive them yourself.

20. A 10 GHz parallel polarization plane wave with a peak amplitude of 10 mV/m is propagating in free space and is incident on a lossless dielectric half space at an angle of 40°. If the transmitted wave propagates in the dielectric medium at an angle of 20° with respect to the z-axis, determine the following.

    (a) The relative permittivity of the dielectric medium.

    (b) Expressions for the incident and reflected electric fields in free space.

    (c) An expression for the total electric field in the dielectric medium.

    Note: make sure to fully evaluate the expressions in (b) and (c); i.e., don't leave symbols like $\Gamma_\parallel$ in your expressions if they can be numerically evaluated.

21. A 10 GHz circularly polarized plane wave given by

    $$\mathbf{E}_{\text{inc}}(y, z, t) = |E_0| \Big[\hat{x} \cos(2\pi \times 10^{10}\, t + 66.28y - 198.83z)$$

    $$- (\hat{y}\, 0.9487 + \hat{z}\, 0.3162) \sin(2\pi \times 10^{10}\, t + 66.28y - 198.83z) \Big]$$

    is propagating in free space and is incident at the Brewster angle on a lossless dielectric half space located at $y = 0$. If the relative permittivity of the dielectric half space is $\epsilon_r = 9$, answer the following questions.

(a) What is the sense of rotation for the incident plane wave?

(b) What is the polarization of the reflected wave?

(c) Find an expression for the transmitted wave.

(d) What is the polarization of the transmitted wave?

22. A 14 GHz plane wave given by

$$\mathbf{E}_{\text{inc}}(x, y, t) = -E_0(\hat{x}\,0.7660 + \hat{y}\,0.6428)e^{j(342.62x - 408.31y)}$$

is propagating in a lossless dielectric and is incident on another lossless dielectric half space that begins at $x = 0$. If the relative permittivity of the dielectric half space the wave will be transmitted into is $\epsilon_{r2} = 16.5$, answer the following questions.

(a) What is the angle of incidence for this wave?

(b) What is the relative permittivity of the dielectric medium the incident wave is propagating in?

(c) What is the transmission angle for this problem?

(d) Find an expression for the reflected magnetic field.

(e) Find an instantaneous expression for the transmitted electric field.

23. A 5 GHz plane wave propagating in free space is given by

$$\mathbf{E}_{\text{inc}} = \hat{y}5e^{-j(85.840x - 60.106z)}\ [\text{V/m}].$$

This wave is incident on a dielectric half space that begins at $z = 0$. If the relative permittivity of the dielectric half space is $\epsilon_r = 20$, answer the following questions.

(a) What is the angle of incidence of this wave (measured relative to the z-axis)?

(b) What will the angle of transmission be?

(c) Find an expression for the incident magnetic field.

(d) Find an instantaneous expression for the reflected magnetic field.

(e) Find an expression for the transmitted electric field.

Chapter 2

Transmission Lines

2.1   Introduction to Transmission Lines

Although the most accurate solution to a problem will always come from performing a complete field analysis of the system, this is often impractical at many stages of a design. In certain cases, we can apply some significant approximations to simplify Maxwell's equations down to a "0-dimensional" kind of circuit theory. However, there are many practical applications where this kind of circuit theory is no longer sufficient. In particular, this occurs when the sizes of the circuit begin to approach the size of the wavelength of the electromagnetic fields existing in the system. Transmission line theory provides us with a set of tools to bridge the gap between circuit theory and a full field analysis of systems. As a result, it is significantly simpler than performing a complete field analysis, but still allows us to model the wave behavior of voltages and currents that are propagating through our circuitry. This forms a kind of one-dimensional wave theory that is extremely useful in analyzing and designing a wide range of practical systems. This transmission line theory is an indispensable tool in the fields of radio frequency (RF) and microwave engineering.

   Before we begin our discussions on transmission line theory, it is useful to look at a few examples of the different kinds of geometries used to guide waves between circuit components. In Fig. 2.1, we show some common kinds of non-planar transmission lines. Similarly, Fig. 2.2 shows some of the most common kinds of planar transmission line structures. Transmission line theory will provide us with a way to analyze this diverse range of geometries using a unified theoretical framework.

2.2   Telegrapher's Equations

We now turn our attention to seeing how wave behavior can begin to manifest itself in a transmission line. This can typically be done in two ways. One approach is to begin with Maxwell's equations and simplify them for a specific geometry. Alternatively, one can start with a simplified model and more quickly derive the wave behavior. We take the second approach here, and will see aspects of the first approach later in the course.

   To derive general equations to describe transmission line behavior, we must introduce the concept of a *distributed parameter or distributed element network*. The idea is that because
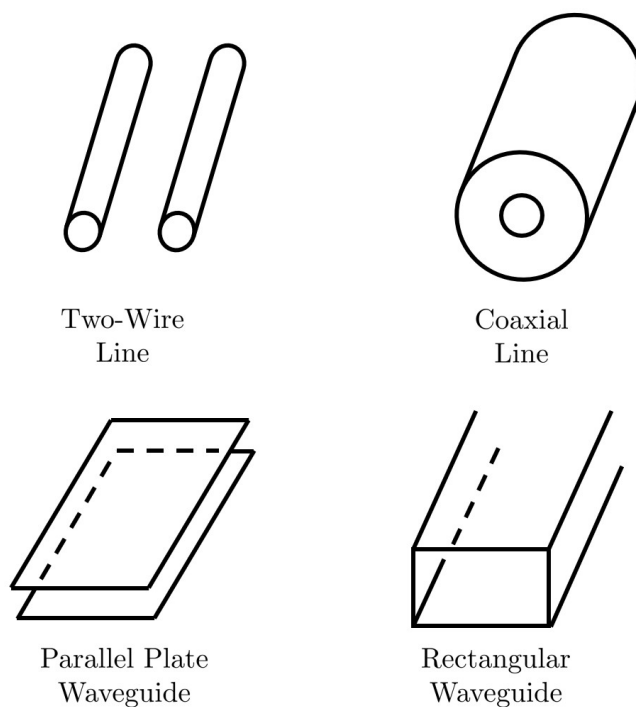
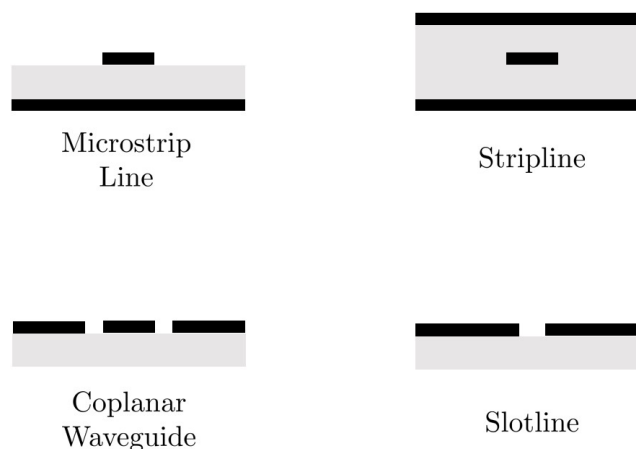Figure 2.1: Examples of non-planar transmission lines.

Figure 2.2: Examples of planar transmission lines. Dark boxes are conductors and lighter boxes are dielectric substrates.

the physical dimensions of the transmission line are of a relevant size to the wavelength of the fields, the voltages and currents that flow along the transmission line can change in magnitude and phase in appreciable ways before reaching other components the transmission line connects. Due to this, we need a more general description for how voltages and currents change as they move along the transmission line.

The strategy to model this behavior is to initially focus on an infinitesimal portion of a transmission line. We would typically draw this small section of transmission line in the
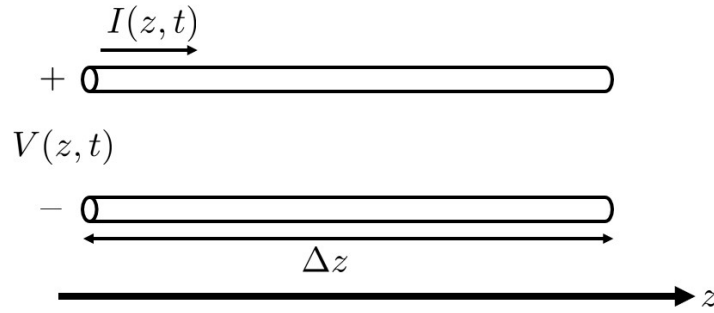
Figure 2.3: Schematic of an infinitesimal portion of a continuous transmission line.
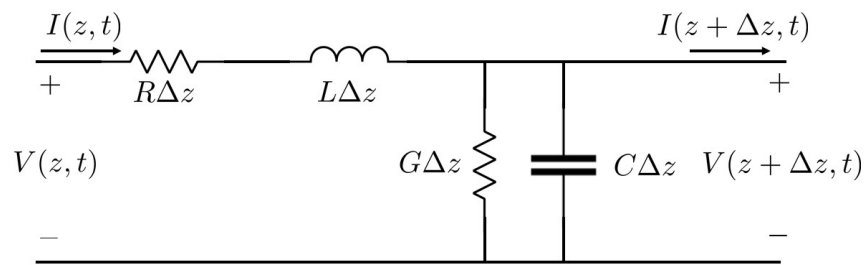

Figure 2.4: Lumped element model of an infinitesimal portion of a continuous transmission line.

manner shown in Fig. 2.3. Now, because the portion of the transmission line is very small compared to the wavelength we can assume that we may describe the small changes that will occur in the voltages and currents using *lumped elements*. That is, a standard circuit theory expression of various components. A general lumped element model for the infinitesimal portion of transmission line is shown in Fig. 2.4.

This circuit is composed of various *per unit length* quantities that characterize the behavior of the transmission line. These are:

- R = series resistance per unit length, for all conductors present, in Ω/m,

- L = series inductance per unit length, for all conductors present, in H/m,

- G = shunt conductance per unit length, in S/m,

- C = shunt capacitance per unit length, in F/m.

As we will see later, these quantities can be computed for a specific geometry from a field-based analysis of a cross-section of the transmission line using Maxwell's equations. Briefly, R accounts for the loss of power due to the finite conductivity of the conductors, while G accounts for loss in the material potentially filling the transmission line geometry. Meanwhile, L represents the total self-inductance of the geometry and C is a capacitance due to the close proximity of the conductive surfaces forming the transmission line.

Since we are only considering a very small portion of the overall transmission line, we can use the basic techniques of circuit theory to determine equations for the voltages and currents in Fig. 2.4. We begin with Kirchhoff's voltage law (KVL) to get

$$V(z,t) - R\Delta z I(z,t) - L\Delta z \partial_t I(z,t) - V(z + \Delta z, t) = 0. \tag{2.1}$$

We can rearrange terms and divide by Δz to get

$$\frac{V(z + \Delta z, t) - V(z,t)}{\Delta z} = -RI(z,t) - L\partial_t I(z,t). \tag{2.2}$$

We recognize the left-hand side as being a finite difference approximation to a derivative. Hence, if we take the limit of $\Delta z \to 0$, we will have that

$$\partial_z V(z,t) = -RI(z,t) - L\partial_t I(z,t). \tag{2.3}$$

Similarly, we can use Kirchhoff's current law (KCL) to get

$$I(z,t) - G\Delta z V(z + \Delta z, t) - C\Delta z \partial_t V(z + \Delta z, t) - I(z + \Delta z, t) = 0. \tag{2.4}$$

We rearrange terms to get

$$\frac{I(z + \Delta z, t) - I(z,t)}{\Delta z} = -GV(z + \Delta z, t) - C\partial_t V(z + \Delta z, t), \tag{2.5}$$

which becomes

$$\partial_z I(z,t) = -GV(z,t) - C\partial_t V(z,t) \tag{2.6}$$

after taking the limit of $\Delta z \to 0$.

In summary, the telegrapher's equations are "Maxwell's equations" for a simple 1D transmission line system. By this, we mean that they are the governing partial differential equations for the voltage and current that exist on the transmission line. The telegrapher's equations in the time domain are

$$\text{KVL}: \quad \partial_z V(z,t) = -RI(z,t) - L\partial_t I(z,t), \tag{2.7}$$

$$\text{KCL}: \quad \partial_z I(z,t) = -GV(z,t) - C\partial_t V(z,t). \tag{2.8}$$

2.2.1 Lossless Line Wave Equation

We can combine these to derive a wave equation. This is easiest if we begin by assuming a *lossless case* where $R = G = 0$. Then, we have

$$\partial_z V(z,t) = -L\partial_t I(z,t) \tag{2.9}$$

$$\partial_z I(z,t) = -C\partial_t V(z,t). \tag{2.10}$$

We take a derivative with respect to z on (2.9) to get

$$\partial_z^2 V(z,t) = -L\partial_t \partial_z I(z,t). \tag{2.11}$$

We then use (2.10) to substitute for $\partial_z I(z,t)$. This gives us

$$\partial_z^2 V(z,t) = LC\partial_t^2 V(z,t). \tag{2.12}$$

We can rearrange to finally have

$$\boxed{\partial_z^2 V(z,t) - LC\partial_t^2 V(z,t) = 0.} \tag{2.13}$$

At this point, we can easily recognize this as a one-dimensional wave equation.

To proceed, let's now convert to the frequency domain to have

$$\partial_z^2 V(z) + \beta^2 V(z) = 0, \tag{2.14}$$

where the phase constant is

$$\beta = \omega\sqrt{LC}, \tag{2.15}$$

the wavelength is

$$\lambda = \frac{2\pi}{\beta}, \tag{2.16}$$

and the phase velocity can be seen to be

$$v_p = \frac{\omega}{\beta} = \frac{1}{\sqrt{LC}}. \tag{2.17}$$

Since L and C depend on the geometry of the transmission line, we will find that all of these parameters characterizing the propagation of waves along a transmission line are also very much a function of its geometry.

We know that solutions of (2.14) will be of the form

$$V(z) = V_0^+ e^{-j\beta z} + V_0^- e^{j\beta z}. \tag{2.18}$$

We can compute what $I(z)$ will be by using

$$\partial_z V(z) = -j\omega L I(z). \tag{2.19}$$

Evaluating the derivative gives us that

$$I(z) = \frac{\beta}{\omega L}V_0^+ e^{-j\beta z} - \frac{\beta}{\omega L}V_0^- e^{j\beta z}. \tag{2.20}$$

We can simplify this expression by defining a *characteristic impedance*. This is the ratio between the voltage and current (with appropriate signs adopted to achieve a positive result). For the lossless case, the characteristic impedance is defined as

$$Z_0 = \sqrt{\frac{L}{C}}. \tag{2.21}$$

We can use this to see that (2.20) is

$$I(z) = \frac{V_0^+}{Z_0}e^{-j\beta z} - \frac{V_0^-}{Z_0}e^{j\beta z}. \tag{2.22}$$

It should be noted that when using the characteristic impedance to quickly find a voltage or current from the other, it is vital to remember the need to flip the sign if the wave if propagating in the negative direction. This is similar to the sign changes that happened between electric and magnetic fields when plane waves were propagating in opposite directions. In the plane wave case, we had the built-in check on our work to avoid sign errors by making sure that $\mathbf{E} \times \mathbf{H}$ pointed in the correct propagation direction. We no longer have this built-in check for transmission line problems, and so we must be very careful to not forget to account for this when we are working through a problem!

2.2.2 Lossy Line Wave Equation

We will now consider the full case of lossy transmission lines. For this situation, we have $R, G \neq 0$. In the frequency domain, the telegrapher equations become

$$\partial_z V(z) = -(R + j\omega L)I(z), \tag{2.23}$$

$$\partial_z I(z) = -(G + j\omega C)V(z). \tag{2.24}$$

We can combine these to form a wave equation, which is

$$\partial_z^2 V(z) - \gamma^2 V(z) = 0, \tag{2.25}$$

where

$$\gamma = \alpha + j\beta = \sqrt{(R + j\omega L)(G + j\omega C)}. \tag{2.26}$$

For this situation, we will have that $\lambda = 2\pi/\beta$, where β is calculated from (2.26). The characteristic impedance must also be updated for lossy lines. It now becomes

$$Z_0 = \sqrt{\frac{R + j\omega L}{G + j\omega C}}. \tag{2.27}$$

Considering this definition in comparison to the lossless case of (2.21), we see that if the characteristic impedance is complex-valued we have a lossy line. Further, we see that it is the imaginary part of the characteristic impedance that accounts for the loss in the transmission line.

2.2.3 Computing R, L, G, and C

In order for the transmission line theory to be useful, we are going to need some way to calculate all of the different per unit length transmission line parameters that we were using in the previous section. We now turn our attention to doing this.
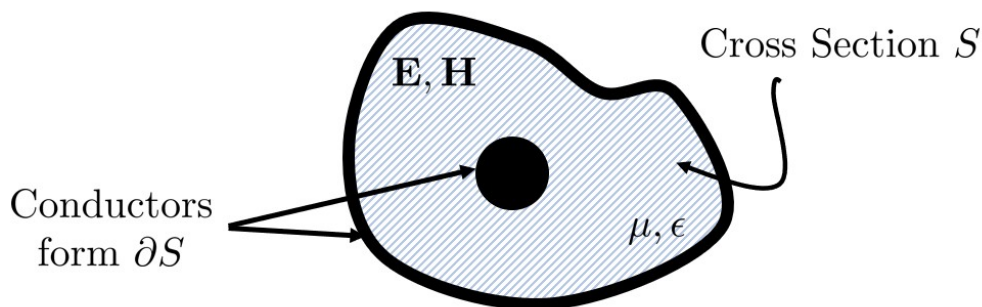
Figure 2.5: Illustration of a cross section of a transmission line where the per unit length transmission line parameters can be computed from.

We assume that we already know the description of the fields that exist within a cross section of the transmission line geometry (e.g., that illustrated in Fig. 2.5). Although we won't concern ourselves with this now, we can compute these fields by first solving Maxwell's equations for the transmission line geometry. We will revisit this later in the course.

We assume that we have a voltage $V_0 e^{\pm j\gamma z}$ and current $I_0 e^{\pm \gamma z}$ defined on the transmission line. We will define the different per unit length parameters so that the definitions of power and energy densities will give the same results between Poynting's theorem and what we know from circuit theory. Starting with L, we note that the time-averaged magnetic energy density stored in the cross section of the transmission line will be equal to

$$w_m = \frac{\mu'}{4} \int_S \mathbf{H} \cdot \mathbf{H}^* dS. \tag{2.28}$$

Meanwhile, the circuit theory result would be that

$$w_m = L|I_0|^2/4. \tag{2.29}$$

Hence, we can define the inductance per unit length to be

$$\boxed{L = \frac{\mu'}{|I_0|^2} \int_S \mathbf{H} \cdot \mathbf{H}^* dS.} \tag{2.30}$$

We may define the capacitance per unit length in a similar manner. We first note that the time-average electric energy density stored in the cross section of the transmission line will be

$$w_e = \frac{\epsilon'}{4} \int_S \mathbf{E} \cdot \mathbf{E}^* dS. \tag{2.31}$$

The circuit theory result for a capacitor would be

$$w_e = C|V_0|^2/4. \tag{2.32}$$

Hence, we can define the capacitance per unit length to be

$$\boxed{C = \frac{\epsilon'}{|V_0|^2} \int_S \mathbf{E} \cdot \mathbf{E}^* dS.} \tag{2.33}$$

121

The power loss due to finite conductivity that is accounted for by the per unit length resistance R can be calculated in a few different ways. One common approach to account for this is to make the approximation that even though the conductor has finite conductivity, the current on the conductor still flows as a surface current. We can then introduce a *surface resistivity* to account for the loss that occurs in reality due to the finite penetration depth of the fields and corresponding current flow in a real conductor. When we do this, we can then compute the corresponding power loss in terms of the magnetic field that is tangential to the cross section of the transmission line. These fields are proportional to the surface current densities flowing along the conductors (think about the boundary condition). Hence, we can write the power loss as

$$P_C = \frac{R_s}{2} \int_{\partial S} \mathbf{H} \cdot \mathbf{H}^* d\ell, \tag{2.34}$$

where R_s is the surface resistivity of the metal, ∂S denotes that the integral should be taken over the boundary of the surface S, and dl is simply the one dimensional integration along those bounding curves. The circuit theory result would have

$$P_C = R|I_0|^2/2, \tag{2.35}$$

so the resistance per unit length is

$$\boxed{R = \frac{R_s}{|I_0|^2} \int_{\partial S} \mathbf{H} \cdot \mathbf{H}^* d\ell.} \tag{2.36}$$

Finally, the per unit length conductance G needs to be defined. This is due to loss in the dielectric material filling the medium. We would typically have this be

$$P_d = \frac{\omega \epsilon''}{2} \int_S \mathbf{E} \cdot \mathbf{E}^* dS \tag{2.37}$$

from Poynting's theorem. The circuit theory result is

$$P_d = G|V_0|^2/2, \tag{2.38}$$

so the conductance per unit length is

$$\boxed{G = \frac{\omega \epsilon''}{|V_0|^2} \int_S \mathbf{E} \cdot \mathbf{E}^* dS.} \tag{2.39}$$

If magnetic materials are used that contribute magnetic loss, these formulas can be modified as needed to account for these additional loss effects. However, this is a fairly uncommon situation to encounter in typical practice, so this is not considered here.

Once these per unit length transmission line parameters have been computed for a geometry, we can continue with our analysis of a problem using the simplified one-dimensional transmission line theory that we will continue to study. By breaking the analysis of a complete problem into these smaller parts, we are able to greatly simplify the overall analysis, leading to a very useful design and analysis tool that is still used today.
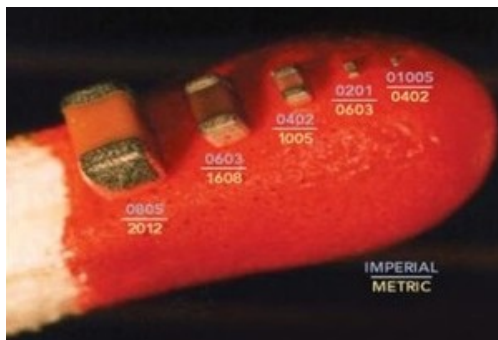
Figure 2.6: Example of common SMT component sizes sitting on the head of a matchstick (image from Surface Mount Process).

2.3   Terminated Lossless Transmission Lines

One of the fundamental problems that we need to be able to analyze using transmission line theory is what happens when a transmission line is terminated in some kind of load impedance. For this kind of problem, we assume that the load impedance can be represented as a *lumped element.* Implicit to this assumption is that the load impedance is small compared to the wavelength of the voltage/current waves on the transmission line.

From a practical perspective, we tend to be able to use this kind of representation fairly pervasively in microwave engineering when we use surface mount technology (SMT). These are very small components that can be soldered completely on a single surface of a printed circuit board (i.e., they do not have leads that poke through the board like larger components you likely used in circuits labs). There are many standard sizes of SMT parts, but you can readily buy ones that are only 20 mils long by 10 mils wide (see Fig. 2.6)! These often have to be handled with tweezers and soldered under a microscope. However, even with components that are this small, the physical nature of the devices that our waves interact with can become important at microwave frequencies. As a result, it is often a certain kind of "balancing act" in including enough approximate detail into our description of a lumped element load impedance for our transmission line equations to provide sufficiently accurate solutions for practical engineering design. We won't be able to focus on this here, but it is something that you will likely become quite familiar with in other advanced courses or in your career if you work as a microwave engineer. Generally, many of these non-ideal effects are described as being due to *parasitic reactances* that occur because of the physical features of a practical component.

2.3.1   Derivation of Reflection Coefficient

We now return to considering the case of a transmission line terminated with a load impedance. We illustrate the setup of our problem in Fig. 2.7. In it, we have our transmission line characterized by Z_0 and β. At $z = 0$, we have a load impedance Z_L connecting the signal line of the transmission line to ground. As is convention in solving transmission line problems, we typically consider the transmission line to have some finite length ℓ that measures into negative z values. This won't be important for this particular problem, but will be used in
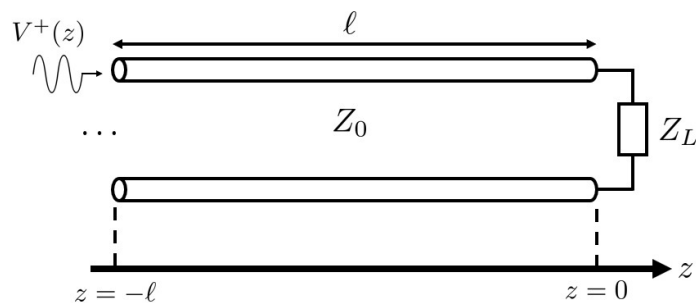
Figure 2.7: Illustration of a transmission line terminated with a load impedance.

future scenarios. We further assume that we have some voltage wave propagating toward the load impedance as part of our problem statement.

We can approach this problem using the same three step process we learned about for solving Maxwell's equations with plane wave solutions. In Step 1, we identify the regions of the problem. Here, we simply have our transmission line for $z < 0$ and then at $z = 0$ we have our lumped element Z_L. As an excitation to the problem, we have an incident voltage wave given by

$$V^+(z) = V_0^+ e^{-j\beta z}. \tag{2.40}$$

The corresponding current is

$$I^+(z) = \frac{V_0^+}{Z_0} e^{-j\beta z}. \tag{2.41}$$

In Step 2, we propose a suitable general solution. Much like with plane waves encountering different kinds of materials, the load impedance acts as a discontinuity to our voltage wave. Hence, we will need to have some reflected voltage wave be generated to ensure that the boundary conditions of our problem are met. We will express this wave as

$$V^-(z) = V_0^- e^{j\beta z}, \tag{2.42}$$

$$I^-(z) = -\frac{V_0^-}{Z_0} e^{j\beta z}. \tag{2.43}$$

Note that the minus sign comes about for the current due to the spatial derivative on z in the telegrapher's equations. *It is essential that you remember to include this minus sign, or you are likely to get incorrect results in future problems!*

We can now move to Step 3 where we use boundary conditions to solve for our unknown quantities. For this problem, we need to solve for V_0^-. Now, what should our boundary condition be? We don't have anything clearly appropriate from the boundary conditions we derived for fields previously, so we have to think a bit harder about this. It turns out that our load impedance is actually our boundary condition. This is because the load impedance

by definition expresses the ratio of the voltage and current at the load impedance through Ohm's law. Hence, we will need to have the total voltage and current at the load impedance give us

$$Z_L = \frac{V(0)}{I(0)}, \tag{2.44}$$

where V and I are the total voltage and current on the transmission line. By recalling that

$$V(z) = V_0^+ e^{-j\beta z} + V_0^- e^{j\beta z} \tag{2.45}$$

and

$$I(z) = \frac{1}{Z_0}\left(V_0^+ e^{-j\beta z} - V_0^- e^{j\beta z}\right), \tag{2.46}$$

we can fill in these expressions in (2.44) to get

$$Z_L = Z_0 \frac{V_0^+ + V_0^-}{V_0^+ - V_0^-}. \tag{2.47}$$

We can rearrange this to solve for V_0^-, giving

$$V_0^- = \frac{Z_L - Z_0}{Z_L + Z_0} V_0^+. \tag{2.48}$$

This formula should look rather familiar to you. In particular, the relationship between the impedances looks almost identical to the reflection coefficient that we had when considering a plane wave normally incident on a half space of some homogeneous medium. This commonality is no accident, and will happen time and time again when analyzing transmission lines. It turns out that the analysis of transmission lines and plane waves normally incident on different regions of space are mathematically equivalent, since both are essentially one-dimensional wave equations.

Now, considering this similarity, we often define a *voltage reflection coefficient* at the load impedance as

$$\Gamma_L = \frac{V^-(0)}{V^+(0)} = \frac{Z_L - Z_0}{Z_L + Z_0}. \tag{2.49}$$

It is important to point out that by convention, the reflection coefficient is always defined with respect to the voltages.

2.3.2   Other Characterizations of Reflections

Characterizing the reflections on a transmission line is such an important concept that a number of different conventions have arisen over the years to describe this behavior. In addition to the reflection coefficient, engineers will also often refer to the *return loss (RL)* of a system. This is simply a logarithmic representation of the reflection coefficient as

$$\mathrm{RL} = -20 \log |\Gamma_L| \ [dB]. \tag{2.50}$$

It is referred to as return loss because it helps characterize the "loss" in our overall system due to power being reflected at a load rather than being absorbed into it. *The minus sign is in the expression to make the return loss always be a positive number.* Oftentimes, engineers will accidentally refer to something as a return loss without including this minus sign so that they report a "negative loss". This is unphysical for any passive system (i.e., it doesn't have any amplifiers/gain), and is an error that you should not make. Considering the range of $|\Gamma_L|$, we see that $0 < \text{RL} < \infty$. Further, it should be noted that a larger return loss is a better value (i.e., it represents smaller reflections). As a rule of thumb, a return loss of 10 dB is generally considered a minimally acceptable value for "complex" parts (e.g., an antenna or a phase shifter), while a value in the range of 20 to 40 dB is typical for simpler components (e.g., an attenuator).

To see why the return loss characterizes the power "lost" to reflections, we can consider the average power on our transmission line. Considering that

$$V(z) = V_0^+ \left(e^{-j\beta z} + \Gamma_L e^{j\beta z} \right) \tag{2.51}$$

and

$$I(z) = \frac{V_0^+}{Z_0} \left(e^{-j\beta z} - \Gamma_L e^{j\beta z} \right), \tag{2.52}$$

the average power is

$$P_{\text{avg}} = \frac{1}{2} \text{Re}\{V(z)I(z)^*\} = \frac{|V_0^+|^2}{2Z_0} \text{Re}\{1 - \Gamma_L^* e^{-2j\beta z} + \Gamma_L e^{2j\beta z} - |\Gamma_L|^2\}. \tag{2.53}$$

The middle two terms combine to be a purely imaginary quantity so that the average power is

$$P_{\text{avg}} = \frac{|V_0^+|^2}{2Z_0} (1 - |\Gamma_L|^2). \tag{2.54}$$

Hence, we see that the reflected power that lowers the average power delivered to a load is proportional to the return loss given in (2.50).

There is yet another way that engineers will characterize the reflections on a transmission line. This is with the *voltage standing wave ratio (VSWR)*, which is the ratio of the maximum and minimum voltage anywhere on the transmission line. This is also sometimes simply referred to as the *standing wave ratio (SWR)*. We can look at the total expression for the voltage on our line to quickly determine what these values will be. In particular, we end up having

$$\text{VSWR} = \frac{V_{max}}{V_{min}} = \frac{1 + |\Gamma_L|}{1 - |\Gamma_L|}. \tag{2.55}$$

This quantity can range anywhere from $1 < \text{VSWR} < \infty$, with larger values indicating a worse system (i.e., larger reflections). A VSWR of $2 : 1$ is roughly equivalent to a 10 dB return loss. An illustration of the VSWR for a situation where $Z_L = R_L < Z_0$ is shown in Fig. 2.8.
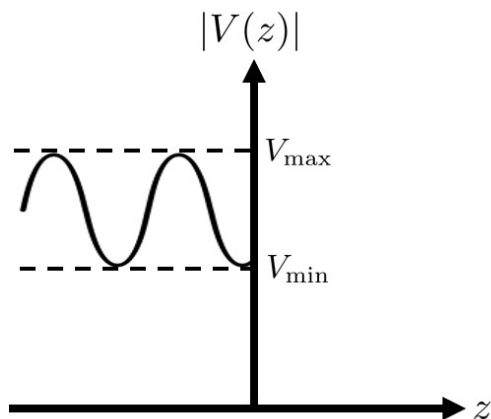
Figure 2.8: Illustration of the voltage standing wave ratio.

2.3.3 Input Impedance and Impedance Transformation

Previously, we discussed how to incorporate a load on a transmission line into our transmission line analysis. We followed a very general purpose approach that was reminiscent of how we solved certain plane wave problems earlier in the course. Although this is a solid general purpose approach, it is rather laborious and quickly becomes impractical when we want to analyze a complicated microwave network. We will now discuss certain useful "shortcuts" that are invaluable in analyzing transmission line problems.

For now, we will return to considering the case of a transmission line terminated with a load impedance. We illustrate the setup of our problem in Fig. 2.9. In it, we have our transmission line characterized by Z_0 and β. At $z = 0$, we have a load impedance Z_L connecting the signal line of the transmission line to ground. As is convention in solving transmission line problems, we typically consider the transmission line to have some finite length ℓ that measures into negative z values. We further assume that we have some voltage wave propagating toward the load impedance as part of our problem statement.

At this point, we have seen that a terminated transmission line will often be accompanied with some amount of reflected waves existing on the transmission line. Due to the presence of these reflected waves, the total voltage becomes an oscillatory function of position, similar to what we have seen previously in many plane wave and transmission line problems. This brings us to the idea of an *input impedance*. You likely encountered this concept at some point in your circuits courses. Basically, there are many practical situations where it is useful to represent some complex network of circuit components by an effective impedance that properly gives the relationship between the voltage and current that would exist at the input terminals to this complex network. This helps us have a very simple problem to consider from the viewpoint of a source/generator.

What happens with transmission lines is that the input impedance of some transmission line network will vary as a function of position due to the oscillations in the total voltage and current as functions of position. Hence, the effect that some load impedance will have on a source/generator will depend on how far the load is situated away from the source. Importantly, we can characterize complex combinations of transmission lines and
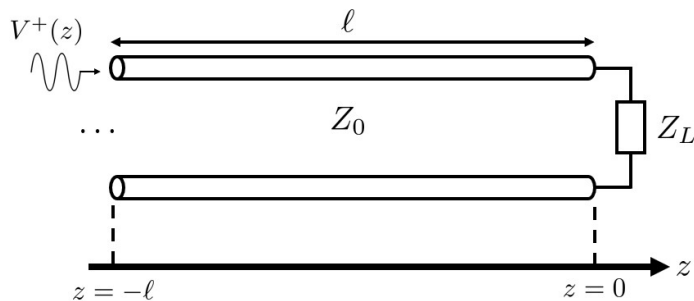
Figure 2.9: Illustration of a transmission line terminated with a load impedance.

load impedances as a simple *lumped element* input impedance. This provides us with a very powerful strategy for solving transmission line problems. We can often remove the presence of the transmission lines (and the corresponding wave behavior) by computing appropriate input impedances and then simply use circuit theory to solve for the voltages and currents that are of interest.

To see the basic way this works, we will again consider the simple case of Fig. 2.9. We will now look at the input impedance of the system at some position $-\ell$. For this, we have

$$Z_{\text{in}}(-\ell) = \frac{V(-\ell)}{I(-\ell)} = Z_0 \frac{V_0^+(e^{j\beta\ell} + \Gamma_L e^{-j\beta\ell})}{V_0^+(e^{j\beta\ell} - \Gamma_L e^{-j\beta\ell})} = Z_0 \frac{e^{j\beta\ell} + \Gamma_L e^{-j\beta\ell}}{e^{j\beta\ell} - \Gamma_L e^{-j\beta\ell}}. \tag{2.56}$$

A more commonly used form can be found by replacing Γ_L with

$$\Gamma_L = \frac{Z_L - Z_0}{Z_L + Z_0} \tag{2.57}$$

and simplifying. Initially, this gives us

$$Z_{\text{in}}(-\ell) = Z_0 \frac{(Z_L + Z_0)e^{j\beta\ell} + (Z_L - Z_0)e^{-j\beta\ell}}{(Z_L + Z_0)e^{j\beta\ell} - (Z_L - Z_0)e^{-j\beta\ell}}. \tag{2.58}$$

We can regroup terms to write this expression using trigonometric functions. This gives us

$$Z_{\text{in}}(-\ell) = Z_0 \frac{Z_L \cos(\beta\ell) + jZ_0 \sin(\beta\ell)}{Z_0 \cos(\beta\ell) + jZ_L \sin(\beta\ell)}. \tag{2.59}$$

Typically, the cosine function is factored out so that we only have to deal with one trigonometric function in the numerator and denominator. This gives us

$$\boxed{Z_{\text{in}}(-\ell) = Z_0 \frac{Z_L + jZ_0 \tan(\beta\ell)}{Z_0 + jZ_L \tan(\beta\ell)}.} \tag{2.60}$$

This is an incredibly useful formula. It allows us to quickly take a load impedance and transform its effect into a new lumped element load impedance located at a more convenient position in our transmission line network. This is often referred to as *impedance transformation*. Once the impedance transformation has been completed, we can treat the new load

128

as a lumped element in the rest of our transmission line analysis because the impedance transformation has accounted for all of the complex wave interference effects that occur in the transmission line section "transformed" out of the analysis. This is a significant simplification, and allows us to rapidly solve complex problems that would be extremely tedious to consider if we were to try and apply the full 3-step solution process we have used previously at all sections of the transmission line system. *Final Note: ℓ should be plugged into the RHS of this equation as a positive value!*

Before moving on, we briefly mention that this derivation may be repeated quite easily for a lossy transmission line. In this case, the impedance transformation equation becomes

$$\boxed{Z_{\text{in}}(-\ell) = Z_0 \frac{Z_L + Z_0 \tanh(\gamma\ell)}{Z_0 + Z_L \tanh(\gamma\ell)}.} \tag{2.61}$$

In this equation, $\tanh(x)$ is a hyperbolic tangent function.

2.3.4 Special Cases of Lossless Terminated Lines

It is useful to consider a few special cases of a terminated transmission line that occur rather frequently in microwave engineering. The first case we will consider is what happens when a transmission line is terminated with a short circuit. For this case, $Z_L = 0$, so the impedance transformation equation (2.60) becomes

$$Z_{\text{in}}(-\ell) = jZ_0 \tan(\beta\ell). \tag{2.62}$$

This is purely imaginary and can take on values between $+j\infty$ and $-j\infty$. Hence, a short-circuited transmission line can effectively look like any purely reactive impedance we can imagine! This is quite useful for designing different kinds of transmission line systems, especially *matching networks*. The voltage, current, and effective reactance of the short-circuited transmission line is shown in Fig. 2.10. Looking at the variation of the voltage and current in Fig. 2.10, we see that this short circuit boundary is similar to a PEC boundary condition for plane waves.

The next case we will consider is what happens when a transmission line is terminated with an open circuit. For this case, "$Z_L = \infty$", so the impedance transformation equation (2.60) becomes

$$Z_{\text{in}}(-\ell) = -jZ_0 \cot(\beta\ell). \tag{2.63}$$

This is also purely imaginary and can take on values between $+j\infty$ and $-j\infty$. Hence, we get quite similar behavior to the short circuit case. The voltage, current, and effective reactance of the open-circuit terminated transmission line is shown in Fig. 2.11. Looking at the variation of the voltage and current in Fig. 2.11, we see that this open circuit boundary is similar to a PMC boundary condition for plane waves.

In practice, it depends on the type of transmission line structure being used as to whether it is convenient to implement a short or open circuit termination. Additionally, because the type of reactance achieved closer to the line is opposite for open vs. short circuit terminations,
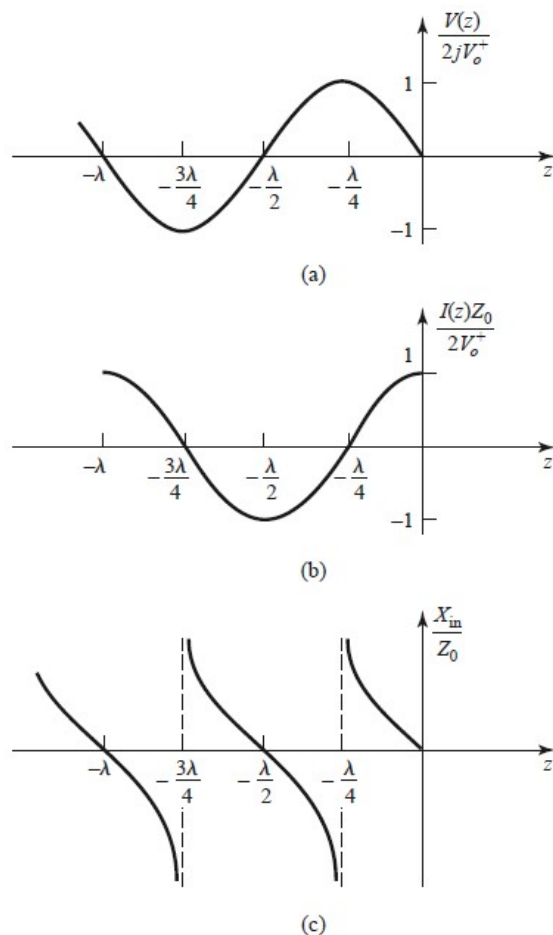
Figure 2.10: Illustration of the (a) voltage, (b) current, and (c) reactance as a function of position for a short circuit terminated line (image from [13]).

it can often be more practical to use whichever case will give you a shorter overall transmission line in practice.

There are also some "special" lengths of transmission lines that are useful to be aware of. The first is what happens if $\ell = \lambda/2$ (or some multiple of $\lambda/2$). For this case, the tangent function becomes 0 so that the input impedance becomes

$$Z_{\text{in}} = Z_L. \tag{2.64}$$

That is, it simply looks like the load impedance.

Next, we also often want to consider *quarter-wavelength* sections of transmission line. These are useful because they make the input impedance

$$Z_{\text{in}} = \frac{Z_0^2}{Z_L}. \tag{2.65}$$

To see why this is the result, we note that for a quarter-wavelength line we will have that $\beta\ell = (2\pi/\lambda)(\lambda/4) = \pi/2$. We can't plug this into (2.60) directly because $\tan(\pi/2) \to \infty$.
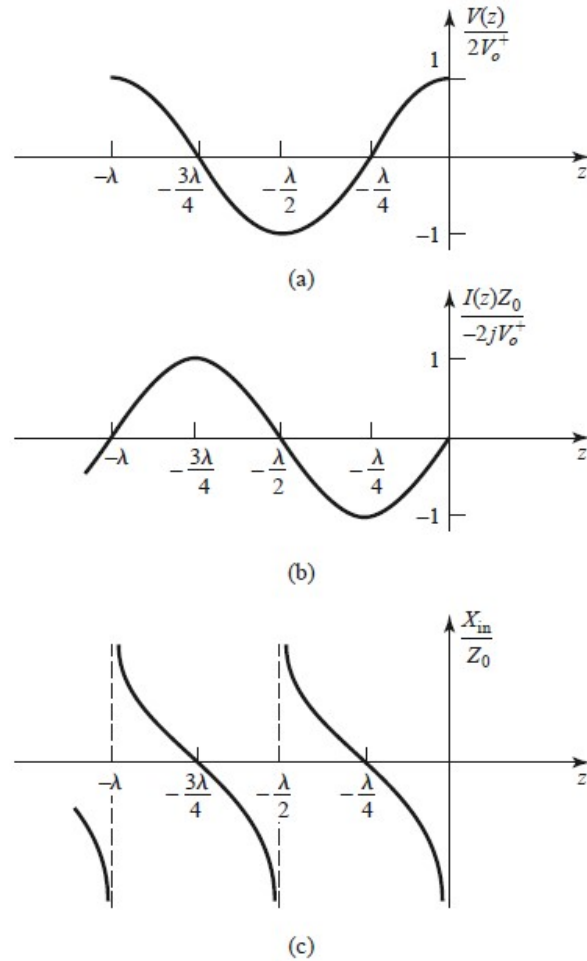
Figure 2.11: Illustration of the (a) voltage, (b) current, and (c) reactance as a function of position for a open circuit terminated line (image from [13]).

To evaluate the result, we can divide the numerator and denominator of (2.60) by $\tan(\beta\ell)$ so that we get

$$Z_{\text{in}}(-\ell) = Z_0 \frac{\dfrac{Z_L}{\tan(\beta\ell)} + jZ_0}{\dfrac{Z_0}{\tan(\beta\ell)} + jZ_L}. \tag{2.66}$$

By now taking the limit as $\beta\ell \to \pi/2$, we can readily see that we recover the result given in (2.65). This type of line is significant because it can be very useful in developing simple matching networks. It is often referred to as a *quarter-wave transformer*.

2.4 Quarter-Wave Transformer Matching

Let's now consider a simple design example to see how a quarter-wave transformer can be used in practice. For this example, consider that we have a known load impedance that
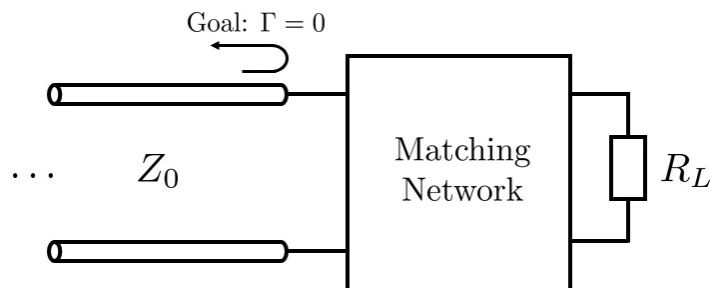
Figure 2.12: Illustration of inserting a matching network between a load and transmission line to eliminate reflections from the perspective of the connected transmission line.
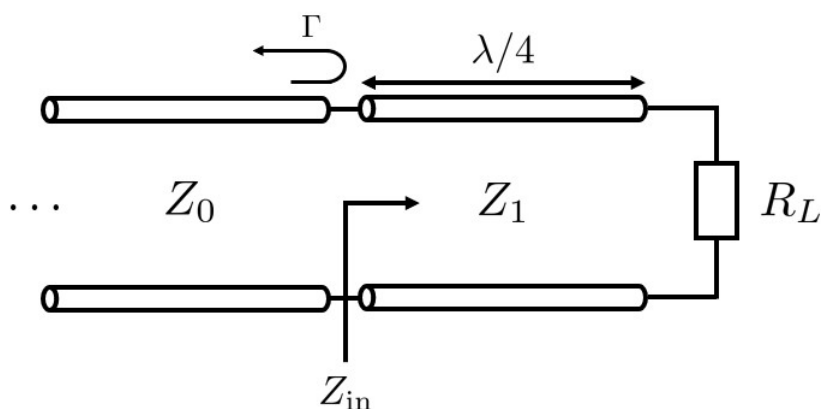

Figure 2.13: Illustration of a quarter-wave transformer for matching a real load impedance.

is purely resistive (and thus is also purely real) that is different from the characteristic impedance Z_0 of the transmission lines we are using in the rest of our system. We want to develop a simple *matching network* that we can insert in between the load resistance and the transmission line in our system so that from the view of our transmission line there will be no reflection due to the load resistance. This is schematically illustrated in Fig. 2.12.

One simple matching network we can use in this case is a quarter-wave transformer. This solution is illustrated in Fig. 2.13. Our goal now becomes choosing an appropriate value for the characteristic impedance of the quarter-wave transformer (Z_1) so that the input impedance at the output of this line matches the characteristic impedance of our system (Z_0).

From our previous analysis, we know that the input impedance of this quarter-wave transformer will be

$$Z_{\text{in}} = \frac{Z_1^2}{R_L}. \tag{2.67}$$

To find the necessary value of Z_1 to achieve a perfect match (i.e., no reflections), we need to
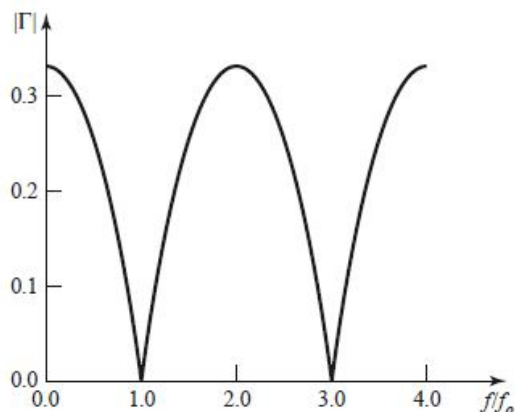
Figure 2.14: Frequency dependence of the reflection coefficient for a quarter-wave transformer for a load resistance of $R_L = 100\,\Omega$ and a system characteristic impedance of $Z_0 = 50\,\Omega$ (image from [13]).
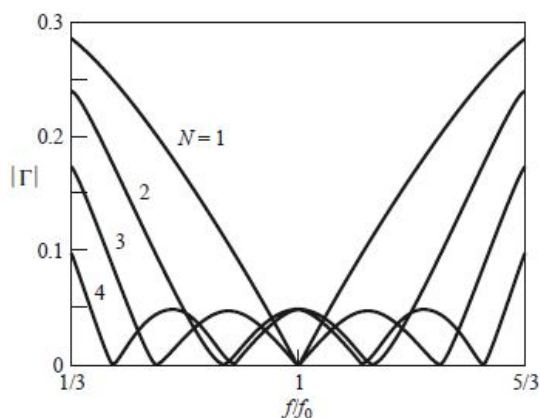
Figure 2.15: Frequency dependence of the reflection coefficient for Chebyshev multisection matching transformers with varying numbers of sections for a load resistance of $R_L = 100\,\Omega$ and a system characteristic impedance of $Z_0 = 50\,\Omega$ (image from [13]).

set $Z_{\text{in}} = Z_0$. Doing this, we can solve for Z_1 to find that

$$Z_1 = \sqrt{Z_0 R_L}. \tag{2.68}$$

This solution is very straightforward, and comes with some drawbacks. One of the primary drawbacks is that the "perfectly matched" condition will only occur when the quarter-wave transformer is a quarter-wave long (or an odd multiple of this) – i.e., at a single frequency. Hence, for practical systems that must operate over a band of frequencies, this simple solution may not be sufficient since the performance can reduce quite quickly as a function of frequency. This effect is illustrated in Fig. 2.14, where the frequency dependence of the reflection coefficient is shown for the case of $R_L = 100\,\Omega$ and $Z_0 = 50\,\Omega$.

One simple way to overcome the bandwidth limitation is to utilize a more sophisticated matching network design. Many examples exist, but they can often require the use

Figure 2.16: Illustration of a typical transmission line problem involving a generator and load that are not matched to the transmission line connecting the two.
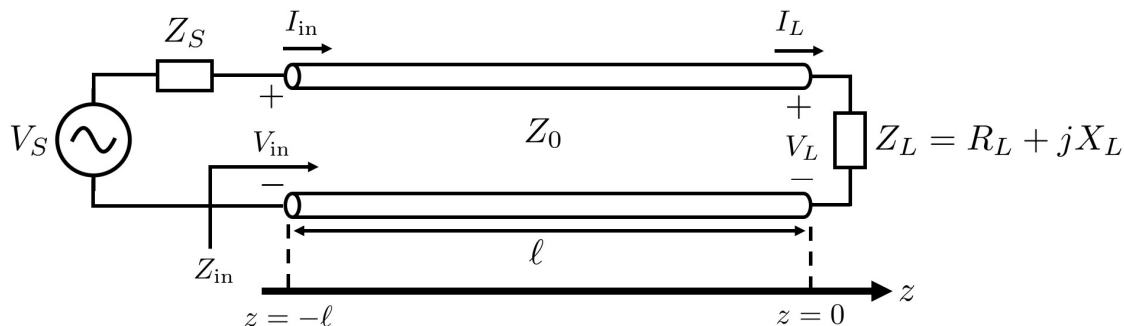
of additional sections of transmission lines with different characteristic impedances (called *multisection matching transformers*). The example of the bandwidth improvement possible with different *Chebyshev multisection matching transformers* is shown in Fig. 2.15. Although better results are possible, this comes at the cost at a more complex design and often requires the use of more "real estate" (i.e., it takes up more area in the physical design).

2.5   Generator and Load Mismatches

In solving practical transmission line problems, it is often necessary to consider a source or generator being connected to a transmission line. The simplest example is shown in Fig. 2.16. In this problem, we have some known source with a source impedance Z_S connected to a transmission line of length ℓ that is terminated in a load impedance Z_L. In general, Z_S may not exactly equal Z_0, so we will have the possibility of the voltage wave that was originally reflected at Z_L also be reflected at Z_S. This leads to an "infinite" number of reflections that can make evaluating certain aspects of the problem somewhat involved. We already saw this same kind of case when we considered a plane wave normally incident on a finite thickness slab of a dielectric material. At that time, we formulated our problem but didn't go through the process of solving it due to the tedious nature. Here, we use what we have just learned about transmission lines to see how we can very quickly extract the important information from the problem without going through the full tedious process.

   To begin, we note that our general solution of the total voltage wave in the transmission line will still need to consist of a combination of a forward and backward traveling waves. We know from the boundary condition at the load that we can express the total voltage wave as

$$V(z) = V^+(e^{-j\beta z} + \Gamma_L e^{j\beta z}). \tag{2.69}$$

The catch here is that we don't know what V^+ is because it depends on the voltage source and the reflections that can be produced at $z = -\ell$. To solve for V^+ we can use a boundary condition at $z = -\ell$. The easiest way to do this is to transform our load impedance back a distance ℓ so that our entire circuit can be represented as a simple lumped element circuit. This gives us the illustration shown in Fig. 2.17.

$$Z_{\text{in}} = Z_0 \frac{Z_L + jZ_0 \tan \beta \ell}{Z_0 + jZ_L \tan \beta \ell}$$
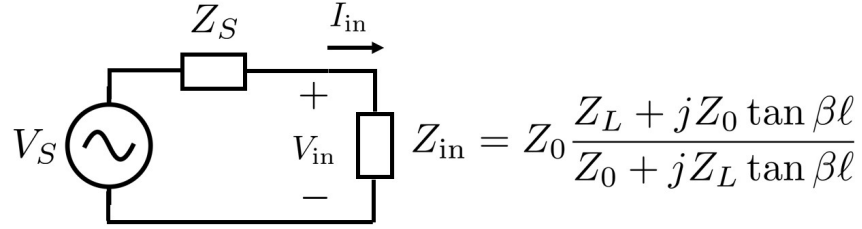
Figure 2.17: Illustration of a typical transmission line problem after transforming the impedance back along the length of the transmission line.
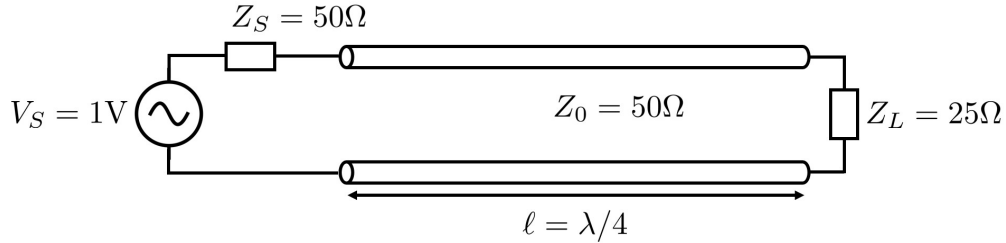
Figure 2.18: Example circuit to look at the effects of generator and load mismatches.

We can easily see from Fig. 2.17 that V_{in} can be found from a simple voltage divider circuit. This gives us

$$V_{\text{in}} = \frac{Z_{\text{in}}}{Z_S + Z_{\text{in}}} V_S. \tag{2.70}$$

We must have that $V(z)$ equals V_{in} at $z = -\ell$. Hence, we see that

$$V_{\text{in}} = V^+(e^{j\beta \ell} + \Gamma_L e^{-j\beta \ell}). \tag{2.71}$$

Equating (2.70) and (2.71), we can solve for V^+. We get that

$$V^+ = \frac{Z_{\text{in}}}{Z_S + Z_{\text{in}}} \frac{V_S}{e^{j\beta \ell} + \Gamma_L e^{-j\beta \ell}}. \tag{2.72}$$

It is helpful to see this full process in action, so we now conclude with an example calculation for the circuit shown in Fig. 2.18. Our goal is to determine expressions for the voltage and current at any point in the transmission line. To begin, we need to evaluate the different parts of (2.69) that will be needed to determine V^+ later. We will first evaluate the load reflection coefficient, which for this circuit is

$$\Gamma_L = \frac{Z_L - Z_0}{Z_L + Z_0} = \frac{25 - 50}{25 + 50} = -\frac{1}{3}. \tag{2.73}$$

Next, to determine V^+ we will need to use the boundary condition at $z = -\lambda/4$. Hence, we will need to know what $\beta \ell$ is equal to for this position to evaluate (2.69). We can quickly find that

$$\beta \ell \big|_{\ell = \lambda/4} = \frac{2\pi}{\lambda} \times \frac{\lambda}{4} = \frac{\pi}{2}, \tag{2.74}$$

Figure 2.19: Example circuit after using an impedance transformation to simplify.
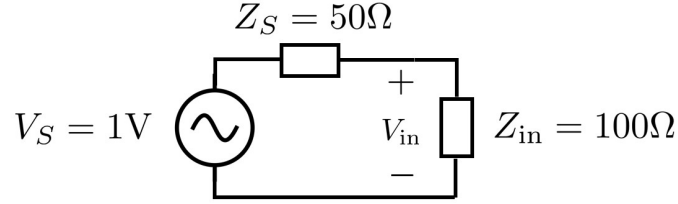
where we have used the fact that $\beta = 2\pi/\lambda$ to simplify the evaluation. Plugging this into (2.69) and simplifying gives us

$$
\begin{aligned}
V(z = -\lambda/4) &= V^+\left(e^{j\pi/2} - \frac{1}{3}e^{-j\pi/2}\right) \\
&= V^+\left(j + \frac{j}{3}\right) \\
&= j\frac{4}{3}V^+.
\end{aligned}
\tag{2.75}
$$

Now, to determine V^+, we need to find what the total voltage $V(z = -\lambda/4)$ should be equal to. We can do this by reducing our transmission line circuit into an equivalent lumped element circuit (i.e., all the wave portions have already been accounted for elsewhere). This is done by finding the input impedance of the circuit from the position $\ell = -\lambda/4$. We use the impedance transformation formula to find this. We write this in a slightly different form of

$$
Z_{\text{in}}(\ell) = Z_0 \frac{\dfrac{Z_L}{\tan(\beta\ell)} + jZ_0}{\dfrac{Z_0}{\tan(\beta\ell)} + jZ_L}
\tag{2.76}
$$

to simplify the analysis. The reason that this helps is because when we plug in for $\beta\ell \to \pi/2$ we have that $\tan(\beta\ell) \to \infty$. With the rearranged form of (2.76) we can quickly evaluate this limit to find that

$$
Z_{\text{in}} = \frac{Z_0^2}{Z_L},
\tag{2.77}
$$

which is the result that we have already discussed for a quarter-wave transformer. Plugging in our numbers for Z_0 and Z_L we find that $Z_{\text{in}} = (50)^2/25 = 100\,\Omega$. Our simplified circuit after this impedance transformation is now shown in Fig. 2.19. This simple voltage divider tells us that $V_{\text{in}} = 2/3$.

Considering this, we can now find V^+ by recognizing that $V_{\text{in}} = V(z = -\lambda/4)$. Using our result from (2.75) and that $V_{\text{in}} = 2/3$ we find that

$$
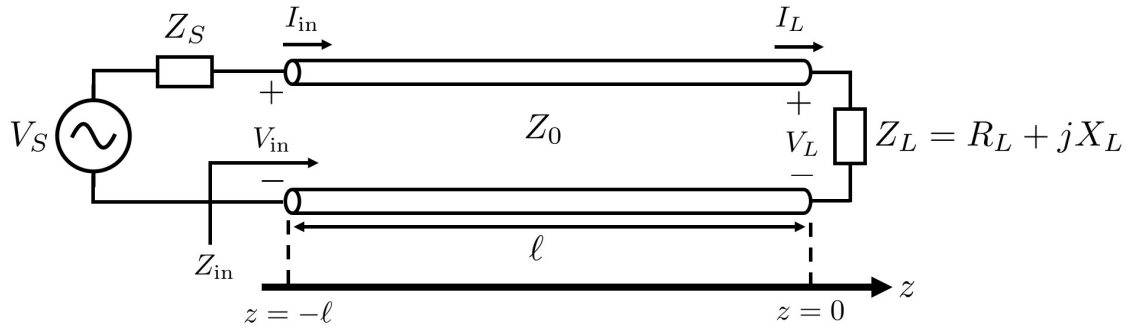j\frac{4}{3}V^+ = \frac{2}{3} \quad \Longrightarrow \quad V^+ = -\frac{j}{2}.
\tag{2.78}
$$

Figure 2.20: Circuit diagram for a generator and load connected by a transmission line.

We can now use this in (2.69) to find that

$$V(z) = -\frac{j}{2}\left(e^{-j\beta z} - \frac{1}{3}e^{j\beta z}\right). \tag{2.79}$$

We can also evaluate the current quickly by remembering that we simply need to divide the total voltage by the characteristic impedance of the line and switch the sign on the portion of the wave propagating in the $-z$-direction. Considering this, we have that

$$I(z) = -\frac{j}{2(50)}\left(e^{-j\beta z} + \frac{1}{3}e^{j\beta z}\right). \tag{2.80}$$

2.6 Power Transfer

We now begin to consider general expressions for how much power is transferred to a load of a transmission line network. Our presentation will follow that given in [13]. Typically, we are interested in maximizing the power that we deliver to a load. For instance, we can view an antenna as a load to a transmission line system. The more power we deliver to the antenna the more power the antenna can correspondingly radiate to achieve different applications (e.g., communication or radar systems).

To analyze the relevant scenario, we will again consider the case where both the generator and load impedance may not be matched to the transmission line connecting the two devices. This is illustrated in Fig. 2.20. We won't go through the detailed derivation again to solve this problem. However, recall that the process that we went through started by transforming the load impedance back along the length of transmission line. Once we did this, we had a simple circuit shown in Fig. 2.21 that we could solve using standard circuit theory equations since the wave aspects of the problem had been accounted for already.

After solving for all the necessary unknown variables, we found that

$$V_{\text{in}} = \frac{Z_{\text{in}}}{Z_S + Z_{\text{in}}}V_S, \tag{2.81}$$
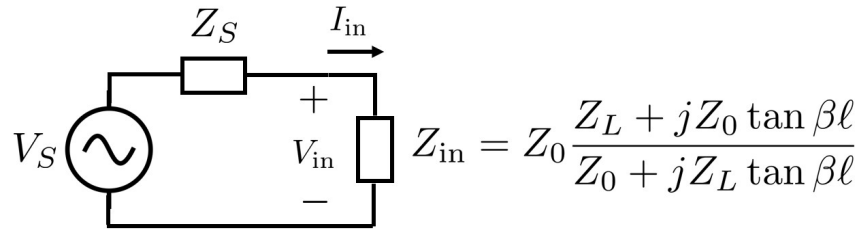
Figure 2.21: Circuit diagram for a generator and load connected by a transmission line after the load has been transformed back along the length of the transmission line.

where the input impedance was given by

$$Z_{\text{in}} = Z_0 \frac{Z_L + jZ_0 \tan(\beta\ell)}{Z_0 + jZ_L \tan(\beta\ell)}. \tag{2.82}$$

For the purposes of this section, it will be much easier for us to just consider Z_{in} in a general sense and not worry about its explicit expression in terms of this formula. It is simply important to remember that whatever we determine about Z_{in} can be eventually related back to Z_L following some appropriate math.

We now want to determine how much power is transferred to the load under different operating conditions that commonly occur. This analysis can be simplified by noting that because we have a lossless transmission line (since Z_0 is assumed to be real for this problem) all of the real power delivered to Z_{in} in our simplified circuit schematic will have to be absorbed in the load. Hence, we have that the power delivered to the load is

$$P_L = \frac{1}{2}\text{Re}\{V_{\text{in}}I_{\text{in}}^*\} = \frac{1}{2}|V_{\text{in}}|^2\text{Re}\{Z_{\text{in}}^{-1}\}. \tag{2.83}$$

We can use our expression for V_{in} from (2.81) to get

$$P_L = \frac{1}{2}|V_S|^2 \left|\frac{Z_{\text{in}}}{Z_{\text{in}} + Z_S}\right|^2 \text{Re}\{Z_{\text{in}}^{-1}\}. \tag{2.84}$$

If we now express $Z_S = R_S + jX_S$ and $Z_{\text{in}} = R_{\text{in}} + jX_{\text{in}}$, we can eventually simplify our above expression to be

$$P_L = \frac{1}{2}|V_S|^2 \frac{R_{\text{in}}}{(R_{\text{in}} + R_S)^2 + (X_{\text{in}} + X_S)^2}. \tag{2.85}$$

We will now consider three different scenarios. First, when the load impedance is matched to the transmission line impedance. Second, when the generator is matched to the loaded line (i.e., matched to Z_{in}). Finally, what is known as conjugate matching.

Load Matched to the Transmission Line

Mathematically, having the load *matched* to the transmission line means that no reflections will be produced at the interface between the load and the transmission line. Recalling that the reflection coefficient at the load is

$$\Gamma_L = \frac{Z_L - Z_0}{Z_L + Z_0}, \tag{2.86}$$

we see that for $\Gamma_L = 0$ we will need $Z_L = Z_0$. We can also see from our impedance transformation formula (2.82) that $Z_{\text{in}} = Z_0$ for this case as well. Hence, the power delivered to the load given in (2.85) becomes

$$P_L = \frac{1}{2}|V_S|^2 \frac{Z_0}{(Z_0 + R_S)^2 + X_S^2}. \tag{2.87}$$

Generator Matched to the Loaded Line
For the generator to be matched to the loaded line means we should have $Z_{\text{in}} = Z_S$. From a practical perspective, if we want to achieve this we can typically do so by making appropriate selections of $\beta\ell$ and Z_0. That is, by changing properties of the transmission line connecting the generator and load.

For this situation, we can consider there to be no overall reflection coefficient from the perspective of the generator because

$$\Gamma_S = \frac{Z_{\text{in}} - Z_S}{Z_{\text{in}} + Z_S} = 0. \tag{2.88}$$

However, this does not mean that there are no reflections in the overall system! In particular, it is likely that we will have a standing wave exist on the transmission line since Z_0 and Z_L will in general not be equal to each other for this case. We are simply adjusting the parameters of the system so that this standing wave "looks" like a perfectly matched line from the view of the generator.

Now, the power transferred to the load becomes

$$P_L = \frac{1}{2}|V_S|^2 \frac{R_S}{4(R_S^2 + X_S^2)}. \tag{2.89}$$

If we compare this to (2.87), we can see that the power transferred to the load for the current scenario may actually be less than the case where only the load was matched to the transmission line. Hence, to determine under what conditions we will transfer a maximum amount of power to a load will require us to be more clever in our approach. We will consider this next.

Conjugate Matching
To determine the maximum power that can be transferred to a load, we will assume that the generator impedance is fixed. This is often reasonable for many practical applications since our sources will typically be a commercial-of-the-shelf (COTS) part that we don't have much design control over. Depending on your application, the load impedance may very well be fixed as well. However, because we will often have some freedom over how we connect the generator and load to one another, we can still consider Z_{in} to be something that is reasonably under our control as designers.

With this in mind, we will determine how to maximize power transfer to the load by differentiating (2.85) with respect to the real and imaginary parts of Z_{in} and setting the results equal to 0. Beginning with the real part, we have that

$$\frac{\partial P_L}{\partial R_{\text{in}}} = \frac{1}{2}|V_S|^2 \left[\frac{1}{(R_{\text{in}} + R_S)^2 + (X_{\text{in}} + X_S)^2} + \frac{-2R_{\text{in}}(R_{\text{in}} + R_S)}{[(R_{\text{in}} + R_S)^2 + (X_{\text{in}} + X_S)^2]^2} \right]. \tag{2.90}$$

If we set this equal to 0 and do some algebra, we can eventually find that this will require

$$R_S^2 - R_{\text{in}}^2 + (X_{\text{in}} + X_S)^2 = 0. \tag{2.91}$$

We can similarly differentiate with respect to the imaginary part of the input impedance to get

$$\frac{\partial P_L}{\partial X_{\text{in}}} = \frac{1}{2}|V_S|^2 \frac{-2R_{\text{in}}(X_{\text{in}} + X_S)}{[(R_{\text{in}} + R_S)^2 + (X_{\text{in}} + X_S)^2]^2}. \tag{2.92}$$

If we set this equal to 0, we will need to have

$$R_{\text{in}}(X_{\text{in}} + X_S) = 0. \tag{2.93}$$

We now need to solve (2.91) and (2.93) together to get

$$X_{\text{in}} = -X_S, \quad R_{\text{in}} = R_S. \tag{2.94}$$

We can write this compactly as

$$Z_{\text{in}} = Z_S^*, \tag{2.95}$$

which is why this is referred to as *conjugate matching*. For this case, the power delivered to the load becomes

$$P_L = \frac{1}{2}|V_S|^2 \frac{1}{4R_S}, \tag{2.96}$$

which is larger than the previous cases we considered.

From an intuitive perspective, this case can be seen to potentially have many multiple reflections occurring since the load is not matched to the transmission line and the input impedance is also not matched to the generator. However, we arrive at a larger overall power transfer than the previous two cases we considered because we have found a case where the multiple reflections can *add in phase constructively*. The end result is a larger amount of power going to the load.

It should also be mentioned that often times it is desirable to make the reactive part of the generator impedance as low as possible. If this happens, we end up with an intuitive result that maximum power transfer will happen if the input impedance of the loaded line is matched to the generator impedance, i.e., $Z_{\text{in}} = Z_S = R_S$. Many types of loads may have some amount of reactance to them (an antenna is an excellent example where this is the case), which can make having $Z_{\text{in}} = R_S$ over a large bandwidth somewhat difficult if we only rely on the impedance transformation formula to modify Z_{in}. In these cases, it is often useful to design an *impedance matching circuit* that sits in between the final load impedance and the transmission line that connects the load to the generator. This circuit often serves the purpose of eliminating the reactive part of the load impedance and leaving the real part of the impedance as something that can be reasonably connected to the transmission line or generator without causing many reflections. We will learn a bit about a simple impedance matching circuit later in this course.
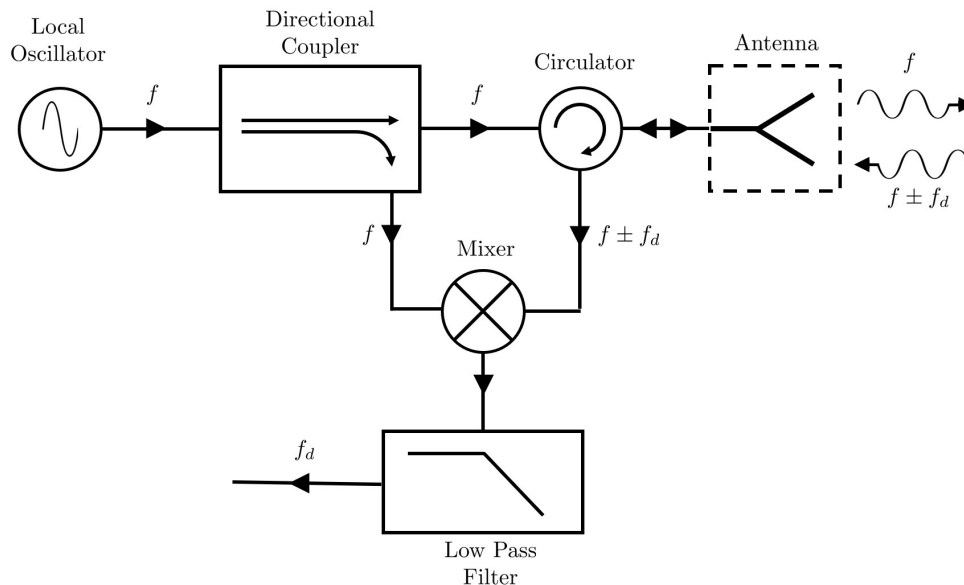
Figure 2.22: Simplified circuit diagram of a Doppler radar.

2.7 Transmission Line Ports and Impedances

Designing practical microwave systems often involves connecting a number of components together to achieve a desired goal. For instance, when we discussed the basic idea of how a Doppler radar worked we needed to use a number of different components to generate our desired signals, route them through our network, and eventually process them. As a refresher, we include again the simplified schematic of the Doppler radar in Fig. 2.22.

Looking at this as a block diagram, we see that the different blocks/components often have a different number of "inputs" and "outputs" depending on the purpose of the device. In a microwave engineering context, we will refer to each of the locations where we can connect to a device as a *terminal* or *port*. To make analysis of large systems tractable, we often only want to consider how a particular device behaves at its ports (i.e., we don't want to worry about finding the voltage and current at every point within the device). This is very much like specifying "input" and "output" relationships of a block diagram like you would see in a signal processing or computer engineering course. The complication here is that because we are dealing with waves that can reflect at various points in a practical device, we must consider every port to simultaneously be an "input" and "output" port! As a result, we will need to use appropriate analysis tools to understand how to design microwave systems. We now begin to learn about how to do this by introducing a few common ways systems are described in terms of ports and what are referred to as *port parameters*. This overall theory is typically referred to as *microwave network analysis*.

2.7.1 Voltage and Current – A Closer Look

To begin, we need to take a harder look at some of the concepts we have been using in describing transmission lines up to this point. In particular, we have been somewhat vague

about what exactly a voltage and current wave look like and how they should be explicitly defined for a given transmission line. The unfortunate reason for this is that with the exception of a particular set of special cases, the definitions of voltage and current for a transmission line are somewhat arbitrary, i.e., there is no unique definition that works for every situation of interest.

We won't dwell on this complication in detail in this course. We will simply mention that a relatively standard way of dealing with this ambiguity is to follow a few guiding principles that often work well for practical situations. First, voltage and current are usually defined to be proportional to the transverse electric and magnetic fields of the transmission line, respectively. Further, we will want to define the voltage and current in a manner so that their product gives us the correct power flow on the transmission line for the particular set of fields that exist (i.e., we want our circuit power flow to match the magnitude of the Poynting vector). Finally, we want the ratio of the voltage to the current to be equal to the characteristic impedance of the transmission line. Unfortunately, because the voltage and current are ambiguous, the characteristic impedance of a general transmission line is also somewhat arbitrary. Typically, we either choose the characteristic impedance to be equal to the wave impedance of the transmission line fields or to be equal to 1.

Now, for simple kinds of transmission lines known as transverse electromagnetic (TEM) lines, we can actually uniquely define voltages and currents. These kinds of lines always must have at least two distinct conductors that can be held at different potentials, as illustrated in Fig 2.23.

Now, for this kind of system the transverse electric field is electrostatic in nature. As a result, we can define a unique voltage as

$$V = \int_+^- \mathbf{E} \cdot d\boldsymbol{\ell}. \tag{2.97}$$

The "uniqueness" is in the sense that the voltage does not depend on the integration path taken because the electric field can be expressed as the gradient of some scalar potential function. To determine the total current of the transmission line, we can use Ampere's law to get

$$I = \oint_C \mathbf{H} \cdot d\boldsymbol{\ell}. \tag{2.98}$$

The integration path is usually taken to be any closed path that encompasses the entire "positive" conductor. We can then determine the characteristic impedance of the line as

$$Z_0 = \frac{V}{I}. \tag{2.99}$$

If we have also determined the propagation constant of the line, we can proceed with using the circuit theory of transmission lines that we have considered previously.

2.7.2 One-Port Network

Let's now consider an arbitrary one-port network to understand how we can use concepts of impedance to make a connection between the full electromagnetic field analysis problem and
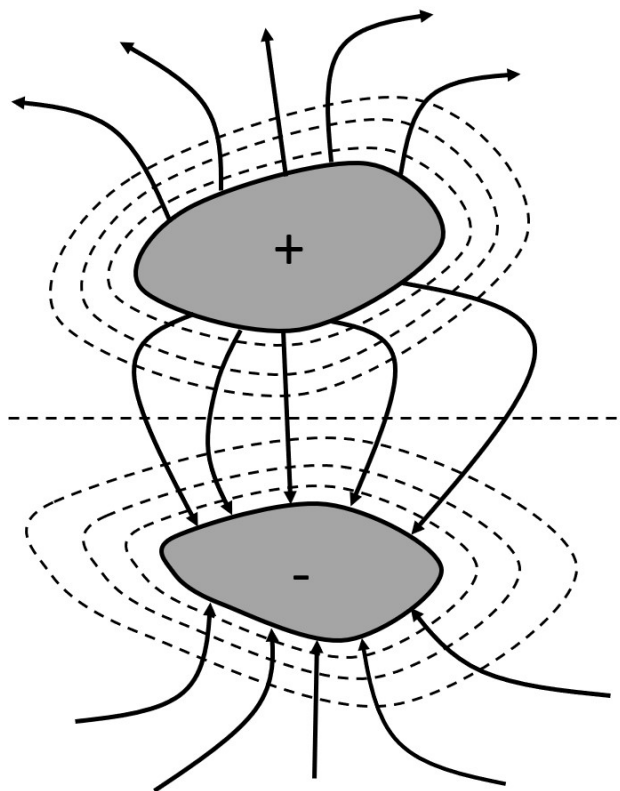
Figure 2.23: Illustration of a two conductor transmission line that can support a TEM mode. The solid (dashed) lines are the electric (magnetic) field lines.
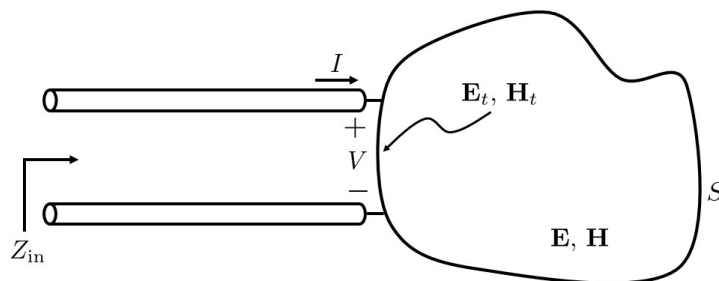
Figure 2.24: Illustration of an arbitrary one-port network from the perspective of Poynting's theorem.

a simpler transmission line network one. We will consider the problem illustrated in Fig. 2.24. We assume that the only way that power and energy can enter/exit the arbitrary one-port network is through a terminal plane. Attached to the terminal plane is a transmission line whose cross section exactly matches the configuration of conductors and materials at the terminal plane of the one-port network.

Considering this, the complex power delivered or extracted from the one-port network is

given by

$$P = \frac{1}{2} \iint_S \mathbf{E} \times \mathbf{H}^* \cdot d\mathbf{S} = P_\ell + 2j\omega(W_m - W_e), \tag{2.100}$$

where P_ℓ is the power dissipated in the one-port network and W_m (W_e) is the total magnetic (electric) energy density in the one-port network. Now, let's assume that we can express the electric and magnetic fields over the terminal plane as

$$\mathbf{E}_t(x, y, z) = V(z)\mathbf{e}(x, y) \tag{2.101}$$

$$\mathbf{H}_t(x, y, z) = I(z)\mathbf{h}(x, y) \tag{2.102}$$

and that the transverse fields $\mathbf{e}$ and $\mathbf{h}$ are normalized such that

$$\iint_S \mathbf{e} \times \mathbf{h}^* \cdot d\mathbf{S} = 1. \tag{2.103}$$

Then we can use these expressions to write the complex power at the terminal as

$$P = \frac{1}{2} \iint_S V(z)I^*(z)\mathbf{e} \times \mathbf{h}^* \cdot d\mathbf{S} = \frac{1}{2}V(z)I^*(z), \tag{2.104}$$

where z would be set at the terminal plane.

We can use this relationship to consider a definition of an overall input impedance for the arbitrary one-port network. In particular, we will have that

$$Z_{\text{in}} = R_{\text{in}} + jX_{\text{in}} = \frac{V}{I}. \tag{2.105}$$

We can work this into a form that relates to the power quantities that we had in (2.104) by multiplying (2.105) by I^*/I^*. This gives us

$$Z_{\text{in}} = \frac{VI^*}{|I|^2} = \frac{2P}{|I|^2}. \tag{2.106}$$

We can now substitute for P from our original statement of Poynting's theorem in (2.100) to get

$$Z_{\text{in}} = \frac{2P_\ell + 4j\omega(W_m - W_e)}{|I|^2}. \tag{2.107}$$

This provides us with a very important result. We have taken a potentially very complicated one-port device and simplified it down to an effective lumped element impedance from the perspective of our transmission line. We can incorporate this lumped element impedance into our transmission line analysis in the same way as we had been doing previously for what we were treating as some kind of arbitrary impedance that was just "there" as part of the problem. This gives us a way to start using our transmission line theory for practical problems involving complex microwave devices.

Now, from (2.107) we can see that the impedance of our network will only have losses if it has a non-zero real part. We also see that if $W_m > W_e$ then the network will be inductive and if $W_e > W_m$ the network will look like a capacitive load. Although we won't go into detail on this, it should be noted that the effective resistance and reactance of the one-port network will generally be a function of frequency (sometimes a strong one at that).

It is very important to remember what an impedance represents in order to properly interpret it. For instance, we have just seen that when we characterize an arbitrary microwave network in terms of a lumped impedance it will only have a real part if the microwave network contains some amount of loss. To make things more complicated, this is the exact opposite behavior to what we had when discussing the characteristic impedance of a transmission line. For transmission lines, the characteristic impedance for a lossless line had to be purely real. We must be careful to keep these things straight. To help with this, we summarize in the next section some details about the different kinds of impedances that we have encountered throughout the course. With time you will become adept at recalling these different properties, but starting out it is essential to make sure you are careful and don't accidentally make a mistake due to the many different kinds of impedance concepts we use in electromagnetic engineering.

2.7.3   Summary of Impedances

1. Intrinsic impedance of a medium

   - Formula: $\eta = \sqrt{\dfrac{\mu}{\epsilon}}$.

   - Purpose: this impedance characterizes how a material medium will affect the relationship between the electric and magnetic fields in an unbounded medium. It is only a function of the material properties of a material.

   - Real/Imaginary/Complex: purely real in a lossless medium, becomes a complex number in a lossy medium.

2. Wave impedance

   - Formula: $Z_w = \dfrac{E_{\text{tan}}}{H_{\text{tan}}}$.

   - Purpose: this impedance is the ratio of the tangential components of the electric and magnetic fields. It depends on the type of wave that exists in a medium, and as a result can be dependent on the type of waveguide supporting the wave. Generally, it will also be a function of frequency. We saw how this could be useful when studying the reflection and transmission coefficients for oblique incidence plane waves on the boundary between homogeneous materials. It is also useful in defining the characteristic impedance of complicated waveguides that don't support TEM waves.

   - Real/Imaginary/Complex: purely real for a propagating wave in a lossless medium. The wave impedance often has η somewhere in its definition, so it will inherit the complex-valued nature of the intrinsic impedance in a lossy medium.

3. Characteristic impedance

- Formula: $Z_0 = \dfrac{V^+}{I^+}$.

- Purpose: this impedance characterizes the ratio of voltage and current of a traveling wave on a transmission line. It depends on the geometry of the transmission line, the materials that may exist around the transmission line, as well as the kind of wave supported by the transmission line.

- Real/Imaginary/Complex: purely real for a lossless transmission line, becomes a complex number for a lossy line.

4. Terminal/Port/Network impedance

- Formula: Z (explicit formula depends on properties of the system it is characterizing).

- Purpose: this impedance characterizes the ratio of the total voltage and current at the terminal or port of an arbitrary transmission line or microwave network. It depends on all of the properties of the system it is characterizing. The input impedance of a terminated transmission line can be viewed as a simple example of this kind of impedance. This kind of impedance is useful for building equivalent circuit models of complex devices.

- Real/Imaginary/Complex: purely imaginary for a lossless system, becomes a real or complex number if the system is lossy.

2.8 Introduction to Port Parameters

Now that we have an idea of what a port is, we can begin to look at how systems with many ports are typically described. We will begin by discussing the impedance and admittance matrix approach, followed by the very important scattering matrix. Often, people will refer to these quantities as being *port parameters*. Due to this terminology, one will often refer to the *impedance parameters* or the *scattering parameters*, for instance, of a system when discussing the properties of different microwave components.

For all of these parameters, the general setup that we are considering is shown in Fig. 2.25. In it, we assume that we have some arbitrary electromagnetic system that eventually transitions into N-ports at which we can imagine connecting a long section of a constant cross-section transmission line. At some location along the transmission line, we consider that we have a *reference* or *terminal plane*. We will define all of our voltages and currents with respect to these reference planes.

We should note before moving on that a single physical "port" of a device may be able to support *multiple kinds of waves*. As a result, we will have to introduce additional *effective ports* to our description of our arbitrary system. In general, each impedance/admittance/scattering parameter will only account for one kind of wave at each physical port a device may have.
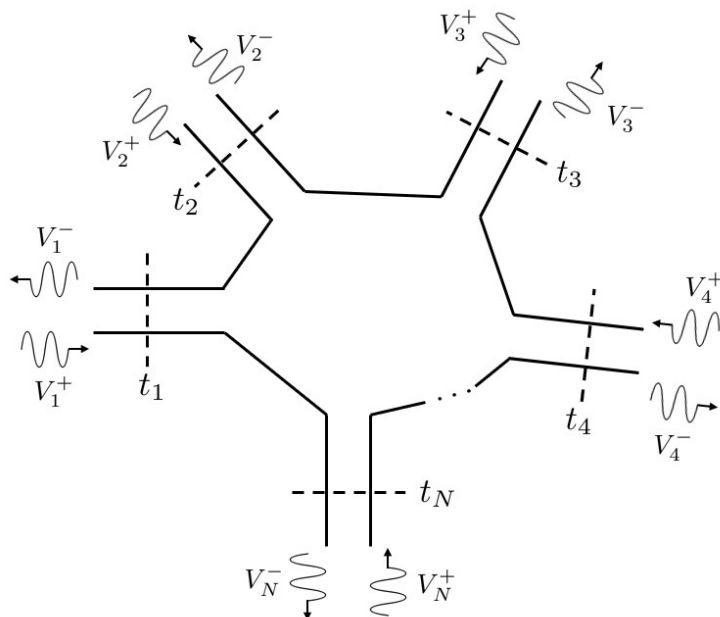
Figure 2.25: Diagram of an arbitrary N-port microwave network.

2.8.1 Impedance Parameters

The impedance parameters of an N-port network relate the total voltage and currents that exist at each terminal plane of a device to each other. At the nth terminal plane, we have that the total voltage and current will be given by

$$V_n = V_n^+ + V_n^-, \tag{2.108}$$

$$I_n = I_n^+ - I_n^-, \tag{2.109}$$

where the right hand sides of each of these equations are written in terms of the forward- and backward-propagating voltages and currents that make up the total waves on the transmission line.

The impedance parameters can then be used to relate these voltages and currents to each other. A matrix notation is typically best for this, so we write the impedance matrix as $[Z]$ and column vectors of voltages and currents as $\{V\}$ and $\{I\}$. Within this notation, our relationship between voltages and currents is compactly written as

$$\{V\} = [Z]\{I\}. \tag{2.110}$$

This can be expanded out explicitly as

$$\begin{Bmatrix} V_1 \\ V_2 \\ \vdots \\ V_N \end{Bmatrix} = \begin{bmatrix} Z_{11} & Z_{12} & \dots & Z_{1N} \\ Z_{21} & Z_{22} & \dots & Z_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ Z_{N1} & \dots & \dots & Z_{NN} \end{bmatrix} \begin{Bmatrix} I_1 \\ I_2 \\ \vdots \\ I_N \end{Bmatrix}. \tag{2.111}$$

Figure 2.26: Circuit diagram of a two-port T-network.

How do we go about calculating or measuring one of these impedance parameters in practice? We know that we need to be careful abo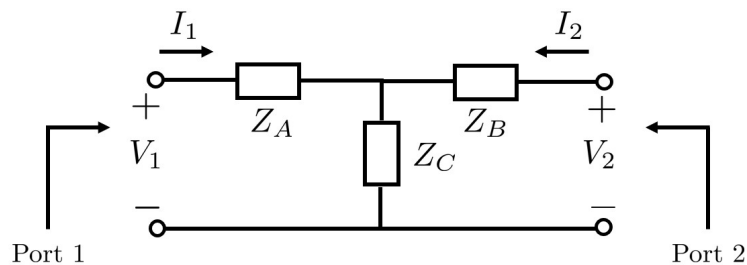ut this due to the wave characteristics that are going to exist in our device. To figure out the correct procedure, we can simply look at the structure of (2.111). In particular, we see that

$$Z_{ij} = \left. \frac{V_i}{I_j} \right|_{I_k = 0, \ \forall k \neq j} . \tag{2.112}$$

In words, we can find Z_{ij} by driving port j with a current I_j and then measuring the total voltage at port i. For this to work properly, we need to make sure that we leave all other ports of the device open-circuited so that no current can exist at the other ports. Due to the complexity in defining voltages for some transmission lines, the measurement of impedance parameters is not extremely common. However, they can be computed from other commonly performed measurements and there are certain kinds of numerical methods that can be used to calculate them directly. As a result, the impedance parameters remain an important diagnostic tool for understanding the properties of a particular microwave network.

Of particular interest are the Z_{ii} entries of the impedance matrix. These are the input impedances of each port when the other ports are open-circuited. This information can be particularly useful when trying to design matching circuits for a particular microwave network. As we will touch on later in the course, designing a matching circuit often involves "canceling" out certain undesirable impedance behaviors as a function of frequency (e.g., some excess reactance, similar to what we had for conjugate matching). Hence, understanding the impedance parameters of a device can be a fundamental step in this design process.

Impedance Parameter Derivation Example

Let's look at how to derive the impedance parameters for a two-port T-network. This system is shown in Fig. 2.26. T-networks find a number of uses in microwave engineering. They can be used in simplifying complex circuits, in providing a reasonable approximation for the behavior of many simple two-port devices, as well as in forming a useful configuration for implementing certain kinds of attenuators.

Let's start by deriving the input impedances, Z_{11} and Z_{22}. Beginning with Z_{11}, we will need to determine the ratio of V_1 and I_1 when port 2 is left open-circuited. We see that for this situation the total impedance seen by port 1 will simply be the series combination of

Z_A and Z_C. Hence, we have that

$$Z_{11} = \left.\frac{V_1}{I_1}\right|_{I_2=0} = Z_A + Z_C. \tag{2.113}$$

A similar process may be followed for Z_{22} to quickly yield

$$Z_{22} = \left.\frac{V_2}{I_2}\right|_{I_1=0} = Z_B + Z_C. \tag{2.114}$$

Now, for the transfer impedance Z_{12} we need to measure the open-circuit voltage at port 1 when port 2 is driven by a current I_2. Since port 1 is open-circuited, no current will flow through Z_A, so the port 1 voltage will simply equal the voltage drop over Z_C. This can be given by a voltage divider to be

$$V_1 = \frac{Z_C}{Z_B + Z_C} V_2. \tag{2.115}$$

The trouble here is that we don't know V_2 because it is the voltage drop over an ideal current source. This dilemma can be quickly solved by noting that if we plug (2.115) into the definition for Z_{12} for V_1 we will get

$$Z_{12} = \left.\frac{V_1}{I_2}\right|_{I_1=0} = \frac{Z_C}{Z_B + Z_C} \frac{V_2}{I_2}. \tag{2.116}$$

We can recognize V_2/I_2 as being equivalent to Z_{22}. Hence, if we plug (2.114) into the above expression for V_2/I_2 we will get

$$Z_{12} = \left.\frac{V_1}{I_2}\right|_{I_1=0} = Z_C. \tag{2.117}$$

Alternatively, we could have also noted that since the ideal current source will cause current I_2 to flow through the series combination of Z_B and Z_C (since port 1 is open-circuited) the actual current in Z_C will be I_2. Hence, the voltage drop will be given by $V = Z_C I_2$, which is equal to V_1 due to the open circuit at port 1. As a result, $V_1/I_2 = Z_C$. Now, by the symmetry of the system, we can quickly recognize that $Z_{21} = Z_C$ as well. We could verify this easily following the same process as what we used to derive the expression for Z_{12} given in (2.117).

Consolidating our results together, we have that

$$[Z] = \begin{bmatrix} Z_A + Z_C & Z_C \\ Z_C & Z_B + Z_C \end{bmatrix}. \tag{2.118}$$

This is a symmetric matrix, which is a characteristic that indicates that this system is *reciprocal*. That is, the transfer characteristics of the device do not depend on which direction we apply a signal from. Most practical microwave components are reciprocal. However, some very important non-reciprocal devices play key roles in building communication and radar systems. For instance, the circulator that we saw in the simplified schematic of a Doppler radar was a kind of non-reciprocal component.

2.8.2 Admittance Parameters

For certain types of systems/circuits, it can be more convenient to work with a set of admittance parameters. These are very similar to the impedance parameters, and are in some sense their "inverse" representation. In particular, we have that

$$\{I\} = [Y]\{V\}, \tag{2.119}$$

where $[Y]$ is the admittance matrix. The individual elements of the admittance matrix can be computed as

$$Y_{ij} = \left. \frac{I_i}{V_j} \right|_{V_k=0, \ \forall k \neq j}. \tag{2.120}$$

In words, we can find Y_{ij} by driving port j with a voltage V_j and then measuring the total current at port i. For this to work properly, we need to make sure that we leave all other ports of the device short-circuited so that no voltage can exist at the other ports. From the similarity of the matrix equations relating voltages and currents in terms of $[Z]$ and $[Y]$, we can quickly see that $[Y] = [Z]^{-1}$.

2.8.3 Scattering Parameters

Although the impedance and admittance parameters can be a useful tool, they are not always convenient to measure, calculate, or use. One of the main difficulties comes from the fact that working with the total voltage and current at a port is not typically how we want to think of a device's operation. Generally, it is far more intuitive to think of the device in terms of a set of input and output relationships. The scattering parameters, often abbreviated as S-parameters, provide us with such a tool.

Introduction

In particular, the S-parameters relate an incident set of voltage waves to the voltage waves leaving each port of an N-port microwave network. In matrix form, known as the scattering matrix, we have

$$\{V^-\} = [S]\{V^+\}, \tag{2.121}$$

which can be expanded out to be

$$\begin{Bmatrix} V_1^- \\ V_2^- \\ \vdots \\ V_N^- \end{Bmatrix} = \begin{bmatrix} S_{11} & S_{12} & \dots & S_{1N} \\ S_{21} & S_{22} & \dots & S_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ S_{N1} & \dots & \dots & S_{NN} \end{bmatrix} \begin{Bmatrix} V_1^+ \\ V_2^+ \\ \vdots \\ V_N^+ \end{Bmatrix}. \tag{2.122}$$

A specific S-parameter can be calculated as

$$S_{ij} = \left. \frac{V_i^-}{V_j^+} \right|_{V_k^+=0, \ \forall k \neq j}. \tag{2.123}$$
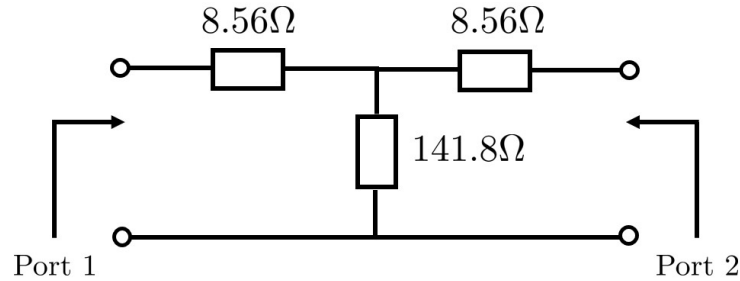
Figure 2.27: Example of a 3 dB attenuator matched to 50 Ω.

In words, we need to measure the voltage wave leaving port i when there is only an incident wave coming in at port j. For all the incident waves to be zero except at port j we need to terminate each port of the device not actively being used with a *matched load*. A matched load perfectly absorbs all of the energy incident on it from the port so that there are no reflections that can become an unintended V^+ at a particular port.

Considering these definitions, we can interpret the different S-parameters in the following manner. Each S_{ii} is the reflection coefficient at a particular port when all other ports have been terminated with matched loads. Similarly, S_{ij} is a transmission coefficient from port j to port i when all other ports have been terminated with matched loads. It is important to remember the ordering of the indices for this relationship, the second index is the source and the first index is the receiver.

Derivation Example

Let's now consider an example of how to derive the S-parameters for a particularly simple device. In particular, we will consider a kind of 3 dB attenuator shown in Fig. 2.27. This device is designed to dissipate half of the power transmitted through it (hence, 3 dB). Although it may initially seem odd for us to want to simply remove power from our system, there are many practical scenarios where it is desirable to use attenuators. A few examples are providing protection to sensitive circuits that can be damaged if too much power is incident on them, equalizing the powers coming from different sources, and improving the impedance match of certain devices. Further, it is often much easier to generate a tunable/variable attenuator rather than a source that has a tunable output power, so variable attenuators can also be useful in generating a desired variable output power level.

To begin, let's find S_{11}. By definition, this will be the reflection coefficient at port 1 when port 2 is terminated with a matched load. To determine this, we need to know what impedance this device is matched to. Due to the symmetry of the device, we can comfortably conclude that the characteristic impedance that port 1 and port 2 should be matched to will be equivalent. Considering this, we can conclude that $Z_{\text{in},1} = Z_0$, where $Z_{\text{in},1}$ is the input impedance at port 1 when port 2 is terminated in Z_0. Considering this, we can expand $Z_{\text{in},1}$ as

$$8.56 + \frac{141.8(8.56 + Z_0)}{141.8 + (8.56 + Z_0)}. \tag{2.124}$$
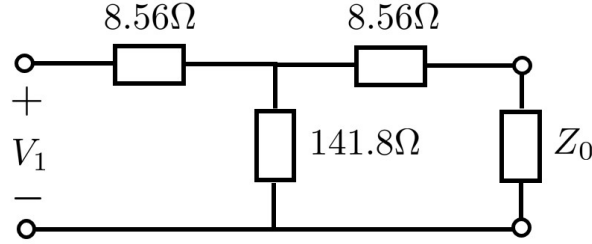
Figure 2.28: Computing S_{21} for the 3 dB attenuator example.

Setting this equal to Z_0, and then solving for Z_0, gives us $Z_0 = 50\,\Omega$. This is actually a standard characteristic impedance that almost all practical microwave engineering circuits operate at. Almost all measurement instruments and COTS devices are matched to 50 Ω.

Now, since we know that $Z_{\text{in},1} = Z_0 = 50\,\Omega$ we can return to calculating S_{11}. To do this, we need to compute the reflection coefficient at port 1 when we attach a transmission line with a 50 Ω characteristic impedance to it and have port 2 terminated in Z_0 as well. This tells us that

$$S_{11} = \frac{Z_{\text{in},1} - Z_0}{Z_{\text{in},1} + Z_0} = 0. \tag{2.125}$$

Due to the symmetry of the device, we see that $S_{22} = 0$ as well.

Let's now compute S_{21}, while also noting that due to the symmetry $S_{21} = S_{12}$. By definition, to calculate S_{21} we need to inject an incident voltage wave V_1^+ at port 1 and measure the output voltage wave V_2^- at port 2. We must do this when $V_2^+ = 0$, which requires that we imagine connecting an infinite length transmission line of impedance Z_0 at port 2. We are then trying to compute the transmission coefficient into this impedance. Now, from the perspective of port 1 there is no difference between having an infinite length transmission line of Z_0 and a lumped impedance of Z_0 at port 2. Hence, we really just need to compute the voltage over a lumped element impedance of Z_0 at port 2 to determine S_{21}. Our corresponding circuit is shown in Fig. 2.28.

We begin finding the voltage over the Z_0 impedance in Fig. 2.28 by first finding the voltage over the parallel combination of 141.8 Ω and $Z_0 + 8.56 = 58.56\,\Omega$. This parallel combination is 41.44 Ω, and by a simple voltage divider we can find the voltage over it as

$$V_{eq} = \frac{41.44}{41.44 + 8.56} V_1 = 0.8288 V_1. \tag{2.126}$$

We then find $V_2^- = V_2$ as another voltage divider of V_{eq} over the 8.56 Ω and Z_0 impedances in Fig. 2.28. This gives us that

$$V_2^- = \frac{50}{50 + 8.56} V_{eq} = 0.707 V_1^+ \approx \frac{1}{\sqrt{2}} V_1^+. \tag{2.127}$$

Hence, we see that $S_{21} \approx 1/\sqrt{2}$. If we recognize that the power transmitted through the device is proportional to $|S_{21}|^2$, we see that half of the input power will emerge at port 2. In

dB, we have the output power is proportional to $20 \log_{10}(|S_{21}|) = -3$ dB. Since there are no reflections at port 1, the remaining power is dissipated in the resistors in the device.

This computation of S_{21} should hopefully help you understand why it is so convenient for us to try and make sure that all of our devices are matched to a single common impedance reference. If this were not the case, we would have significantly more reflected waves occurring in our devices and it would be far more difficult to compute the S-parameters. In reality, no device is matched perfectly and so all of these reflections will still occur and it is often very important for us to computationally account for them for many practical designs. However, if we can keep the reflections low we can typically come up with a reasonable intuition for how a system will operate without requiring these fully detailed analyses. When the full analysis is needed, there are various analytical tools to solve these problems, as well as powerful numerical methods available to us. We won't learn about these more general methods in this class, but if you take a course on microwave engineering you will have an opportunity to learn more about these different analysis methods.

Measuring S-Parameters

Accurately measuring the S-parameters of a microwave device is often a very important task in the development of a design. From the definition of the S-parameters, we see that this kind of measurement is certainly going to require a much more sophisticated piece of technology than what we are used to working with in typical circuits labs. The particular instrument that is most often used to measure S-parameters now is known as a *vector network analyzer (VNA)*. These are truly amazing instruments that automate a large amount of the work involved in measuring S-parameters. A basic block diagram of how these instruments work is included in Fig. 2.29.

One of the key components in the VNA are its *directional couplers*. These are the components that are directly connected to the device under test in Fig. 2.29. The special features of these devices are that they allow us to separately sample the forward and backward traveling waves that go through the directional coupler, which is obviously a vital step in measuring the S-parameters of a device.

Before moving on, it should be noted that VNAs are finely tuned instruments and that measuring the S-parameters *accurately* requires significantly more care than a standard circuit measurement, even when using something as powerful as a VNA. We won't cover the practical aspects of measurements in this course, but we will mention that it is vital to regularly calibrate the VNA (each time you do a new set of measurements, essentially) so that it can utilize its sophisticated error models to provide highly accurate results.

2.8.4   Scattering Parameters of Common Components

We will now briefly discuss a number of common passive components (i.e., they do not provide gain) that are frequently used in the design of microwave systems. To keep our conversation brief, we will only be able to report the ideal S-parameters for these components rather than actually deriving them. The derivation of many of these S-parameters requires the use of analytical techniques that are outside of the scope of this course, but would often be covered in a course dedicated to the study of microwave engineering. We will also briefly discuss
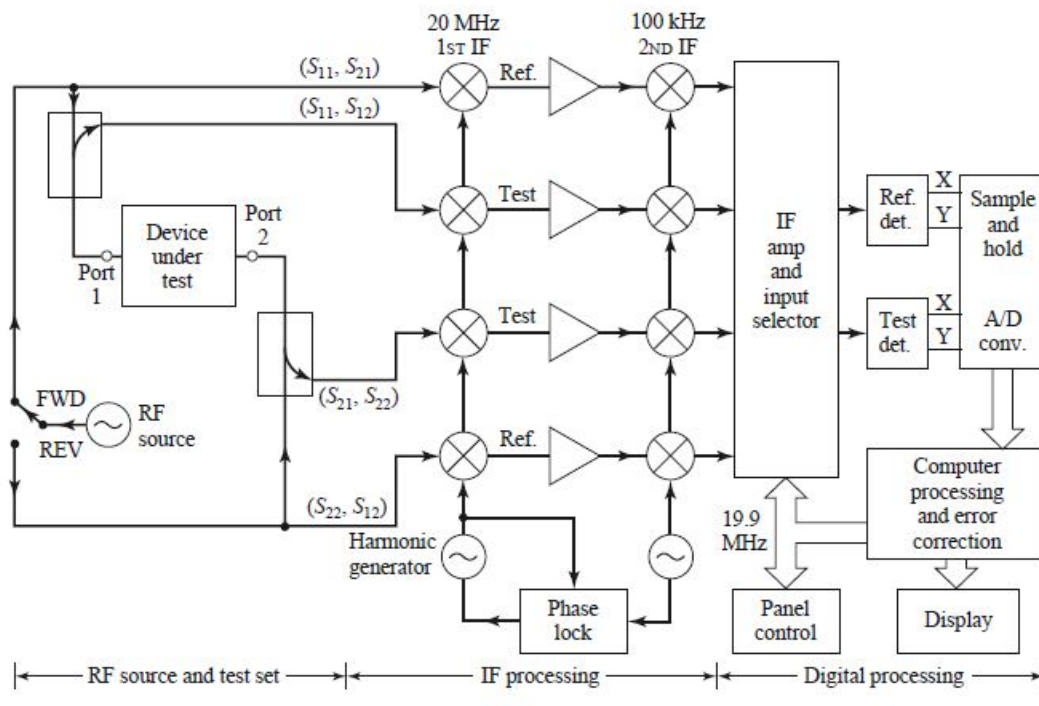
Figure 2.29: Simplified block diagram of a vector network analyzer (image from [13]).

a few of the uses of the different components that we will introduce. We will organize our discussion in terms of the number of ports that the devices have.

2-Port Devices

We will begin our discussion with 2-port devices. As a result, these are all described by 2x2 scattering matrices.

Attenuators – We have already discussed the basic concept of an attenuator as an example of deriving the S-parameters for a simple circuit. Depending on the circumstances, a wide range of different attenuation values may be needed in the design and testing of various microwave systems. As a result, it is possible to buy many different attenuators that have fairly standard values (e.g., 3 dB, 6 dB, 10 dB, 20 dB, etc.) – these are known as *fixed attenuators*. More complicated devices also exist that can provide a variable amount of attenuation based off of external settings – these are known as *variable attenuators*. These can sometimes be digitally controlled, but in older technology they often had to be mechanically controlled (e.g., rotating a part of the component).

The ideal scattering parameters for an attenuator are

$$[S] = \begin{bmatrix} 0 & |\alpha| \\ |\alpha| & 0 \end{bmatrix}, \tag{2.128}$$

where $|\alpha| < 1$ and sets the desired value of attenuation. In reality, the S_{11} and S_{22} values will not be perfectly matched (i.e., will not equal 0). However, because the attenuator is

already providing loss to the system, it is often possible for them to be able to achieve very low values of $|S_{11}|$ and $|S_{22}|$ (e.g., less than -30 dB).

Phase Shifters – Another very useful component that is typically a 2-port device (but doesn't necessarily have to be) is a phase shifter. The goal of a phase shifter is to apply a well-characterized shift in the phase of the signal as it propagates through the device. This can be a very important feature in a number of different microwave systems. An excellent example that we will discuss briefly toward the end of the course is an antenna array. The overall radiation pattern of an antenna array can be controlled by modifying the amplitude and phase of the signals reaching each antenna element within an array. Phase shifters can be a versatile technology to help control the phase part of this engineering problem. Furthermore, it is common to use variable phase shifters that can have the amount of phase shift applied to the signal entering the device changed *in situ*. Digital phase shifters that provide a relatively constant phase shift over a band of frequencies with 6 bits of resolution over the full 360° range are widely available. However, these components are typically rather lossy (e.g., about 1 to 1.5 dB of loss per bit of resolution), so they have to be used carefully in system design.

The ideal scattering parameters for a phase shifter are

$$[S] = \begin{bmatrix} 0 & |\alpha|e^{j\varphi} \\ |\alpha|e^{j\varphi} & 0 \end{bmatrix}, \tag{2.129}$$

where φ sets the desired value of phase shift and $|\alpha|$ describes the (typically unavoidable) loss that occurs in the device. In reality, the S_{11} and S_{22} values will not be perfectly matched, and it is much harder to achieve a good match with a wideband constant-phase phase shifter (i.e., it provides the same phase shift over a band of frequencies). As a result, typical values of $|S_{11}|$ or $|S_{22}|$ may be closer to -10 to -15 dB.

Isolators – Isolators are a kind of non-reciprocal 2-port device that only allows propagation through the device in one direction. This is often useful in protecting certain sensitive components from reflections that may be produced in other nearby components – e.g., preventing the output of an oscillator or amplifier from perturbing the bias point of these devices. Traditionally, isolators were built using magnetic materials to provide the non-reciprocity. More recent designs have found other clever engineering approaches to avoid the typically high loss and heavy weight associated with using magnetic materials (e.g., using time-varying or nonlinear systems).

The ideal S-parameters of an isolator that can only allow transmission from port 1 to port 2 would be

$$[S] = \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}. \tag{2.130}$$

In reality, the isolator may not be perfectly matched, will result in some loss so that $|S_{21}| < 1$, and will not be able to have a perfect amount of isolation (so that $|S_{12}| \neq 0$). Depending on the type of isolator being used, the isolation may be somewhere in the 20 to 30 dB range (i.e., $|S_{12}|$ will be around -20 to -30 dB).

Filters – Filters are often 2-port devices that are designed to provide a particular frequency response in S_{21} (and likewise S_{12} since most filters will be reciprocal) while also

being matched well. However, because the frequency response will be seen as a function of frequency, looking at the scattering matrix at a single frequency isn't particularly helpful. However, ideally, the loss in the filter will be small so that in the passband of the filter the $|S_{21}| \approx 1$, while in the stopband the $|S_{21}| \approx 0$.

3-Port Devices

We will now turn our discussion to 3-port devices, which are all described by 3x3 scattering matrices.

   Power Dividers/Combiners – The first component we discuss will be power dividers. These devices are designed to take a single input and divide it into output signals with equal phase but amplitudes set by the power division ratio of the device. For instance, one of the most common power dividers would be an equal one that splits the input signal into two equal amplitude outputs. For some applications, such as phased array antennas, it can be desirable to utilize an uneven amplitude division to help achieve a desired array weighting function. Since power dividers are reciprocal devices, they can also be used equally well as *power combiners*, where they are now viewed as having multiple inputs that are combined (according to the amplitude ratio of the device) to a single output. It is also possible to design power dividers/combiners that have more than 3-ports, however, in their most basic form they will be 3-port devices.

   As an example, the ideal S-parameters of an *equal division power divider* would be

$$[S] = -j \begin{bmatrix} 0 & 1/\sqrt{2} & 1/\sqrt{2} \\ 1/\sqrt{2} & 0 & 0 \\ 1/\sqrt{2} & 0 & 0 \end{bmatrix}. \tag{2.131}$$

This kind of power divider is known as a *Wilkinson power divider*. A detailed analysis can show that it is impossible for a power divider to be matched, reciprocal, and lossless at the same time. The Wilkinson power divider is matched and reciprocal, and thus incorporates some loss in the system (typically through a carefully located resistance). This loss is also used in maintaining isolation between the two "output ports" (ports 2 and 3). This is why $|S_{23}| = |S_{32}| = 0$ in the given S-parameters.

   Circulators – We discussed the basic idea of a circulator when we discussed the basic functionality of a Doppler radar. We saw that a circulator was a kind of non-reciprocal device that could be useful in separating signals propagating in opposite directions; e.g., the signal traveling to/from an antenna to separate the transmit and receive signals. Like isolators, circulators were traditionally designed using magnetic materials that were heavy, bulky, and potentially more lossy than desirable. As a result, designing compact and efficient circulators is still and important component design task (e.g., this is still a pain point in designing quantum computers that operate at microwave frequencies).

   The ideal S-parameters of a circulator would be

$$[S] = \begin{bmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}. \tag{2.132}$$

It can also be possible to design circulators with more ports. These generalize the above S-parameters in a simple fashion.

4-Port Devices

We will now turn our discussion to 4-port devices, which are all described by 4x4 scattering matrices.

   Directional Couplers – The first 4-port device we will discuss is a directional coupler. We discussed these in the contexts of Doppler radar and vector network analyzers, where in each instance the directional coupler was used to route a small portion of a signal into another direction to be used elsewhere in the device (typically to be compared against another signal). The output port that has the main signal is referred to as the *through port*, while the port with the smaller signal is referred to as the *coupled port*. The amount of power that is routed to the coupled port is given in terms of the *coupling factor*. The fourth port of a directional coupler is known as the *isolated port*, and ideally should have no power exiting it. In reality, it is arbitrary which port we label as the "input port", and so any port can become the through port, the coupled port, or the isolated port depending on how the device is operated.

   The ideal S-parameters of a directional coupler would be

$$[S] = \begin{bmatrix} 0 & \tau & \kappa & 0 \\ \tau & 0 & 0 & \kappa \\ \kappa & 0 & 0 & \tau \\ 0 & \kappa & \tau & 0 \end{bmatrix}, \tag{2.133}$$

where τ is the transmission coefficient (to the through port) and κ is the coupling coefficient (to the coupled port). These are, in general, complex numbers with magnitudes less than 1. In reality, the isolation will not be perfect, so additional terms in the S-matrix will not be 0.

   Hybrid Couplers – A hybrid coupler is a special kind of directional coupler that equally divides the power between the through and coupled ports. It also provides a 90° relative phase shift between the signals exiting the through and coupled ports. As a result, this kind of coupler can be very useful in connecting to two ports of a dual linearly polarized antenna to generate circular polarization. It is also useful in *I-Q* mixers that require decomposing signals into their "real" and "imaginary" components.

   The ideal S-parameters of a hybrid coupler would be

$$[S] = \frac{1}{\sqrt{2}} \begin{bmatrix} 0 & 1 & j & 0 \\ 1 & 0 & 0 & j \\ j & 0 & 0 & 1 \\ 0 & j & 1 & 0 \end{bmatrix}. \tag{2.134}$$

Just like with the regular directional couplers, the isolation will not be perfect so that additional terms in the S-matrix will be non-zero. Also, a completely lossless system is not possible, so there will be some amount of insertion loss lowering the amplitudes of the signals exiting the through and coupled ports. However, these losses can be made small and be balanced so that the desired equal amplitude relationship is still achieved for the signals exiting the through and coupled ports.

2.9 Introduction to the Smith Chart

The Smith chart was created by P. H. Smith in the 1940's at Bell Labs. It is one of the most useful graphical tools for building intuition about many transmission line problems, especially impedance matching problems. Most VNAs and commercial software that is used to analyze microwave engineering designs have a built-in capability for plotting data on a Smith chart.

The Smith chart can often be a daunting tool to learn how to use, so we will introduce it in stages as opposed to jumping directly into looking at the entire chart at once. We will simply mention that there is a reason one of the most popularly used Smith charts available online has the subtitle "Black Magic Design" on it!

Now, the key to understanding the Smith chart is to first recognize that it is based on a polar plot of the reflection coefficient on the complex plane. For a passive device (which is what the standard Smith chart is applicable to), we know that $|\Gamma| \leq 1$. As a result, the Smith chart is completely contained in the unit circle of the complex plane. A particular reflection coefficient is then plotted as $\Gamma = |\Gamma|e^{j\theta}$, where θ is measured from the positive real axis in a counterclockwise direction. If we have a matched system, then $|\Gamma| = 0$, and we will be at the center of the Smith chart. The goal of impedance matching problems is to get to the "center of the Smith chart", i.e., make $|\Gamma| = 0$.

The key realization that Smith had was that for transmission line problems we can make a one-to-one correspondence between a particular reflection coefficient and *normalized impedances or normalized admittances*. To see this, we recall that

$$\Gamma = \frac{Z_L - Z_0}{Z_L + Z_0}. \tag{2.135}$$

To use the Smith chart, we will work with normalized quantities, where the normalization is often taken to be the characteristic impedance of the transmission line. We will write normalized quantities as lowercase letters, which is a common convention. Hence, we have that

$$\Gamma = \frac{z_L - 1}{z_L + 1}, \tag{2.136}$$

where $z_L = Z_L/Z_0$. The Smith chart provides us with a very quick way to change back and forth between these representations to quickly solve problems without needing to use any complex formulas. This is why it is so valuable for gaining intuition about solving transmission line problems.

To see how this is done, we first solve for z_L in (2.136) to get that

$$z_L = \frac{1 + \Gamma}{1 - \Gamma}. \tag{2.137}$$

We can write this in terms of real and imaginary parts as

$$r_L + jx_L = \frac{(1 + \Gamma_r) + j\Gamma_i}{(1 - \Gamma_r) - j\Gamma_i}. \tag{2.138}$$

We can separate this into real and imaginary parts to get

$$r_L = \frac{1 - \Gamma_r^2 - \Gamma_i^2}{(1 - \Gamma_r)^2 + \Gamma_i^2} \tag{2.139}$$

$$x_L = \frac{2\Gamma_i}{(1 - \Gamma_r)^2 + \Gamma_i^2}. \tag{2.140}$$

Both of these equations can be rearranged to get equations for two families of circles in the (Γ_r, Γ_i)-plane. Doing this gives us,

$$\left(\Gamma_r - \frac{r_L}{1 + r_L}\right)^2 + \Gamma_i^2 = \left(\frac{1}{1 + r_L}\right)^2 \tag{2.141}$$

$$(\Gamma_r - 1)^2 + \left(\Gamma_i - \frac{1}{x_L}\right)^2 = \left(\frac{1}{x_L}\right)^2. \tag{2.142}$$

Let's start by considering the family of r_L-circles given by (2.141). The first thing to recognize is that all of these circles have their centers on the real axis. Next, we see that as r_L increases, the circles will get smaller and will have their centers move progressively toward the $(1, i0)$ point on the complex plane. This point corresponds to a real impedance of ∞, which we can recognize as being equivalent to an open circuit. At the opposite side of the Smith Chart (i.e., the leftmost side), we have the $(-1, i0)$ point, which corresponds to a short circuit. Finally, we can note that all of the r_L-circles pass through the "open circuit" point at the rightmost side of the Smith chart. A plot with a few of these circles drawn on it is shown in Fig. 2.30.

Next, we consider the family of x_L-circles given by (2.142). These circles have centers on the $\Gamma_r = 1$ line at the rightmost edge of the Smith chart. All circles with $x_L > 0$ have points above the Γ_r axis and those with $x_L < 0$ have points below the Γ_r axis. Hence, inductive reactances are always above the real line and capacitive reactances are always below the real line. Similar to the r_L-circles, as x_L increases the circles become smaller and move closer to the real line, eventually merging with the open circuit point on the Smith chart. Further, we can note that all of the x_L-circles pass through the "open circuit" point as well. Some of these circles are also shown in Fig. 2.30.

Not drawn on most Smith charts are constant $|\Gamma|$ circles. These are circles that are centered on the origin of the complex plane. Every point on one of these circles has a radius that is equal to the magnitude of the reflection coefficient, i.e., $|\Gamma|$. Further, whatever angle a particular normalized impedance point makes with the positive real axis is also equal to the angle of the reflection coefficient. Although these circles don't come standard on the Smith chart, we often draw them ourselves as part of solving transmission line problems. To do this accurately by hand requires the use of a drawing compass. These circles are very useful in determining the input impedance of a particular loaded transmission line some length forward or backward along a transmission line.
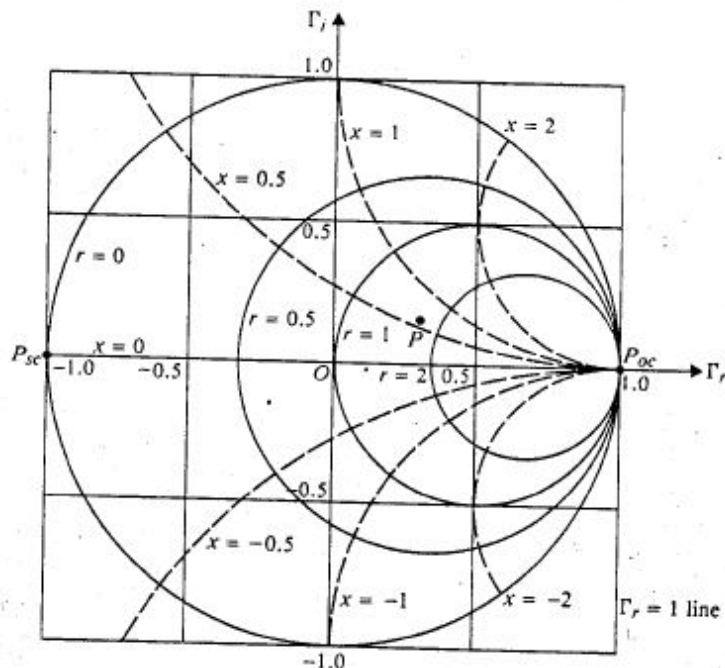
Figure 2.30: Development of the Smith chart. A few constant resistance circles are shown in solid lines and a few constant reactance circles are shown in dashed lines. Point labeled by P is $z_L = 1.7 + j0.6$, which corresponds to a $\Gamma = 0.33 \angle 28.07°$ (image from [14]).

To see why this is the case, we need to recall that the input impedance of a loaded transmission line at location $-\ell$ from the load is

$$Z_{\text{in}} = \frac{V(-\ell)}{I(-\ell)} = Z_0 \frac{e^{j\beta\ell} + \Gamma_L e^{-j\beta\ell}}{e^{j\beta\ell} - \Gamma_L e^{-j\beta\ell}} = Z_0 \frac{1 + \Gamma_L e^{-2j\beta\ell}}{1 - \Gamma_L e^{-2j\beta\ell}}. \tag{2.143}$$

For a normalized input impedance, we will have

$$z_{\text{in}} = \frac{1 + \Gamma_L e^{-2j\beta\ell}}{1 - \Gamma_L e^{-2j\beta\ell}}. \tag{2.144}$$

We can consider the term $\Gamma_L e^{-2j\beta\ell}$ to simply be a new reflection coefficient referenced to a different position along the line (in particular, the position is $-\ell$). If we write $\Gamma_L = |\Gamma_L| \angle \theta_{\Gamma_L}$, we can recognize that our new reflection coefficient $\Gamma(-\ell)$ will have the same magnitude as Γ_L. Since only the phase of $\Gamma(-\ell)$ changes, we can find the corresponding z_{in} on the Smith chart that would be associated with a reflection coefficient $\Gamma(-\ell)$ by simply rotating in a clockwise direction along a constant $|\Gamma|$ circle.

To facilitate this operation, the Smith chart has a set of "scales" measured in electrical length along the perimeter of the Smith chart (by electrical length, we mean a physical distance normalized by the operating wavelength). You determine the point to rotate to by first drawing a radial line from the origin and through the point z_L out to the perimeter of the Smith chart. You read the point on the scale you are at, move along the scale marked "wavelengths toward generator" an appropriate amount, and then draw a new radial line

through the point on the perimeter you are at to the origin of the Smith chart. Where this line intersects the constant $|\Gamma|$ circle you had been on originally marks your new z_{in}!

At this point, we can now appreciate a full Smith chart, which is shown in Fig. 2.31. There are many other tricks that can be performed with the Smith chart to rapidly determine other quantities of interest, like the standing wave ratio or return loss of a loaded line. We won't go into detail on these methods in this course, however, you can learn about some of them from your textbook and other online references if you are interested. Instead, we will be more interested in having an intuitive understanding of the Smith chart and how it can be "traversed" in the course of practical design problems like impedance matching.

2.10   Single-Stub Impedance Matching

To get an idea of how the Smith chart can be used in practical design scenarios, we will look at an example of what is known as a single-stub matching circuit. The goal of this circuit is to take a generic load impedance Z_L and match it to a lossless transmission line so that looking into the single-stub matching circuit there are no reflections (i.e., $Z_{\text{in}} = Z_0$). The configuration of a single shunt-stub matching circuit is shown in Fig. 2.32.

Due to the stub being in parallel to the transmission line with the load impedance, it is easiest to think of the operation of this matching circuit in terms of admittances. Considering this, the basic idea of the circuit is to transform the admittance $Y_L = 1/Z_L$ back along a length of transmission line until it has an admittance with its real part equal to $Y_0 = 1/Z_0$ and its susceptance equal to some arbitrary number B. A "stub" is then connected in parallel to this part of the circuit, and consists of either an open- or short-circuited length of transmission line. We learned previously that by choosing an appropriate length of line, this stub will be able to achieve an input impedance that is equal to any possible susceptance. Hence, we simply need to choose a susceptance that will cancel out the susceptance B of the transformed load admittance due to the parallel combination of the transformed load admittance and the stub. Since admittances add in parallel, we can determine that the needed susceptance of the stub will simply be $-B$.

It is certainly possible to determine analytical expressions for what ℓ and d (see Fig. 2.32 for definitions) need to be to achieve the desired match. However, this is fairly unintuitive and rather tedious. Solving these problems on a Smith chart is quite easy, and can provide better intuition as to how certain circuit components can be used to transform a load impedance around the Smith chart to a new desired point.

As mentioned earlier, actually solving this kind of problem is most easily done in terms of admittances (since the parallel combination of the stub and loaded line becomes a simple addition). One can follow some simple steps to use the Smith chart (which is written in terms of normalized impedances) to be used in terms of normalized admittances instead. We won't go through the details here, but to determine the corresponding normalized admittance (denoted as y_L) from a particular normalized impedance one only needs to draw a constant $|\Gamma|$ circle through the z_L point. The y_L point is then the point on the constant $|\Gamma|$ circle that is diametrically opposite to the z_L point.

The Complete Smith Chart

Black Magic Design
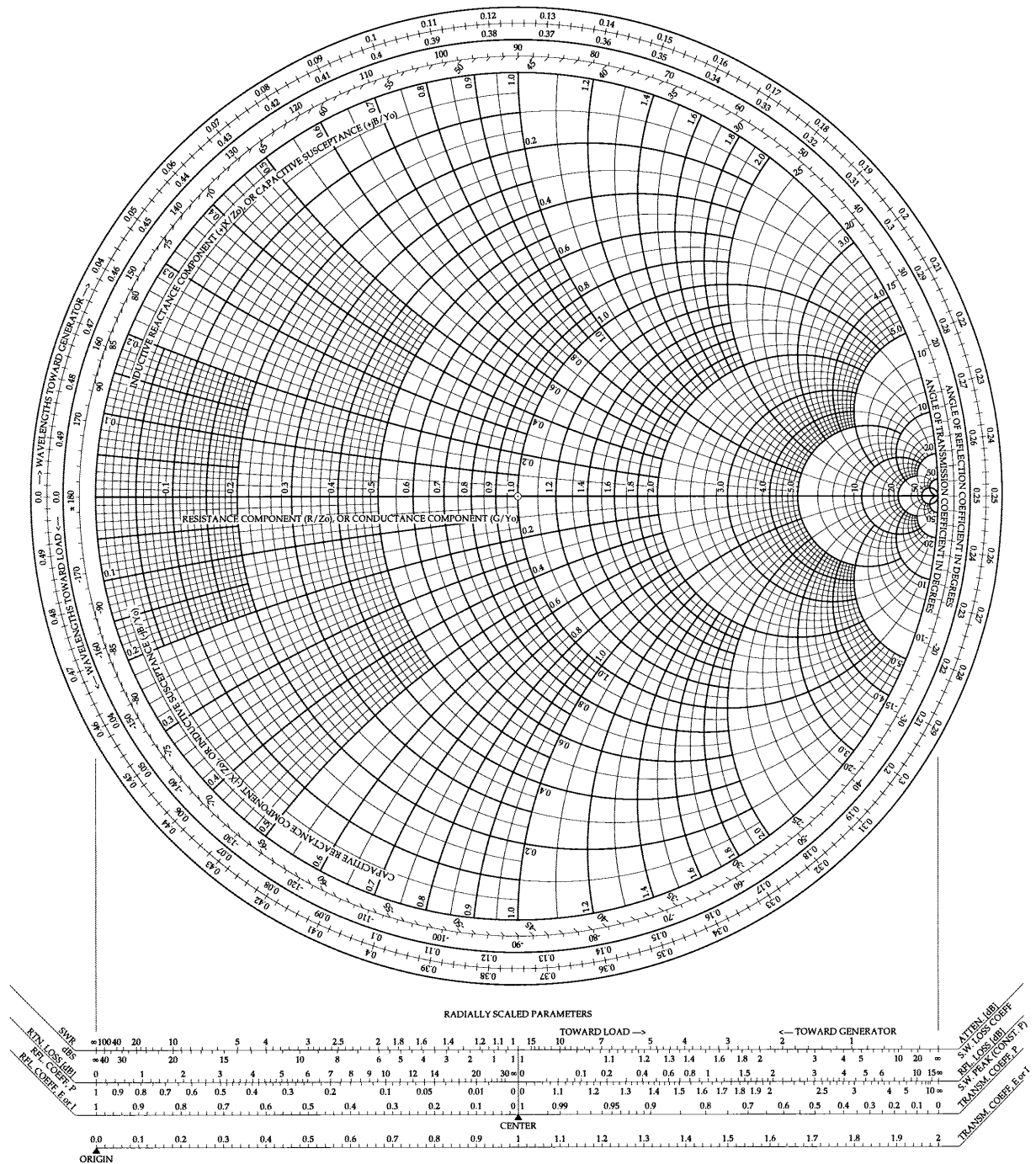

RADIALLY SCALED PARAMETERS

Figure 2.31: Full Smith chart that can be used in solving transmission line problems.
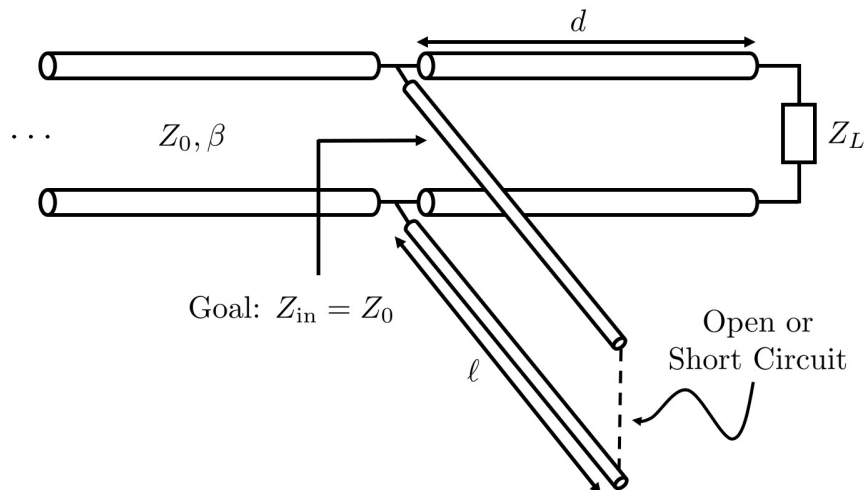
Figure 2.32: Circuit schematic of a single-stub matching circuit. The "stub" is the portion coming off of the main transmission line sections in parallel to Z_L and the section of line with length d.

2.10.1 Example

Let's now work out an example. The filled in Smith chart is shown in Fig. 2.33. We will list out the steps in order here. The problem that we are tasked with is matching a normalized load impedance of $z_L = 0.55 - j0.3$.

1. Determine d:

   (a) Mark the normalized load impedance on the Smith chart. This is the point labeled z_L in Fig. 2.33.

   (b) Draw the constant $|\Gamma|$ circle that intersects z_L. This is the smaller blue circle shown in Fig. 2.33.

   (c) We want to work with admittances, since this will make the parallel combination of the loaded line and stub easier to calculate. To find the normalized load admittance, we "reflect" our z_L point through the origin to the opposite side of the constant $|\Gamma|$ circle. The result is marked with the point y_L in Fig. 2.33.

   (d) Later on we will need to know how long d should be. To find this we will need to know where y_L lies on the "wavelength toward generator" scale on the perimeter of the Smith chart. We have extended the line through the origin, z_L, and y_L far enough so that we can read the desired number off the scale as 0.188λ.

   (e) We now need to determine the points on our constant $|\Gamma|$ circle that will intersect the "normalized conductance equals 1" circle. This is the larger blue circle in Fig. 2.33. We see that our constant $|\Gamma|$ circle intersects this new circle at two points, marked as S_1 and S_2. We can read these points off of the Smith chart to see that they are $S_1 = 1 - j0.75$ and $S_2 = 1 + j0.75$.

(f) We can now "rotate" our normalized admittance around the constant $|\Gamma|$ circle by extending the length of line that connects it to the point where we add in the stub (i.e., by changing d). This will rotate us in a clockwise direction, so we see that we are going to reach the point S_1 far sooner than we can reach S_2. Correspondingly, the length of line to get to S_1 will be smaller, which is usually the more practical design from an engineering perspective. We have extended a line through the origin and S_1 out to the edge of the Smith chart to read off what this point is on the "wavelength toward generator" scale. The result we find is 0.347λ.

(g) We can determine d as the difference between the two spots we marked on the "wavelength toward generator" scale. That is, $d = 0.347\lambda - 0.188\lambda = 0.159\lambda$.

2. Determine ℓ

(a) We can determine ℓ on the Smith chart as well, however, this leads to a rather crowded figure so we have opted to do this using formulas instead. Since we are only dealing with a single open- or short-circuited stub, the formulas are quite simple. What we need to do is find the length of line that gives us a normalized susceptance that cancels the susceptance at the point S_1. Recalling that $S_1 = 1 - j0.75$, we see that we need a susceptance of $b = +j0.75$.

(b) Now, we need to choose whether to use a short- or open-circuited stub. For many transmission line geometries, it may only be possible to do a good job implementing either a short circuit or an open circuit. However, some planar transmission lines can allow either short or open circuits given the right conditions. All things equal, it is usually preferable to choose whichever solution will lead to a smaller stub length. For this problem, we can see that the $j0.75$ circle will intersect the $|\Gamma| = 1$ circle closer to the leftmost side of the Smith chart, so we will want to use an open circuit stub. To see this, remember that the open circuit *impedance* point is the rightmost spot on the Smith chart, but as an admittance this gets reflected through the origin which makes the *leftmost* spot the open circuit admittance point.

(c) We can now use an impedance transformation formulas we encountered earlier to compute ℓ. For an open-circuited line we had that the normalized input impedance would be

$$z_{\text{in}} = -j\cot(\beta\ell). \tag{2.145}$$

In terms of admittances, this will become

$$y_{\text{in}} = j\tan(\beta\ell). \tag{2.146}$$

We can solve for ℓ to get $y_{\text{in}} = j0.75$. This will tell us that $\ell = 0.1024\lambda$.
Note: we could also have read this distance off of the Smith chart quite easily. Try and do this yourself to test your understanding.

(d) We have graphically marked what this shunt stub does in Fig. 2.33 in green. In particular, as we add a compensating susceptance we rotate in a clockwise direction along the "normalized conductance equals 1" circle toward the origin. The reason we rotate along this circle is because we are modifying the susceptance of our system without changing the conductance. As a result, we stay on the same constant conductance circle as we had started on.
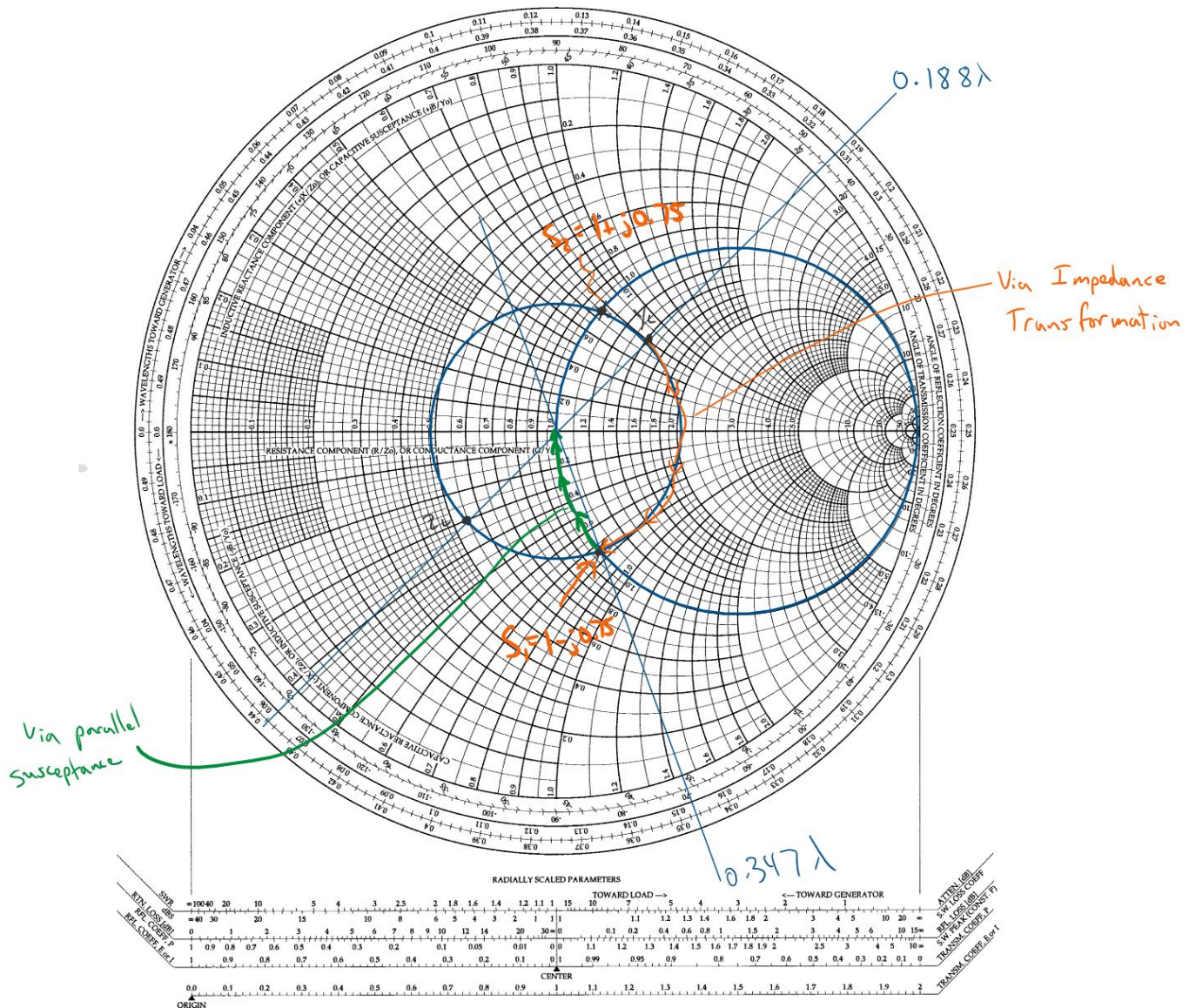

Figure 2.33: Smith chart solution of a single-stub matching circuit design.

165

2.11 Practice Problems

1. A general TEM wave propagating in the $+z$-direction will have $E_z = H_z = 0$ by definition. Considering this, find the relations between $E_x(x,y)$, $E_y(x,y)$, $H_x(x,y)$, and $H_y(x,y)$ by using the source-free Maxwell's equations and noting that $E_z = H_z = 0$. You will further want to note that all field components have an exponential phase factor of $\exp[-j\beta z]$.

2. Beginning from the scalar Helmholtz wave equation, show that a general TEM wave propagating in the z-direction will have all four scalar field components satisfy the two-dimensional Laplace's equation.

3. Assume that the fields within a coaxial cable are given by

$$\mathbf{E} = \frac{V_0\hat{\rho}}{\rho \ln(b/a)} e^{-\gamma z}, \tag{2.147}$$

$$\mathbf{H} = \frac{I_0\hat{\phi}}{2\pi\rho} e^{-\gamma z}, \tag{2.148}$$

where a is the radius of the inner conductor and b is the radius to the beginning of the outer conductor. Inside the region $a < \rho < b$, the transmission line is filled with a material characterized by $\epsilon = \epsilon' - j\epsilon''$ and $\mu = \mu'$. The conductors in this problem can also be characterized with a surface resistivity of R_s. Considering this, answer the following questions:

   (a) Find the per unit length inductance,
   (b) Find the per unit length capacitance,
   (c) Find the per unit length resistance,
   (d) Find the per unit length conductance,
   (e) Find the characteristic impedance of this line assuming that the losses can be neglected.

4. For a purely reactive load impedance and a purely real characteristic impedance, show that the magnitude of the reflection coefficient is always one.

5. A lossless transmission line with characteristic impedance $Z_0 = 50\,\Omega$ and length $\ell = 0.7\lambda$ is terminated with a load impedance $Z_L = 45 + j30\,\Omega$. Find the following quantities.

   (a) The reflection coefficient at the load.
   (b) The standing wave ratio on the line.
   (c) The input impedance of the system.
   (d) If the end of this transmission line (i.e., the part that is 0.7λ away from the load) were connected to another transmission line with characteristic impedance of $100\,\Omega$, what would be the resulting reflection coefficient at the connection point?

6. Beginning with the formula for the propagation constant in a lossy transmission line, show that if

$$\frac{R}{L} = \frac{G}{C} \tag{2.149}$$

the group velocity on the transmission line will equal the phase velocity.
Point of Interest: This means that there will be no dispersion on the line. This kind of line is referred to as a *distortionless line*, and often requires special engineering of the transmission line to achieve this behavior if it is needed.

7. For the circuit shown in Fig. 2.34, calculate the time-average power propagating forward in the finite-length transmission line (P_i), the time-average power propagating backward in the finite-length transmission line (P_r), and the time-average power propagating in the forward direction in the infinitely-long transmission line (P_t). Note that Z_0 is the characteristic impedance of the half-wavelength section of transmission line and Z_1 is the characteristic impedance of the infinitely-long transmission line.
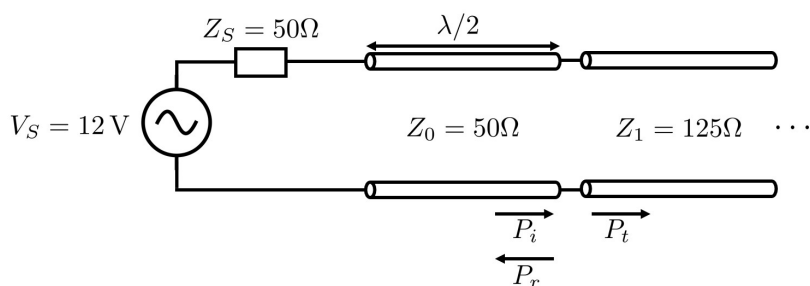

Figure 2.34

Hint: think about a simple way to represent an infinitely-long section of transmission line from the perspective of the rest of the circuit that the infinitely-long line is attached to.

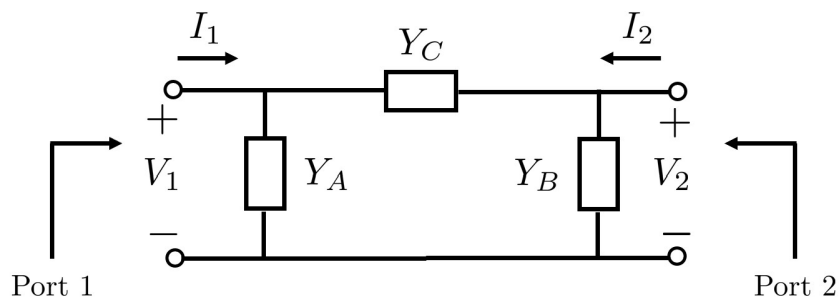8. Derive the admittance parameters for the π-network shown in Fig. 2.35.

Figure 2.35: Example of a two-port π-network.

Note: Pay attention to the direction convention for the currents when evaluating the admittance parameters. It is easy to accidentally be off by a minus sign if you aren't careful. You will not receive full credit for the problem if you do an incorrect procedure but get the correct answer due to errors canceling out.

9. Determine the Y_{11} and Y_{22} elements of the admittance matrix for the transmission line circuit shown in Fig. 2.36.
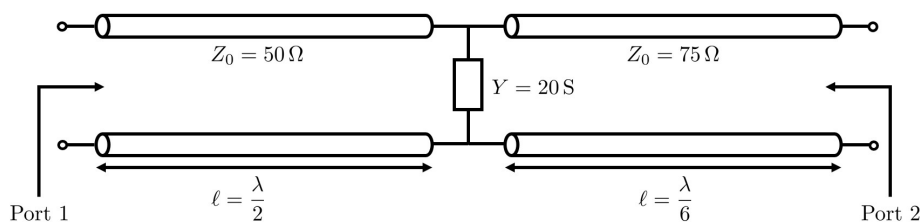
Figure 2.36

10. Assuming that the characteristic impedance for every port of an arbitrary n-port network is equal to 1, show that the scattering matrix can be computed from the impedance matrix as

$$[S] = ([Z] + [I])^{-1}([Z] - [I]) \qquad (2.150)$$

where $[I]$ is the n-dimensional identity matrix.

11. Use the circuit shown in Fig. 2.37 to answer the following questions. For this problem, all of the transmission lines have a characteristic impedance of 50 Ω and $\beta = 200\pi$.
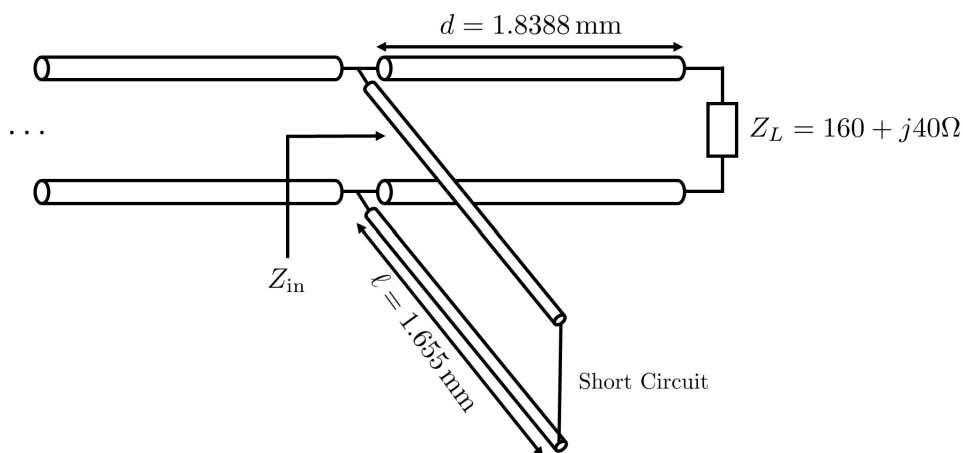
Figure 2.37

(a) What is the input impedance of the circuit at the location marked by Z_{in}?

Note: This input impedance must account for the combination of **both** of the transmission line segments shown (i.e., the one that is d long and the one that is ℓ long).

(b) From (a), you have found that this circuit is not matched to the $50\,\Omega$ transmission line it is connected to. You have been tasked with changing the length of **one** of the sections of transmission line (**either** the 1.8388 mm line or the 1.655 mm line, but not both) so that this circuit functions correctly as a single stub matching circuit. Determine which line length should be adjusted and find the correct length it should be.

Hint: Review the lecture material on how a single stub matching circuit works. Determine a test you can make to see if the line of length d is accomplishing its purpose for a properly functioning single stub matching circuit. If it is not working than d must be fixed. If it is working then ℓ must be fixed.

Chapter 3

Waveguides and Resonators

3.1   TEM, TE, and TM Waveguide Modes

In the last portion of the course, we discussed how a one-dimensional transmission line theory could be used to solve many practical problems in microwave engineering. We also saw that in order to use this simplified theory we would need to be able to derive certain characteristic values about our transmission line, such as its propagation constant and characteristic impedance. At the time, we simply assumed that we already knew these values so that we could focus on learning the fundamentals of transmission line theory. We now revisit this problem by performing a more complete field analysis of common transmission line systems. From this field analysis, we will be able to calculate the properties needed to apply transmission line theory for an "arbitrary" geometry. Of course, we will only be able to perform a full analytical analysis of fairly simple geometries. Numerical methods or judicious approximations are typically necessary to handle more complex transmission line geometries.

To begin this process, we will consider different kinds of solutions to Maxwell's equations in a general waveguide geometry. In particular, we will consider the properties of three different kinds of solutions; namely, transverse electromagnetic (TEM), transverse electric (TE), and transverse magnetic (TM) modes. In this section, we will focus on deriving the general equations that these different kinds of waves will need to obey for a particular geometry. We will then consider solving these equations for a rectangular waveguide later.

3.1.1   General Equations

To determine the general equations that will be useful for analyzing waveguide geometries, we will assume that we have a geometry that is oriented along the z-axis, is infinitely long, and has a constant cross sectional shape over the entirety of the waveguide. Under this assumption, we can assume that we will have a simple propagating wave characteristic for the z-dependence of the electric and magnetic fields contained in the waveguide. Due to the orientation of the geometry, it makes since to break our electric and magnetic fields into their *transverse* (i.e., in the cross section of the waveguide) and *longitudinal* (i.e., along the

length of the waveguide) components. Considering this, we can write our fields as

$$\mathbf{E}(x, y, z) = [\mathbf{E}_t(x, y) + \hat{z}E_z(x, y)]e^{-j\beta z}, \tag{3.1}$$

$$\mathbf{H}(x, y, z) = [\mathbf{H}_t(x, y) + \hat{z}H_z(x, y)]e^{-j\beta z}, \tag{3.2}$$

where $\mathbf{E}_t$ and $\mathbf{H}_t$ contain the transverse components (i.e., x- and y-components) of the electric and magnetic fields, respectively.

For a general waveguide analysis, we will be considering the electromagnetic fields that exist in the source-free region contained inside of a particular waveguide geometry. Hence, we can use the source-free form of Maxwell's equations to derive the Helmholtz wave equation for $\mathbf{E}$ and $\mathbf{H}$ following our standard process. This gives us for the electric field

$$\nabla^2 \mathbf{E} + k^2 \mathbf{E} = 0, \tag{3.3}$$

where $k = \omega\sqrt{\mu\epsilon}$ is, as usual, the wavenumber. If we write

$$\nabla^2 = \nabla_t^2 + \partial_z^2, \tag{3.4}$$

we can simplify our wave equation given the known z-dependence of our field given in (3.1). In particular, we will get that

$$\nabla_t^2 \mathbf{E} + (k^2 - \beta^2)\mathbf{E} = 0. \tag{3.5}$$

This looks similar to some of the expressions we saw when discussing the separation of variables in the context of arbitrarily-directed plane waves. This is no accident, and it is advantageous to define a new kind of wavenumber suggested by (3.5) as

$$k_c^2 = k^2 - \beta^2. \tag{3.6}$$

We refer to k_c as the *cutoff wavenumber*. The reason for this terminology can be seen by rearranging this into the form of a *dispersion relation* as

$$\beta^2 = k^2 - k_c^2. \tag{3.7}$$

We see that if $k^2 < k_c^2$, our propagation constant will become imaginary. Recalling the assumed z-dependence of our fields in (3.1), we see that this will correspond to an exponentially attenuating set of fields as a function of z. Hence, we will not have *wave propagation* unless our frequency is such that our wavenumber is *above* the cutoff wavenumber. For the kinds of geometries we have been considering in class up to this point, we haven't typically had to worry about this notion because it turned out that the cutoff wavenumber was 0. However, we will consider shortly cases where the cutoff wavenumber is greater than 0, which causes these particular waveguide geometries to act as a kind of high pass filter (i.e., only signals above a certain frequency will be able to be transmitted through them).

Now, to see a useful strategy for actually solving practical waveguide problems it can be more advantageous to return to the source-free form of Maxwell's equations. If we evaluate Faraday's law in terms of Cartesian coordinates for (3.1) and (3.2), we can get

$$\partial_y E_z + j\beta E_y = -j\omega\mu H_x, \tag{3.8}$$

$$-j\beta E_x - \partial_x E_z = -j\omega\mu H_y, \tag{3.9}$$

$$\partial_x E_y - \partial_y E_x = -j\omega\mu H_z. \tag{3.10}$$

A similar set of equations can be derived from Ampere's law. It turns out that all of these equations can be rearranged to express the transverse components of **E** and **H** in terms of the longitudinal components of **E** and **H**. In particular, we get that

$$H_x = \frac{j}{k_c^2}(\omega\epsilon\partial_y E_z - \beta\partial_x H_z), \tag{3.11}$$

$$H_y = -\frac{j}{k_c^2}(\omega\epsilon\partial_x E_z + \beta\partial_y H_z), \tag{3.12}$$

$$E_x = -\frac{j}{k_c^2}(\beta\partial_x E_z + \omega\mu\partial_y H_z), \tag{3.13}$$

$$E_y = \frac{j}{k_c^2}(-\beta\partial_y E_z + \omega\mu\partial_x H_z). \tag{3.14}$$

The important result is that if we can solve for E_z and H_z we can calculate the rest of the components without returning to solving the wave equation from scratch. This turns out to be very useful for waveguide problems. It also suggests a useful way to decompose the different wave solutions that may exist into different "families" based on their characteristics in terms of E_z and H_z. In particular, we will consider TEM modes characterized by $E_z = H_z = 0$, TE modes characterized by $E_z = 0$, $H_z \neq 0$, and TM modes characterized by $E_z \neq 0$, $H_z = 0$. We will now consider some general properties about each of these different kinds of waves before explicitly finding these kinds of solutions for a rectangular waveguide.

3.1.2 TEM Waves, $E_z = 0$, $H_z = 0$

The first thing to note about these waves is that if we try to use (3.11) to (3.14) we won't get very far because everything will look like a trivial solution. However, we can actually notice from these equations an important detail about k_c for TEM waves. In particular, there were no undue assumptions in the derivation of (3.11) to (3.14), so these can be considered to be valid equations. As a result, we see that the only way that all of our transverse components of **E** and **H** won't be trivial will be if (3.11) to (3.14) produce some kind of indeterminate result. This will only be possible if $k_c = 0$, which is a property of TEM waves (i.e., they have no cutoff frequency).

Although this is a useful conclusion, it doesn't help us to determine the spatial structure of the transverse field components. To do this, we will need to return to the Helmholtz wave equation and simplify it for $\mathbf{E} = \mathbf{E}_t e^{-j\beta z}$ (since $E_z = 0$). You actually have already done this as a homework problem, where you saw that both $\mathbf{E}_t$ and $\mathbf{H}_t$ will need to satisfy Laplace's equation in the transverse cross section of the waveguide geometry for a TEM wave. That is,

$$\nabla_t^2 \mathbf{E}_t = 0, \tag{3.15}$$

$$\nabla_t^2 \mathbf{H}_t = 0. \tag{3.16}$$

We won't consider in detail solving these equations for waveguide geometries in this class. The important concept to note is that because these equations are the equations that govern static fields, we can conclude that the spatial structure of the transverse fields of a TEM wave are the same as the static fields that can exist between the conductors. Hence, the concepts you learned in your earlier electromagnetics courses can be valuable for intuitively picturing what the fields will look like for a particular TEM wave.

A few other points of interest for TEM waves are the following.

1. Due to the static field behavior, TEM waves can only exist if we have multiple conductors in our geometry (like in a coaxial cable or a parallel plate waveguide).

2. Since $k_c = 0$, we have that $\beta = k = \omega\sqrt{\mu\epsilon}$. This tells us that the phase velocity of our TEM waves will be constant. As a result, the only dispersion that will exist for our TEM waves is due to the dispersion inherent in the material filling the waveguide itself.

3. The wave impedance is equal to the intrinsic impedance of the materials filling the transmission line, that is

$$Z_{\text{TEM}} = \frac{E_x}{H_y} = -\frac{E_y}{H_x} = \eta. \tag{3.17}$$

4. The transverse fields can be related to each other as

$$\mathbf{H}_t = \frac{1}{Z_{\text{TEM}}} \hat{z} \times \mathbf{E}_t. \tag{3.18}$$

3.1.3   TE Waves, $E_z = 0$, $H_z \neq 0$

Let's now consider TE waves. We will see that the process for solving for TE waves is somewhat different from solving for TEM waves. However, the approach for TE waves will also apply to TM waves (and other more complicated kinds of waves) so that it can be viewed as a more general process, while solving for TEM waves is somewhat of a special case.

Now, from (3.11) to (3.14) with $E_z = 0$ we can see that if we are able to calculate H_z we can find all the remaining transverse field components from this single scalar component. Hence, it will serve us well to find the equation for H_z. Following an identical process to what gave us the wave equation only involving transverse derivatives in (3.5) for the magnetic field will give us

$$\nabla_t^2 \mathbf{H} + k_c^2 \mathbf{H} = 0, \tag{3.19}$$

where the cutoff wavenumber is still given by $k_c^2 = k^2 - \beta^2$. In Cartesian coordinates, it is quite easy for us to simplify this to only consider the H_z components. This will give us the equation

$$\nabla_t^2 H_z + k_c^2 H_z = (\partial_x^2 + \partial_y^2)H_z + k_c^2 H_z = 0. \tag{3.20}$$

We must solve this equation subject to the boundary conditions of a particular waveguide geometry. For most geometries we can handle by hand, we will be able to use the separation of variables approach to solve this equation.

Despite not being able to solve this equation yet, we can still conclude some useful properties of TE waves. These are summarized briefly here.

1. TE waves can exist in single conductor systems, such as a hollow metal "pipe".

2. $k_c > 0$ for all TE waves, so the waveguide geometries that support TE waves can often be looked at as being like a high pass filter

3. The wave impedance for a TE wave is

$$Z_{\text{TE}} = \frac{E_x}{H_y} = -\frac{E_y}{H_x} = \frac{k\eta}{\beta}. \tag{3.21}$$

   Since $\beta = \sqrt{k^2 - k_c^2}$ we see that Z_{TE} will be a function of frequency.

4. The transverse fields can be related to each other as

$$\mathbf{H}_t = \frac{1}{Z_{\text{TE}}} \hat{z} \times \mathbf{E}_t. \tag{3.22}$$

5. The guided wavelength for a TE wave is given by $\lambda_g = 2\pi/\beta$. Hence, we have that

$$\lambda_g = \frac{2\pi}{\sqrt{k^2 - k_c^2}}. \tag{3.23}$$

   From this, we can see that the guided wavelength will in general not be equal to the "free-space" or "homogeneous medium" wavelength of $\lambda = c/f$. This is important to remember when dealing with non-TEM waves.

3.1.4 TM Waves, $E_z \neq 0$, $H_z = 0$

For TM waves, we will be able to use a very similar process to what we did for TE waves. However, we now have from (3.11) to (3.14) with $H_z = 0$ that we will need to calculate E_z to quickly find the remaining transverse field components from this single scalar component. Hence, it will serve us well to find the equation for E_z. Here, we will have

$$\nabla_t^2 \mathbf{E} + k_c^2 \mathbf{E} = 0, \tag{3.24}$$

where the cutoff wavenumber is still given by $k_c^2 = k^2 - \beta^2$. In Cartesian coordinates, it is quite easy for us to simplify this to only consider the E_z components. This will give us the equation

$$\nabla_t^2 E_z + k_c^2 E_z = (\partial_x^2 + \partial_y^2) E_z + k_c^2 E_z = 0. \tag{3.25}$$

We must solve this equation subject to the boundary conditions of a particular waveguide geometry. For most geometries we can handle by hand, we will be able to use the separation of variables approach to solve this equation.

Despite not being able to solve this equation yet, we can still conclude some useful properties of TE waves. These are summarized briefly here.

1. TM waves can exist in single conductor systems, such as a hollow metal "pipe".

2. $k_c > 0$ for all TM waves, so the waveguide geometries that support TM waves can often be looked at as being like a high pass filter

3. The wave impedance for a TM wave is

$$Z_{\text{TM}} = \frac{E_x}{H_y} = -\frac{E_y}{H_x} = \frac{\beta \eta}{k}. \tag{3.26}$$

   Since $\beta = \sqrt{k^2 - k_c^2}$ we see that Z_{TM} will be a function of frequency. We can also see that $Z_{\text{TM}} \neq Z_{\text{TE}} \neq Z_{\text{TEM}}$, in general.

4. The transverse fields can be related to each other as

$$\mathbf{H}_t = \frac{1}{Z_{\text{TM}}} \hat{z} \times \mathbf{E}_t. \tag{3.27}$$

5. The guided wavelength for a TM wave is given by $\lambda_g = 2\pi/\beta$. Hence, we have that

$$\lambda_g = \frac{2\pi}{\sqrt{k^2 - k_c^2}}. \tag{3.28}$$

   From this, we can see that the guided wavelength will in general not be equal to the "free-space" or "homogeneous medium" wavelength of $\lambda = c/f$. This is important to remember when dealing with non-TEM waves.

3.2 Rectangular Waveguide Modes

We will now consider how to solve for the waves that can exist for the rectangular waveguide geometry shown in Fig. 3.1. From this figure, we see that this waveguide is a single conductor system, and as a result will not be able to support TEM waves. In our solution process, we will find that there are actually many different TE and TM waves that can exist in this geometry. To differentiate between all of the different valid solutions, we will refer to each of them as *modes* of the waveguide. Each mode will have its own unique set of propagation characteristics (e.g., guided wavelength, propagation constant, cutoff wavenumber, wave impedance), which can make the analysis of systems involving these kinds of waveguide geometries much more complicated than simple plane wave situations we considered earlier in the course. Although things are more complex here, we often build practical systems that limit the number of waveguide modes that need to be simultaneously considered so that the conceptual analysis of these systems is still intuitive and possible in practice.

3.2.1 TE Modes

Recall that for TE modes, we have that $E_z = 0$, $H_z \neq 0$. Hence, we need to solve the equation

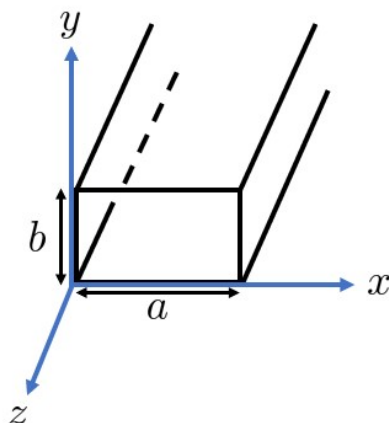$$(\partial_x^2 + \partial_y^2) H_z + k_c^2 H_z = 0, \tag{3.29}$$

Figure 3.1: Geometry of a rectangular waveguide made from a perfect electric conductor.

under the assumption that

$$H_z(x, y, z) = H_z(x, y)e^{-j\beta z}. \tag{3.30}$$

Considering the structure of this partial differential equation, we see that if we follow a separation of variables approach we will be able to simplify the solution of this equation. In particular, we will assume that

$$H_z(x, y, z) = X(x)Y(y)e^{-j\beta z}. \tag{3.31}$$

Substituting this into (3.29) will, after some rearranging, give us

$$\frac{1}{X}\frac{d^2}{dx^2}X + \frac{1}{Y}\frac{d^2}{dy^2}Y + k_c^2 = 0. \tag{3.32}$$

By the standard separation of variables argument, we can conclude that each term on the LHS of this equation will need to individually equal a constant for this equation to be satisfied. Hence, we can define separation constants k_x and k_y to get that

$$\frac{d^2}{dx^2}X + k_x^2 X = 0, \tag{3.33}$$

$$\frac{d^2}{dy^2}Y + k_y^2 Y = 0. \tag{3.34}$$

We will also have the modified dispersion relation from (3.32) that

$$k_c^2 = k_x^2 + k_y^2. \tag{3.35}$$

We can now follow our standard three step process for solving Maxwell's equations. We have already completed *Step 1* in Fig. 3.1. To be more explicit, we can identify Region 1 as being inside the waveguide that is filled with a material characterized by μ and ϵ. The walls

of the waveguide are made from a perfect electric conductor (PEC). We have walls at $x = 0$ and $x = a$ that range continuously from $0 \leq y \leq b$. Similarly, we have another two walls at $y = 0$ and $y = b$ which range continuously over the range $0 \leq x \leq a$. Region 2 would be the rest of space, which can be assumed to be fully filled with PEC material.

Now, for *Step 2*, we need to propose a general solution in each region of our problem. We already know that there will be no fields in Region 2 because it is filled with PEC material, so we can ignore this region for the rest of the problem. In Region 1, we can look at the two separated 1D wave equations in (3.33) and (3.34) and see that our standard set of wave solutions will be applicable here for $X(x)$ and $Y(y)$. However, because we are working with a finite-sized region it will now be more advantageous to use trigonometric functions as our wave solutions as opposed to exponential functions (this makes enforcing the boundary conditions easier algebraically). Hence, we will have as our general wave solution that

$$H_z(x, y) = X(x)Y(y) = \big(X_c \cos(k_x x) + X_s \sin(k_x x)\big)\big(Y_c \cos(k_y y) + Y_s \sin(k_y y)\big). \quad (3.36)$$

We can now begin *Step 3* to solve for the unknown constants based on the boundary conditions of our problem. For a PEC boundary, we would typically use the fact that the tangential component of the electric field will need to vanish as our boundary condition. Since we have our proposed solution in terms of H_z, one solution approach would be to use H_z to compute E_x and E_y and then apply the boundary condition to these field components. This is facilitated by using (3.13) and (3.14) and by noting that for a TE wave $E_z = 0$. This then gives us that

$$E_x = -\frac{j}{k_c^2}\omega\mu\partial_y H_z, \quad (3.37)$$

$$E_y = \frac{j}{k_c^2}\omega\mu\partial_x H_z. \quad (3.38)$$

Alternatively, we can use (3.37) and (3.38) to recast the boundary condition on the tangential components of **E** into an equivalent boundary condition on **H**. For a PEC boundary, this will tell us that the *normal derivative* of H_z at the PEC boundary should be 0 (i.e., at a surface with $\hat{n} = \hat{x}$, we would take $\partial_x H_z = 0$ as our boundary condition). Hence, we can determine that the four boundary conditions we should enforce to solve our current problem are

$$\partial_x H_z = 0 \text{ (i.e., } E_y = 0) \text{ at } x = 0, \quad (3.39)$$

$$\partial_x H_z = 0 \text{ (i.e., } E_y = 0) \text{ at } x = a, \quad (3.40)$$

$$\partial_y H_z = 0 \text{ (i.e., } E_x = 0) \text{ at } y = 0, \quad (3.41)$$

$$\partial_y H_z = 0 \text{ (i.e., } E_x = 0) \text{ at } y = b. \quad (3.42)$$

We can now plug (3.36) into these boundary conditions. Beginning with (3.39), we see that we will get

$$\left[-X_c k_x \sin(k_x\, 0) + X_s k_x \cos(k_x\, 0) \right] Y(y) = X_s k_x Y(y) = 0. \tag{3.43}$$

This tells us that $\boxed{X_s = 0,}$ since we can't in general have $k_x = 0$ or $Y(y) = 0$ as a completely general solution. Similarly, we can use (3.40) to get that

$$-X_c k_x \sin(k_x a) Y(y) = 0 \tag{3.44}$$

after taking into account that $X_s = 0$. We can't simply say that $X_c = 0$ since this would completely remove the $X(x)$ dependence in our separation of variables expression. Hence, we will need to have that

$$\sin(k_x a) = 0, \tag{3.45}$$

from which we can determine that

$$\boxed{k_x = \frac{m\pi}{a}, \; m \in \mathbb{Z}.} \tag{3.46}$$

This tells us that there are an infinite number of values of m for which this equation will be able to be satisfied. This is our first "taste" of why there will be many different possible TE modes that can exist in a particular waveguide geometry.

Continuing, we move on to enforcing our boundary conditions in (3.41) and (3.42). Starting with (3.41), we get that

$$\left[-Y_c k_y \sin(k_y\, 0) + Y_s k_y \cos(k_y\, 0) \right] X(x) = Y_s k_y X(x) = 0. \tag{3.47}$$

By similar logic to what we saw earlier, we can conclude that $\boxed{Y_s = 0.}$ Now, we can use (3.42) to get that

$$\sin(k_y b) X(x) = 0, \tag{3.48}$$

which tells us that

$$\boxed{k_y = \frac{n\pi}{b}, \; n \in \mathbb{Z}.} \tag{3.49}$$

This tells us we can also have an infinite number of values of n for which our equation may be able to be solved.

Putting these results together, we get that

$$\boxed{H_z(x,y) = H_0 \cos\left(\frac{m\pi}{a}x\right) \cos\left(\frac{n\pi}{b}y\right), \; m, n \in \mathbb{Z}.} \tag{3.50}$$

We can compute the remaining field components of H_x, H_y, E_x, and E_y using (3.11) to (3.14) and noting that $E_z = 0$. As a result, by finding (3.50) we have completed our solution of this problem. Before moving on, it is important to consider the integers m and n more closely. For simplicity of notation, we have written in (3.50) that these may be any integer. However, the values of m and n do loosely depend on each other. In particular, we need to make sure that the combination of m and n don't cause H_z to either be equal to 0 or a constant. We can't have these cases because they either invalidate our assumption for this kind of wave that $H_z \neq 0$ or represent a trivial case that is not of interest. Considering this, we see that the only constraint on m and n with respect to each other is that they cannot be simultaneously equal to 0. However, they can be 0 individually and still give us a valid solution.

We can now look at the values of some important propagation characteristics for these TE modes. We will begin with the propagation constant, which becomes

$$\beta = \sqrt{k^2 - k_c^2} = \sqrt{k^2 - (k_x^2 + k_y^2)} = \sqrt{k^2 - \left(\frac{m\pi}{a}\right)^2 - \left(\frac{n\pi}{b}\right)^2}. \tag{3.51}$$

Considering this, we can ask ourselves the question of what will be the first TE mode that can propagate in our particular waveguide (i.e., which mode will have the lowest k_c). To help organize ourselves, we will keep track of modes by labeling them as TE_{mn}. Since $a > b$, we will find that the cutoff frequency for the TE_{10} mode will be the first mode that can propagate in our waveguide. We will typically refer to the first mode that can propagate in a waveguide as being the *dominant mode*. After analyzing the TM mode case next, we will find that the TE_{10} mode is in fact the dominant mode of a rectangular waveguide (i.e., all TM modes have higher cutoff frequencies).

3.2.2 TM Modes

Recall that for TM modes, we have that $H_z = 0$, $E_z \neq 0$. Hence, we need to solve the equation

$$(\partial_x^2 + \partial_y^2)E_z + k_c^2 E_z = 0, \tag{3.52}$$

under the assumption that

$$E_z(x, y, z) = E_z(x, y)e^{-j\beta z}. \tag{3.53}$$

Once we have solved for E_z, we can use this to compute all other non-zero field components in the waveguide by recalling that we had previously derived from Maxwell's equations that

$$H_x = \frac{j}{k_c^2}(\omega\epsilon\partial_y E_z - \beta\partial_x H_z), \tag{3.54}$$

$$H_y = -\frac{j}{k_c^2}(\omega\epsilon\partial_x E_z + \beta\partial_y H_z), \tag{3.55}$$

$$E_x = -\frac{j}{k_c^2}(\beta \partial_x E_z + \omega\mu \partial_y H_z), \tag{3.56}$$

$$E_y = \frac{j}{k_c^2}(-\beta \partial_y E_z + \omega\mu \partial_x H_z). \tag{3.57}$$

Further, by considering the structure of (3.52), we see that if we follow a separation of variables approach similar to what we did for TE modes we will be able to simplify the solution of this equation in the same way. This will give us the same modified dispersion relation

$$k_c^2 = k_x^2 + k_y^2, \tag{3.58}$$

and the different separation equations will also be the same (i.e., 1-D wave equations).

We can now follow our standard three step process for solving Maxwell's equations. We have the same geometry as for TE modes so Step 1 doesn't change. Similarly, we have the same separated equations (although, the boundary conditions do change), so Step 2 doesn't need to really change either. Hence, we will have as our general wave solution that

$$E_z(x, y) = X(x)Y(y) = \left(X_c \cos(k_x x) + X_s \sin(k_x x)\right)\left(Y_c \cos(k_y y) + Y_s \sin(k_y y)\right). \tag{3.59}$$

We can now begin Step 3 to solve for the unknown constants based on the boundary conditions of our problem. For a PEC boundary, we can use the fact that the tangential component of the electric field will need to vanish as our boundary condition. Hence, we have that

$$E_z = 0 \text{ at } x = 0 \text{ and } x = a, \tag{3.60}$$

$$E_z = 0 \text{ at } y = 0 \text{ and } y = b. \tag{3.61}$$

We can now plug (3.59) into these boundary conditions. Beginning with (3.60) at $x = 0$, we see that we will get

$$\left[X_c \cos(k_x \, 0) + X_s \sin(k_x \, 0)\right]Y(y) = 0 \implies \boxed{X_c = 0.} \tag{3.62}$$

Similarly, we can use (3.60) at $x = a$ to get that

$$X_s \sin(k_x a)Y(y) = 0 \tag{3.63}$$

after taking into account that $X_c = 0$. We can't simply say that $X_s = 0$ since this would completely remove the $X(x)$ dependence in our separation of variables expression. Hence, we will need to have that

$$\sin(k_x a) = 0, \tag{3.64}$$

181

from which we can determine that

$$\boxed{k_x = \frac{m\pi}{a}, \ m \in \mathbb{Z}.} \tag{3.65}$$

This tells us that there are an infinite number of values of m for which this equation will be able to be satisfied, just as we had for TE modes.

Continuing, we move on to enforcing our boundary conditions in (3.61). Starting with (3.61) at $y = 0$, we get that

$$\left[Y_c \cos(k_y\, 0) + Y_s \sin(k_y\, 0) \right] X(x) = 0 \implies \boxed{Y_c = 0.} \tag{3.66}$$

Now, we can use (3.61) at $y = b$ to get that

$$Y_s \sin(k_y b) X(x) = 0, \tag{3.67}$$

which tells us that

$$\boxed{k_y = \frac{n\pi}{b}, \ n \in \mathbb{Z}.} \tag{3.68}$$

This tells us we can also have an infinite number of values of n for which our equation may be able to be solved.

Putting these results together, we get that

$$\boxed{E_z(x, y) = E_0 \sin\left(\frac{m\pi}{a}x\right) \sin\left(\frac{n\pi}{b}y\right), \ m, n \in (\mathbb{Z} - \emptyset).} \tag{3.69}$$

All of the transverse field components can be easily computed from this expression for E_z using (3.54) to (3.57) and noting that $H_z = 0$. These transverse fields can be found in your book in (10-134) to (10-137) [14]. Now, we have labeled that m and n can be any integers that aren't zero (this is what we mean by $m, n \in (\mathbb{Z} - \emptyset)$). We can't allow m or n to equal 0 because this would cause the entire $E_z = 0$, which violates the assumption of a TM mode. Hence, the first TM mode that can propagate in a rectangular waveguide is the TM_{11} mode.

We can now look at the values of some important propagation characteristics for these TM modes. We will begin with the propagation constant, which becomes

$$\beta = \sqrt{k^2 - k_c^2} = \sqrt{k^2 - (k_x^2 + k_y^2)} = \sqrt{k^2 - \left(\frac{m\pi}{a}\right)^2 - \left(\frac{n\pi}{b}\right)^2}, \tag{3.70}$$

which matches the result we had for TE modes. However, because our first TM mode is the TM_{11}, we see that the dominant mode of the overall waveguide will be the TE_{10} mode since that will have the lowest cutoff frequency.
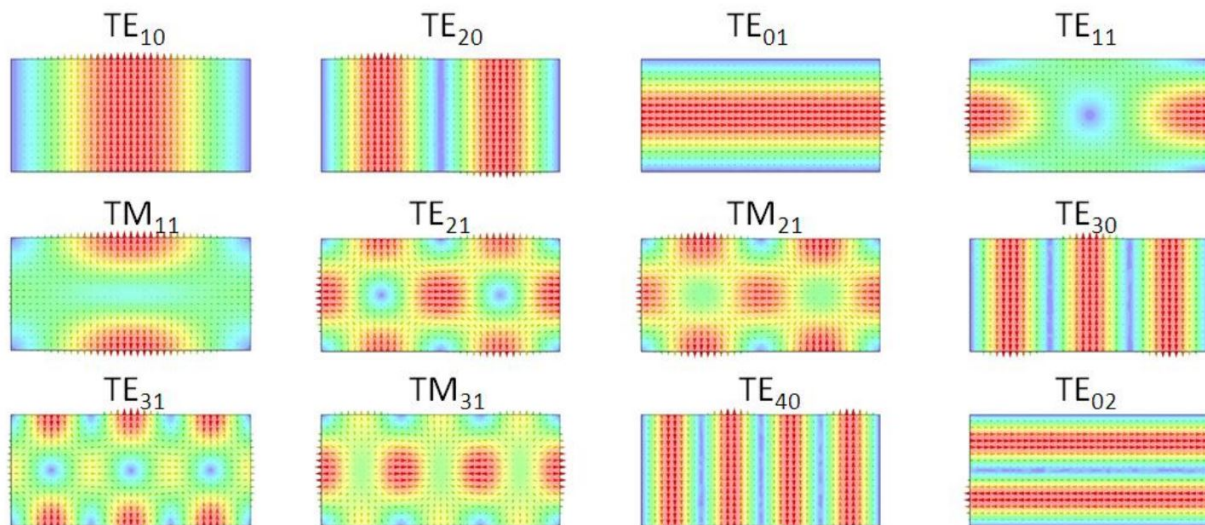
Figure 3.2: Examples of the transverse electric field for low-order modes in a rectangular waveguide (images from [15]).

3.2.3 Examples of Waveguide Modes

To help better understand the field structure of rectangular waveguide modes, it is useful to plot the fields in various cross sections of the waveguide. The transverse electric field for a number of low-order modes are shown in Fig. 3.2. Other cross sections can be found online and in your book. The figures of interest in your book are Figs. 10-11 to 10-13. In general, each mode number represents the number of half-sinusoidal variations that the field solution will take over a particular dimension of the waveguide. For example, the TE_{10} mode means that there will be a single half-sinusoidal variation along the x-dimension of the waveguide and no variation along the y-dimension. Similarly, TE_{21} would have a full sinusoidal variation (i.e., one period) along the x-dimension of the waveguide and a single half-sinusoidal variation along the y-dimension.

3.2.4 Practical Waveguide Operation

Before moving on, it is useful to comment on a few remaining details with regard to how rectangular waveguides are used in practice. The first point of note is that there are many cases that can occur when two different modes will have the same cutoff frequency (e.g., the TE_{11} and TM_{11} modes as a trivial example). These are referred to as *degenerate modes*. Dealing with these can sometimes be inconvenient, so it is typical to design waveguides that minimize the number of degenerate modes. One common way to do this is to make practical waveguides have cross-sectional dimensions of $a \approx 2b$, but avoiding the case of $a = 2b$. However, for certain applications, it is actually *desirable* to have two degenerate orthogonal polarized modes (for instance, feeding certain circularly polarized antennas). In this case, one might use a square waveguide that has many degenerate modes. However, this is typically the exception, rather than the rule.
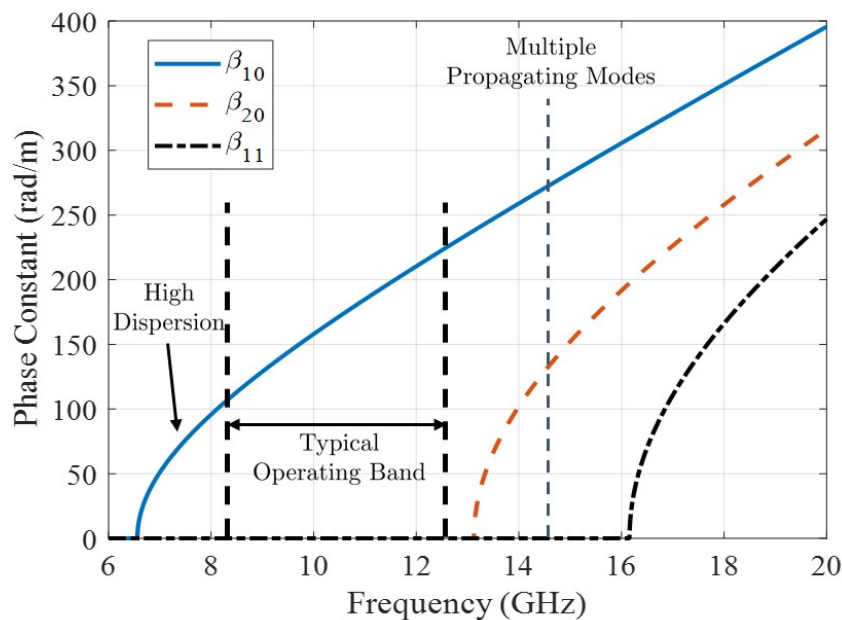
Figure 3.3: Dispersion diagram for a WR-90 rectangular waveguide illustrating the typical operating bandwidth.

Another important point about the use of waveguides is to define their typical operating band. This is illustrated in Fig. 3.3. We see that the beginning of the operating band is generally taken to be about 1.25 times the cutoff frequency of the waveguide. This is done to avoid the highly nonlinear portion of the dispersion curve near the cutoff frequency because this region has significant dispersion that can lead to distortion of practical signals that encompass a band of frequencies. The upper limit for the operating band is usually set to be slightly lower than the cutoff frequency of the second propagating mode in the waveguide. This is done to ensure that only a single mode can propagate in the waveguide, which makes the analysis and operation of a waveguide much smoother in practical situations.

As an example of a standard waveguide, we can consider the WR-90 waveguide that is designed to correspond to the X-band of frequencies (8.2 – 12.4 GHz). The "WR" stands for "waveguide rectangular" and the "90" denotes that the "a" dimension of the waveguide is 0.9" wide. It is a straightforward exercise to compute that the cutoff frequency of the TE_{10} mode is approximately 6.557 GHz for this waveguide. We can see that $8.2/6.557 = 1.25$, as mentioned earlier.

3.3 Introduction to Resonators

Microwave resonators are a valuable tool in many practical applications. For instance, they can form an essential part of many kinds of filters, oscillators, and amplifiers. At many microwave frequencies, it is not achievable to form resonators using lumped element components. As a result, alternative approaches using distributed resonators (e.g., formed by some configuration of transmission line structure) are often needed.
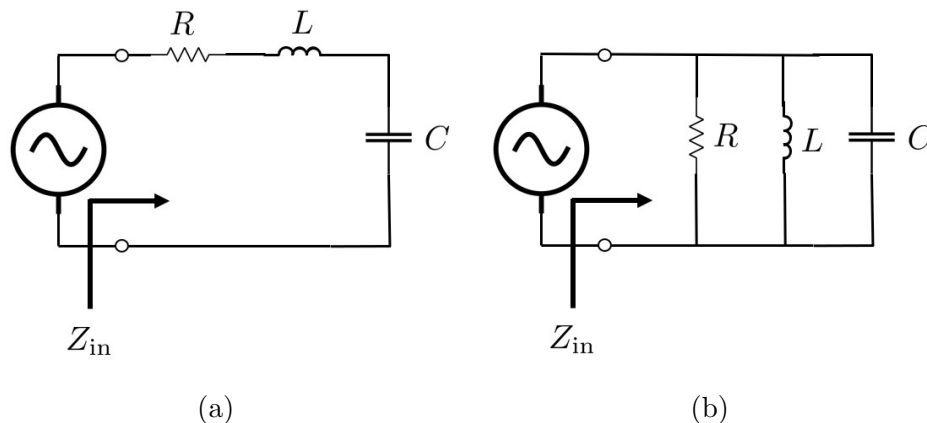
Figure 3.4: Examples of lumped element resonant circuits. (a) Series case and (b) parallel case.

3.3.1 Lumped Element Resonators: A Short Review

Before discussing distributed resonators, it will be useful to recall some basic properties of lumped element resonators. The simplest examples of resonant circuits are series and parallel RLC networks. These each support a single resonance frequency, which occurs when the reactance or susceptance of the inductor and capacitor completely cancel each other out. If we consider the series resonant circuit in Fig. 3.4(a), we can see that our resonance condition will require

$$j\omega_0 L + \frac{1}{j\omega_0 C} = 0. \tag{3.71}$$

We can solve this to find that our resonance frequency will be given by

$$\omega_0 = \frac{1}{\sqrt{LC}}. \tag{3.72}$$

If we repeat these steps for the parallel resonant circuit we arrive at an identical result. For both cases, we typically define the bandwidth of a resonant circuit to occur at the "half-power" point, or when the input impedance has changed by a factor of $\sqrt{2}$ from its value at resonance.

Another important quantity that is used to characterize resonators is its Q (also called *quality factor*). This is defined as

$$Q = \omega \frac{\text{average energy stored}}{\text{energy loss / second}} = \omega \frac{W_m + W_e}{P_{\text{loss}}}. \tag{3.73}$$

This helps us understand how long a resonator will continue to resonate after the excitation has been removed. In general, the total Q of a resonator will depend on what other circuitry is connected to it (e.g., a source or a load). When we only consider the "intrinsic" Q of a resonator, we refer to it as the *unloaded Q*. If we take into account the loading due to other

circuits, we refer to the new Q as the *loaded Q*. For the series resonance case we have that the unloaded Q, denoted as Q_0, is

$$Q_0 = \frac{1}{R}\sqrt{\frac{L}{C}} = \frac{\omega_0 L}{R} = \frac{1}{\omega_0 RC}. \tag{3.74}$$

Similarly, the unloaded Q of the parallel resonator is

$$Q_0 = R\sqrt{\frac{C}{L}} = \frac{R}{\omega_0 L} = \omega_0 RC. \tag{3.75}$$

More generally, the total Q of a system can be found as

$$\frac{1}{Q} = \frac{1}{Q_0} + \sum_i \frac{1}{Q_i} \tag{3.76}$$

where the Q_i characterize the quality factors of additional components, such as extra load components. For simple resonators, we can relate the total Q to the bandwidth of the resonator as BW $= 1/Q$.

Although these various definitions are usually introduced in the context of analyzing simple lumped element circuits, we use the same terminology to analyze more complicated microwave resonators. Of course, the evaluation of the different quantities involved in the definitions will become more complex for this case. However, much of the intuition about simpler circuits can be useful in understanding the more complicated cases of microwave resonators. Depending on how the input impedance of a microwave resonator varies as a function of frequency, certain distributed resonators will also be referred to as supporting "series" or "parallel" resonances in analogy to the lumped element case. We won't use this terminology here, but you may encounter it in your future career.

3.3.2 Transmission Line Resonators

To achieve a resonance with a transmission line, we typically must terminate both ends of the transmission line with a particular boundary condition (if part of the transmission line were to extend to infinity we would just end up with the cases we have considered previously that weren't resonators). We can typically think of our resonant solutions as being a standing wave solution that "fits" into our particular geometry perfectly. To see this, let's consider a simple example of a lossless transmission line of length ℓ that is terminated with a short circuit at both ends (shown in Fig. 3.5).

Now, we want to determine the solutions to the wave equation that we got from the telegrapher's equations; i.e.,

$$\frac{d^2}{dz^2}V(z) + \beta^2 V(z) = 0, \tag{3.77}$$

where we had that $\beta = \omega\sqrt{LC}$. Since our geometry has a finite length, we know that it will often be quickest to determine a valid solution if we write our general solution in terms of sine and cosine as opposed to exponential functions (which are traveling wave solutions).
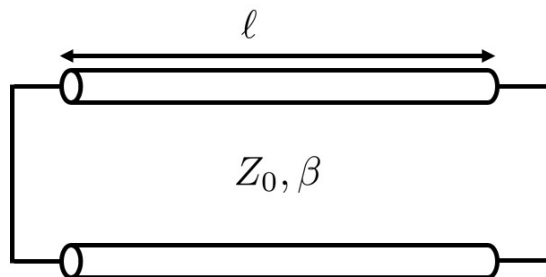
Figure 3.5: Example of a short-circuited transmission line resonator. The boundary conditions on the line will only be able to be satisfied for particular values of ℓ considering the fixed transmission line parameters Z_0 and β.

Considering this, our proposed general solution for the voltage on the transmission line can be

$$V(z) = V_s \sin(\beta z) + V_c \cos(\beta z). \tag{3.78}$$

We now need to enforce our boundary conditions at $z = 0$ and $z = \ell$. Since we have a short circuit at both of these locations, we will need $V(z)$ to vanish at these positions on the line. From this, we can readily find out that $V_c = 0$ and that

$$\beta = \frac{m\pi}{\ell}, \quad m \in \mathbb{Z}. \tag{3.79}$$

However, as mentioned previously, we also know that $\beta = \omega\sqrt{LC} = 2\pi/\lambda_g$. Considering this, we can rearrange our equation to see that we will only have a valid solution when

$$\ell = m\frac{\lambda_g}{2}, \quad m \in \mathbb{Z}. \tag{3.80}$$

That is, our transmission line will only be able to support a valid solution when the length of the line is exactly a multiple of half a guided wavelength. This exercise can be repeated for a transmission line that is terminated in open-circuits at both ends as well, and the same result will be achieved. The unique aspects of how these resonances differ can be tied back to notions of "series" and "parallel" resonances, but we will not delve into this in this course.

Consolidating these results, we have that valid solutions of the voltage will be

$$V_m(z) = V_{0,m} \sin(\beta_m z), \tag{3.81}$$

with $\beta_m = m\pi/\ell$. If we wish to find the current that goes with these voltage solutions, we can use the telegrapher's equations. In particular, we can recall that

$$\partial_z V = -j\omega L I \tag{3.82}$$

to find that

$$I_m(z) = -\frac{V_{0,m}}{j\omega L}\beta_m \cos(\beta_m z). \tag{3.83}$$

Recalling that β_m will still need to equal $\omega\sqrt{LC}$ at the particular resonant frequency, we can simplify this expression into terms of the characteristic impedance $Z_0 = \sqrt{L/C}$. In particular, we get that

$$I_m(z) = -\frac{V_{0,m}}{jZ_0}\cos(\beta_m z). \tag{3.84}$$

Now, it should be understood that finding the resonant solutions of a system in this way has a few deficiencies to it. First, what we have seen is that the resonance only exists at a discrete set of frequencies when the transmission line is a multiple of half a guided wavelength. In reality, the loss of a line will cause the resonance to have a finite bandwidth. Further, we haven't accounted for any loading effects that may exist due to nearby circuitry. These kinds of effects can usually be accounted for in different ways as perturbations to the results we have just found. Overall, the idealization of a lossless system is often mathematically convenient for quickly finding general ideas about resonators, such as how long they should be to achieve resonance at a particular frequency and that they will have equally-spaced resonance frequencies. However, a more careful consideration is needed to successfully design a properly functioning resonator for a particular application, and is something that can be learned about in more advanced microwave engineering courses.

3.3.3 Rectangular Waveguide Resonators

We won't consider these in depth in lecture, but you will solve some problems related to these in your homework. However, it should hopefully be clear to you that if we take a rectangular waveguide of finite length and terminate both ends of it with a perfectly conducting wall (to end up with a hollow metal box) we will find a similar result to what we just found for a transmission line resonator. In this case, we can still support the same set of transverse modes, but the finite size along the longitudinal direction now forces us to find standing wave solutions along this axis as well. The end result is that we can only support a non-trivial solution at discrete frequencies. We typically number the resulting modes as TE_{mnp} and TM_{mnp} modes, where the index p is associated with the longitudinal dimension of the waveguide. The first mode that exists will typically be the TE_{101} mode, since the b dimension of the waveguide cavity is generally taken to be the smallest dimension by convention.

3.4 Practice Problems

1. A general guided wave solution can be written as

$$\mathbf{E} = \big[\mathbf{E}_{\text{T}}(x,y) + \hat{z}E_z(x,y)\big]e^{-\gamma z}, \tag{3.85}$$

$$\mathbf{H} = \big[\mathbf{H}_{\text{T}}(x,y) + \hat{z}H_z(x,y)\big]e^{-\gamma z}, \tag{3.86}$$

where a subscript T denotes the quantities involved are purely transverse to the direction of propagation. Hence, $\mathbf{E}_T$ is a shorthand for the E_x and E_y components of the

electric field. Now, by substituting these expressions in Faraday and Ampere's laws, show that

$$\mathbf{E}_T = -\frac{1}{k_c^2}\left(\gamma\nabla_T E_z - \hat{z}j\omega\mu \times \nabla_t H_z\right) \tag{3.87}$$

$$\mathbf{H}_T = -\frac{1}{k_c^2}\left(\gamma\nabla_T H_z + \hat{z}j\omega\epsilon \times \nabla_t E_z\right). \tag{3.88}$$

To arrive at these expressions, you will need to solve for E_x in terms of H_z and E_z, and similar for the other transverse components E_y, H_x, and H_y. Then, combine the equations for E_x and E_y into the vector expressions given above.

2. Calculate and list in ascending order the cutoff frequencies of a rectangular waveguide with width a and height b for the following modes: TE_{01}, TE_{10}, TE_{11}, TE_{02}, TE_{20}, TM_{11}, TM_{12}, and TM_{22}. Do this for

   (a) if $a = 2b$, and

   (b) if $a = b$.

   Compute the cutoff frequencies relative to that of the dominant mode of the waveguide.

3. An air-filled rectangular waveguide has $a = 6.5$ cm and $b = 3.5$ cm. If the operating frequency is set to 3 GHz, calculate β, v_p, λ_g, and the wave impedance of the dominant mode.

   For the following questions, assume that we have an air-filled rectangular waveguide whose cross section is defined by $0 < x < 22.86$ mm and $0 < y < 10.16$ mm.

   (a) Compute the time-average Poynting vector for the TM_{11} mode of this waveguide at 16.5 GHz.

   (b) Compare the wavelengths of the TE_{10} and TM_{11} modes to that of a plane wave propagating in free space at 16.5 GHz.

4. A lossless transmission line of length 2.145 mm is terminated with a short circuit at $z = 0$ mm and an open circuit at $z = 2.145$ mm. If the transmission line has per-unit-length parameters of $L = 0.7125\,\mu H/m$ and $C = 285\,pF/m$, answer the following questions.

   (a) Using the 3-step process to solving partial differential equations, determine the expressions for the voltage and current solutions that can exist for this geometry.

   (b) At what frequency will the first resonance occur?

5. Write a computer program to compute the input admittance at the position marked in the circuit shown in Fig. 3.6 for frequencies ranging from 2 to 16 GHz. Bold lines with lengths marked denote transmission line segments, while thin lines with no lengths given are regular circuit connections. To perform the calculations, you may assume that

β can be computed using formulas for a homogeneous dielectric medium with relative permittivity $\epsilon_r = 6.325$. Plot the real part of the input admittance to recreate Fig. 3.7. Further, provide a simple explanation for why the bandwidth of each resonance increases as the frequency goes up.
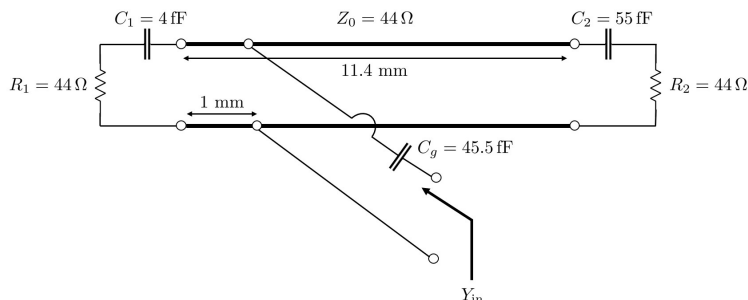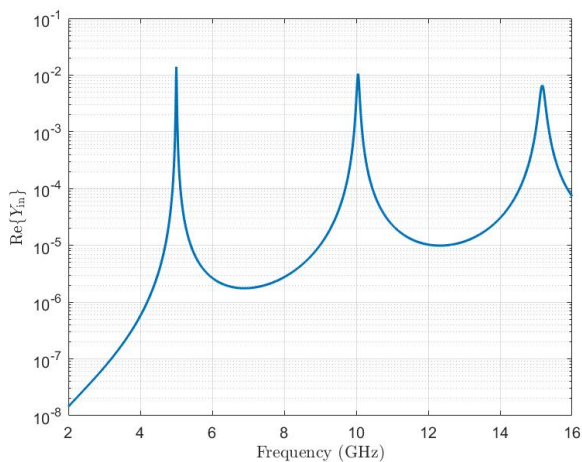

Figure 3.6

Figure 3.7

6. Using the 3-step process to solving Maxwell's equations and the separation of variables approach, derive expressions for all relevant components of the electric and magnetic fields for a general TM_{mnp} mode in a rectangular cavity resonator. Assume all walls of the resonator are made from perfect electric conductors and that the interior region of the cavity is specified by $0 < x < a$, $0 < y < b$, and $0 < z < d$. Given the field expressions, what is the lowest order TM_{mnp} mode that can be supported in this geometry?

Hint: Start with a general solution for E_z that is a generalization of the one we used for solving for the TM modes of a waveguide. The application of the boundary conditions at $x = 0$, $x = a$, $y = 0$, and $y = b$ will be identical to what we did in the lecture notes previously. At $z = 0$ and $z = d$, you will need to apply a boundary condition on either E_x or E_y. So, you will need to determine a process to compute E_x and E_y

from E_z. Later, you will need to also compute H_x and H_y from E_z for this current scenario. Think about how you can use the results from Problem 1 to quickly take care of this rather than re-using Maxwell's equations to relate E_z to all the transverse field components (it has to do with "replacing" $-\gamma$ in the expressions with something else).

Chapter 4

Antennas

4.1 Introduction to Antenna Concepts

Earlier in this class we spent a significant amount of time learning about the properties of plane waves: how they propagate in an unbounded medium, what happens when they encounter certain kinds of discontinuities, and in a crude sense one way that they could be generated (i.e., from an infinite current sheet). We have also spent time learning about transmission lines and how to engineer systems to try and maximize the amount of power to a "load", which we mentioned in passing could often be an antenna for many applications. We are now at the point where we can begin to put these concepts together. We will see how antennas are a practical method for generating plane waves, and will also consider how we can represent an antenna from the perspective of a transmission line network problem.

Antenna engineering is a broad topic that continues to garner a significant amount of research interest. Part of its longevity has come from the need to develop more and more sophisticated kinds of antennas to support new wireless communication and other wireless transmission applications. The many kinds of applications that exist have lead to a widely divergent set of requirements for antennas to meet. As a result, many different kinds of antennas have been invented over the year to provide unique sets of properties that can be valuable for different applications. Despite the many differences, the basic purpose of an antenna remains the same. That is, to take a "tightly bound" wave contained on some kind of transmission line structure and find an efficient way to "separate" the fields from the physical system so that they can begin to propagate away from the antenna. The basic process of this for a small dipole antenna is shown in Fig. 4.1. In general, there are many different kinds of structures that have been found that can efficiently produce radiated fields with different properties. Some common types of antennas are shown in Figs. 4.2 to 4.6.
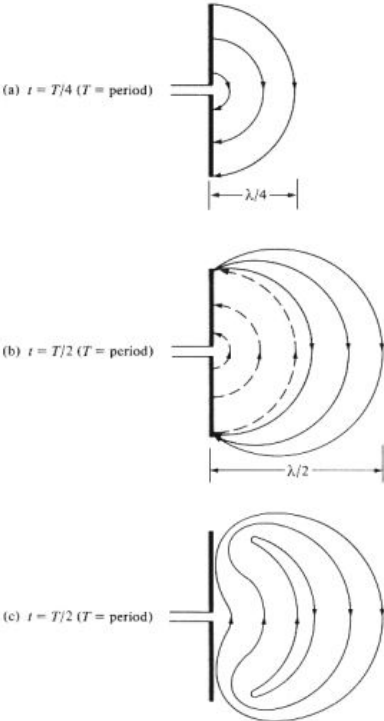
Figure 4.1: Example of electric fields detaching from a small dipole antenna and beginning to radiate away from the antenna (images from [16]).

Figure 4.2: Common examples of wire antennas (images from [16]).

194

(a) Pyramidal horn

(b) Conical horn

(c) Rectangular waveguide

Figure 4.3: Common examples of aperture antennas (images from [16]).
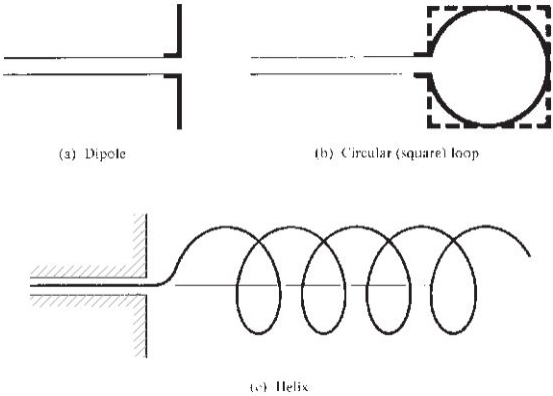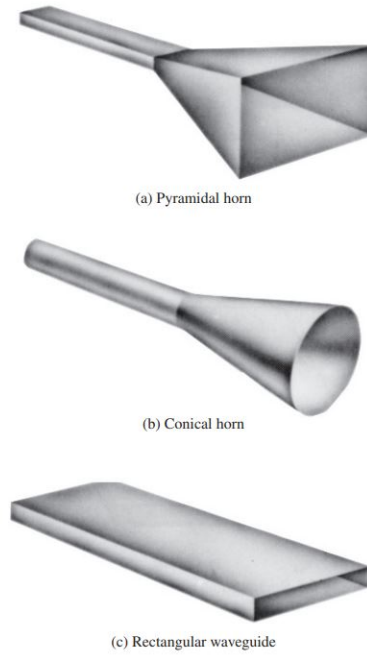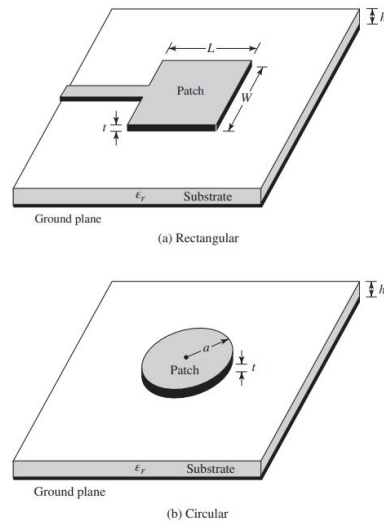

(a) Rectangular

(b) Circular

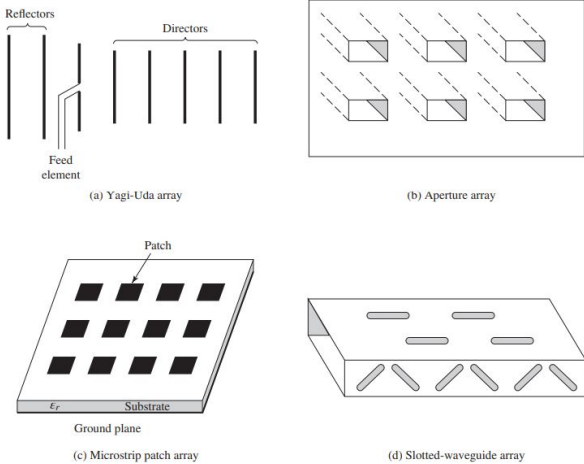Figure 4.4: Common examples of patch antennas (images from [16]).

195

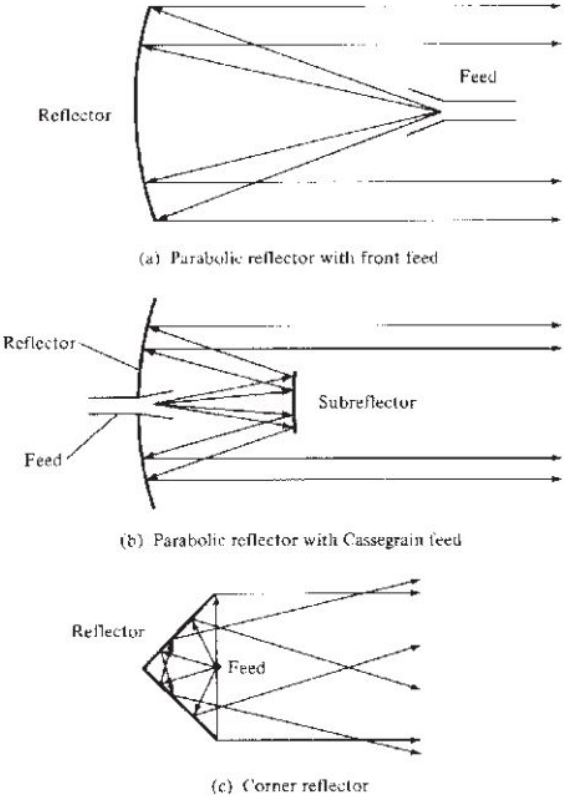Figure 4.5: Common examples of array antennas (images from [16]).

Figure 4.6: Common examples of reflector antennas (images from [16]).

4.2 Electromagnetic Potentials

In many cases, it can be difficult to attempt to directly solve for $\mathbf{E}$, $\mathbf{H}$, $\mathbf{D}$, and $\mathbf{B}$. In these situations, it can be advantageous to define a set of electromagnetic potentials to help us go about solving a problem. This is especially true in antenna problems. You should already be familiar with this in solving electrostatic and magnetostatic problems. We will now consider what this looks like in a full electromagnetic picture.

To begin, we start with Maxwell's equations, which are

$$\nabla \times \mathbf{H} = \partial_t \mathbf{D} + \mathbf{J}, \tag{4.1}$$

$$\nabla \times \mathbf{E} = -\partial_t \mathbf{B}, \tag{4.2}$$

$$\nabla \cdot \mathbf{D} = \rho, \tag{4.3}$$

$$\nabla \cdot \mathbf{B} = 0. \tag{4.4}$$

We note that from Gauss' law of magnetism (4.4) we can safely say that $\mathbf{B}$ can be given by the curl of some vector field (since the divergence of a curl is always 0). Hence, we will introduce a *magnetic vector potential*, denoted as $\mathbf{A}$, that is related to $\mathbf{B}$ by

$$\boxed{\mathbf{B} = \nabla \times \mathbf{A}.} \tag{4.5}$$

We can move forward with this magnetic vector potential by using it to rewrite Faraday's law. This gives us

$$\begin{aligned}\nabla \times \mathbf{E} &= -\partial_t \mathbf{B} \\ &= -\partial_t (\nabla \times \mathbf{A}).\end{aligned} \tag{4.6}$$

We can consolidate terms on the left-hand side to get

$$\nabla \times \left(\mathbf{E} + \partial_t \mathbf{A}\right) = 0 \tag{4.7}$$

Recalling that the curl of a gradient is always equal to 0, we see that we can set $\mathbf{E} + \partial_t \mathbf{A}$ equal to a gradient vector for this equation to always be satisfied. That is

$$\mathbf{E} + \partial_t \mathbf{A} = -\nabla \Phi, \tag{4.8}$$

where Φ is the *electric scalar potential*. We can consolidate our potentials together to see that

$$\boxed{\mathbf{E} = -\partial_t \mathbf{A} - \nabla \Phi.} \tag{4.9}$$

Next, lets see what happens when we use these potentials in Ampere's law. In a homogeneous medium, Ampere's law is

$$\nabla \times \mu^{-1} \mathbf{B} = \partial_t \epsilon \mathbf{E} + \mathbf{J}. \tag{4.10}$$

197

We can substitute in our potentials using (4.38) and (4.40) to get

$$\nabla \times \mu^{-1} \nabla \times \mathbf{A} = -\epsilon \left(\partial_t^2 \mathbf{A} + \partial_t \nabla \Phi \right) + \mathbf{J}. \tag{4.11}$$

We can rearrange terms to get

$$\nabla \times \nabla \times \mathbf{A} + \mu\epsilon \partial_t^2 \mathbf{A} = -\mu\epsilon \partial_t \nabla \Phi + \mu \mathbf{J}. \tag{4.12}$$

We see that the left-hand side looks like one of our familiar wave equations with a propagation speed matching the speed of light.

What happens if we apply our vector identity to rewrite the $\nabla \times \nabla \times$ operator? This will give us

$$\nabla^2 \mathbf{A} - \mu\epsilon \partial_t^2 \mathbf{A} = \nabla \left(\nabla \cdot \mathbf{A} \right) + \mu\epsilon \partial_t \nabla \Phi - \mu \mathbf{J}. \tag{4.13}$$

Have we accomplished anything? Not yet, but if you recall the Helmholtz decomposition theorem, we realize that we have some freedom in specifying what $\nabla \cdot \mathbf{A}$ should be equal to. This freedom exists because we have introduced these auxiliary potential functions to help us solve our problem, they don't already come fully specified like our fields and fluxes did. It is this freedom that we can exploit to help us simplify the solution of certain problems.

To move forward, we need to specify what we are going to force $\nabla \cdot \mathbf{A}$ to equal. This is called setting a *gauge condition*. Although electromagnetic theory was one of the first areas where gauge conditions were widely used, it has become a very prevalent concept in many areas of modern physics to describe different kinds of forces. For instance, gauge conditions are a very prevalent and important part of the Standard Model of particle physics, which in addition to electromagnetism includes theories for the weak and strong nuclear forces. These additional forces obey a set of equations that can be viewed as a generalization of Maxwell's equations (albeit, this takes some fairly sophisticated mathematics to see).

For our current purposes, let's use our gauge condition to try and simplify our wave equation. We can do this by making the two terms involving $\mathbf{A}$ and Φ cancel on the right-hand side of (4.13). That is, by setting

$$\nabla \cdot \mathbf{A} = -\mu\epsilon \partial_t \Phi. \tag{4.14}$$

This gauge condition is used frequently enough that it has its own name. It is the *Lorenz gauge condition*. Note that many people erroneously will call this the Lorentz gauge condition (unfortunately, your textbook does this). This is incorrect and should not be done. Both Lorenz and Lorentz made important contributions to electromagnetic theory, and they should be appropriately commended for their contributions.

Now, after setting this gauge condition, our vector potential wave equation in the Lorenz gauge becomes

$$\boxed{\nabla^2 \mathbf{A} - \mu\epsilon \partial_t^2 \mathbf{A} = -\mu \mathbf{J}.} \tag{4.15}$$

What does the equation for Φ look like? We can find this out by substituting our electromagnetic potentials into Gauss' law of electricity. This takes us from having

$$\nabla \cdot \mathbf{E} = \rho/\epsilon, \tag{4.16}$$

to having

$$\nabla \cdot \left(\nabla \Phi + \partial_t \mathbf{A} \right) = -\rho/\epsilon. \tag{4.17}$$

We can use the Lorenz gauge condition to rewrite this as

$$\boxed{\nabla^2 \Phi - \mu\epsilon\partial_t^2 \Phi = -\rho/\epsilon.} \tag{4.18}$$

We see that we get a *scalar Helmholtz equation* for Φ. This is one of the advantages of the Lorenz gauge. It allows us to arrive at two decoupled wave equations for our potentials that immediately show that they both propagate at the speed of light. This is part of the reason why this gauge condition is frequently utilized in the study of special relativity (where the fact that electromagnetic waves propagate at the speed of light is of paramount importance).

Another gauge condition that is frequently used in certain areas of physics is the *Coulomb gauge*. This is a gauge characterized by

$$\nabla \cdot \mathbf{A} = 0. \tag{4.19}$$

We can sometimes view this gauge as being an approximation to the Lorenz gauge where the time variation is so small relative to the spatial variation that we can treat the $\partial_t \Phi$ term as being equal to 0. This kind of approximation is frequently used in studying the interaction of an electromagnetic field with an atomic system in optics and quantum optics. It should be noted, however, that it is possible to use this gauge without making any approximations.

Although we have freedom in how we specify our potentials, it is important to note that regardless of our gauge condition $\mathbf{A}$ and Φ will always work together to produce the correct fields and fluxes to solve a particular problem. Hence, we see that the potentials are in some sense arbitrary. Due to this, it has long been said that the potentials are not *physical quantitites*, but that they are simply a useful *mathematical tool*. However, the advent of quantum theory has shown us that the electromagnetic potentials are very much physical quantities that can impact the dynamics of other systems. The *Aharonov-Bohm effect* is an excellent example of this. This measured effect shows that the trajectory of an electron is affected by the potentials even when the electric and magnetic fields are identically zero in the region the electron travels through.

4.3   Potentials Produced by Known Source Distributions

One of the fundamental tasks in antenna analysis and design is to determine the fields that are produced from a particular "source configuration", i.e., some set of materials and currents that exist on the antenna. We could in principle include the current distributions in our derivation of a wave equation for $\mathbf{E}$ and $\mathbf{H}$. However, this often leads to us needing to compute the derivatives of the source distributions, which is often mathematically complicated by the ways that we typically represent source distributions (e.g., as infinitely thin sheets). In this situation, it can often be more efficient to first compute the electromagnetic

potentials from the source distribution and then compute the fields from the potentials. Although this may seem "roundabout", it is in practice much easier than trying to directly compute the fields.

We will now investigate how the electromagnetic potentials can be computed from a known distribution of a source function, i.e., from $\mathbf{J}$ and ρ. To build up to this, we will first recall some basic principles from electrostatics. We know that Coulomb's law told us that the electric field produced by a point charge was given by

$$\mathbf{E}(\mathbf{r}) = \hat{r}\frac{q}{4\pi\epsilon|\mathbf{r}|^2}. \tag{4.20}$$

Recognizing that $\mathbf{E} = -\nabla\Phi$, we can conclude that the potential would be

$$\Phi(\mathbf{r}) = \frac{q}{4\pi\epsilon|\mathbf{r}|}. \tag{4.21}$$

To be more general in our description, it is best that we allow our point charge to be located at a position other than 0. This generalization is possible by writing the electric field and potential as

$$\mathbf{E}(\mathbf{r}) = \frac{q(\mathbf{r}')}{4\pi\epsilon|\mathbf{r} - \mathbf{r}'|^2}\left[\frac{\mathbf{r} - \mathbf{r}'}{|\mathbf{r} - \mathbf{r}'|}\right], \tag{4.22}$$

$$\Phi(\mathbf{r}) = \frac{q(\mathbf{r}')}{4\pi\epsilon|\mathbf{r} - \mathbf{r}'|}, \tag{4.23}$$

where $\mathbf{r}$ and $\mathbf{r}'$ are position vectors pointing to the observation and source points, respectively.

If instead of having a single point charge we had a distribution, you will recall that our formula for the potential generalizes to

$$\Phi(\mathbf{r}) = \iiint \frac{1}{4\pi|\mathbf{r} - \mathbf{r}'|}\rho(\mathbf{r}')/\epsilon \, dV', \tag{4.24}$$

where the prime on dV' denotes that we are integrating with respect to the primed variables. We can interpret this formula as being a superposition of the potentials produced by a collection of point charges. A common terminology is to refer to the

$$\frac{1}{4\pi|\mathbf{r} - \mathbf{r}'|} \tag{4.25}$$

portion of this integral as an *integration kernel*. The reason for this is because this is the *point source response* that we are going to integrate against our distribution to determine the potential at any other location. In the study of physical systems, this point source response is often referred to as a *Green's function*.

We can view our Green's function as helping us to "invert" differential equations by recalling that Poisson's equation is

$$\nabla^2\Phi = -\rho/\epsilon. \tag{4.26}$$

The "point source response" or Green's function for this equation would be defined as

$$\nabla^2 g(\mathbf{r}, \mathbf{r}') = -\delta(\mathbf{r} - \mathbf{r}'), \tag{4.27}$$

where $\delta(\mathbf{r} - \mathbf{r}')$ is the Dirac delta function. This PDE has as solution the Green's function

$$g(\mathbf{r}, \mathbf{r}') = \frac{1}{4\pi|\mathbf{r} - \mathbf{r}'|}. \tag{4.28}$$

We can use this Green's function to "invert" Poisson's equation by multiplying it by $g(\mathbf{r}, \mathbf{r}')$ on both sides and integrating over the volume of the problem. The result is

$$\iiint \frac{1}{4\pi|\mathbf{r} - \mathbf{r}'|} \nabla'^2 \Phi(\mathbf{r}') dV' = -\iiint \frac{1}{4\pi|\mathbf{r} - \mathbf{r}'|} \rho(\mathbf{r}')/\epsilon \, dV'. \tag{4.29}$$

We can simplify our result by transferring the derivatives from Φ in (4.29) onto the Green's function using various kinds of integration by parts. You may recall that these forms of "integration by parts" are typically referred to as Green's identities from your other courses. Once we have transferred the derivatives appropriately, we can replace the operations on the Green's function using (4.27) to simply be $-\delta(\mathbf{r} - \mathbf{r}')$. At that point, evaluating the remainder of the integration is trivial, and gives Coulomb's law as

$$\Phi(\mathbf{r}) = \iiint \frac{1}{4\pi|\mathbf{r} - \mathbf{r}'|} \rho(\mathbf{r}')/\epsilon \, dV'. \tag{4.30}$$

Hence, we can see that in some sense, Coulomb's law is just the inverse of Poisson's equation. This process will generally work for PDEs; however, it is often very difficult to determine the Green's function of a problem. Regardless, there are many situations in antenna theory where a known and simple Green's function can be used. We will consider this particular Green's function now.

Let's now turn our attention back to the electromagnetic case. For this situation, we want to solve the equation

$$\nabla^2 \Phi - \mu\epsilon \partial_t^2 \Phi = -\rho/\epsilon. \tag{4.31}$$

We can first transform to the frequency domain to simplify things. This gives us

$$\nabla^2 \Phi + k^2 \Phi = -\rho/\epsilon. \tag{4.32}$$

We will want to "invert" this equation using a Green's function that is the solution of

$$\nabla^2 g(\mathbf{r}, \mathbf{r}') + k^2 g(\mathbf{r}, \mathbf{r}') = -\delta(\mathbf{r} - \mathbf{r}'). \tag{4.33}$$

Comparing this to (4.27), we see that the equations are very similar. Hence, we expect that the solution may also be able to be a similar function. It turns out that the solution is

$$g(\mathbf{r}, \mathbf{r}') = \frac{1}{4\pi|\mathbf{r} - \mathbf{r}'|} e^{-jk|\mathbf{r} - \mathbf{r}'|}. \tag{4.34}$$

This looks very much like our plane wave solution multiplied by our Green's function for static problems. Hence, we have a phase changing as we spherically move away from our point source augmented by a reduction in magnitude. This particular Green's function is typically referred to as the *free-space Green's function* (if $\epsilon = \epsilon_0$ and $\mu = \mu_0$), or, more generally, as the *homogeneous-medium Green's function* (if one of ϵ or μ does not equal the corresponding free space value).

We can use it in a similar fashion to find that the potential produced by a charge distribution is

$$\Phi(\mathbf{r}) = \iiint \frac{1}{4\pi|\mathbf{r} - \mathbf{r}'|} e^{-jk|\mathbf{r}-\mathbf{r}'|} \rho(\mathbf{r}')/\epsilon \, dV'. \tag{4.35}$$

Considering that our vector potential wave equation was

$$\nabla^2 \mathbf{A} - \mu\epsilon\partial_t^2 \mathbf{A} = -\mu\mathbf{J}, \tag{4.36}$$

we see that the same Green's function can be useful here as well. In particular, we have that

$$\mathbf{A}(\mathbf{r}) = \iiint \frac{1}{4\pi|\mathbf{r} - \mathbf{r}'|} e^{-jk|\mathbf{r}-\mathbf{r}'|} \mu\mathbf{J}(\mathbf{r}')dV'. \tag{4.37}$$

Hence, if we know our distributions of $\mathbf{J}$ and ρ in a homogeneous medium it becomes a somewhat "simple" exercise to compute the $\mathbf{A}$ and Φ that are produced (granted, this may need the help of numerical integration routines if the distributions of the sources are complicated).

Once we know the potentials we can compute the fields from them. For instance, we know from the definition of the magnetic vector potential that

$$\mathbf{H} = \mu^{-1}\nabla \times \mathbf{A}. \tag{4.38}$$

We can then use Ampere's law to find the electric field at any point in space away from the source distribution. Alternatively, we can use the potentials to compute the electric field directly. In general, we have that

$$\mathbf{E} = -j\omega\mathbf{A} - \nabla\Phi. \tag{4.39}$$

Often we only want to work with a single source distribution, so we will use our gauge condition to write Φ in terms of $\mathbf{A}$. For antenna problems, the Lorenz gauge is typically used so that one will often see the electric field written in terms of $\mathbf{A}$ as

$$\mathbf{E} = -j\omega\mathbf{A} + \frac{1}{j\omega\mu\epsilon}\nabla(\nabla \cdot \mathbf{A}). \tag{4.40}$$

Next, we will be able to begin computing the fields produced by an *infinitesimal dipole* using these concepts.

4.4 Fields of an Infinitesimal Dipole

Assume that we have an infinitesimal linear wire that is positioned symmetrically at the origin of the coordinate system and is oriented along the z-axis. Due to the extremely small dimension of the wire, we assume that we can represent the current along the wire as a linear current density given by

$$\mathbf{J}(\mathbf{r}') = \begin{cases} \hat{z}I_0, & |z| \leq \ell/2, \\ 0, & \text{elsewhere,} \end{cases} \tag{4.41}$$

where I_0 is the amplitude of the current and ℓ is the length of the infinitesimal current density. In reality, it is not possible for the current density to have a perfectly uniform density along the length of a short wire (it must go to zero at the endpoints). However, this is a reasonable approximation for extremely small wires (compared to the wavelength of radiation) and greatly simplifies the analysis, making it a useful starting point for analyzing antennas. Overall, this kind of antenna is typically referred to as an *infinitesimal dipole* or a *Hertzian dipole*.

Now, we need to evaluate the integration of this current density with the free-space Green's function to determine the magnetic vector potential produced by the current. This will give us

$$\mathbf{A}(x, y, z) = \hat{z}\mu I_0 \int_{-\ell/2}^{\ell/2} \frac{e^{-jkR}}{4\pi R} dz', \tag{4.42}$$

where

$$R = |\mathbf{r} - \mathbf{r}'| \tag{4.43}$$

and a primed (un-primed) coordinate corresponds to a source (observation) location. In general, this integral can be difficult to evaluate. To avoid the complications, it is assumed here that because $\ell \ll \lambda$ that the functions involving R don't change appreciably over the integration region. This allows us to then get

$$\mathbf{A}(x, y, z) = \hat{z}\frac{\mu I_0}{4\pi r}e^{-jkr} \int_{-\ell/2}^{\ell/2} dz' = \hat{z}\frac{\mu I_0 \ell}{4\pi r}e^{-jkr}, \tag{4.44}$$

where $r = |\mathbf{r}|$.

For antenna problems, it is often advantageous to express the radiated fields in a spherical coordinate system. We can use standard coordinate system conversions to find that

$$A_r = A_z \cos\theta = \frac{\mu I_0 \ell}{4\pi r}e^{-jkr} \cos\theta, \tag{4.45}$$

$$A_\theta = -A_z \sin\theta = -\frac{\mu I_0 \ell}{4\pi r}e^{-jkr} \sin\theta, \tag{4.46}$$

$$A_\phi = 0 \times A_z = 0. \tag{4.47}$$

The electric and magnetic fields may now be computed from the magnetic vector potential using

$$\mathbf{E} = -j\omega\mathbf{A} + \frac{1}{j\omega\mu\epsilon}\nabla(\nabla \cdot \mathbf{A}), \tag{4.48}$$

$$\mathbf{H} = \mu^{-1}\nabla \times \mathbf{A}. \tag{4.49}$$

This can be done rather easily, however, it is essential to remember that the forms of the various vector differential operators are coordinate system dependent. As a result, we can't use the simple identities that we have used throughout the course for Cartesian coordinates.

Evaluating the various derivatives gives us

$$H_r = H_\theta = 0 \tag{4.50}$$

$$H_\phi = j\frac{kI_0\ell\sin\theta}{4\pi r}\left[1 + \frac{1}{jkr}\right]e^{-jkr} \tag{4.51}$$

for the magnetic field, and

$$E_r = \eta\frac{I_0\ell\cos\theta}{2\pi r^2}\left[1 + \frac{1}{jkr}\right]e^{-jkr} \tag{4.52}$$

$$E_\theta = j\eta\frac{kI_0\ell\sin\theta}{4\pi r}\left[1 + \frac{1}{jkr} - \frac{1}{(kr)^2}\right]e^{-jkr} \tag{4.53}$$

$$E_\phi = 0 \tag{4.54}$$

for the electric field. Being able to evaluate the fields like this by hand for all regions in space (away from the source region) is an extremely uncommon occurrence for an antenna problem. Typically, we are only able to determine the fields that exist far away from the antenna, which we will comment on more later.

Although we didn't dwell much on evaluating the derivatives, it should hopefully now make more sense why using the magnetic vector potential as part of our process is advantageous. If we were to try and directly calculate the electric or magnetic fields, we would need to compute the derivative of the current density given in (4.41). Although this is possible using somewhat more sophisticated forms of mathematics than is often taught in introductory calculus classes, it does complicate the procedure. By first integrating the current density against the free-space Green's function, we end up with a function that is much easier to compute derivatives of, leading to the fields given in (4.50) to (4.54).
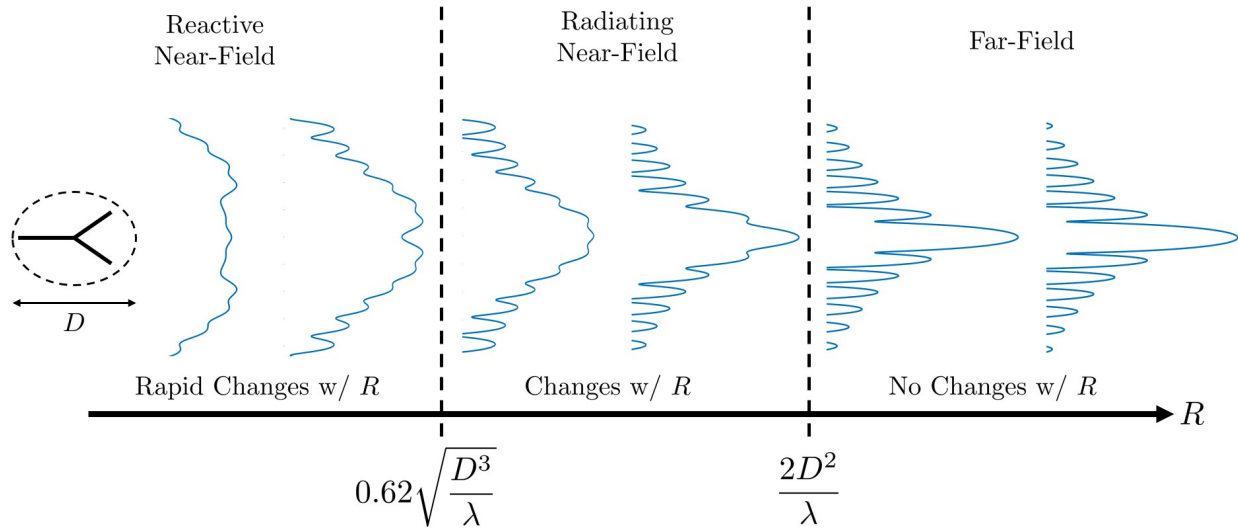
Figure 4.7: Illustration of different field regions around an antenna. Only in the far-field does the pattern stabilize as a function of distance away from the antenna.

4.4.1 Field Regions

Due to the complexity of the expressions for the fields produced by an antenna, it is typically useful to define a few different regions where various terms of the overall field expressions will "dominate" the other ones. This helps build intuition related to the operation of antennas, and hints at some important properties of analyzing antenna systems. The boundaries between regions are somewhat arbitrary and can need to be revisited for certain applications. However, they are often all given in terms of the wavelength and a dimension D, which is taken to be the diameter of a sphere that can completely contain the antenna being analyzed. This dimension and the various field regions are illustrated in Fig. 4.7. We will now discuss the various regions and their characteristics in more detail.

Now, the first region is often called the *near-field*, or sometimes the *reactive near-field*. This is the region that is immediately surrounding the antenna, where terms with higher inverse powers of r will dominate the overall field expressions. It can be shown that these terms do not contribute to a real outward power flow (which would be radiation), but instead contribute to imaginary power that is stored near the antenna. Due to the energy being stored in this way, we can think of these fields as being "reactive", much like a reactive load influences a circuit or transmission line problem. Due to the dependence on higher-order powers, the fields in this region of space typically rapidly change as a function of position. The typical boundary for this region is given as $R < 0.62\sqrt{D^3/\lambda}$.

The next region is often referred to as the *radiating near-field* or *Fresnel region*. This is an intermediate region where the fields are transitioning from being predominantly reactive to being predominantly radiative. The overall shape of the fields as a function of radial distance away from the antenna continue to change in this region. The typical boundary for this region is given as $0.62\sqrt{D^3/\lambda} < R < 2D^2/\lambda$.

The final region that is most often considered in antenna problems is the *far-field* or

Fraunhofer region. In this region, the fields are purely radiative and propagate radially away from the antenna. Further, the fields are "fully formed" in the sense that the distribution of radiation over the θ and ϕ coordinates no longer changes as a function of r. This is an essential property that greatly simplifies the analysis of antennas. Typically, we only care about the fields far away from the antenna. We can evaluate these fields directly by making certain approximations in the evaluation of the radiation integrals of the form shown in (4.42). Further, in this region we are often far enough away from the antenna that we can consider the fields to be plane waves locally. We can then use the tools we learned earlier in the course for analyzing plane wave problems. The typical boundary for this region is given as $R > 2D^2/\lambda$. You should memorize this distance. When we simplify fields to be in the far-field, we drop all terms that have an inverse power of R larger than 1. That is, terms that have dependence like R^{-2} or R^{-3} are assumed to be 0 in the far-field.

For the infinitesimal dipole, the far-field radiated fields would be

$$H_\phi = j\frac{kI_0\ell\sin\theta}{4\pi r}e^{-jkr} \tag{4.55}$$

$$E_\theta = j\eta\frac{kI_0\ell\sin\theta}{4\pi r}e^{-jkr}. \tag{4.56}$$

Note that if we were to calculate the Poytning vector for these fields, we would end up with a single component pointed in the radial direction, which matches our intuition for how fields should radiate away from a source. We also can note that there are no radial components to the field. This is a general property that always happens for fields in the far-field region of an antenna. This also matches our intuition about antennas producing plane waves (which are a special type of TEM wave) far away from the antenna.

4.5   Fundamental Parameters of Antennas

Now that we have a basic understanding of how to determine the fields of an antenna, we can begin to understand some of the fundamental parameters that are used to describe the performance of antennas. The first parameter we will discuss is the *radiation pattern* of an antenna. This is a mathematical function that describes the spatial distribution of radiation produced by an antenna, often as a function of the spatial coordinates θ and ϕ. Many different kinds of radiation patterns can be plotted for a single antenna to characterize a variety of properties. However, most often the radiation pattern is plotted in the far-field zone of the antenna so that the spatial distribution will be independent of the distance from the antenna (as long as it is far away from the antenna). Examples of radiation patterns would be *field* and *power* patterns, which plot a particular polarization of the radiated wave in either field or power units (e.g., V/m or W/m², respectively). Examples of a typical pattern for a directional antenna is shown in Fig. 4.8.

We typically refer to each region of the pattern between zeros as being a particular *lobe* of the radiation pattern. The primary lobe of the antenna where most power is radiated is referred to as the *mainlobe*. All other lobes will typically be referred to as *sidelobes*. For many applications, it is desirable to have a narrow mainlobe and as small of sidelobes as
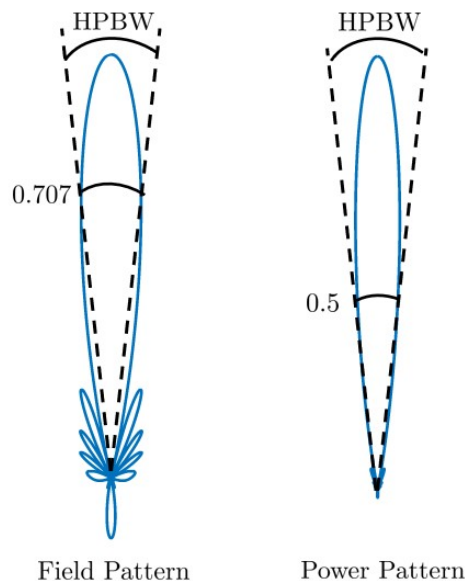
Figure 4.8: Example polar plots of field and power patterns of a directive antenna. The half-power beamwidth (HPBW) is also marked on both plots.

possible. This helps ensure the power predominantly goes toward the intended location and a minimal amount "spills" into undesirable directions. As an example, many modern radar applications require antennas with sidelobes that are approximately 30 to 40 dB smaller (i.e., 3 to 4 orders of magnitude smaller) than the peak magnitude of the mainlobe. This is typically a stringent requirement, which necessitates clever and precise engineering to be achievable.

Another common characteristic of an antenna is the *half-power beamwidth*. This is defined on Fig. 4.8, and is the angular separation between the two points on the mainlobe that are half the value of the mainlobe peak power. This beamwidth is one common parameter that is used to helps us understand how narrow the mainlobe is.

Often, we are interested in characterizing the radiated power associated with a particular antenna. To get to the parameters of most interest for characterizing antennas, we will need to go through a cascade of different power definitions. Our starting point is the time-average Poynting vector, which from earlier in the course we know is defined as

$$\mathbf{S}_{\text{avg}} = \frac{1}{2}\text{Re}\{\mathbf{E} \times \mathbf{H}^*\}. \tag{4.57}$$

In the far-field, we only are concerned with the *radiated power density* which we sometimes denote as $\mathbf{W}_{\text{rad}} = \mathbf{S}_{\text{avg}}$. Since $\mathbf{W}_{\text{rad}}$ is assumed to be in the far-field, we can safely assume that its vector direction is purely in the radial direction, i.e., $\hat{r}$.

We can determine the total power radiated from the antenna by integrating the outward directed radiated power density over a large sphere that encloses the entire antenna. This will give us

$$P_{\text{rad}} = \oiint_S \mathbf{W}_{\text{rad}} \cdot \hat{r}\, dS. \tag{4.58}$$

It is typically useful for us to have some convenient reference antenna to compare all other antennas to in specifying certain radiation properties. The usual reference antenna is an *isotropic radiator*, which is an ideal source that radiates equally in all directions. For this type of antenna, we can determine a useful correspondence between the radiation density and total radiated power of an antenna. We do this by noting that for W_0, the radiation density of an isotropic source, we will have the total radiated power given by

$$P_{\text{rad}} = \oiint_S \hat{r} W_0 \cdot \hat{r} dS = \oiint_S W_0 r^2 \sin\theta d\theta d\phi = 4\pi r^2 W_0. \tag{4.59}$$

Hence, we have that

$$\mathbf{W}_0 = \hat{r} W_0 = \hat{r} \frac{P_{\text{rad}}}{4\pi r^2} \tag{4.60}$$

which shows the even spreading of the power density over an increasingly large sphere as r increases.

We are often interested in removing the dependence on r in our far-field expressions altogether. The form of (4.60) suggests that we can define what is known as the *radiation intensity* as

$$U = r^2 |\mathbf{W}_{\text{rad}}|. \tag{4.61}$$

The radiation intensity helps measure the power that is radiated from an antenna *per unit solid angle*. Solid angles are measured in units of *steradians*, which are a generalization of the concept of a radian (a one-dimensional relation on a circle) to a two-dimensional area on a sphere. The exact definition isn't important for our purposes here, the main idea is that the far-field radiation intensity of an antenna will not depend on how far the observation distance is from the antenna. As a result, it is a useful reference property that doesn't have any ambiguities related to how far away the antenna is, which would always persist if we tried to specify things in terms of the radiation power *density*.

We can now define one of the most commonly used ways to characterize the performance of an antenna. Known as the *directivity*, this parameter is equal to the ratio of the antenna's radiation intensity in a given direction normalized by the radiation intensity of an isotropic source. Its mathematical definition is

$$D = \frac{U}{U_0} = \frac{4\pi U}{P_{\text{rad}}}, \tag{4.62}$$

where we have also used the fact that the equivalent isotropic radiation intensity is equal to $U_0 = P_{\text{rad}}/4\pi$. Often, the directivity pattern of the antenna will be plotted and the maximum directivity will also be reported to characterize an antenna.

Let's now calculate the directivity for the Hertzian dipole we considered earlier. Since the directivity is a far-field property of the antenna, we can use as our beginning fields

$$H_\phi = j\frac{kI_0\ell \sin\theta}{4\pi r} e^{-jkr} \tag{4.63}$$

$$E_\theta = j\eta \frac{kI_0\ell \sin\theta}{4\pi r}e^{-jkr}. \tag{4.64}$$

Looking at the definition of the directivity in (4.62), our first step will be finding the radiation intensity. Evaluating the quantities in (4.61) will give us the radiation intensity of a Hertzian dipole as

$$U = r^2|\mathbf{W}_{\text{rad}}| = |I_0|^2\frac{\eta}{2}\left(\frac{k\ell}{4\pi}\right)^2 \sin^2\theta = |I_0|^2\frac{\eta}{8}\left(\frac{\ell}{\lambda}\right)^2 \sin^2\theta \tag{4.65}$$

From the definition in (4.62), we see that we also need to determine the total radiated power. We can find this by integrating the radiation intensity over all directions. That is,

$$P_{\text{rad}} = \int_0^{2\pi}\int_0^\pi U\sin\theta d\theta d\phi = \eta|I_0|^2\frac{\pi}{3}\left(\frac{\ell}{\lambda}\right)^2 \tag{4.66}$$

We can now substitute the results of (4.65) and (4.66) into the directivity definition to get that

$$\boxed{D = \frac{3}{2}\sin^2\theta.} \tag{4.67}$$

From this, we can conclude that the maximum directivity will occur at $\theta = \pi/2$ and will be equal to

$$\boxed{D_0 = \frac{3}{2}.} \tag{4.68}$$

A plot of the Hertzian dipoles directivity pattern is shown in Fig. 4.9. From this, we can also see that the there will be no far-field radiation along the axis of the dipole itself. This makes intuitive sense, as we know that the electric field produced by the dipole near its endpoints will be oriented in the same direction as the $\hat{r}$ vector at these points, and so won't be able to produce a TEM wave that can propagate away from the antenna.

Although the directivity can often be computed using numerical techniques, it can be difficult to measure. The difficulty is that any real antenna will have a number of different loss mechanisms associated with it: such as conductor losses, dielectric losses, input matching losses, and so on. It can be difficult to determine all of these different contributing factors accurately, and they invariably impact the total power received during any antenna measurement. As a result, the loss will bias the results of any measurements of the directivity of an antenna. To accommodate this unavoidable issue, we most often will measure the *antenna gain*. This is simply the directivity of the antenna multiplied by its efficiency, i.e.,

$$G = e_0 D \tag{4.69}$$

where e_0 is the total efficiency of the antenna and ranges from 0 to 1. The gain is a quantity that we can readily measure to characterize an antenna, and is also the quantity of interest for many system-level analyzes (e.g., a link budget between two communication system antennas).
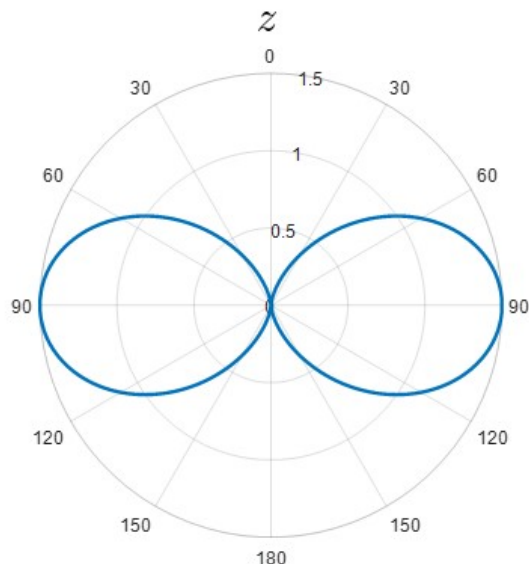
Figure 4.9: Example polar plot of the directivity pattern for a Hertzian dipole.

Often, both the directivity and gain will be reported in terms of decibels. In this case, one can convert either the gain or directivity into decibel units by taking $10 \log_{10}(X)$, where X should either be the directivity D or the gain G. Likewise, if a directivity or gain is given in terms of decibels, its linear scale units can be easily found by inverting the earlier relationship. Although it is typical for directivity or gain to be reported in decibels, it is important to emphasize that all equations given in these lecture notes are only valid for the directivity or gain when expressed on a linear scale.

Another important property of an antenna is its *input impedance*. This is the same concept as we dealt with when discussing how a general electromagnetic network could be viewed as an equivalent input impedance from the perspective of a transmission line network. As a result, we can conclude that the real part of the antenna's input impedance will characterize the effective loss of the antenna from the perspective of the transmission line. However, this loss can generally be broken into two parts, a loss due to actual power dissipation and a "loss" due to radiation. Since the fields propagate away from the antenna, this effectively looks like lost power to the transmission line network. To characterize this, we will sometimes define a *radiation resistance*. This is a fictitious resistance value that properly characterizes how much power is "lost" to radiation for a given voltage or current flowing into the antenna.

We can also conclude that the imaginary part of the antenna's input impedance will largely characterize the reactive near-fields of the antenna. The form of this impedance will depend on the type of antenna we are dealing with. For instance, the Hertzian dipole will generally be expected to have a capacitive reactance. This makes sense intuitively, because the extremely small dipole can be thought of as being similar to a slightly modified "open circuit" termination to a transmission line. We know that a small distance away from an open circuit will look like a capacitive reactance, which is the case here. This also clues us in on one of the difficulties with Hertzian dipoles as practical antennas. In general, they

have a very large reactance and a very small radiation resistance. As a result, it is difficult to have them efficiently radiate energy. This is a common difficulty for any antenna that is *electrically small.* There is currently a large amount of research interest in determining ways to more efficiently radiate power from electrically small antennas.

To address the issues with a small radiation resistance and a large reactance, we often need to make our antennas electrically larger. A classic example is a *half-wavelength dipole antenna.* By increasing the length of the arms of the dipole, it can begin to act more like a resonator. If we operate the antenna at the right frequencies, we can achieve a higher radiation resistance and a reactance that can be more easily addressed with realistic matching circuits. For instance, a half-wavelength dipole has an input impedance of approximately $Z_{\text{in}} = 73 + j42.5\Omega$, which can be readily matched using various techniques. Another effect that happens when we make our antennas electrically larger is that they tend to be better able to focus radiation into a more directive pattern. A great example of this is a reflector antenna. To achieve a narrower antenna pattern we need to make the reflector larger so that it can collect and focus energy from a larger "sampling area".

4.6   Friis Transmission Equation

One of the fundamental purposes of an antenna is to transmit power wirelessly between two points in space. The *Friis transmission equation* provides us with a simple tool to compute how much power will be received by a particular antenna given a certain amount of power that is transmitted by another antenna. We will consider a simple derivation of this equation now that is valid if both antennas are located in each other's far-field region. Before getting into the details, there is an important fact about the *reciprocity* of antennas that we should introduce. The basic principle is that due to the reciprocity of electromagnetic systems (this is a general electromagnetic theorem you will learn about in advanced EM courses), antennas will typically be reciprocal devices under many practical circumstances. One of the main consequences of this is that we don't need to use different antenna gain patterns to describe the antenna in transmission or receive mode, the two patterns will be identical due to reciprocity.

To begin, we need to determine the power density that will be produced by our transmitter. By definition, we know that the power density for an isotropic source with transmit power P_t will be

$$W_0 = \frac{P_t}{4\pi R^2}. \tag{4.70}$$

We also know that by definition, the power density of our actual transmitter will be equal to W_0 modified by the *gain pattern* of the transmitter. In particular, we will get that

$$W_t = \frac{P_t G_t(\theta_t, \phi_t)}{4\pi R^2}, \tag{4.71}$$

where G_t is the gain of the transmitter.

To determine the power that is received by the receiving antenna, it is useful to introduce another antenna parameter known as the *effective area* or *effective aperture* of the antenna.

This is a fictitious area that when multiplied by an incident power density will give the power delivered to a matched load at the antenna's terminals. If we denote the effective area as A_r, we will get that the power received P_r is

$$P_r = A_r \frac{P_t G_t(\theta_t, \phi_t)}{4\pi R^2}. \tag{4.72}$$

One special property of the effective area is that there is a universal relationship between the ratio of the effective area and gain *for any antenna*. The particular ratio is that

$$\boxed{\frac{A_r}{G_r} = \frac{\lambda^2}{4\pi}.} \tag{4.73}$$

Considering this, we can rewrite (4.72) as

$$P_r = P_t \left(\frac{\lambda}{4\pi R}\right)^2 G_t(\theta_t, \phi_t) G_r(\theta_r, \phi_r). \tag{4.74}$$

It is traditional to write this as a ratio between P_r and P_t to define the link budget of a system. This gives us the simple form of Friis transmission equation to be

$$\boxed{\frac{P_r}{P_t} = \left(\frac{\lambda}{4\pi R}\right)^2 G_t(\theta_t, \phi_t) G_r(\theta_r, \phi_r).} \tag{4.75}$$

The factor in the parentheses is often called the *free-space loss factor*, since it takes into account the spherical spreading of power as it propagates in space. The gains in this equation then serve to help correct for the fact that our antennas don't actually transmit or receive energy into a perfect sphere, but instead focus it in a particular direction.

Before moving on, it should be mentioned that this is one of the simplest forms of Friis transmission equation. Many other factors can need to be taken into account to make the formula more complete. For instance, one simple factor we neglected was the *polarization* of the antennas with respect to each other. If two antennas are oriented so they transmit or receive orthogonal polarizations the amount of power that can be sent between the two will drop dramatically.

4.7 Introduction to Antenna Arrays

For many applications, it is desirable to have a large amount of control over the antenna pattern. For instance, radar or communication systems often want to achieve a large gain in a particular direction to avoid wasting power by allowing it to spill out into undesirable directions. Likewise, for high-performance applications we typically have stringent requirements on the maximum level that sidelobes can reach without spoiling the properties of a system, such as an imaging radar for example.

We have already seen that the radiated fields of the antenna will depend on the current density that exists on the antenna as

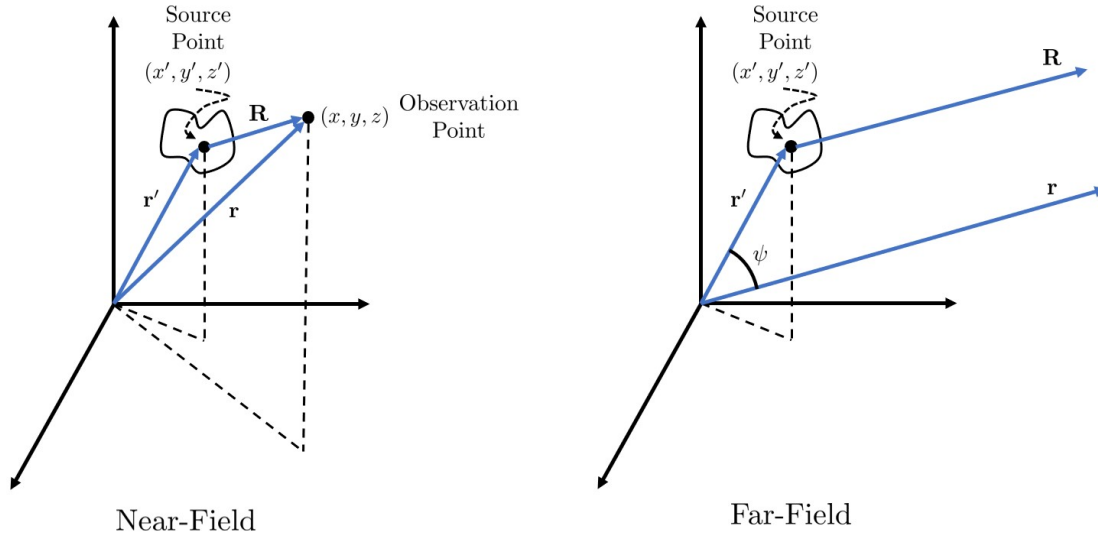$$\mathbf{A}(\mathbf{r}) = \iiint \frac{1}{4\pi R} e^{-jkR} \mu \mathbf{J}(\mathbf{r}') d\mathbf{r}', \tag{4.76}$$

Figure 4.10: Illustration of the far-field approximation. When the observation point is far away from the source point, the vectors **r** and **R** are approximately parallel and certain simplifications can be made to simplify the evaluation of "radiation integrals" that are needed to analyze an antenna.

where $R = |\mathbf{r} - \mathbf{r}'|$. In the far-field, we are able to approximate the electric field as being

$$\mathbf{E} \approx -j\omega\mathbf{A}, \tag{4.77}$$

where **A** will only have non-zero θ- and ϕ-components due to the far-field assumption. Hence, we have that

$$\mathbf{E} \approx \frac{-j\omega\mu}{4\pi} \iiint \frac{e^{-jkR}}{R}\mathbf{J}(\mathbf{r}')d\mathbf{r}'. \tag{4.78}$$

If we only care about finding the fields in the far-field region, we can simplify the integration by assuming that

$$R \approx r \quad \text{for amplitude variations}, \tag{4.79}$$

$$R \approx r - r'\cos\psi \quad \text{for phase variations}, \tag{4.80}$$

where ψ is the angle between the vectors **r** and **r'**. This approximation is illustrated in Fig. 4.10, where it shows that in the far-field region the difference between R and r will often be offset by a small factor.

For amplitude values, this small value often doesn't make a significant difference so we can use the simplification given above. However, because the phase factor depends on the wavelength, this small difference can lead to large changes in behavior. This is why we need to add the additional term given above for accounting for phase variations in the far-field. Although it is still approximate, it does a very good job while still greatly simplifying the overall difficulty of the integrations involved. Considering these simplifications, we have that

$$\mathbf{E} \approx -j\omega\mu\frac{e^{-jkr}}{4\pi r} \iiint \mathbf{J}(\mathbf{r}')e^{-jkr'\cos\psi}d\mathbf{r}'. \tag{4.81}$$

The important point is that if we expand $r' \cos \psi$ in a particular set of coordinates that are convenient for a given problem we can always view this integral as being a spatial Fourier transform of the current density.

To see this, it is helpful to expand $r' \cos \psi$ for the simple case of a planar antenna confined in the $z = 0$ plane. In this case, we find that (4.81) becomes

$$\mathbf{E} \approx -j\omega\mu \frac{e^{-jkr}}{4\pi r} \iint \mathbf{J}(x', y') e^{-jkx' \sin\theta \cos\phi} e^{-jky' \sin\theta \sin\phi} dx' dy'. \tag{4.82}$$

If we define $k_x = k \sin\theta \cos\phi$ and $k_y = k \sin\theta \sin\phi$, we can quickly recognize that the integration is simply a *spatial Fourier transform* of the current density along the x' and y' axes. Each data point in the (k_x, k_y)-plane can then be considered to give us the complex amplitude of the plane wave that would be radiated into a particular (θ, ϕ)-direction, which corresponds to our far-field antenna pattern. As a result, we can use our intuition about Fourier transforms to help determine the amplitude and phase that $\mathbf{J}(\mathbf{r}')$ should have to achieve a desired radiation pattern.

Although this is a powerful concept, it is not typically very easy to implement in practice. The issue is that $\mathbf{J}(\mathbf{r}')$ must satisfy Maxwell's equations, and as a result it is a complicated function of the geometry that the current exists on. Hence, we can't arbitrarily control $\mathbf{J}(\mathbf{r}')$ over a large area, and in general it will be difficult to get it to even approximate a particular structure that may be needed to produce a desired antenna pattern.

The way around this issue is to use *antenna arrays*. Antenna arrays are formed by a combination of discrete antennas that are driven together to achieve some desirable radiation characteristics. Often, each antenna in the array will have the same basic shape, as it greatly simplifies the design, analysis, and manufacturing of the array, although there is not necessarily a requirement for this to be the case. We can think of using an array as being synonymous with approximating a continuous Fourier transform with a discrete Fourier transform. The important practical consequence is that we can achieve a great amount of control over the different excitations of each antenna within an antenna array by designing a suitable transmission line network. This network is often referred to as a *feed network*. Designing a feed network is a very tractable problem, whereas designing a continuous source to have a desired current density is almost impossible. Hence, antenna arrays are incredibly prevalent in practical applications.

The overall fields radiated by an antenna array is the vector superposition of the fields radiated from each individual antenna. As a result, if we already know the radiation pattern of the array element, we can quickly find the far-field pattern of the entire array (under somewhat ideal conditions). To see this, let's consider a simple two-element array oriented along the z-axis. If both antennas have the same complex-valued electric field pattern in the far-field given by $\mathbf{F}(\theta, \phi)$, then for the array shown in Fig. 4.11 we have that the total radiated field will be

$$\mathbf{E}_t = \mathbf{E}_1 + \mathbf{E}_2 = |A_1| \mathbf{F}(\theta, \phi) \frac{e^{-j(kR_1 - \Delta/2)}}{4\pi R_1} + |A_2| \mathbf{F}(\theta, \phi) \frac{e^{-j(kR_2 + \Delta/2)}}{4\pi R_2} \tag{4.83}$$

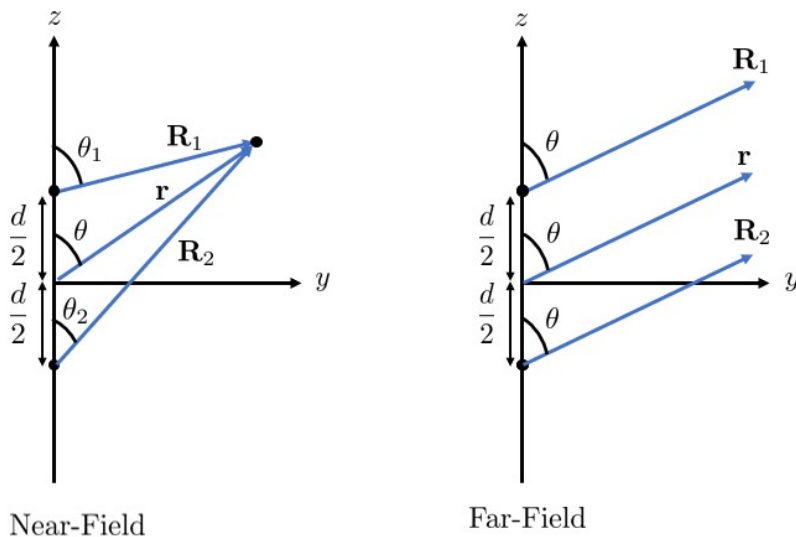where Δ is the phase difference between the excitations of the two antenna elements. We

Figure 4.11: Illustration of the simple two-element array.

can make our far-field approximation by assuming that

$$R_1 \approx R_2 \approx r \quad \text{for amplitude variations} \tag{4.84}$$

$$R_1 \approx r - \frac{d}{2}\cos\theta \quad \text{for phase variations} \tag{4.85}$$

$$R_2 \approx r + \frac{d}{2}\cos\theta \quad \text{for phase variations.} \tag{4.86}$$

We can then write our total electric field as

$$\mathbf{E}_t = \mathbf{F}(\theta, \phi)\frac{e^{-jkr}}{4\pi r}\left[|A_1|e^{j(kd\cos\theta+\Delta)/2} + |A_2|e^{-j(kd\cos\theta+\Delta)/2}\right]. \tag{4.87}$$

If we were to set $|A| = |A_1| = |A_2|$, we could simplify our pattern further to be

$$\mathbf{E}_t = \frac{e^{-jkr}}{4\pi r}\mathbf{F}(\theta, \phi) \times 2|A|\cos\left(\frac{1}{2}(kd\cos\theta + \Delta)\right). \tag{4.88}$$

We have used the $\times$ symbol to emphasize that we can look at the total pattern as being the multiplication between the antenna pattern of the individual elements *times* what is often referred to as the *array factor*. This concept is known as *pattern multiplication*, and is the reason why the analysis of most antenna arrays involves primarily finding the array factor of a proposed array. Before concluding, it is worth noting that we can use Δ to bias the array factor to have a maximum value at a particular value of θ. This allows us to *steer* the mainlobe of an antenna array in a desirable direction by choosing suitable phase values at the different antenna elements. This is referred to as a *phased array*, and is a very common tool in many practical antenna applications.

4.8 Practice Problems

1. Assume that two antennas operating at 4 GHz are in the far-field of each other and are aligned such that there is no loss due to polarization mismatch. Further, assume that the antennas are oriented so that the angles between the two antennas correspond to both antennas maximum gain point in their respective patterns. You may also assume the antennas operate in free space.

   (a) If the antennas are 90 km apart, the transmit power is 500 W, and the gain of the transmitter is 24 dB, what is the power density at the location of the receiver?

   (b) If the received power at the receive antenna is 0.15 μW, what is the gain of the receive antenna in dB?

2. Assume that a two-element array of infinitesimal dipoles are located symmetrically about the z-axis and are a distance $d = \lambda/2$ apart. Further, the antenna excitations are set such that they have the same magnitude and 180° phase difference. For this scenario, roughly sketch the directivity pattern for the the two-element array of infinitesimal dipoles as a polar plot in the (y, z)-plane.

3. Assuming that a two-element array of infinitesimal dipoles are located symmetrically about the z-axis and are a distance $d = \lambda/4$ apart, answer the following questions.

   (a) Assuming that the magnitude of the excitations of both dipoles are identical, what should the relative phase shift between the two antenna excitations be so that the far-field antenna pattern of the array produces no radiation at $\theta = 60°$?

   (b) For the excitation scheme described in (a), roughly sketch the directivity pattern for the the two-element array of infinitesimal dipoles in the (y, z)-plane.

Bibliography

[1] D. Styer, "Calculation of the anomalous magnetic moment of the electron," 2012.

[2] Wikipedia contributors, "Gradient — Wikipedia, the free encyclopedia," https://en.wikipedia.org/w/index.php?title=Gradient&oldid=1196514741, 2024, [Online; accessed 9-February-2024].

[3] W. C. Chew, *Waves and Fields in Inhomogeneous Media.* IEEE Press, 1995.

[4] Wikipedia contributors, "Plane wave — Wikipedia, the free encyclopedia," https://en.wikipedia.org/w/index.php?title=Plane_wave&oldid=1187992025, 2023, [Online; accessed 9-February-2024].

[5] ——, "Over-the-horizon radar — Wikipedia, the free encyclopedia," https://en.wikipedia.org/w/index.php?title=Over-the-horizon_radar&oldid=1195102860, 2024, [Online; accessed 9-February-2024].

[6] I. D. Olin, "Polarization characteristics of coherent waves," *IEEE Antennas and Propagation Magazine*, vol. 56, no. 3, pp. 90–102, 2014.

[7] Wikipedia contributors, "Polarization (waves) — Wikipedia, the free encyclopedia," https://en.wikipedia.org/w/index.php?title=Polarization_(waves)&oldid=1203084697, 2024, [Online; accessed 9-February-2024].

[8] ——, "Turnstile antenna — Wikipedia, the free encyclopedia," https://en.wikipedia.org/w/index.php?title=Turnstile_antenna&oldid=1095640919, 2022, [Online; accessed 9-February-2024].

[9] C.-C. Chang, H. P. Sardesai, and A. M. Weiner, "Dispersion-free fiber transmission for femtosecond pulses by use of a dispersion-compensating fiber and a programmable pulse shaper," *Optics Letters*, vol. 23, no. 4, pp. 283–285, 1998.

[10] S. Ullah, W.-H. Yeo, H. Kim, and H. Yoo, "Development of 60-GHz millimeter wave, electromagnetic bandgap ground planes for multiple-input multiple-output antenna applications," *Scientific Reports*, vol. 10, no. 1, pp. 1–12, 2020.

[11] Wikipedia contributors, "Waveplate — Wikipedia, the free encyclopedia," https://en.wikipedia.org/w/index.php?title=Waveplate&oldid=1196060382, 2024, [Online; accessed 9-February-2024].

[12] ——, "Faraday effect — Wikipedia, the free encyclopedia," https://en.wikipedia.org/w/index.php?title=Faraday_effect&oldid=1200966862, 2024, [Online; accessed 9-February-2024].

[13] D. M. Pozar, *Microwave Engineering.* John Wiley & Sons, 2009.

[14] D. K. Cheng, *Field and Wave Electromagnetics*, 2nd ed. Addison-Wesley Publishing Company, Inc., 1989.

[15] E. Jensen, "Cavity basics," *arXiv preprint arXiv:1201.3202*, 2012.

[16] C. A. Balanis, *Antenna Theory: Analysis and Design.* John Wiley & Sons, 2005.