Lectures on
Theory of Microwave and Optical Waveguides

Weng Cho CHEW [1]

Fall 2015

[1] updated February 15, 2016

Contents

Preface

This monograph is an outgrowth of many years of teaching at the University of Illinois at Urbana-Champaign since Fall 1987. It has been recently updated December, 2015.

Weng Cho CHEW
Fall 2015

Chapter 1

Preliminary Background

1.1 Introduction

Waveguiding phenomena occur naturally or are man made. For instance, waveguides are a fundamental component of radio wave, microwave and optical circuits [1–5]. They are indispensable in modern technology in the radio frequency to the optical frequency range. They are used in the telecommunications as well as in wireless communications, for example, in the design of a cell phone. The purpose of a waveguide is to guide the energy of a wave through a channel or a path with little attenuation. Waveguides are also used to prevent interference between two electromagnetic signals.

The precursor to electromagnetic waveguides were acoustic waveguides as acoustic wave theory, being scalar, was well established before electromagnetic theory [6]. Since acoustic waves are longitudinal waves. They can be guided as a longitudinal mode in a hollow tube for all frequencies. As a result, tubes of acoustic waveguides of different lengths have been used as musical instruments since ancient times. The first analysis of electromagnetic guided wave was probably done by Lord Rayleigh [7].

As we shall see later, a simple way to guide electromagnetic wave for all frequencies, is to use two metallic conductors, usually an inner one and an outer one as in a coaxial cable. As optical fiber guides a mode for all frequencies too, but as shall be shown, when the frequency is very low, the mode's energy is weakly trapped inside the fiber, making it impractical as a waveguide for extremely low frequencies. Since most sources are finite in extent, e.g., antennas, they generate spherical waves in the far-field which decays algebraically. However, waveguides, by confining the energy of the wave to a tube or a line, can cause a wave to traverse great distances with little attenuation. An example is an optical fiber, which can guide a signal with less than 0.3 dB/Km of attenuation [8].

Other emerging waveguiding technologies are plasmonic waveguides at optical frequencies in nano-optics, or guiding waves using a chain of nano particles [28]. As nanoelectronic devices are getting smaller, their dimensions are approaching the wavelengths of electron wave functions. The propagation of electron waves in a channel can be viewed as a waveguiding problem [29].

There are two main types of waveguides: the closed waveguide and the open waveguide. In

a closed waveguide, the electromagnetic energy is completely trapped within metallic walls. The only way to gain access to the energy is to tap holes in the waveguide wall. Hence, it transmits signals with very good shielding and very little interference from other signals. Figure 1.1 shows some examples of closed waveguides. Notice that a closed waveguide can be of one or more conductors. On the other hand, an open waveguide allows its field to permeate all of space, even though most of the energy is still trapped and localized around the guidance structure. As shown in Figure 1.2, an open waveguide is either a multi-conductor waveguide or a dielectric waveguide. It is usually easier to fabricate an open waveguide. However, as a result of their openness, such waveguides usually radiate at discontinuities and bends.

Because open waveguides radiate at discontinuities and bends, some of them are even used as antennas [9]. There are also modes that are weakly guided by an open waveguide, i.e., it radiates as it is being guided. Examples of such modes are the leaky modes. An antenna built using such a mode is known as a leaky wave antenna.

The analysis of waveguides requires a basic understanding of electromagnetic theory. We will review our basic electromagnetic theory in the following section.

1.2 History of Electricity and Magnetism

Humans are exposed to electromagnetic phenomena on a daily basis. Light wave is an electromagnetic phenomenon, so is lightning. Lodestone is probably the first human experience with something magnetic. Ancient Chinese knew about the magnetic properties of lodestones, and made compasses out of them. Static electricity was a phenomenon popularly demonstrated in European courts to entertain the nobilities. But it was not until 1771-1773 that serious experiments were done on static electricity by Henry Cavendish (1731-1810). To this day, the Cavendish Laboratory stands in the University of Cambridge in England to the honor of Cavendish [10].

Faraday's law was formulated by Michael Faraday (1791-1867) to describe the fact that a changing magnetic flux, linked to a metallic loop, will induce a voltage in the loop [12]. This fact can be used to design generators that produce electricity for our homes. A multi-turn coil can be immersed in the magnetic field of a permanent magnet, and rotated rapidly. A voltage is then induced in the coil, which can be tapped to deliver electricity for a large number of applications. Conversely, a DC current in a static magnetic field experiences a force due to Lorentz force law. This idea can be used to design a motor. In fact, a DC motor was invented by William Sturgeon in 1832 [27].

Ampere's law was later formulated by André Marié Ampère (1775-1836) who stipulated that a wire carrying a current produces a magnetic field [13]. Moreover, the magnetic field is produced according to the right-hand rule. (Note: The stipulation that the magnetic field goes from the north pole of a bar magnet to its south pole is entirely by convention. Hence, the right-hand rule in Ampere's law is also entirely by convention. Also, the concept of right-handedness and left-handedness is hard to describe to an extra-terrestrial creature living in another universe who has never seen a human before. Try that for yourself [11]!)

Gauss' law by Carl Friedrich Gauss (1777-1855) describes that if a charge generating an electric field is enclosed by a surface S, the sum of the total flux flowing through the surface is equal to the total charge contained within the surface [15]. If the surface does not enclose

any charge, the sum of the total charge through the surface is equal to zero. Coulomb's law can be derived from Gauss' law.

The above period represented the era during which the understanding of electromagnetism was incomplete. Nevertheless, technology using electricity and magnetism was prevalent. As soon as Alessandro Volta invented the battery, Ampere developed telegraphy in the early 1800s. In fact, submarine cables were laid during a large part of the nineteenth century by the British empire around the world to enable telegraphic communication. So it was quite well known that wave phenomena existed on telegraphic lines before the completion of Maxwell's theory as we shall discuss next.

Electromagnetic theory was completely formulated by the work of James Clerk Maxwell (1831-1879) [16]. In 1864, he put forth the theory that there should be a term, called the displacement current term, to be added to Ampere's law. The work completed electromagnetic theory and it was proven mathematically that electromagnetic wave was a possible electromagnetic phenomenon. Consequently, it was realized that light waves were electromagnetic waves. Because of this important discovery, electromagnetic theory is also known as Maxwell's theory, and the set of equations is also known as Maxwell's equations. However, when Maxwell first wrote down the complete form of electromagnetic theory, it was in some 20 equations. It was Oliver Heaviside who recast those equations in their present succinct form. Rightfully, these equations should be called the Maxwell-Heaviside equations [17].

In 1888, Heinrich Rudolf Hertz (1857-1894) performed an experiment to verify the existence of electromagnetic wave. Two spheres in close proximity to each other were used as capacitors to store electric charges. The charges generate an electric field. A rapid discharge of the electric charge causes the electric field to collapse, producing an electromagnetic wave. The wave has both electric and magnetic field in it. Therefore, a wire loop, via Faraday's law, can be linked to the time varying magnetic flux, producing a voltage. This voltage creates a spark in a gap left in the loop, even when the loop is at a distance from the spheres. To his honor, the unit for frequency, which was cycles/second, is now named Hertz. The term megahertz (MHz), or gigahertz (GHz) now adorns the spec sheets of most computers.

In 1901, Guglielmo Marchese Marconi (1874-1937) successfully transmitted an electromagnetic signal across the Atlantic Ocean from Cornwall, England to Saint John, Newfoundland in North America [18]. Many nay sayers predicted that he would be doomed to failure as the earth surface is curved. Fortunately, the ionosphere in the outer atmosphere acted like a mirror, and the electromagnetic waves bounced back to earth. It was in fact a serendipitous experiment. It was after his experiments that wireless telegraphy was established. However, Marconi never received a patent for his invention. The patent for telecommunication was claimed by Nikola Tesla, who had the idea before Marconi.

Since then, electromagnetic theory has spurred the development of myriads of technologies, many of which are electrical engineering related. Some of the more prominent ones are the development of the radar, various antennas for telecommunication, remote sensing systems, lasers and optics and more recently, wireless communications, computer chip design, and electromagnetic compatibility and electromagnetic interference. The advent of quantum technologies as seen in quantum optics, quantum computers, quantum communications, Casimir force in MEMS/NEMS, quantum transport in electronic devices, photonics, will also dwell on classical electromagnetics in combination with modern physics concepts.

As of this date, electromagnetic theory continues to help in the conception, analysis, and

design of many new technologies. Hence, Maxwell's equations are solved over and again for many of these analysis tasks. As a consequence, much research has gone into developing methods to impact many analyses in science and engineering.

1.3 Maxwell's Equations

Soon after the advent of Maxwell's theory, much analysis was performed with Maxwell's equations. By 1897, Lord Rayleigh had already studied the propagation of electromagnetic waves through tubes. There was then much knowledge on propagation and guidance of acoustic waves. Hence, analogue between acoustic waves and electromagnetic waves were drawn as much as possible, although acoustic waves are scalar while electromagnetic waves are vector in nature.

In vector notation, and MKS units, Maxwell's equations are given as

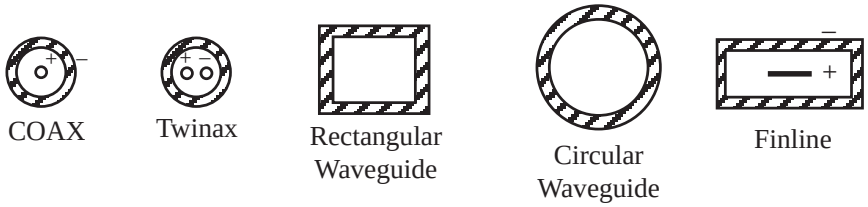


Figure 1.1: Examples of closed waveguides.

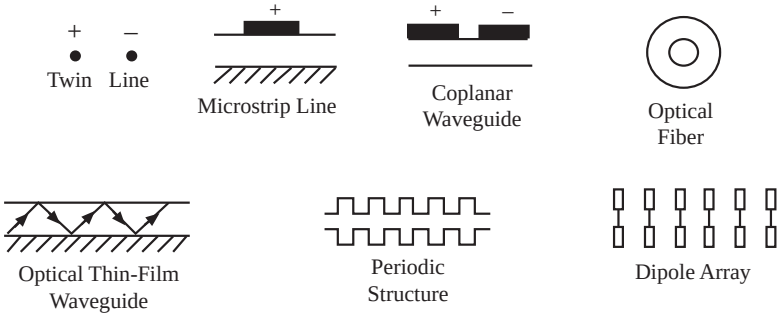


Figure 1.2: Examples of open waveguides.

$$\nabla \times \mathbf{E}(\mathbf{r}, t) = -\frac{\partial}{\partial t}\mathbf{B}(\mathbf{r}, t), \tag{1.3.1}$$

$$\nabla \times \mathbf{H}(\mathbf{r}, t) = \frac{\partial}{\partial t}\mathbf{D}(\mathbf{r}, t) + \mathbf{J}(\mathbf{r}, t), \tag{1.3.2}$$

$$\nabla \cdot \mathbf{B}(\mathbf{r},t) = 0, \tag{1.3.3}$$

$$\nabla \cdot \mathbf{D}(\mathbf{r},t) = \rho(\mathbf{r},t). \tag{1.3.4}$$

where $\mathbf{E}$ is the electric field in volts/m, $\mathbf{H}$ is the magnetic field in amperes/m, $\mathbf{D}$ is the electric flux in coulombs/m^2, $\mathbf{B}$ is the magnetic flux in webers/m^2, $\mathbf{J}(\mathbf{r},t)$ is the current density in amperes/m^2, and $\rho(\mathbf{r},t)$ is the charge density in coulombs/m^3. For time varying electromagnetic fields, only two of the four Maxwell's equations are independent. Equations (1.3.3) and (1.3.4) can be derived from Equations (1.3.1) and (1.3.2) by using the continuity equation:

$$\nabla \cdot \mathbf{J}(\mathbf{r},t) + \frac{\partial \rho(\mathbf{r},t)}{\partial t} = 0. \tag{1.3.5}$$

If we assume that $\mathbf{A}(\mathbf{r},t) = \Re e\left\{\mathbf{A}(\mathbf{r})e^{-i\omega t}\right\}$, for all $\mathbf{E}(\mathbf{r},t)$, $\mathbf{H}(\mathbf{r},t)$, $\mathbf{J}(\mathbf{r},t)$, and $\rho(\mathbf{r},t) = \Re e\left\{\rho(\mathbf{r})e^{-i\omega t}\right\}$; namely, the fields are time harmonic, the above equations become,

$$\nabla \times \mathbf{E}(\mathbf{r}) = i\omega \mathbf{B}(\mathbf{r}), \tag{1.3.6}$$

$$\nabla \times \mathbf{H}(\mathbf{r}) = -i\omega \mathbf{D}(\mathbf{r}) + \mathbf{J}(\mathbf{r}), \tag{1.3.7}$$

$$\nabla \cdot \mathbf{B}(\mathbf{r}) = 0, \tag{1.3.8}$$

$$\nabla \cdot \mathbf{D}(\mathbf{r}) = \rho(\mathbf{r}). \tag{1.3.9}$$

The electric and magnetic fluxes are related to the electric and magnetic fields via the constitutive relations, the most general of which are

$$\mathbf{D} = \overline{\boldsymbol{\epsilon}} \cdot \mathbf{E} + \overline{\boldsymbol{\xi}} \cdot \mathbf{H}, \tag{1.3.10}$$

$$\mathbf{B} = \overline{\boldsymbol{\mu}} \cdot \mathbf{H} + \overline{\boldsymbol{\zeta}} \cdot \mathbf{E}, \tag{1.3.11}$$

where $\overline{\boldsymbol{\epsilon}}$, $\overline{\boldsymbol{\xi}}$, $\overline{\boldsymbol{\mu}}$ and $\overline{\boldsymbol{\zeta}}$ are tensors. It is also the constitutive relations that characterize the medium we are describing. A medium with the above constitutive relations is known as a bianisotropic medium. A more commonly encountered medium is an anisotropic medium with the constitutive relations

$$\mathbf{D} = \overline{\boldsymbol{\epsilon}} \cdot \mathbf{E}, \tag{1.3.12}$$

$$\mathbf{B} = \overline{\boldsymbol{\mu}} \cdot \mathbf{H}. \tag{1.3.13}$$

When $\overline{\boldsymbol{\epsilon}}$, $\overline{\boldsymbol{\xi}}$, $\overline{\boldsymbol{\mu}}$ and $\overline{\boldsymbol{\zeta}}$ are functions of space, the medium is also known as an inhomogeneous medium. When they are functions of frequency, the medium is frequency dispersive. When they are functions of wavelength, it is spatially dispersive. For an isotropic medium, the constitutive relations simply become

$$\mathbf{D} = \epsilon \mathbf{E}, \qquad \mathbf{B} = \mu \mathbf{H}. \tag{1.3.14}$$

In free-space, $\epsilon = \epsilon_0 = 8.854 \times 10^{-12}$ farad/m, $\mu = \mu_0 = 4\pi \times 10^{-7}$ henry/m. The constant $c = \frac{1}{\sqrt{\mu_0 \epsilon_0}}$ is related to the velocity of light, which has been very accurately measured. The unit of meter is defined such that c is exactly equal to 299,792,458 m/s. The value of μ_0 is assigned to be $4\pi \times 10^{-7}$ henry/m while the value of ϵ_0 is calculated from c.

1.4 Wave Equation

For an anisotropic, inhomogeneous medium, Maxwell's equations for time-harmonic fields could be written as

$$\nabla \times \mathbf{E}(\mathbf{r}) = i\omega\overline{\boldsymbol{\mu}} \cdot \mathbf{H}(\mathbf{r}), \tag{1.4.1}$$

$$\nabla \times \mathbf{H}(\mathbf{r}) = -i\omega\overline{\boldsymbol{\epsilon}} \cdot \mathbf{E}(\mathbf{r}) + \mathbf{J}(\mathbf{r}), \tag{1.4.2}$$

$$\nabla \cdot \overline{\boldsymbol{\mu}} \cdot \mathbf{H}(\mathbf{r}) = 0, \tag{1.4.3}$$

$$\nabla \cdot \overline{\boldsymbol{\epsilon}} \cdot \mathbf{E}(\mathbf{r}) = \rho(\mathbf{r}). \tag{1.4.4}$$

If we take the curl of $\overline{\boldsymbol{\mu}}^{-1} \cdot$ (1.4.1), we obtain, via the use of (1.4.2), that

$$\nabla \times \overline{\boldsymbol{\mu}}^{-1} \cdot \nabla \times \mathbf{E}(\mathbf{r}) - \omega^2\overline{\boldsymbol{\epsilon}} \cdot \mathbf{E}(\mathbf{r}) = i\omega\mathbf{J}(\mathbf{r}). \tag{1.4.5}$$

Similarly, we can show that

$$\nabla \times \overline{\boldsymbol{\epsilon}}^{-1} \cdot \nabla \times \mathbf{H}(\mathbf{r}) - \omega^2\overline{\boldsymbol{\mu}} \cdot \mathbf{H}(\mathbf{r}) = \nabla \times \overline{\boldsymbol{\epsilon}}^{-1} \cdot \mathbf{J}(\mathbf{r}). \tag{1.4.6}$$

Equations (1.4.5) and (1.4.6) are two vector wave equations governing the solutions of electromagnetic fields in an inhomogeneous, anisotropic medium. Here, $\overline{\boldsymbol{\mu}}$ and $\overline{\boldsymbol{\epsilon}}$ are functions of positions; hence, they do not commute with the ∇ operator. Also, for time-varying fields, $\mathbf{E}$ and $\mathbf{H}$ are derivable from each other; only one of the two equations (1.4.5) and (1.4.6) is necessary to fully describe the electromagnetic fields.

For an isotropic medium, (1.4.5) and (1.4.6) reduce to

$$\nabla \times \mu^{-1}\nabla \times \mathbf{E}(\mathbf{r}) - \omega^2\epsilon\mathbf{E}(\mathbf{r}) = i\omega\mathbf{J}(\mathbf{r}), \tag{1.4.7}$$

$$\nabla \times \epsilon^{-1}\nabla \times \mathbf{H}(\mathbf{r}) - \omega^2\mu\mathbf{H}(\mathbf{r}) = \nabla \times \epsilon^{-1}\mathbf{J}(\mathbf{r}). \tag{1.4.8}$$

For electrodynamics, either one of the above equations is self-contained. We can derive the phenomena of dynamic electromagnetic fields by just studying one of them. However, when $\omega \to 0$, these equations are not solvable, and we have to invoke all four of Maxwell's equations when solving static problems.

1.5 Boundary Conditions

We cannot find a unique solution to a partial differential equation unless we specify the boundary conditions as well. Equations (1.4.5) to (1.4.8) are vector wave equations whose solutions we will seek over and again. One common method of solving the above equations is to find the solutions in each of the homogeneous regions that constitute the inhomogeneity, provided that the inhomogeneity is piecewise constant. The unique solution is then obtained by matching the boundary conditions at the interface.

Since either Equation (1.4.5) or (1.4.6) is sufficient in describing electromagnetic fields, the boundary conditions must be buried in them. Therefore, we can derive the boundary

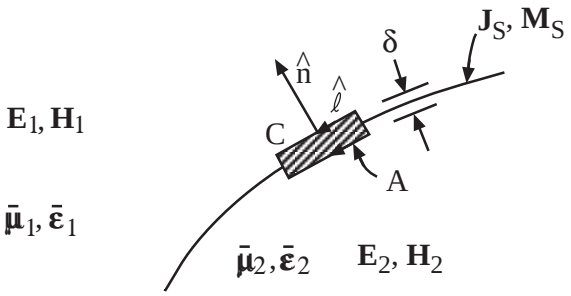


Figure 1.3: Boundary conditions at an interface.

conditions from them. To do this, we integrate (1.4.5) about a small area between the interface of two media. Invoking Stokes' theorem, we have

$$\oint_C d\mathbf{l} \cdot (\overline{\boldsymbol{\mu}}^{-1} \cdot \nabla \times \mathbf{E}) - \omega^2 \int_A d\mathbf{S} \cdot \overline{\boldsymbol{\epsilon}} \cdot \mathbf{E} = i\omega \int_A d\mathbf{S} \cdot \mathbf{J}. \tag{1.5.1}$$

Letting $\delta \to 0$, the surface integral on the left-hand side of the above equation vanishes. Assuming that we have a current sheet $\mathbf{J}_s$, we can show that

$$\hat{n} \times (\overline{\boldsymbol{\mu}}_1^{-1} \cdot \nabla \times \mathbf{E}_1) - \hat{n} \times (\overline{\boldsymbol{\mu}}_2^{-1} \cdot \nabla \times \mathbf{E}_2) = i\omega \mathbf{J}_s. \tag{1.5.2}$$

Since $\nabla \times \mathbf{E} = i\omega \overline{\boldsymbol{\mu}} \cdot \mathbf{H}$, we have

$$\hat{n} \times \mathbf{H}_1 - \hat{n} \times \mathbf{H}_2 = \mathbf{J}_s. \tag{1.5.3}$$

Performing the same analysis for Equation (1.4.6), we arrive at

$$\hat{n} \times \mathbf{E}_1 - \hat{n} \times \mathbf{E}_2 = 0. \tag{1.5.4}$$

Equations (1.5.3) and (1.5.4) are the important boundary conditions we will use over and again.

The boundary condition (1.5.3) can also be gleaned from (1.3.7). If $\mathbf{J}(\mathbf{r})$ is a current sheet at on interface, represented by a delta function singularity, then this singularity must be from the normal derivative of the tangential component of the magnetic field. From this fact we can derive (1.5.3). By the some token, (1.5.4) can be derived from (1.3.8).

1.6 Reciprocity Theorem

If we have two sources $\mathbf{J}_1$ and $\mathbf{J}_2$ radiating in an anisotropic, inhomogeneous medium, and $\mathbf{J}_1$ produces the field $\mathbf{E}_1$, $\mathbf{J}_2$ produces the field $\mathbf{E}_2$, the reciprocity theorem requires that for a reciprocal medium,

$$\langle \mathbf{E}_1, \mathbf{J}_2 \rangle = \langle \mathbf{E}_2, \mathbf{J}_1 \rangle, \tag{1.6.1}$$

where $\langle \mathbf{A}, \mathbf{B} \rangle$ stands for $\int d\mathbf{r} \mathbf{A} \cdot \mathbf{B}$. This theorem is derivable from Equation (1.4.5) with constraints on $\overline{\boldsymbol{\epsilon}}$ and $\overline{\boldsymbol{\mu}}$. When the source $\mathbf{J}_1$ is radiating, the field $\mathbf{E}_1$ satisfies the equation

$$\nabla \times \overline{\boldsymbol{\mu}}^{-1} \cdot \nabla \times \mathbf{E}_1 - \omega^2 \overline{\boldsymbol{\epsilon}} \cdot \mathbf{E}_1 = i\omega \mathbf{J}_1. \tag{1.6.2}$$

When $\mathbf{J}_2$ is radiating, the field $\mathbf{E}_2$ satisfies the equation

$$\nabla \times \overline{\boldsymbol{\mu}}^{-1} \cdot \nabla \times \mathbf{E}_2 - \omega^2 \overline{\boldsymbol{\epsilon}} \cdot \mathbf{E}_2 = i\omega \mathbf{J}_2, \tag{1.6.3}$$

where $\overline{\boldsymbol{\mu}}$ and $\overline{\boldsymbol{\epsilon}}$ in (1.6.2) and (1.6.3) represent the same medium. Dot-multiplying (1.6.2) by $\mathbf{E}_2$ and integrating, and (1.6.3) by $\mathbf{E}_1$ and integrating, we have

$$\langle \mathbf{E}_2, \nabla \times \overline{\boldsymbol{\mu}}^{-1} \cdot \nabla \times \mathbf{E}_1 \rangle - \omega^2 \langle \mathbf{E}_2, \overline{\boldsymbol{\epsilon}} \cdot \mathbf{E}_1 \rangle = i\omega \langle \mathbf{E}_2, \mathbf{J}_1 \rangle, \tag{1.6.4}$$

$$\langle \mathbf{E}_1, \nabla \times \overline{\boldsymbol{\mu}}^{-1} \cdot \nabla \times \mathbf{E}_2 \rangle - \omega^2 \langle \mathbf{E}_1, \overline{\boldsymbol{\epsilon}} \cdot \mathbf{E}_2 \rangle = i\omega \langle \mathbf{E}_1, \mathbf{J}_2 \rangle. \tag{1.6.5}$$

Since

$$\langle \mathbf{E}_2, \nabla \times \overline{\boldsymbol{\mu}}^{-1} \cdot \nabla \times \mathbf{E}_1 \rangle = \int_V d\mathbf{r} \mathbf{E}_2 \cdot \nabla \times \overline{\boldsymbol{\mu}}^{-1} \cdot \nabla \times \mathbf{E}_1, \tag{1.6.6}$$

we can use the identity

$$\nabla \cdot (\mathbf{A} \times \mathbf{B}) = \mathbf{B} \cdot \nabla \times \mathbf{A} - \mathbf{A} \cdot \nabla \times \mathbf{B} \tag{1.6.7}$$

and Gauss' theorem to get

$$\langle \mathbf{E}_2, \nabla \times \overline{\boldsymbol{\mu}}^{-1} \cdot \nabla \times \mathbf{E}_1 \rangle = \int_V d\mathbf{r} (\nabla \times \mathbf{E}_2) \cdot \overline{\boldsymbol{\mu}}^{-1} \cdot (\nabla \times \mathbf{E}_1) \tag{1.6.8}$$

$$+ \int_S dS \hat{n} \cdot (\overline{\boldsymbol{\mu}}^{-1} \cdot \nabla \times \mathbf{E}_1) \times \mathbf{E}_2$$

$$\int_V d\mathbf{r} (\nabla \times \mathbf{E}_2) \cdot \overline{\boldsymbol{\mu}}^{-1} \cdot (\nabla \times \mathbf{E}_1) \tag{1.6.9}$$

$$+ i\omega \int_S dS \hat{n} \cdot (\mathbf{H}_1) \times \mathbf{E}_2, \tag{1.6.10}$$

where V and S are a volume and a surface tending to infinity. When $S \to \infty$, $\overline{\boldsymbol{\mu}}$ becomes isotropic and homogeneous. Furthermore, the solutions to the vector wave equation become plane waves. Hence, $\nabla \to i\mathbf{k}$, and we have

$$(\overline{\boldsymbol{\mu}}^{-1} \cdot \nabla \times \mathbf{E}_1) \times \mathbf{E}_2|_{\mathbf{r}\epsilon s} = i\mu_0^{-1} (\mathbf{k} \times \mathbf{E}_1) \times \mathbf{E}_2 = -i\mu_0^{-1} \mathbf{k} (\mathbf{E}_1 \cdot \mathbf{E}_2). \tag{1.6.11}$$

where we have assumed that $\mathbf{k} \cdot \mathbf{E}_2 = 0$. In this manner, the surface integral in (1.6.10) is symmetric about $\mathbf{E}_1$ and $\mathbf{E}_2$. If $\overline{\boldsymbol{\mu}}^{-1}$ is symmetric, then the first integral on the right-hand side of (1.6.10) is also symmetric about $\mathbf{E}_1$ and $\mathbf{E}_2$. Hence, if $\overline{\boldsymbol{\mu}}^{-1}$ is symmetric, the first term of (1.6.4) and (1.6.5) are equal. If $\overline{\boldsymbol{\epsilon}}$ is also symmetric, then the second term of (1.6.4) and

(1.6.5) are also equal. Therefore, we deduce that (1.6.1) is satisfied or that reciprocity holds when

$$\overline{\boldsymbol{\mu}} = \overline{\boldsymbol{\mu}}^t, \qquad \overline{\boldsymbol{\epsilon}} = \overline{\boldsymbol{\epsilon}}^t. \tag{1.6.12}$$

In other words, $\overline{\boldsymbol{\mu}}$ and $\overline{\boldsymbol{\epsilon}}$ are symmetric (if $\overline{\boldsymbol{\mu}}$ is symmetric, $\overline{\boldsymbol{\mu}}^{-1}$ is symmetric). The condition expressed in Equation (1.6.12) is necessary for an anisotropic medium to be areciprocal medium. It also follows that all isotropic media are reciprocal.

The integral defined in Equation (1.6.1) is also known as a reaction. It could be thought of as a generalized measurement. In words, the reciprocity theorem states that for a reciprocal medium, *the* $\mathbf{E}$*-field due to* $\mathbf{J}_1$ *measured by* $\mathbf{J}_2$ *is the same as the* $\mathbf{E}$*-field due to* $\mathbf{J}_2$ *measured by* $\mathbf{J}_1$.

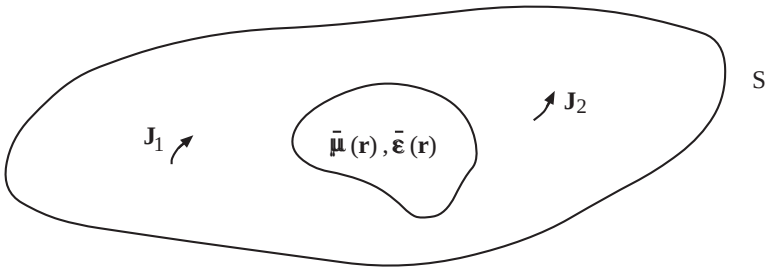


Figure 1.4: Proof of Reciprocity.

Examples of non-reciprocal media are plasma and ferrite media biased by a magnetic field. A medium can be lossy and still be reciprocal.

In electromagnetics, it is customary to add a fictitions magnetic current $\mathbf{M}$ to Faraday's law such that

$$\nabla \times \mathbf{E} = -i\omega\mathbf{D} - \mathbf{M} \tag{1.6.13}$$

A reciprocity theorem that can be derived involving magnetic current is

$$\langle \mathbf{E}_1, \mathbf{J}_2 \rangle - \langle \mathbf{H}_1, \mathbf{M}_2 \rangle = \langle \mathbf{E}_2, \mathbf{J}_1 \rangle - \langle \mathbf{H}_2, \mathbf{M}_1 \rangle \tag{1.6.14}$$

in replacement of (1.6.1).

Reciprocity theorem is deeply related to the symmetry of differential operators related to Maxwell's equations. For example, we can express (1.6.2) and (1.6.3) as

$$\mathcal{D}\mathbf{E}_1 = \mathbf{J}_1 \tag{1.6.15}$$

$$\mathcal{D}\mathbf{E}_2 = \mathbf{J}_2 \tag{1.6.16}$$

Where $\mathcal{D}$ is the pertinent differential operators. Then, $\langle \mathbf{E}_2, \mathbf{J}_1 \rangle = \langle \mathbf{E}_1, \mathbf{J}_2 \rangle$ implies

$$\langle \mathbf{E}_2, \mathcal{D}\mathbf{E}_1 \rangle = \langle \mathbf{E}_1, \mathcal{D}\mathbf{E}_2 \rangle \tag{1.6.17}$$

The above is the analogue of

$$\mathbf{a}^t \cdot \overline{\mathbf{A}} \cdot \mathbf{b} = \mathbf{b}^t \cdot \overline{\mathbf{A}} \cdot \mathbf{a} \tag{1.6.18}$$

which implies the $\overline{\mathbf{A}} = \overline{\mathbf{A}}^t$ or $\overline{\mathbf{A}}$ is symmetric. Hence, (1.6.17) implies that $\mathcal{D}$ is symmetric, and this is possible only if (1.6.12) is satisfied.

The symmetry of $\mathcal{D}$ is the deeper underlying reason for the reciprocity theorem. For media that are reciprocal, the symmetry of the electromagnetic equations will give rise to a number of operators that are also symmetrical such as the impedance and admittance matrices, as we shall learn later.

1.6.1 Lorentz Reciprocity Theorem

If the volume integrals in (1.6.4) and (1.6.5) are taken over a finite volume, then on subtracting the two equations, making use of (1.6.10), and assuming the symmetry of the permeability and permittivity tensors, we arrive at the general case of the reciprocity theorem,

$$\langle \mathbf{E}_2, \mathbf{J}_1 \rangle - \langle \mathbf{E}_1, \mathbf{J}_2 \rangle = \oint_S dS\hat{n} \cdot (\mathbf{E}_1 \times \mathbf{H}_2 - \mathbf{E}_2 \times \mathbf{H}_1) \tag{1.6.19}$$

When the volume does not enclose the sources, we arrive at

$$\oint_S dS\hat{n} \cdot (\mathbf{E}_1 \times \mathbf{H}_2) = \oint_S dS\hat{n} \cdot (\mathbf{E}_2 \times \mathbf{H}_1) \tag{1.6.20}$$

The above is generally known as the Lorentz reciprocity theorem. It is useful in waveguides when sources are not involved.

1.7 Energy Conservation

Energy conservation in electromagectics is defined by the Poynting theorem. Poynting theorem holds for the time domain as well as the frequency domain. The theorem in the time domain is actually quite different from that in the frequency domain. We shall present first the time domain version.

1.7.1 Time Domain Poynting Theorem

The time domain Poynting theorem, sometimes known as the real Poynting theorem, governs the conservation of instantaneous energy for electromagnetic field. To derive it, we start with

$$\begin{aligned} \nabla \cdot [\mathbf{E}(\mathbf{r}, t) \times \mathbf{H}(\mathbf{r}, t)] &= \mathbf{H} \cdot \nabla \times \mathbf{E} - \mathbf{E} \cdot \nabla \times \mathbf{H} \\ &= -\mathbf{H} \cdot \frac{\partial \mathbf{B}}{\partial t} - \mathbf{E} \cdot \frac{\partial \mathbf{D}}{\partial t} - \mathbf{E} \cdot \mathbf{J} \end{aligned} \tag{1.7.1}$$

Defining the Poynting vector

$$\mathbf{S}(\mathbf{r}, t) = \mathbf{E}(\mathbf{r}, t) \times \mathbf{H}(\mathbf{r}, t) \tag{1.7.2}$$

we have

$$\nabla \cdot \mathbf{S}(\mathbf{r}, t) = -\left(\mathbf{H} \cdot \frac{\partial \mathbf{B}}{\partial t} + \mathbf{E} \cdot \frac{\partial \mathbf{D}}{\partial t} \right) - \mathbf{E} \cdot \mathbf{J} \tag{1.7.3}$$

For free space where $\mathbf{B} = \mu_0\mathbf{H}$, $\mathbf{D} = \epsilon_0\mathbf{E}$, we can show that

$$\mathbf{H} \cdot \frac{\partial \mathbf{B}}{\partial t} = \mathbf{H} \cdot \mu_0 \frac{\partial \mathbf{H}}{\partial t} = \frac{1}{2}\mu_0 \frac{\partial}{\partial t}\mathbf{H} \cdot \mathbf{H} \tag{1.7.4}$$

and similarly, for the electric flux term, we have

$$\nabla \cdot \mathbf{S}(\mathbf{r}, t) = -\frac{\partial}{\partial t}\frac{1}{2}(\mu_0\mathbf{H} \cdot \mathbf{H} + \epsilon_0\mathbf{E} \cdot \mathbf{E}) - \mathbf{E} \cdot \mathbf{J} \tag{1.7.5}$$

Then term

$$W_T = \frac{1}{2}(\mu_0\mathbf{H} \cdot \mathbf{H} + \epsilon_0\mathbf{E} \cdot \mathbf{E}) \tag{1.7.6}$$

corresponds to the total energy stored in the magnetic field and the electric field. When $\frac{\partial}{\partial t}W_T$ is positive, it corresponds or contributing the negative term to $\nabla \cdot \mathbf{S}$ implying a influx of power at a point. The last term $\mathbf{E} \cdot \mathbf{J}$ corresponds to power absorbed or generated by the current $\mathbf{J}$. When $\mathbf{E} \cdot \mathbf{J}$ is positive, the current $\mathbf{J}$ is absorptive. This is true of a conductive medium where $\mathbf{J} = \sigma\mathbf{E}$.

Please note that the above derivation that leads to expression (1.7.6) is not valid for material media. All material media have to be frequency dispersive, and hence, in the time domain, the constitutive relations are denoted by time convolutions. As a curious fact, the above can be generalized to inhomogeneous, anisotropic, reciprocal media.

1.7.2 Frequency Domain Poynting Theorem

The frequency domain Poynting theorem governs energy conservation for complex power. Hence, it is also known as the complex Poynting theorem. We start with

$$\begin{aligned} \nabla \cdot (\mathbf{E} \times \mathbf{H}^*) &= i\omega\mathbf{H}^* \cdot \mathbf{B} - i\omega\mathbf{E} \cdot \mathbf{D}^* - \mathbf{E} \cdot \mathbf{J}^* \\ &= i\omega\mathbf{H}^* \cdot \overline{\mu} \cdot \mathbf{H} - i\omega\mathbf{E} \cdot \overline{\epsilon}^* \cdot \mathbf{E}^* - \mathbf{E} \cdot \mathbf{J}^* \end{aligned} \tag{1.7.7}$$

Defining the complex Poynting vector

$$\tilde{\mathbf{S}}(\mathbf{r}) = \mathbf{E}(\mathbf{r}) \times \mathbf{H}^*(\mathbf{r}) \tag{1.7.8}$$

the above becomes

$$\nabla \cdot \tilde{\mathbf{S}} = i\omega(\mathbf{H}^* \cdot \overline{\mu} \cdot \mathbf{H} - \mathbf{E} \cdot \overline{\epsilon}^* \cdot \mathbf{E}^*) - \mathbf{E} \cdot \mathbf{J}^* \tag{1.7.9}$$

For a source-free region, this becomes

$$\nabla \cdot \tilde{\mathbf{S}} = i\omega(\mathbf{H}^* \cdot \overline{\mu} \cdot \mathbf{H} - \mathbf{E} \cdot \overline{\epsilon} \cdot \mathbf{E}^*) \tag{1.7.10}$$

If

$$\mathbf{H}^* \cdot \overline{\mu} \cdot \mathbf{H} = \mathbf{E} \cdot \overline{\epsilon} \cdot \mathbf{E}^* \tag{1.7.11}$$

in a region, then the right-hand side is zero, and there is no net power flux into or out of the region. This occurs at resonance in a cavity.

1.7.3 Complex Power

The complex Poynting theorem is quite different from the real Poynting theorem. It can be shown that half the real part of the complex Poynting vector is the time average of the instantaneous Poynting vector, viz., [see problem 1.2]

$$\langle \mathbf{S}(\mathbf{r},t)\rangle = \frac{1}{2}\Re e\left[\tilde{\mathbf{S}}(\mathbf{r})\right] \tag{1.7.12}$$

The part that corresponds to the stored energy in the complex Poynting theorem is the difference of the magnetic energy and electric energy stored, whereas that in the real Poynting theorem is the sum of the two. This is because the imaginary part of the complex power is reactive power [see problem 1.3]. Reactive power in a time harmonic system corresponds to power that flows into a system, and later flows out of a system. Hence, its time average is zero.

Notice that in a resonance system or circuit such as the LC tank circuit, the stored magnetic energy and electric energy are equal to each other, and they exchange with each other. Since the reactive power is the difference in the store magnetic and electric energy, it is zero in this case. Therefore, when an LC tank circuit is at resonance, there is no need for an external supply of reactive power.

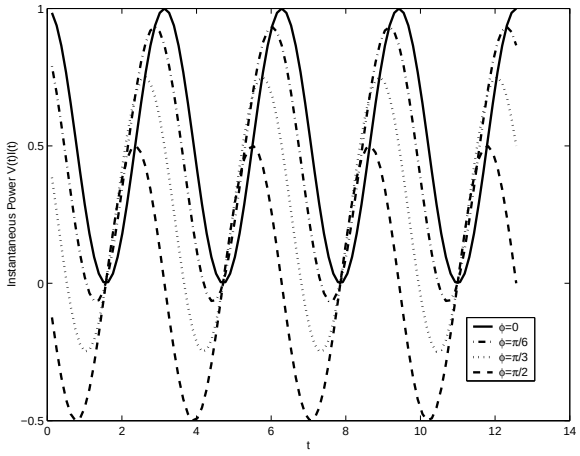


Figure 1.5: Plots of $P(t) = V(t)I(t)$ versus time t where $V(t) = \cos(t)$, and $I(t) = \cos(t+\phi)$ for various values of ϕ.

Even though no net power is delivered in the reactive power, a power utility company will still charge its customers for the use of this power for two reasons: First, not all reactive power is retrievable as it has to be sent over power lines that have conductive losses. Second, the power company has to maintain a generator that can absorb the oscillation in the total power caused by the presence of reactive power. Figure 1.5 shows that the instantaneous power can be negative as well as positive when there is a phase shift between the voltage and the current in a circuit.

1.7.4 Lossless Conditions

For an isotropic medium, the conditions for it to be lossless are that $\Im m(\mu) = 0$, and $\Im m(\epsilon) = 0$ where "$\Im m$" implies "imaginary part." However, the condition for an anisotropic medium is quite different. We can derive the general lossless condition from energy conservation.

For a lossless medium, for energy conservation, and from the complex Poynting theorem, we require that

$$\Re e \int_V dV \nabla \cdot (\mathbf{E} \times \mathbf{H}^*) = \Re e \oint_S d\mathbf{S} \cdot (\mathbf{E} \times \mathbf{H}^*) = 0, \tag{1.7.13}$$

since $\Re e[\mathbf{E} \times \mathbf{H}^*]$ corresponds to time average power flow. The above implies that

$$\Re e \left\{ i\omega \int_V dV (\mathbf{H}^* \cdot \overline{\boldsymbol{\mu}} \cdot \mathbf{H} - \mathbf{E} \cdot \overline{\boldsymbol{\epsilon}}^* \cdot \mathbf{E}^*) \right\} = 0. \tag{1.7.14}$$

A sufficient condition for arbitrary V is to require that $\mathbf{H}^* \cdot \overline{\boldsymbol{\mu}} \cdot \mathbf{H}$ and $\mathbf{E} \cdot \overline{\boldsymbol{\epsilon}}^* \cdot \mathbf{E}^*$ to be purely real or their conjugates to be themselves, i.e.,

$$(\mathbf{H}^* \cdot \overline{\boldsymbol{\mu}} \cdot \mathbf{H})^* = \mathbf{H} \cdot \overline{\boldsymbol{\mu}}^* \cdot \mathbf{H}^* = \mathbf{H}^* \cdot \overline{\boldsymbol{\mu}}^\dagger \cdot \mathbf{H} = \mathbf{H}^* \cdot \overline{\boldsymbol{\mu}} \cdot \mathbf{H}. \tag{1.7.15}$$

Therefore, $\overline{\boldsymbol{\mu}}^\dagger = \overline{\boldsymbol{\mu}}$. Similarly the condition on $\mathbf{E} \cdot \overline{\boldsymbol{\epsilon}}^* \cdot \mathbf{E}^*$ to be purely real is $\overline{\boldsymbol{\epsilon}}^\dagger = \overline{\boldsymbol{\epsilon}}$. Consequently, the lossless conditions for an anisotropic medium is

$$\overline{\boldsymbol{\epsilon}} = \overline{\boldsymbol{\epsilon}}^\dagger, \qquad \overline{\mu} = \overline{\boldsymbol{\mu}}^\dagger. \tag{1.7.16}$$

In other words, the permittivity tensor and the permeability tensor have to be Hermitian.

1.8 Energy Density in Dispersive Medium

In the following derivation, we assume that $\mathbf{E}$, $\mathbf{H}$, and $\mathbf{J}$ have $e^{-i\omega t}$ time dependence, where ω is a complex frequency [30]. Then

$$\begin{aligned}
\nabla \cdot [\mathbf{E}(t) \times \mathbf{H}^*(t)] &= \mathbf{H}^*(t) \cdot \nabla \times \mathbf{E}(t) - \mathbf{E}(t) \cdot \nabla \times \mathbf{H}^*(t) \\
&= \mathbf{H}^*(t) \cdot [i\omega \overline{\mu} \cdot \mathbf{H}(t)] - \mathbf{E}(t) \cdot [i\omega^* \overline{\epsilon}^* \cdot \mathbf{E}^*(t) + \mathbf{J}^*(t)] \\
&= i\omega \mathbf{H}^*(t) \cdot \overline{\mu} \cdot \mathbf{H}(t) - i\omega^* \mathbf{E}(t) \cdot \overline{\epsilon}^* \cdot \mathbf{E}^*(t) - \mathbf{E}(t) \cdot \mathbf{J}^*(t)
\end{aligned} \tag{1.8.1}$$

Next, we let $\omega = \omega' + i\omega''$ where ω' and ω'' are real numbers. Then

$$\begin{aligned}
\nabla \cdot [\mathbf{E}(t) \times \mathbf{H}^*(t)] &= i(\omega' + i\omega'') \mathbf{H}^*(t) \cdot \overline{\mu}(\omega' + i\omega'') \cdot \mathbf{H}(t) \\
&\quad - i(\omega' - i\omega'') \mathbf{E}(t) \cdot \overline{\epsilon}^*(\omega' + i\omega'') \cdot \mathbf{E}^*(t) - \mathbf{E}(t) \cdot \mathbf{J}^*(t)
\end{aligned} \tag{1.8.2}$$

Ordinarily, if ω is pure real, the time dependence would have canceled in the above, but because ω is complex, each of the above terms has time dependence of $\exp(2\omega'' t)$. Assuming that $\omega'' \ll \omega'$, we can Taylor expand the right-hand side to get

$$\begin{aligned}
\nabla \cdot (\mathbf{E} \times \mathbf{H}^*) \doteq\ & i(\omega' + i\omega'') \mathbf{H}^* \cdot \left[\overline{\mu}(\omega') + i\omega'' \frac{\partial}{\partial \omega'} \overline{\mu}(\omega') \right] \cdot \mathbf{H} \\
& - i(\omega' - i\omega'') \mathbf{E} \cdot \left[\overline{\epsilon}^*(\omega') - i\omega'' \frac{\partial}{\partial \omega'} \overline{\epsilon}^*(\omega') \right] \cdot \mathbf{E}^* - \mathbf{E} \cdot \mathbf{J}^*
\end{aligned} \tag{1.8.3}$$

Collecting leading order and first order terms, we have

$$\begin{aligned}\nabla\cdot(\mathbf{E}\times\mathbf{H}^*) &\doteq i\omega'[\mathbf{H}^*\cdot\overline{\mu}(\omega')\cdot\mathbf{H}-\mathbf{E}\cdot\overline{\epsilon}^*(\omega')\cdot\mathbf{E}^*]\\ &\quad-\omega''\left\{\mathbf{H}^*\cdot\left[\overline{\mu}(\omega')+\omega'\frac{\partial}{\partial\omega'}\overline{\mu}(\omega')\right]\cdot\mathbf{H}\right.\\ &\quad\left.+\,\mathbf{E}\cdot\left[\overline{\epsilon}^*(\omega')+\omega'\frac{\partial}{\partial\omega'}\overline{\epsilon}^*(\omega')\right]\cdot\mathbf{E}\right\}-\mathbf{E}\cdot\mathbf{J}^*\\ &= i\omega'[\mathbf{H}^*\cdot\overline{\mu}(\omega')\cdot\mathbf{H}-\mathbf{E}\cdot\overline{\epsilon}^*(\omega')\cdot\mathbf{E}^*]\\ &\quad-\omega''\left\{\mathbf{H}^*\cdot\frac{\partial}{\partial\omega'}[\omega'\overline{\mu}(\omega')]\cdot\mathbf{H}+\mathbf{E}\cdot\frac{\partial}{\partial\omega'}[\omega'\overline{\epsilon}^*(\omega')]\cdot\mathbf{E}\right\}-\mathbf{E}\cdot\mathbf{J}^*\end{aligned}\tag{1.8.4}$$

For lossless media, the first term is purely imaginary, while the second term is purely real. The first term corresponds to reactive power. The second term comes about because for a complex exponential of the form $e^{-i(\omega'+i\omega'')t}=e^{-i\omega'+\omega''t}$, the field strength is growing with $e^{2\omega''t}$ time dependence. If we take the real part of (1.8.4), and focussing on the part of space where $\mathbf{J}=0$, we have

$$\begin{aligned}\nabla\cdot\frac{1}{2}\Re e(\mathbf{E}\times\mathbf{H}^*) &= -\frac{1}{2}\omega''\left\{\mathbf{H}^*\cdot\frac{\partial}{\partial\omega'}\omega'\overline{\mu}(\omega')\cdot\mathbf{H}+\mathbf{E}\cdot\frac{\partial}{\partial\omega'}\omega'\overline{\epsilon}^*(\omega')\cdot\mathbf{E}^*\right\}\\ &= -\frac{\partial}{\partial t}W_T\end{aligned}\tag{1.8.5}$$

The above has the physical meaning that the divergence of the time-average real power flow on the left-hand side is due to the time variation of the energy density on the right-hand side. The energy density has a time dependence of $e^{2\omega''t}$. Consequently, we identify the energy density for dispersive media as

$$W_T=\frac{1}{4}\left\{\mathbf{H}^*\cdot\frac{\partial}{\partial\omega'}\omega'\overline{\mu}(\omega')\cdot\mathbf{H}+\mathbf{E}\cdot\frac{\partial}{\partial\omega'}\omega'\overline{\epsilon}^*(\omega')\cdot\mathbf{E}^*\right\}\tag{1.8.6}$$

When the medium is free space, we have

$$W_T=\frac{1}{4}\left\{\mathbf{H}^*\cdot\mu_0\mathbf{H}+\mathbf{E}\cdot\epsilon_0\mathbf{E}^*\right\}\tag{1.8.7}$$

which agrees with what we have derived from time-domain Poynting theorem.

1.9 Symmetries in Electromagnetics

Symmetries play an important role in the solutions of Maxwell's equations. They can be used for simplifying solutions to Maxwell's equations, or they can be used to derive new solutions. Alternatively, they can be used to predict how solutions behave once symmetry is broken.

In solid state physics, symmetry is used to understand the propagation of electronic waves in crystalline structures which have a high degree of symmetry. Due to the symmetries, group theory can be used to analyze the physical characteristics of waves in crystalline structures

[20, 21]. Unlike electronic waves, electromagnetic waves, for most applications, propagate in in nonsymmetric structures. Hence, the exploitation of symmetry in electromagnetics has not reached the level in solid state physics. However, there are still a few symmetries we can exploit, especially in waveguides and resonators which usually have some degrees of symmetry associated with them.

Some of the obvious symmetries are translational symmetry and rotational symmetry. Translational symmetry exploits the fact that Maxwell's equations are invariant after a translation in space. Rotational symmetry implies that a solution of Maxwell's equations remains a solution after rotation. Other symmetries are time-reversal symmetry and reflection symmetry that we shall discuss next.[1]

1.9.1 Time Reversal Symmetry

In a lossless environment, solutions to Maxwell's equations are time reversible in the same medium. That is if a solution is found, and if we change t to $-t$, the solution remains a valid solution to Maxwell's equations within the same medium. This is like playing a movie backward. However, the right-hand rule becomes the left-hand rule in the movie playback.

It is clear that the solution to the wave equation is time reversible. However, when we have a lossy wave equation, the solution decays forward in time, but grows backward in time. Therefore, the solution is not time reversible, namely, the reverse-time solution is a solution to an active medium (amplifying medium like a laser cavity) but the forward-time solution corresponds to a lossy medium. For the same reason, solutions to Maxwell's equations are not time reversible in a lossy medium.

To obtain a time-reversed field, we let $t \to -t$. For example, a time-reversed $\mathbf{B}$ field is $\mathbf{B}(\mathbf{r}, -t)$. Then the time derivative of this time-reversed $\mathbf{B}$ field is

$$\frac{\partial}{\partial t}\mathbf{B}(\mathbf{r}, -t) = -\frac{\partial}{\partial t'}\mathbf{B}(\mathbf{r}, t') \tag{1.9.1}$$

which is the negative of the original time derivative. Consequently, when time-reversed fields are substituted back into Maxwell's equations, they can be written as

$$\nabla \times \mathbf{E}(\mathbf{r}, t') = \frac{\partial}{\partial t'}\mathbf{B}(\mathbf{r}, t'), \tag{1.9.2}$$

$$\nabla \times \mathbf{H}(\mathbf{r}, t') = -\frac{\partial}{\partial t'}\mathbf{D}(\mathbf{r}, t') + \mathbf{J}(\mathbf{r}, t'), \tag{1.9.3}$$

$$\nabla \cdot \mathbf{B}(\mathbf{r}, t') = 0, \tag{1.9.4}$$

$$\nabla \cdot \mathbf{D}(\mathbf{r}, t') = \rho(\mathbf{r}, t'). \tag{1.9.5}$$

We will retrieve the original Maxwell's equations if the signs of $\mathbf{H}$, $\mathbf{B}$, and $\mathbf{J}$ are reversed. The need to reverse these quantities is also necessary for energy conservation in Poynting theorem. The sign of $\mathbf{E} \times \mathbf{H}$ and has to change for time-reversed solution to reflect that the energy flow has to change direction. Note that the sources $\mathbf{J}$ and ρ are also time reversed.

[1] Some of these symmetries have been used successfully in computational electromagnetics to expedite numerical solutions of Maxwell's equations [22]. See also discussions in [23, p. 268].

Alternatively, we can change the sign of $\mathbf{E}$, $\mathbf{D}$, and ρ. But the convention is to change the signs of $\mathbf{H}$, $\mathbf{B}$, and $\mathbf{J}$. A positive charge, when moving through space, remains a positive charge when time-reversed. However, it produces a current of opposite polarity.

Since time always occurs as $\exp(-i\omega t)$ in the frequency domain, replacing t with $-t$ is the same as replacing i with $-i$. Hence, a time reversed field is obtained by conjugating the frequency-domain field. If a time-harmonic field is represented by its phasor, then the conjugate of the phasor represents a time-reversed solution as it can be easily shown that if

$$\mathbf{E}(\mathbf{r},t) = \Re e\left[\mathbf{E}(\mathbf{r},\omega)e^{-i\omega t}\right] \tag{1.9.6}$$

then

$$\Re e\left[\mathbf{E}^*(\mathbf{r},\omega)e^{-i\omega t}\right] = \Re e\left[\mathbf{E}(\mathbf{r},\omega)e^{i\omega t}\right] = \mathbf{E}(\mathbf{r},-t) \tag{1.9.7}$$

The design of phase-conjugate mirror was in vogue in optics to create a time-reversed optical field [31–33].

1.9.2 Reflection Symmetry

It was believed once that all laws of physics can be replicated in the mirror world, namely, laws of physics remain the same under reflection. This is known as the conservation of parity. However, it is now known that some laws of physics do not satisfy parity conservation [24]. However, the law of electromagnetics satisfies parity conservation. We just need to replace a right-hand rule with a left-hand rule for the reflected solution, namely, the solution in the mirror world.

A symmetry closely related to reflection symmetry is inversion symmetry [21]. In inversion, we let $\mathbf{r} \to -\mathbf{r}$, or in detail, $x \to -x$, $y \to -y$, and $z \to -z$. A reflected function or object can always be obtained from an inverted function or object by a rotation. For instance, if we have a mirror in the xy plane, and we put an object in front of the mirror, the reflected object will have $z \to -z$ with its xy coordinates unchange. However, this can also be obtained by first inverting the object, followed by a 180 degree rotation about the z axis. Since a rotation of a solution is still a solution to Maxwell's equations, we will just discuss what inversion does to a solution.

When we have a vector field such as $\mathbf{E}(\mathbf{r},t)$, we assume that the direction of the field also change after inversion by replacing $\hat{x} \to -\hat{x}$, $\hat{y} \to -\hat{y}$, and $\hat{z} \to -\hat{z}$. Hence, a vector field under inversion becomes $-\mathbf{E}(-\mathbf{r},t)$. If we take the curl of this inverted field, we have

$$\nabla \times [-\mathbf{E}(-\mathbf{r},t)] = \nabla' \times \mathbf{E}(\mathbf{r}',t) \tag{1.9.8}$$

after we let $\mathbf{r}' = -\mathbf{r}$, and under this change of variables, $\nabla = -\nabla'$. We can substitute these inverted fields into Maxwell's equations to see if these equations retain their original forms.

Consequently, Maxwell's equations under substitution of inverted fields, and with a change of variables $\mathbf{r}' = -\mathbf{r}$, become

$$\nabla' \times \mathbf{E}(\mathbf{r}',t) = \frac{\partial}{\partial t}\mathbf{B}(\mathbf{r}',t), \tag{1.9.9}$$

$$\nabla' \times \mathbf{H}(\mathbf{r}',t) = -\frac{\partial}{\partial t}\mathbf{D}(\mathbf{r}',t) - \mathbf{J}(\mathbf{r}',t), \tag{1.9.10}$$

$$\nabla' \cdot \mathbf{B}(\mathbf{r}', t) = 0, \tag{1.9.11}$$

$$\nabla' \cdot \mathbf{D}(\mathbf{r}', t) = \rho(\mathbf{r}', t). \tag{1.9.12}$$

However, the above is not the original Maxwell's equations. The original Maxwell's equations can be retrieved if we can change the signs of $\mathbf{B}$ and $\mathbf{H}$. This is understandable, since the right-hand rule becomes a left-hand rule in the mirror or reflected world (since the reflected world is related to the inverted world by just a rotation). Hence, a change of the signs of $\mathbf{B}$ and $\mathbf{H}$ will convert the left-hand rule back to the right-hand rule.

1.9.3 Polar Vectors and Pseudovectors

A word is in order about polar vectors versus pseudovectors (also known as axial vectors) [23]. A polar vector (or vector) changes sign under inversion, but a pseudovector does not. For instance, if $\mathbf{A}$ and $\mathbf{B}$ are polar vectors, they will change sign under inversion. However, a vector $\mathbf{C} = \mathbf{A} \times \mathbf{B}$ will not change sign under inversion if the cross product is defined with with the right-hand rule in the original right-handed coordinate system. It will change sign if the left-hand rule is used. Hence, $\mathbf{C}$ is a pseudovector because it does not follow the sign-change rule of the polar vectors under inversion. In electromagnetics, we can regard $\mathbf{H}$ and $\mathbf{B}$ as pseudovectors that do not change sign under inversion. In this case, Maxwell's equations are invariant under inversion or reflection.

By the same token, pseudoscalars exist. The scalar $\mathbf{a} \cdot \mathbf{b} \times \mathbf{c}$, where $\mathbf{a}$, $\mathbf{b}$, and $\mathbf{c}$ are polar vectors, is a pseudoscalar which changes sign under inversion.

If we define $\mathbf{B}$ and $\mathbf{H}$ to be pseudovectors instead, and they do not change sign under inversion, then we do not have to change the sign of $\mathbf{B}$ and $\mathbf{H}$.

1.10 Green's Function

The Green's function to a wave equation is the solution when the source is a point source [25, 26]. When we know the solution to the wave equation due to a point source, the solution due to a general source can be obtained by the principle of linear superposition. This is because the wave equation is linear, and a general source could be thought of as a superposition of point sources.

For example, if we need to find the solution to the following equation,

$$(\nabla^2 + k^2)\psi(\mathbf{r}) = S(\mathbf{r}), \tag{1.10.1}$$

we can first find the Green's function which is the solution to the following equation,

$$(\nabla^2 + k^2)g(\mathbf{r} - \mathbf{r}') = -\delta(\mathbf{r} - \mathbf{r}'). \tag{1.10.2}$$

If we know $g(\mathbf{r} - \mathbf{r}')$, $\psi(\mathbf{r})$ can be found formally. Multiplying (1.10.1) by $g(\mathbf{r} - \mathbf{r}')$ and (1.10.2) by $\psi(\mathbf{r})$, and integrating over volume, and subtracting, we obtain

$$\int_V d\mathbf{r}[g(\mathbf{r} - \mathbf{r}')\nabla^2\psi(\mathbf{r}) - \psi(\mathbf{r})\nabla^2 g(\mathbf{r} - \mathbf{r}')] = \int_V g(\mathbf{r} - \mathbf{r}')S(\mathbf{r})d\mathbf{r} + \psi(\mathbf{r}'). \tag{1.10.3}$$

Noting that $g\nabla^2\psi - \psi\nabla^2 g = \nabla\cdot(g\nabla\psi - \psi\nabla g)$, we can rewrite the left-hand side, using Gauss' divergence theorem as

$$\oint_S d\mathbf{S}\cdot(g\nabla\psi - \psi\nabla g) = \int_V g(\mathbf{r}-\mathbf{r}')S(\mathbf{r})d\mathbf{r} + \psi(\mathbf{r}'). \tag{1.10.4}$$

When $S \to \infty$, all fields look like plane waves, and we can replace $\nabla \to i\mathbf{k}$. The left-hand side of (1.10.4) then vanishes, and we have

$$\psi(\mathbf{r}) = -\int_V d\mathbf{r}' g(\mathbf{r}-\mathbf{r}')S(\mathbf{r}'). \tag{1.10.5}$$

Hence the solution of (1.10.1) can be written as an integral superposition of the solution of (1.10.2). In fact, we can invoke the principle of linear superposition to arrive at the above as well.

To find the solution of Equation (1.10.2), we solve it in spherical coordinates with the origin at $\mathbf{r}'$. Then, it becomes

$$(\nabla^2 + k^2)g(\mathbf{r}) = -\delta(\mathbf{r}) = -\delta(x)\delta(y)\delta(z). \tag{1.10.6}$$

For $r \neq 0$, the homogeneous, spherically symmetric solution to (1.10.6) is

$$g(\mathbf{r}) = C\frac{e^{+ikr}}{r} + D\frac{e^{-ikr}}{r}. \tag{1.10.7}$$

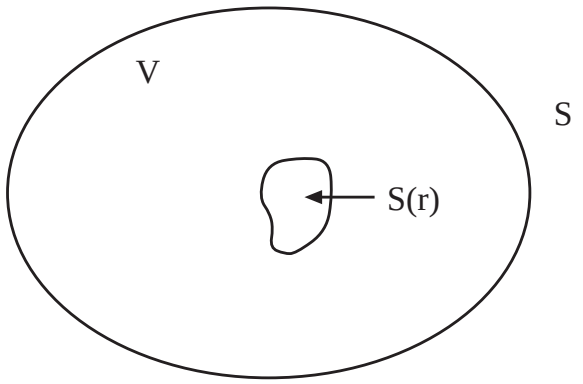


Figure 1.6: The radiation of a source $S(\mathbf{r})$ in a volume V.

Physical grounds require that we have only outgoing solutions; hence,

$$g(\mathbf{r}) = C\frac{e^{ikr}}{r}. \tag{1.10.8}$$

We can match the constant C to the singularity at the origin by substituting (1.10.8) into (1.10.6), and integrating Equation (1.10.6) over a small volume about the origin.

$$\int_{\Delta V} dV \nabla \cdot \nabla \frac{Ce^{ikr}}{r} + \int_{\Delta V} dV k^2 \frac{Ce^{ikr}}{r} = -1. \tag{1.10.9}$$

The second integral vanishes when $\Delta V \to 0$, because $dV = 4\pi r^2 dr$. We can convert the first integral in (1.10.9) into a surface integral using Gauss' theorem, and obtain

$$\lim_{r\to 0} 4\pi r^2 \frac{\partial}{\partial r} C \frac{e^{ikr}}{r} = -1, \tag{1.10.10}$$

or that $C = 1/4\pi$. Therefore, in general

$$g(\mathbf{r} - \mathbf{r}) = \frac{e^{ik|\mathbf{r}-\mathbf{r}'|}}{4\pi|\mathbf{r} - \mathbf{r}'|} \tag{1.10.11}$$

The solution to (1.10.1), from Equation (1.10.5) is then

$$\psi(\mathbf{r}) = -\int_V d\mathbf{r}' \frac{e^{ik|\mathbf{r}-\mathbf{r}'|}}{4\pi|\mathbf{r} - \mathbf{r}'|} S(\mathbf{r}'). \tag{1.10.12}$$

Equation (1.10.12) is a convolutional integral, a consequence of the principle of linear superposition. The above Green's function is the one that satisfies the radiation condition. Hence, the linearly superposed solution also satisfies the radiation condition.

For the vector wave equation in a homogeneous, isotropic medium, the equation is

$$\nabla \times \nabla \times \mathbf{E}(\mathbf{r}) - k^2 \mathbf{E}(\mathbf{r}) = i\omega\mu \mathbf{J}(\mathbf{r}). \tag{1.10.13}$$

By using the fact that $\nabla \times \nabla \times \mathbf{E} = -\nabla^2 \mathbf{E} + \nabla\nabla \cdot \mathbf{E}$, and that $\nabla \cdot \mathbf{E} = \frac{1}{\epsilon}\rho = \frac{1}{i\omega\epsilon}\nabla \cdot \mathbf{J}$, we can rewrite (1.10.13) as

$$\nabla^2 \mathbf{E}(\mathbf{r}) + k^2 \mathbf{E}(\mathbf{r}) = -i\omega\mu \left(\bar{\mathbf{I}} + \frac{\nabla\nabla}{k^2}\right) \cdot \mathbf{J}(\mathbf{r}). \tag{1.10.14}$$

There are three scalar wave equations embedded in the above equation. We can solve each of them in the manner of Equation (1.10.5), and we have

$$\mathbf{E}(\mathbf{r}) = i\omega\mu \int_V d\mathbf{r}' g(\mathbf{r}' - \mathbf{r}) \left(\bar{\mathbf{I}} + \frac{\nabla'\nabla'}{k^2}\right) \cdot \mathbf{J}(\mathbf{r}'). \tag{1.10.15}$$

It can be shown that

$$\int_V d\mathbf{r}' g(\mathbf{r} - \mathbf{r}') \nabla' f(\mathbf{r}') = \nabla \int_V d\mathbf{r}' g(\mathbf{r} - \mathbf{r}') f(\mathbf{r}'), \tag{1.10.16}$$

$$\int_V d\mathbf{r}' g(\mathbf{r}-\mathbf{r}')\nabla'\cdot\mathbf{F}(\mathbf{r}') = \nabla\cdot\int_V d\mathbf{r}' g(\mathbf{r}-\mathbf{r}')\mathbf{F}(\mathbf{r}'), \tag{1.10.17}$$

by using the vector identities $\nabla gf = f\nabla g + g\nabla f$, $\nabla\cdot g\mathbf{F} = g\nabla\cdot\mathbf{F} + (\nabla g)\cdot\mathbf{F}$, and that $\nabla' g(\mathbf{r}-\mathbf{r}') = -\nabla g(\mathbf{r}-\mathbf{r}')$. Hence, Equation (1.10.15) can be rewritten as

$$\mathbf{E}(\mathbf{r}) = i\omega\mu\left(\bar{\mathbf{I}} + \frac{\nabla\nabla}{k^2}\right)\cdot\int_V d\mathbf{r}' g(\mathbf{r}-\mathbf{r}')\mathbf{J}(\mathbf{r}'). \tag{1.10.18}$$

Sometimes, Equation (1.10.18) is written as

$$\mathbf{E}(\mathbf{r}) = i\omega\mu\int_V d\mathbf{r}'\overline{\mathbf{G}}(\mathbf{r},\mathbf{r}')\cdot\mathbf{J}(\mathbf{r}'), \tag{1.10.19}$$

where

$$\overline{\mathbf{G}}(\mathbf{r},\mathbf{r}') = \left(\bar{\mathbf{I}} + \frac{\nabla\nabla}{k^2}\right) g(\mathbf{r}-\mathbf{r}'), \tag{1.10.20}$$

is a dyad known as the dyadic Green's function. It has to be used with caution, since Equation (1.10.19), with the $\nabla\nabla$ operator inside the integration, has to be clarified since it does not converge uniformly when $\mathbf{r}$ is also in the source region occupied by $\mathbf{J}(\mathbf{r})$. Hence, it is only a convenient notation when the observation point is outside the source region.

1.11 Uniqueness Theorem

The uniqueness theorem provides conditions under which the solution to the wave equation is unique. This is especially important because the solutions to a problem should not be indeterminate. These conditions under which a solution to a wave equation is unique are the boundary conditions and the radiation condition. Uniqueness also allows one to construct solutions by inspections; if a candidate solution satisfies the conditions of uniqueness, it is the unique solution. Because of its simplicity, the scalar wave equation shall be examined first for easier insight into this problem.

1.11.1 Scalar Wave Equation

Given a scalar wave equation with a source term on the right-hand side, we shall derive the conditions under which a solution is unique. First, assume that there are two different solutions to the scalar wave equation, namely,

$$[\nabla^2 + k^2(\mathbf{r})]\,\phi_1(\mathbf{r}) = s(\mathbf{r}), \tag{1.11.1}$$

$$[\nabla^2 + k^2(\mathbf{r})]\,\phi_2(\mathbf{r}) = s(\mathbf{r}), \tag{1.11.2}$$

where $k^2(\mathbf{r})$ includes inhomogeneities of finite extent. Then, on subtracting the two equations, we have

$$[\nabla^2 + k^2(\mathbf{r})]\,\delta\phi(\mathbf{r}) = 0, \tag{1.11.3}$$

where $\delta\phi(\mathbf{r}) = \phi_1(\mathbf{r}) - \phi_2(\mathbf{r})$. Note that the solution is unique if and only if $\delta\phi = 0$ for all $\mathbf{r}$.

Then, after multiplying (1.11.3) by $\delta\phi^*$, integrating over volume, and using the vector identity $\nabla \cdot \psi\mathbf{A} = \mathbf{A} \cdot \nabla\psi + \psi\nabla \cdot \mathbf{A}$, we have

$$\int_S \hat{n} \cdot (\delta\phi^* \nabla\delta\phi)\, dS - \int_V |\nabla\delta\phi|^2 dV + \int_V k^2 |\delta\phi|^2 dV = 0, \tag{1.11.4}$$

where $\hat{n}$ is a unit normal to the surface S. Then, the imaginary part of the above equation is

$$\Im m \int_S \hat{n} \cdot (\delta\phi^* \nabla\delta\phi)\, dS + \int_V \Im m(k^2) |\delta\phi|^2 dV = 0. \tag{1.11.5}$$

Hence, if $\Im m[k^2(\mathbf{r})] \neq 0$ in V, and

(i) $\delta\phi = 0$ or $\hat{n} \cdot \nabla\delta\phi = 0$ on S,

(ii) $\delta\phi = 0$ on part of S and $\hat{n} \cdot \nabla\delta\phi = 0$ on the rest of S, or

(iii) $\delta\phi + \alpha\hat{n} \cdot \nabla\delta\phi = 0$ on S, where α is real, [2]

then the first integral above vanishes, and we have

$$\int_V \Im m[k^2(\mathbf{r})] |\delta\phi|^2 dV = 0. \tag{1.11.6}$$

Since $|\delta\phi|^2$ is positive definite for $\delta\phi \neq 0$, and $\Im m(k^2) \neq 0$ in V,[3] the above is only possible if $\delta\phi = 0$ everywhere inside V. Also, in the third case above, α can vary on the surface S. It can also be chosen so that the first two cases are the special cases of the third case.

Therefore, in order to guarantee uniqueness, so that $\phi_1 = \phi_2$ in V, the above conditions are equivalent to either

(i) $\phi_1 = \phi_2$ on S or $\hat{n} \cdot \nabla\phi_1 = \hat{n} \cdot \nabla\phi_2$ on S,

(ii) $\phi_1 = \phi_2$ on one part of S, and $\hat{n} \cdot \nabla\phi_1 = \hat{n} \cdot \nabla\phi_2$ on the rest of S, or

(iii) $\phi_1 + \alpha\hat{n} \cdot \nabla\phi_1 = \phi_2 + \alpha\hat{n} \cdot \nabla\phi_2$, where α is real.

The specification of ϕ on S is also known as the ***Dirichlet*** boundary condition, while the specification of $\hat{n} \cdot \nabla\phi$, namely, the normal derivative, is also known as the ***Neumann*** boundary condition. The third is the reactive impedance boundary condition. In words, the uniqueness theorem says that if two solutions satisfy the same Dirichlet or Neumann boundary condition or a mixture thereof on S, or the reactive impedance boundary condition, the two solutions must be identical.

Notice that the difference solution, $\delta\phi$ satisfies the boundary conditions above (1.11.6) are all lossless (non-dissipative or non-gain) boundary conditions. When $\Im m[k^2(\mathbf{r})] \neq 0$, and when such boundary conditions are satisfied by the difference solution, (1.11.6) implies that

[2] The author is grateful to J. Mamou for pointing out this case.

[3] More specifically, $\Im m[k^2(\mathbf{r})] > 0$, $\forall\ \mathbf{r} \in V$, or $\Im m[k^2(\mathbf{r})] < 0$, $\forall\ \mathbf{r} \in V$.

only trivial solution $\delta\phi = 0$ exists. In other words, no time-harmonic difference solution can exist in such media with loss or gain.

When $\Im m(k^2) = 0$, i.e., when k^2 is real, the condition $\delta\phi = 0$ or $\hat{n} \cdot \nabla\delta\phi = 0$ on S in (1.11.4) does not necessarily lead to $\delta\phi = 0$ in V, or uniqueness. The reason is that solutions for $\delta\phi = \phi_1 - \phi_2$ where

$$\int_V |\nabla\delta\phi|^2 \, dV = \int_V k^2 |\delta\phi|^2 \, dV \tag{1.11.7}$$

can exist. These are the resonance solutions in the volume V. These resonance solutions are the homogeneous solutions[4] to the wave Equation (1.11.1) at the real resonance frequencies of the volume V. Because the medium is lossless, they are time harmonic solutions which satisfies the boundary conditions, and hence, can be added to the particular solution of (1.11.1). In fact, the particular solution usually becomes infinite at these resonance frequencies if $S(\mathbf{r}) \neq 0$.

Equation (1.11.7) implies the balance of two energies. In the case of acoustic waves, for example, it represents the balance of the kinetic energy and the potential energy in a volume V. When $\Im m(k^2) \neq 0$, however, the resonance solutions of the volume V are exponentially decaying with time for a lossy medium $\left[\Im m(k^2) > 0\right]$, and they are exponentially growing with time for an active medium $\left[\Im m(k^2) < 0\right]$. But if only time harmonic solutions ϕ_1 and ϕ_2 are permitted in (1), these resonance solutions are automatically eliminated from the class of permissible solutions. Hence, for a lossy medium $[\Im m(k^2) > 0]$ or an active medium $[\Im m(k^2) < 0]$, the uniqueness of the solution is guaranteed if we consider only time harmonic solutions where ω is real, namely, two solutions will be identical if they have the same boundary conditions for ϕ and $\hat{n} \cdot \nabla\phi$ on S.[5]

When $S \to \infty$ or $V \to \infty$, the number of resonance frequencies of V becomes denser. In fact, when $S \to \infty$, the resonance frequencies of V become a continuum implying that any real frequency could be the resonant frequency of V. Hence, if the medium is lossless, the uniqueness of the solution is not guaranteed at any frequency, even with appropriate boundary conditions on S at infinity, as a result of the presence of the continuum of resonance frequencies. One remedy then is to introduce a small loss. With this small loss $[\Im m(k) > 0]$, the solution is either exponentially small when $r \to \infty$ (if a solution corresponds to an outgoing wave, e^{ikr}), or exponentially large when $r \to \infty$ (if a solution corresponds to an incoming wave, e^{-ikr}). Now, if the solution is exponentially small, namely, keeping only the outgoing wave solutions, it is clear that the surface integral term in (1.11.5) vanishes when $S \to \infty$, and the uniqueness of the solution is guaranteed. This manner of imposing the outgoing wave condition at infinity is also known as the ***Sommerfeld radiation condition*** [34, p. 188]. This radiation condition can be used in the limit of a vanishing loss for an unbounded medium to guarantee uniqueness.

The uniqueness of the solution to the Helmholtz wave equation is similar to the uniqueness

[4] "Homogeneous solutions" is a mathematical parlance for solutions to (1.11.1) without the source term.

[5] The nonuniqueness associated with the resonance solution for a lossless medium can be eliminated if we consider time domain solutions. In the time domain, we can set up an initial value problem in time, e.g., by requiring all fields be zero for $t < 0$; thus, the nonuniqueness problem can be removed via the causality requirement. The resonance solution, being time harmonic, is noncausal.

of the solution to the the matrix equation

$$\overline{\mathbf{A}} \cdot \mathbf{x} = \mathbf{b} \tag{1.11.8}$$

If a solution to the equation

$$\overline{\mathbf{A}} \cdot \mathbf{x}_N = 0 \tag{1.11.9}$$

exists, then the solution to the first equation is not unique. x_N is the null-space solution to the matrix.

The right-hand side of (1.11.8) is the driving term. If the driving term to Helmholtz wave equation is zero, and yet, a solution exists, it is usually called the resonance solution.[6] The resonance solution is equivalent to the null-space solution in matrix theory.

1.11.2 Vector Wave Equation

Similar to the uniqueness conditions for the scalar wave equation, analogous conditions for the vector wave equation can also be derived. First, assume that there are two different solutions to a vector wave Equation, i.e.,

$$\nabla \times \overline{\boldsymbol{\mu}}^{-1} \cdot \nabla \times \mathbf{E}_1(\mathbf{r}) - \omega^2 \, \overline{\boldsymbol{\epsilon}} \cdot \mathbf{E}_1(\mathbf{r}) = \mathbf{S}(\mathbf{r}), \tag{1.11.10}$$

$$\nabla \times \overline{\boldsymbol{\mu}}^{-1} \cdot \nabla \times \mathbf{E}_2(\mathbf{r}) - \omega^2 \, \overline{\boldsymbol{\epsilon}} \cdot \mathbf{E}_2(\mathbf{r}) = \mathbf{S}(\mathbf{r}), \tag{1.11.11}$$

where $\mathbf{S}(\mathbf{r}) = i\omega \, \mathbf{J}(\mathbf{r}) - \nabla \times \overline{\boldsymbol{\mu}}^{-1} \cdot \mathbf{M}(\mathbf{r})$ corresponds to a source of finite extent. Similarly, $\overline{\boldsymbol{\mu}}$ and $\overline{\boldsymbol{\epsilon}}$ correspond to an inhomogeneity of finite extent. Subtracting (1.11.10) from (1.11.11) then yields

$$\nabla \times \overline{\boldsymbol{\mu}}^{-1} \cdot \nabla \times \delta\mathbf{E} - \omega^2 \, \overline{\boldsymbol{\epsilon}} \cdot \delta\mathbf{E} = 0, \tag{1.11.12}$$

where $\delta\mathbf{E} = \mathbf{E}_1 - \mathbf{E}_2$. The solution is unique if and only if $\delta\mathbf{E} = 0$. Next, on multiplying the above by $\delta\mathbf{E}^*$, integrating over volume V, and using the vector identity $\mathbf{A} \cdot \nabla \times \mathbf{B} = -\nabla \cdot (\mathbf{A} \times \mathbf{B}) + \mathbf{B} \cdot \nabla \times \mathbf{A}$, we have

$$\begin{aligned} -\int_S \hat{n} \cdot (\delta\mathbf{E}^* \times \overline{\boldsymbol{\mu}}^{-1} \cdot \nabla \times \delta\mathbf{E}) \, dS + \int_V \nabla \times \delta\mathbf{E}^* \cdot \overline{\boldsymbol{\mu}}^{-1} \cdot \nabla \times \delta\mathbf{E} \, dV \\ -\omega^2 \int_V \delta\mathbf{E}^* \cdot \overline{\boldsymbol{\epsilon}} \cdot \delta\mathbf{E} \, dV = 0. \end{aligned} \tag{1.11.13}$$

Since $\nabla \times \delta\mathbf{E} = i\omega \, \overline{\boldsymbol{\mu}} \cdot \delta\mathbf{H}$, the above can be rewritten as

$$i\omega \int_S \hat{n} \cdot (\delta\mathbf{E}^* \times \delta\mathbf{H}) \, dS + \omega^2 \int_V (\delta\mathbf{H}^* \cdot \overline{\boldsymbol{\mu}}^{\dagger} \cdot \delta\mathbf{H} - \delta\mathbf{E}^* \cdot \overline{\boldsymbol{\epsilon}} \cdot \delta\mathbf{E}) \, dV = 0. \tag{1.11.14}$$

[6] This is called the homogeneous solution in mathematical parlance. The solution that corresponds to the driving term on the right-hand side is called the inhomogeneous solution.

Then, taking the imaginary part of (1.11.14) yields

$$\Im m\left\{ i\omega \int_S \hat{n}\cdot(\delta\mathbf{E}^*\times\delta\mathbf{H})\,dS \right\}$$

$$-\frac{i\omega^2}{2}\int_V [\delta\mathbf{H}^*\cdot(\overline{\boldsymbol{\mu}}^\dagger-\overline{\boldsymbol{\mu}})\cdot\delta\mathbf{H}+\delta\mathbf{E}^*\cdot(\overline{\boldsymbol{\epsilon}}^\dagger-\overline{\boldsymbol{\epsilon}})\cdot\delta\mathbf{E}]\,dV=0. \tag{1.11.15}$$

(1.11.16)

But if the medium is not lossless (either lossy or active), then $\overline{\boldsymbol{\mu}}^\dagger \neq \overline{\boldsymbol{\mu}}$ and $\overline{\boldsymbol{\epsilon}}^\dagger \neq \overline{\boldsymbol{\epsilon}}$, and the second integral in (1.11.16) may not be zero. Moreover, if

(i) $\hat{n}\times\delta\mathbf{E}=0$ or $\hat{n}\times\delta\mathbf{H}=0$ on S,

(ii) $\hat{n}\times\delta\mathbf{E}=0$ on one part of S and $\hat{n}\times\delta\mathbf{H}=0$ on the rest of S, or

(iii) $\delta\mathbf{H}-i\zeta\hat{n}\times\delta\mathbf{E}=0$ on S, where ζ is a real number,

then the first integral in (1.11.16) vanishes. The above corresponds to lossless boundary conditions for the difference field. The third case corresponds to a lossless reactive impedance boundary condition. [7] Again, ζ can vary on S, and the first two cases can be made special cases of the third case.

The above implies that,

$$\frac{\omega^2}{2}\int_V [\delta\mathbf{H}^*\cdot i(\overline{\boldsymbol{\mu}}^\dagger-\overline{\boldsymbol{\mu}})\cdot\delta\mathbf{H}+\delta\mathbf{E}^*\cdot i(\overline{\boldsymbol{\epsilon}}^\dagger-\overline{\boldsymbol{\epsilon}})\cdot\delta\mathbf{E}]\,dV=0. \tag{1.11.17}$$

In the above, $i(\overline{\boldsymbol{\mu}}^\dagger-\overline{\boldsymbol{\mu}})$ and $i(\overline{\boldsymbol{\epsilon}}^\dagger-\overline{\boldsymbol{\epsilon}})$ are Hermitian matrices. Moreover, the integrand will be positive definite if both $\overline{\boldsymbol{\mu}}$ and $\overline{\boldsymbol{\epsilon}}$ are lossy, and the integrand will be negative definite if both $\overline{\boldsymbol{\mu}}$ and $\overline{\boldsymbol{\epsilon}}$ are active. Hence, the only way for (1.11.17) to be satisfied is for $\delta\mathbf{E}=0$ and $\delta\mathbf{H}=0$, or that $\mathbf{E}_1=\mathbf{E}_2$ and $\mathbf{H}_1=\mathbf{H}_2$, implying uniqueness.

Consequently, in order for uniqueness to be guaranteed, either

(i) $\hat{n}\times\mathbf{E}_1=\hat{n}\times\mathbf{E}_2$ on S or $\hat{n}\times\mathbf{H}_1=\hat{n}\times\mathbf{H}_2$ on S,

(ii) $\hat{n}\times\mathbf{E}_1=\hat{n}\times\mathbf{E}_2$ on a part of S while $\hat{n}\times\mathbf{H}_1=\hat{n}\times\mathbf{H}_2$ on the rest of S, or

(iii) $\mathbf{H}_1-i\zeta\hat{n}\times\mathbf{E}_1=\mathbf{H}_2-i\zeta\hat{n}\times\mathbf{E}_2$ on S.

In other words, if two solutions satisfy the same boundary conditions for tangential $\mathbf{E}$ or tangential $\mathbf{H}$, or a mixture thereof on S, or the same reactive boundary condition, the two solutions must be identical.

Again, the requirement for a nonlossless condition is to eliminate the real resonance solutions which could otherwise be time harmonic, homogeneous solutions to (1.11.10) satisfying

[7] A more complicated boundary condition for the third case may be designed.

the boundary conditions. For example, if the appropriate boundary conditions for $\delta\mathbf{E}$ and $\delta\mathbf{H}$ are imposed so that the first term of (1.11.14) is zero, then

$$\int_V (\delta\mathbf{H}^* \cdot \overline{\boldsymbol{\mu}}^\dagger \cdot \delta\mathbf{H} - \delta\mathbf{E}^* \cdot \overline{\boldsymbol{\epsilon}} \cdot \delta\mathbf{E})\, dV = 0. \tag{1.11.18}$$

The above does not imply that $\delta\mathbf{E}$ or $\delta\mathbf{H}$ equals zero, because at resonances, a perfect balance between the energy stored in the electric field and the energy stored in the magnetic field is maintained. As a result, the left-hand side of the above could vanish without having $\delta\mathbf{E}$ and $\delta\mathbf{H}$ be zero, which is necessary for uniqueness. But away from the resonances of the volume V, the energy stored in the electric field is not equal to that stored in the magnetic field. Hence, in order for (1.11.18) to be satisfied, $\delta\mathbf{E}$ and $\delta\mathbf{H}$ have to be zero since each term in (1.11.18) is positive definite for lossless media due to the Hermitian nature of $\overline{\boldsymbol{\mu}}$ and $\overline{\boldsymbol{\epsilon}}$.

When $V \to \infty$, as in the scalar wave equation case, some loss has to be imposed to guarantee uniqueness. This is the same as requiring the wave to be outgoing at infinity, namely, the radiation condition. Again, the radiation condition can be imposed for an unbounded medium with vanishing loss to guarantee uniqueness.

1.12 Transformation Matrices for Microwave Circuits

1.12.1 Impedance and Admittance Matrices

A general microwave circuit consists of many ports. A convenient way to characterize an N-port network is to describe the network in terms of impedance matrices or admittance matrices [5]. For example, if the N-port network can be characterized by a pair of voltage and current at each port, then a column vector of voltages can be defined and also a column vector of currents. We can write down a relationship between the voltages and the currents as

$$\mathbf{V} = \begin{bmatrix} V_1 \\ V_2 \\ \vdots \\ V_N \end{bmatrix} = \begin{bmatrix} Z_{11} & Z_{12} & \cdots & Z_{1N} \\ Z_{21} & Z_{22} & \cdots & Z_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ Z_{N1} & Z_{N2} & \cdots & Z_{NN} \end{bmatrix} \begin{bmatrix} I_1 \\ I_2 \\ \vdots \\ I_N \end{bmatrix} = \overline{\mathbf{Z}} \cdot \mathbf{I} \tag{1.12.1}$$

By the same token, we can express

$$\mathbf{I} = \overline{\mathbf{Y}} \cdot \mathbf{V} \tag{1.12.2}$$

where $\overline{\mathbf{Y}}$ is the admittance matrix. For reciprocal circuits, it can be shown that $\overline{\mathbf{Z}}$ and $\overline{\mathbf{Y}}$ are symmetric matrices. For lossless circuits, it can be shown that these matrices have pure imaginary elements.

For a two port network which is reciprocal, there are three independent matrix elements. Therefore, a two-port network can often be modeled by a T or a Π equivalent circuit.

1.12.2 Scattering Matrices

For high frequencies, it is more pertinent to think about waves. Then at each port, we can define an incident and a reflected wave. For instance, we can define an incident voltage wave

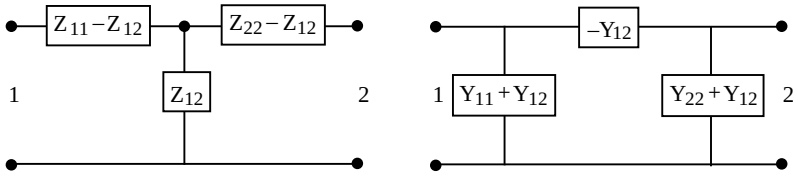


Figure 1.7: The T and Π equivalent circuits for a two-port circuit.

V^+ and a reflected voltage wave V^-. A relationship can then be written between the reflected waves at all the ports to the incident waves at all the ports.

$$\mathbf{V} = \begin{bmatrix} V_1^- \\ V_2^- \\ \vdots \\ V_N^- \end{bmatrix} = \begin{bmatrix} S_{11} & S_{12} & \cdots & S_{1N} \\ S_{21} & S_{22} & \cdots & S_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ S_{N1} & S_{N2} & \cdots & S_{NN} \end{bmatrix} \begin{bmatrix} V_1^+ \\ V_2^+ \\ \vdots \\ V_N^+ \end{bmatrix} = \overline{\mathbf{S}} \cdot \mathbf{V}^+ \tag{1.12.3}$$

It can be proved that $\overline{\mathbf{S}}$ has to be symmetric for reciprocal circuits, and that it has to be unitary if the circuit is lossless.

1.12.3 Chain Matrices

When one needs to cascade a series of two port networks, it is more convenient to work with chain matrices or transmission matrices. A voltage and current transmission matrix relates the voltage and current at one port to the voltage and current at the second port.

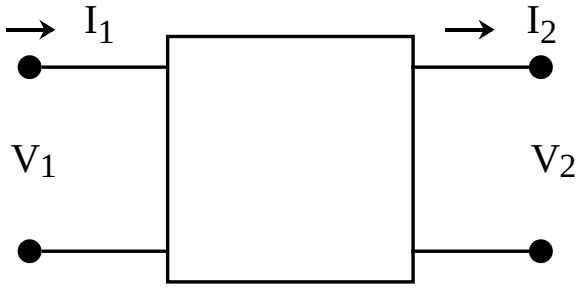


Figure 1.8: A diagram for defining the voltage and current transmission matrix.

Written explicitly, we have

$$\begin{bmatrix} V_1 \\ I_1 \end{bmatrix} = \begin{bmatrix} A_1 & B_1 \\ C_1 & D_1 \end{bmatrix} \begin{bmatrix} V_2 \\ I_2 \end{bmatrix} \tag{1.12.4}$$

Notice that the current at Port 2 is flowing out of the port rather than into the port. In this manner, if we have a second transmission matrix of a second network that relates V_2, I_2 to

V_3, I_3, viz.,

$$\begin{bmatrix} V_2 \\ I_2 \end{bmatrix} = \begin{bmatrix} A_2 & B_2 \\ C_2 & D_2 \end{bmatrix} \begin{bmatrix} V_3 \\ I_3 \end{bmatrix} \tag{1.12.5}$$

Hence, when these two networks are cascaded together, the resultant transmission matrix is the product of the two matrices

$$\begin{bmatrix} V_2 \\ I_2 \end{bmatrix} = \begin{bmatrix} A_1 & B_1 \\ C_1 & D_1 \end{bmatrix} \begin{bmatrix} A_2 & B_2 \\ C_2 & D_2 \end{bmatrix} \begin{bmatrix} V_3 \\ I_3 \end{bmatrix} \tag{1.12.6}$$

For a two port network, it can be shown that

$$A = Z_{11}/Z_{12}, \quad B = (Z_{11}Z_{22} - Z_{12}^2)/Z_{12}, \tag{1.12.7a}$$

$$C = 1/Z_{12}, \quad D = Z_{22}/Z_{12}, \tag{1.12.7b}$$

for a reciprocal network. It is also readily verified that

$$AD - BC = 1 \tag{1.12.8}$$

for this case. Hence, the determinant of a chain matrix is always 1 for a reciprocal network.

Exercises for Chapter 1

Problem 1-1: The fundamental units in electromagnetics can be considered to be meter, kilogram, second, and coulomb.

(a) Show that 1 volt, which is 1 watt/amp, has the dimension of (kilogram meter^2)/(coulomb sec^2).

(b) From Maxwell's equations, show that μ_0 has the dimension of (second volt)/(meter amp), and hence, its dimension is (kilogram meter)/coulomb^2 in the more fundamental units.

(c) If we assign the value of μ_0 to be 4π instead of $4\pi \times 10^{-7}$, what would be the unit of coulomb in this new assignment compared to the old unit? What would be the present value of 1 volt and 1 amp in this new assignment?

Problem 1-2: Show that for two time-harmonic functions,

$$\langle A(\mathbf{r},t)B(\mathbf{r},t)\rangle = \frac{1}{2}\Re e[A(\mathbf{r})B^*(\mathbf{r})], \tag{1.12.9}$$

where $A(\mathbf{r})$ and $B(\mathbf{r})$ are the phasors of $A(\mathbf{r},t)$ and $B(\mathbf{r},t)$.The angular brackets above imply time averaging.

Problem 1-3: Assume that a voltage is time harmonic, i.e., $V(t) = V_0 \cos\omega t$, and that a current $I(t) = I_I \cos\omega t + I_Q \sin\omega t$, i.e., it consist of an in-phase and a quadrature component.

(a) Find the instantaneous power due to this voltage and current, viz.,$V(t)I(t)$.

(b) Find the phaser representations of the voltage and current, and hence the complex power due to this voltage and current.

(c) Establish a relationship between the real part and reactive part of the complex power to the instantaneous power.

(d) Show that the reactive power is due to the quadrature component of the current, which is related to a time-varying part of the instantaneous power with zero-time average.

Problem 1-4: For a scalar-wave equation, $\nabla \cdot \epsilon^{-1}(\mathbf{r})\nabla\phi(\mathbf{r}) + k^2\phi(\mathbf{r}) = S(\mathbf{r})$:

(a) Show that a reciprocal relationship $\langle \phi_1(\mathbf{r}), S_2(\mathbf{r})\rangle = \langle \phi_2(\mathbf{r}), S_1(\mathbf{r})\rangle$ exists.

(b) What is the boundary condition satisfied by ϕ at an interface where $\epsilon(\mathbf{r})$ has a step discontinuity?

Problem 1-5:

(a) Prove that for reciprocal circuits, the impedance matrix and the admittance matrix are symmetric.

(b) Prove that for lossless circuits, the impedance matrix and the admittance matrix have imaginary elements.

Problem 1-6:

(a) Prove that for reciprocal circuits, the scattering matrix is symmetric.

(b) Prove that for lossless circuits, the scattering matrix is unitary.

(c) Prove that for reciprocal circuits, the determinant of the chain matrix is always equal to one.

Bibliography

[1] R.E. Collin, *Field Theory of Guided Waves,* IEEE Press, Piscataway, NJ, 1991.

[2] R. Mittra and S.W. Lee, *Analytical Techniques in the Theory of Guided Waves,* The MacMillan Company, New York, 1971.

[3] L. Levin, *Theory of Waveguides: Techniques for the Solution of Waveguide Problems,* Newnes-Butterworth, London, 1975.

[4] N. Marcuvitz, ed., *Waveguide Handbook,* MIT Radiation Laboratory Series, vol, 10, McGraw-Hill, New York, 1951.

[5] R.E. Collin, *Foundation for Microwave Engineering,* IEEE Press, Piscataway, NJ, 2001.

[6] J. W. Strutt Rayleigh (Lord Rayleigh), *Theory of Sound*, New York: Dover Publ., 1976. (Originally published 1877.)

[7] J. W. Strutt Rayleigh (Lord Rayleigh), "On the passage of electric waves through tubes, or the vibra cylinder," *Phi. Mag.,* vol. 43, pp. 125–132, 1897.

[8] J. Hecht, *City of Light: The Story of Fiber Optics,* Oxford University Press, Oxford, U.K., 1999.

[9] A.A. Oliner, "Leakage from higher modes on microstrip line with application to antennas," *Radio Sci.*, 22(6), pp. 907-912, 1987.

[10] *Encyclopaedia Britanica*, Encyclopaedia Britanica Inc., 2004.

[11] R. Feynman, R.B. Leighton, and M.L. Sands, *The Feynman Lectures on Physics*, vol. I, Chapter 52, Addison-Wesley Publishing Co., 1965.

[12] M. Faraday, "On static electrical inductive action," *Phil. Mag.*, 1843. M. Faraday, Experimental Researches in Electricity and Magnetism. Vol. 1, Taylor & Francis, London, 1839.; Vol. 2, Richard & John E. Taylor, London, 1844; Vol. 3, Taylor and Francis, London, 1855. Reprinted by Dover in 1965. Also see M. Faraday, "Remarks on Static Induction," Proc. Roy. Inst., Feb. 12, 1858.

[13] A. M. Ampère, "Mémoire sur la théorie des phénomènes électrodynamiques," *Mem. Acad. R. Sci. Inst. Fr.*, 6, 228-232, 1823.

[14] C. S. Gillmore, *Charles Augustin Coulomb: Physics and Engineering in Eighteenth Century Frrance*, Princeton, NJ, 1971.

[15] C. F. Gauss, "General theory of terrestrial magnetism," *Scientific Memoirs*, vol. 2, ed. R. Taylor (R & J.E. Taylor, London), pp. 184-251, 1841.

[16] J. C. Maxwell, *A Treatise of Electricity and Magnetism,* 2 vols, Clarendon Press, Oxford, 1873. Also, see P. M. Harman (ed.), *The Scientific Letters and Papers of James Clerk Maxwell, Vol. II, 1862-1873*, Cambridge, U.K.: Cambridge University Press, 1995.

[17] O. Heaviside, "On electromagnetic waves, especially in relation to the vorticity of the impressed forces, and the forced vibration of electromagnetic systems," *Phil. Mag.*, 25, 130-156, 1888. Also, see P. J. Nahin, "Oliver Heaviside," *Scientific American*, pp. 122-129, June 1990.

[18] *Nobel Lectures, Physics 1901-1921*, Elsevier Publishing Company, Amsterdam, 1967.

[19] J. Glenn, ed., *The Complete Patents of Nikola Tesla,* New York: Barnes and Noble Books, 1994.

[20] W. K. Tung, *Group Theory in Physics,* Philadelphia, PA: World Scientific Publ., 1985.

[21] L. M. Falicov, *Group Theory and its Physical Applications,* Chicago: University of Chicago Press, 1966.

[22] W. C. Chew, J. M. Jin, E. Michielssen, and J. M. Song, eds., *Fast and Efficient Algorithms in Computational Electromagnetics*, Artech House, Boston, MA, 2001.

[23] J.D. Jackson, *Classical Electrodynamics,* Third Edition, John Wiley & Sons, Inc., NJ, 1999.

[24] T.D. Lee and C.N. Yang, "Question of parity conservation in weak interaction," *Phys. Rev.*, 104(1), pp. 254-257, 1956.

[25] G. Green, An Essay on the Application of Mathematical Analysis to the Theories of Electricity and Magnetism, T. Wheelhouse, Nottingham, 1828. Also, see L. Challis and F. Sheard, "The green of Green functions," *Physics Today*, pp. 41-46, December 2003.

[26] W.C. Chew, *Waves and Fields in Inhomogeneous Media,* Van Nostrand Reinhold, New York, 1990, reprinted, IEEE Press, Piscataway, NJ, 1995.

[27] K.T. Chau, W.L. Li and C.H.T. Lee, "Challenges and opportunities of electric machines for renewable energy," (invited paper), *Prog. in Electromag. Research B*, vol. 42, 45-74, 2012.

[28] L. Novotny and B. Hecht, *Principles of Nano-Optics*, Cambridge University Press, 2006.

[29] S. Datta, *Electronic Transport in Mesoscopic Systems*, Cambridge University Press, 1995, Paperback Edition 1997.

[30] H.A. Haus, *Electromagnetic Noise and Quantum Optical Measurements*, Springer-Verlag, Berlin, 2000.

[31] A. Yariv, "Phase conjugate optics and real-time holography," *IEEE J. Quantum Electron.,* QE-14, p. 117, 1978.

[32] G.S. Agarwal, A.T. Friberg, and E. Wolf, "Elimination of distortions by phase conjugation without losses or gains," *Opt. Commun.*, 43, p. 446, 1982.

[33] W.C. Chew and T.M. Habashy, "Phase-conjugate mirror and time reversal," *J. Opt. Soc. Am.*, vol. 2, no. 6, p. 808, 1985.

[34] A. Sommerfeld, *Partial Differential Equation.* New York: Academic Press, 1949.

Chapter 2

Hollow Waveguides

Much work has been written on hollow waveguides [1–7]. Theory on guided waves dates back even to earlier dates as alluded to in Chapter 1. Hollow metallic waveguides, when filled with air, allow for high power microwave transmission. However, a hollow electromagnetic waveguide, unlike an acoustic waveguide, has a lower cutoff frequency of operation. The way to counter this is to use a two-conductor waveguide such as a coaxial waveguide. The support of the inner conductor in such waveguides requires the filling of the waveguide with dielectric materials. It will be shown that when a waveguide is homogeneously filled with materials, the theory is essentially the same as that for one which is hollow.

2.1 General Uniform Cylindrical Waveguides

General uniform cylindrical waveguides include transmission lines, and hollow waveguides, as well as multi-conductor waveguides. If the regions between the conductors are filled with a homogeneous material, the waveguide can support purely transverse electric (TE), transverse magnetic (TM) or transverse electromagnetic (TEM) modes. If the wave is TEM, the waveguide is operating in the transmission line mode. It can be shown that a hollow waveguide cannot support a TEM wave. Therefore, a transmission line needs to have at least two conductors.

Because the fields of a general uniform cylindrical waveguide can be decomposed into TE and TM types, we can characterize the TE wave with the $\hat{z}$-component of the magnetic field or H_z, since $H_z \neq 0$ for this type of wave. Similarly, we can characterize the TM wave with the $\hat{z}$-component of the magnetic field or E_z since $E_z \neq 0$ for this type of wave.

We shall derive the equations governing the E_z and H_z components of the fields in a general, uniform cylindrical waveguide filled with a homogeneous material. We shall call such waveguides hollow waveguides. We let [6]

$$\mathbf{E} = \mathbf{E}_s + \hat{z}E_z, \qquad \mathbf{H} = \mathbf{H}_s + \hat{z}H_z, \tag{2.1.1}$$

where the subscript s represents the transverse to z components. Substituting the above

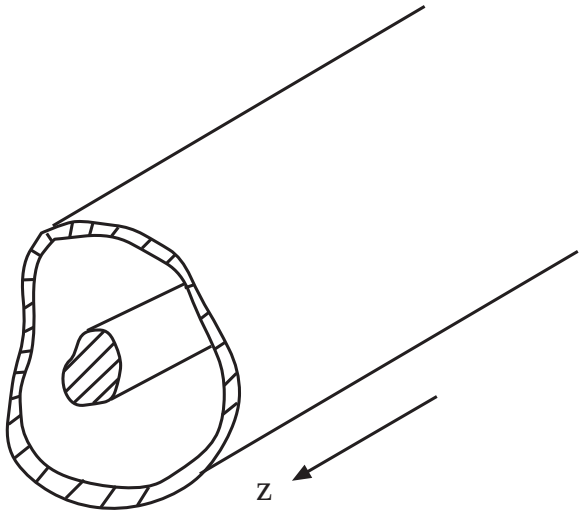


Figure 2.1: Cross-section of a general uniform cylindrical waveguide.

equations into Maxwell's equations, we have

$$\left(\nabla_s + \hat{z}\frac{\partial}{\partial z}\right) \times (\mathbf{E}_s + \hat{z}E_z) = i\omega\mu(\mathbf{H}_s + \hat{z}H_z), \tag{2.1.2}$$

$$\left(\nabla_s + \hat{z}\frac{\partial}{\partial z}\right) \times (\mathbf{H}_s + \hat{z}H_z) = -i\omega\epsilon(\mathbf{E}_s + \hat{z}E_z). \tag{2.1.3}$$

Equating the s components of (2.1.2) and (2.1.3), we have

$$\nabla_s \times \hat{z}E_z + \frac{\partial}{\partial z}\hat{z} \times \mathbf{E}_s = i\omega\mu\mathbf{H}_s, \tag{2.1.4}$$

$$\nabla_s \times \hat{z}H_z + \frac{\partial}{\partial z}\hat{z} \times \mathbf{H}_s = -i\omega\epsilon\mathbf{E}_s. \tag{2.1.5}$$

Substituting for $\mathbf{E}_s$ from (2.1.5) into (2.1.4), we have

$$\omega^2\mu\epsilon\mathbf{H}_s = -i\omega\epsilon\nabla_s \times \hat{z}E_z + \frac{\partial}{\partial z}\hat{z} \times \left(\nabla_s \times \hat{z}H_z + \frac{\partial}{\partial z}\hat{z} \times \mathbf{H}_s\right). \tag{2.1.6}$$

Using the vector identities $\hat{z} \times \nabla_s \times \hat{z} = \nabla_s, \quad \hat{z} \times \hat{z} \times \mathbf{H}_s = -\mathbf{H}_s$, and assuming that the field has $e^{\pm ik_z z}$ dependence, so that $\frac{\partial^2}{\partial z^2} \to -k_z^2$, we can rewrite (2.1.6) as

$$\mathbf{H}_s = \frac{1}{k^2 - k_z^2}\left[\frac{\partial}{\partial z}\nabla_s H_z + i\omega\epsilon\hat{z} \times \nabla_s E_z\right], \tag{2.1.7}$$

where $k^2 = \omega^2\mu\epsilon$. By the same token, we have

$$\mathbf{E}_s = \frac{1}{k^2 - k_z^2}\left[\frac{\partial}{\partial z}\nabla_s E_z - i\omega\mu\hat{z}\times\nabla_s H_z\right]. \tag{2.1.8}$$

Equations (2.1.7) and (2.1.8) allow us to derive all the other components of the field in a waveguide once we know the z components of the field.

If we equate the z components of (2.1.2) and (2.1.3), we have

$$\nabla_s \times \mathbf{E}_s = i\omega\mu\hat{z}H_z, \tag{2.1.9}$$

$$\nabla_s \times \mathbf{H}_s = -i\omega\epsilon\hat{z}E_z. \tag{2.1.10}$$

Substituting (2.1.7) and (2.1.8) into (2.1.9) and (2.1.10), we have

$$(\nabla_s^2 + k_s^2)H_z = 0, \qquad \text{for} \quad \text{TE} \quad \text{waves}, \tag{2.1.11}$$

$$(\nabla_s^2 + k_s^2)E_z = 0, \qquad \text{for} \quad \text{TM} \quad \text{waves}, \tag{2.1.12}$$

where $k_s^2 = k^2 - k_z^2$. Therefore, H_z and E_z satisfy a two-dimensional scalar wave equation also known as the *reduced wave equation*. Once E_z and H_z are solved for from (2.1.11) and (2.1.12), we can derive all the other field components using (2.1.7) and (2.1.8).

In the above, (2.1.11) and (2.1.12) clearly show that H_z and E_z satisfy the wave equation independently. The E_z component satisfies the homogeneous Dirichlet boundary condition, $E_z = 0$ on the waveguide wall. The general boundary condition for the TE field is that $\hat{n}\times\mathbf{E}_s = 0$ on the waveguide wall. From (2.1.8), we can show that this is equivalent to $\hat{n}\cdot\nabla H_z = 0$ on the waveguide wall, which is the homogeneous Neumann boundary condition. The boundary condition does not couple the TE and TM waves implying that they can exist independently of each other. This substantiates our assumption in the very beginning.

It is to be noted that an alternative way of solving the above problem is to define two scalar potentials ψ_h and ψ_e, and let

$$\mathbf{E}^{TE} = \nabla\times\hat{z}\psi_h \tag{2.1.13}$$

$$\mathbf{H}^{TM} = \nabla\times\hat{z}\psi_e \tag{2.1.14}$$

It can be easily shown that H_z from the TE field is proportional to ψ_h and E_z from the TM field is proportional to ψ_e. Hence, this method of solution is completely equivalent to our previous method of solution.

We can envision that a wave is guided in a waveguide because the wave is bouncing around the wall of the waveguide. This is the bouncing wave picture of the wave. In this case, the $\mathbf{k}$ vector is not pointing in the z direction completely, and $k_z < k$. The mode is either TE or TM. However, it is also possible to have a wave directly transmitted through a waveguide, such as in the transmission line. In this case, the $\mathbf{k}$ vector is pointing entirely in the z direction, and $k_z = k$. We shall study this mode for transmission lines.

2.2 Wave Impedance

With the above results, we can define wave impedance concepts in a hollow waveguide. Just as the intrinsic impedance in a homogeneous medium relates the electric field to a magnetic field, we may relate the transverse components of $\mathbf{E}$ and $\mathbf{H}$ in a waveguide by a wave impedance.

For a TE mode, the transverse components are

$$\mathbf{H}_s = \frac{ik_z}{k_s^2}\nabla_s H_z, \qquad \mathbf{E}_s = \frac{-i\omega\mu}{k_s^2}\hat{z}\times\nabla_s H_z. \tag{2.2.1}$$

Therefore, the wave impedance is

$$Z^{TE} = \frac{\hat{z}\times\mathbf{E}_s}{\mathbf{H}_s} = \frac{\omega\mu}{k_z}. \tag{2.2.2}$$

For a TM mode, the transverse components are

$$\mathbf{H}_s = \frac{i\omega\epsilon}{k_s^2}\hat{z}\times\nabla_s E_z, \qquad \mathbf{E}_s = \frac{ik_z}{k_s^2}\nabla_s E_z. \tag{2.2.3}$$

Therefore, the wave impedance is

$$Z^{TM} = \frac{\hat{z}\times\mathbf{E}_s}{\mathbf{H}_s} = \frac{k_z}{\omega\epsilon}. \tag{2.2.4}$$

The above is valid for a general cylindrical waveguide which is homogeneously filled. It is useful in deriving equivalent transmission line models for a waveguide.

2.3 Transmission Line Theory

The propagation of waves on the transmission line was first formulated in terms of telegrapher's equations. Telegraphy was in use in the early 1800s even before the completion of Maxwell's equations in 1864. The telegrahers equations can be derived using circuit theory (see Fig. 2.2), and they are valid even for meandering lines. The transmission line can be thought of as consisting of a sequence of coupled L and C tank circuits. Each tank circuit forms a simple harmonic oscillator. But on a transmission line, these harmonic oscillators are coupled together. It is through the coupling of these harmonic oscillators that a wave can propagate on a transmission line.

2.3.1 TEM Mode of a Transmission Line

In a transmission line, a TEM (transverse electromagnetic) mode can propagate. For TEM waves, both E_z and H_z are zero. By looking at (2.1.7) and (2.1.8), $\mathbf{H}_s$ and $\mathbf{E}_s$ will be non-zero only if $k_z = k$. Therefore, all TEM waves, or TEM modes in a waveguide have e^{ikz} dependence. Furthermore, from (2.1.9) and (2.1.10), we conclude that for TEM waves

$$\nabla_s\times\mathbf{E}_s = 0, \qquad \nabla_s\times\mathbf{H}_s = 0. \tag{2.3.1}$$

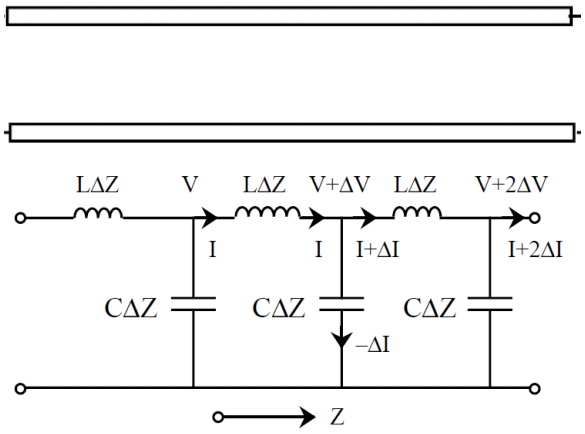


Figure 2.2: Two pieces of parallel wires form a transmission line, even when they are sinuous in nature. A wire has inductance, and capacitance exists between two pieces of metal. Hence, circuit model used to derive the telegrahers equations.

In other words, $\mathbf{E}_s$ is electrostatic in the xy plane while $\mathbf{H}_s$ is magnetostatic in the xy plane. Hence, we can let

$$\mathbf{E}_s = -\nabla_s \phi_s(x,y)e^{ikz}, \qquad \mathbf{H}_s = -\nabla_s \psi_s(x,y)e^{ikz}. \tag{2.3.2}$$

Since $\nabla \cdot \mathbf{E}_s = 0$ and $\nabla \cdot \mathbf{H}_s = 0$ inside the waveguide, ϕ_s and ψ_s satisfy Laplace equations

$$\nabla_s^2 \phi_s(x,y) = 0, \qquad \nabla_s^2 \psi_s(x,y) = 0. \tag{2.3.3}$$

If we have perfect electric conductors, the boundary conditions are $\hat{n} \times \mathbf{E}_s = 0$, and $\hat{n} \cdot \mathbf{H}_s = 0$ on the metallic surfaces. These boundary conditions translate to

$$\phi_s = \text{constant}, \tag{2.3.4}$$

$$\hat{n} \cdot \nabla_s \psi_s = \frac{\partial}{\partial n} \psi_s = 0. \tag{2.3.5}$$

Equation (2.3.4) is known as the *Dirichlet* boundary condition while (2.3.5) is the *Neumann* boundary condition. The constants in (2.3.4) are the potentials on the conductors which may be different for different conductors. ψ_s is a multi-value function because magnetic field always goes in a loop and ends on itself. In order to avoid dealing with multi-value functions, it is a lot easier to solve for ϕ_s. Alternatively, one can solve for the magnetic field using a vector potential.

At this point, it seems that ϕ_s and ψ_s are decoupled, and hence, the electric field and the magnetic field are independent of each other. This could not be true, as the coupling is expressed in Equations (2.1.4) and (2.1.5) (if we set $E_z = H_z = 0$ for discussing TEM modes).

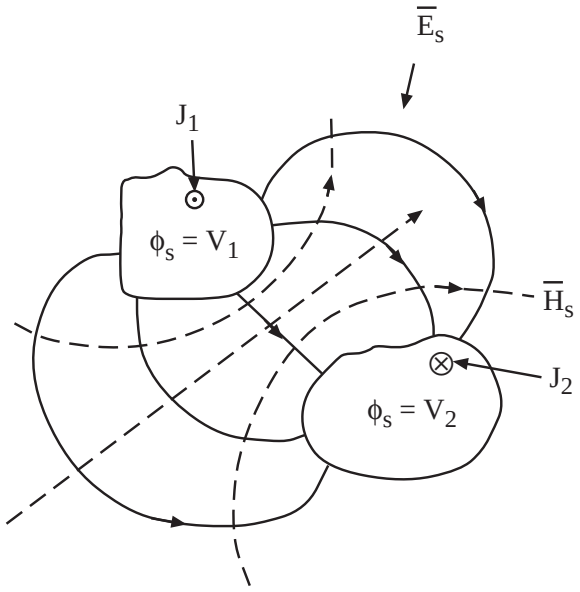


Figure 2.3: TEM mode in a transmission line.

Hence, the coupling of the fields is only through the $\hat{z}$-variation of the fields, i.e.,

$$\frac{\partial}{\partial z}\hat{z} \times \mathbf{E}_s = i\omega\mu\mathbf{H}_s, \tag{2.3.6}$$

$$\frac{\partial}{\partial z}\hat{z} \times \mathbf{H}_s = -i\omega\epsilon\mathbf{E}_s. \tag{2.3.7}$$

Furthermore, from the above, we deduce that $\mathbf{H}_s$ and $\mathbf{E}_s$ are mutually orthogonal in the TEM mode. Since the fields have e^{ikz} dependence, we conclude that

$$\hat{z} \times \mathbf{E}_s = \sqrt{\frac{\mu}{\epsilon}}\mathbf{H}_s = \eta\mathbf{H}_s, \tag{2.3.8}$$

if we assume that the wave is only propagating in one direction. Then η is also known as the *intrinsic impedance* of the medium, and all TEM waves satisfy (2.3.8).

In (2.3.3), the electrostatic and the magnetostatic problems are seemingly decoupled from each other, but the fields are coupled via their z variation, as indicated by the above equations. In (2.3.8), it says that once we know the electrostatic field, the magnetostatic field can be easily derived. Hence, we need only to solve the electrostatic problem to fully characterize the TEM solution.

Derivation of the Telegrapher Equations

We can integrate Equation (2.3.6) about a line contour around one of the conductors to obtain

$$\frac{\partial}{\partial z}\oint_C \hat{z}\times\mathbf{E}_s\cdot d\mathbf{l} = i\omega\mu\oint_C \mathbf{H}_s\cdot d\mathbf{l}. \tag{2.3.9}$$

By Ampere's law, we have $\oint_C \mathbf{H}_s\cdot d\mathbf{l} = I$, the total current on one of the conductors. For the left-hand side, we have

$$\begin{aligned}\oint_C \hat{z}\times\mathbf{E}_s\cdot d\mathbf{l} &= \oint_C d\mathbf{l}\times\hat{z}\cdot\overline{\mathbf{E}}_s\\ &= \frac{Q}{\epsilon},\end{aligned} \tag{2.3.10}$$

Since $d\mathbf{l}\times\hat{z}$ is an outward normal to C and Gauss' theorem can be invoked. Therefore, Equation (2.3.9) becomes

$$\frac{d}{dz}Q = i\omega\mu\epsilon I. \tag{2.3.11}$$

Since the transverse field is purely static, we can define $Q = CV$ where C is the capacitance per unit length and $V = V_1 - V_2$. Hence, (2.3.11) becomes

$$\frac{d}{dz}V = i\omega\frac{\mu\epsilon}{C}I. \tag{2.3.12}$$

Since $\mu\epsilon/C$ has the dimension of henry per meter, we can define $L = \mu\epsilon/C$, an inductance per unit length, and (2.3.12) becomes

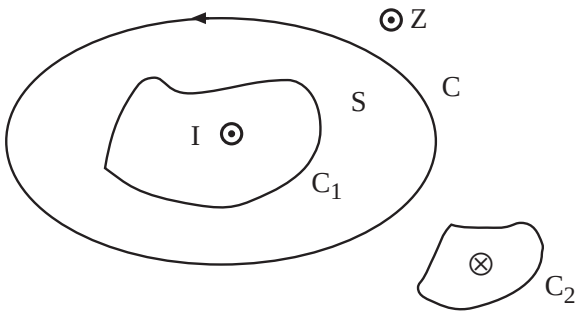


Figure 2.4: Derivation of the telegrapher's equations for a transmission line.

$$\frac{dV}{dz} = i\omega LI. \tag{2.3.13}$$

With similar manipulations to $\hat{z}\times$(2.3.7), we obtain

$$\frac{dI}{dz} = i\omega CV. \tag{2.3.14}$$

Equations (2.3.13) and (2.3.14) are the telegrapher's equations (also known as telegrapher's equations) for a transmission line, which can also be derived from a circuits viewpoint. The velocity of the wave on the line is given by

$$v = \frac{1}{\sqrt{\mu\epsilon}} = \frac{1}{\sqrt{LC}}. \tag{2.3.15}$$

Notice that in the above, we need only to solve the electrostatic problem to obtain C, and L can be derived from C. There is no need to solve the magnetostatic problem.

Characteristic Impedance versus Intrinsic Impedance

Since the fields, and hence the voltage and current, have e^{ikz} dependence,where $k = \omega\sqrt{\mu\epsilon} = \omega\sqrt{LC}$, we deduce either from (2.3.13) or (2.3.14) that

$$\frac{V}{I} = \sqrt{\frac{L}{C}} = Z_0, \tag{2.3.16}$$

if the wave is only propagating in the positive z direction. Z_0 is also known as the *characteristic impedance* of a transmission line. Since C and L are dependent on the geometry of the transmission line, Z_0 is a geometry dependent impedance. This is unlike η, the intrinsic impedance. It can be easily shown from (2.3.13) and (2.3.14) that

$$\frac{d^2V}{dz^2} + \omega^2 LCV = 0, \tag{2.3.17}$$

$$\frac{d^2I}{dz^2} + \omega^2 LCI = 0, \tag{2.3.18}$$

which are one-dimensional scalar wave equations (or Helmholtz wave equations).

Energy Density and Power Flow

The time average energy stored per unit length in a transmission line for a single propagating wave is given by

$$\langle W_e\rangle = \frac{1}{4}\epsilon\int_S \mathbf{E}_s\cdot\mathbf{E}_s^* dS = \frac{1}{4}\epsilon\int_S (\nabla_s\phi_s)^2 dS, \tag{2.3.19}$$

$$\langle W_m\rangle = \frac{1}{4}\mu\int_S \mathbf{H}_s\cdot\mathbf{H}_s^* dS = \frac{1}{4}\mu\int_S (\nabla_s\psi_s)^2 dS, \tag{2.3.20}$$

where $\langle W_e\rangle$ and $\langle W_m\rangle$ are the time average energy stored in the electric field and the magnetic field respectively. Using the fact that $\nabla\cdot(\phi\nabla\phi) = (\nabla\phi)^2 + \phi\nabla^2\phi$, we can write

$$\langle W_e\rangle = \frac{1}{4}\epsilon\oint_{C_1+C_2}\phi\frac{\partial}{\partial n}\phi dl = \frac{1}{4}(V_1 - V_2)Q = \frac{1}{4}CV^2, \tag{2.3.21}$$

where $V = V_1 - V_2$, $\quad Q = CV$. The above could also be derived from circuit theory. As

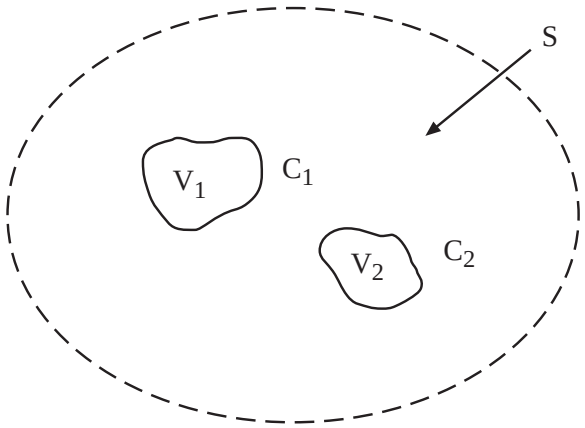


Figure 2.5: Power flow in a transmission line.

we have shown before, $|E_s| = \sqrt{\frac{\mu}{\epsilon}} H_s|$, hence, the time average energy stored in the magnetic field is

$$\langle W_m \rangle = \langle W_e \rangle = \frac{1}{4} C V^2. \tag{2.3.22}$$

Since $V^2 = Z_0^2 I^2 = \frac{L}{C} I^2$, and $\mu\epsilon = LC$, we can also write

$$\langle W_m \rangle = \frac{1}{4} L I^2. \tag{2.3.23}$$

Equation (2.3.23) can also be established by circuit theory. It also establishes our definition of L as an inductance per unit length.

The time average power flow down a transmission line is given by

$$\langle P \rangle = \frac{1}{2} \Re e \int_S dS \hat{z} \cdot (\mathbf{E}_s \times \mathbf{H}_s^*). \tag{2.3.24}$$

Since $\mathbf{H}_s = \sqrt{\frac{\epsilon}{\mu}} \hat{z} \times \mathbf{E}_s$ from (2.3.18), we have

$$\langle P \rangle = \frac{1}{2} \sqrt{\frac{\epsilon}{\mu}} \int_S dS |\mathbf{E}_s|^2 = 2v \langle W_e \rangle = v \langle W_e + W_m \rangle. \tag{2.3.25}$$

Hence, the time average stored energy $\langle W_e + W_m \rangle$ moving at velocity v, contributes to power flow. Equations (2.3.13) and (2.3.14) can also be derived from a circuit model.

2.3.2 Lossy Transmission Lines

Since the telegrapher's equations have strictly circuit theory interpretation, using circuit theory concept, the extension to a lossy transmission line is straight forward: we replace the

series impedance per unit length $-i\omega L$ by $-i\omega L + R$, and the shunt admittance per unit length $-i\omega C$ by $-i\omega C + G$ where R is the series resistance per unit length in the conductor, while G is the shunt conductance per unit length in the insulator. The telegrapher's equations then become

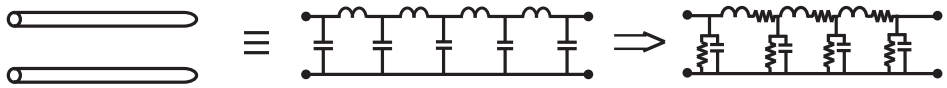


Figure 2.6: Circuits equivalence of a transmission line.

$$\frac{dV}{dz} = (i\omega L - R)I = -ZI, \tag{2.3.26}$$

$$\frac{dI}{dz} = (i\omega C - G)V = -YV. \tag{2.3.27}$$

The characteristic impedance is now

$$Z_0 = \sqrt{\frac{Z}{Y}}, \tag{2.3.28}$$

while the propagation constant becomes

$$k^2 = -(i\omega L - R)(i\omega C - G), \qquad k = \omega\sqrt{LC\left(1 + \frac{iR}{\omega L}\right)\left(1 + \frac{iG}{\omega C}\right)}. \tag{2.3.29}$$

Hence k is complex and the wave e^{ikz} is attenuating.

Strictly speaking, when we have a lossy transmission line due to conductor loss, a pure TEM wave cannot exist. This is because the axial current flow meets a resistance, and hence, an axial component of the electric field is necessary now to drive a current in the conductor. Therefore, the field is only quasi-TEM. However, the conductor loss can be thought of as a small perturbation of the perfect conductor case, and the electromagnetic field in the lossy line will not be vastly different from that of a lossless case.

The shunt conductance G in a lossy line can be found as follows. If the capacitance per unit length between two conductors is given by the formula

$$C = \epsilon K, \tag{2.3.30}$$

where K is a geometry dependent factor, the shunt admittance would be given by $Y = -i\omega C = -i\omega\epsilon K$. If now the dielectric medium is lossy so that $\epsilon = \epsilon' + \frac{i\sigma}{\omega}$, then the shunt admittance is given by

$$Y = -i\omega\epsilon' K + \sigma K. \tag{2.3.31}$$

Hence, we identify $G = \sigma K$. Note that the derivations in (2.3.6) to (2.3.16) hold true even if ϵ is complex. For this reason, Y in (2.3.31) is exact.

The series resistance R can be found by calculating the resistance of the conductor in a perturbative manner when it is lossy. The skin-effect will confine the current to flow only on the surface of the conductor. Since the skin depth in a conductor is $\delta = \sqrt{\frac{2}{\omega\mu\sigma}}$, the current is confined to flow in a thinner region at higher frequencies, hence, increasing this series resistance.

Another way of calculating transmission line loss is via a perturbation argument and the use of energy conservation. If a transmission line is lossy such that $k = k' + ik''$, and

$$V, I \sim e^{ik'z - k''z}, \tag{2.3.32}$$

Then, the power flow in a line, which is proportional to $|V|^2$ or $|I|^2$ is

$$P \sim e^{-2k''z}. \tag{2.3.33}$$

By energy conservation,

$$\frac{dP}{dz} = -P_d = -2k''P, \tag{2.3.34}$$

where P_d is the power dissipated per unit length on the line. Therefore, the attenuation constant k'' can be derived to be

$$k'' = \frac{P_d}{2P}, \tag{2.3.35}$$

if we know P_d. We can assume P to be close to that of a lossless line in using (2.3.35) in a perturbative concept.

Absence of TEM Mode in a Hollow Waveguide

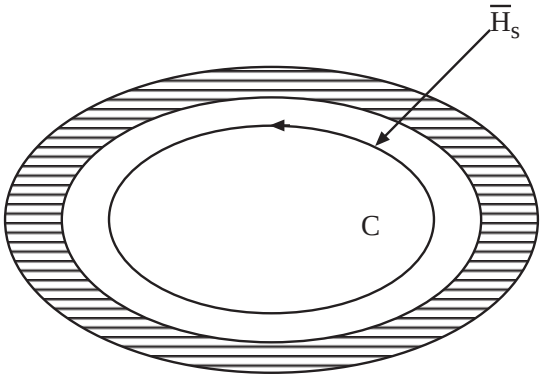


Figure 2.7: Absence of TEM mode in a hollow, enclosed waveguide.

Before ending this section, we would like to prove by contradiction that a hollow waveguide (i.e. without a center conductor) cannot support a TEM mode as follows. If we assume that

it does, then the magnetic field has to end on itself due to the absence of magnetic charges. It is clear that $\oint_C \mathbf{H}_s \cdot d\mathbf{l} \neq 0$ about any closed contour following the magnetic field lines. This is clearly in violation of Equation (2.3.1) for a TEM mode which implies that $\oint \mathbf{H}_s \cdot d\mathbf{l} = 0$ if C does not enclose any conducting current. These two results are contradictory implying the absence of a TEM mode in a hollow waveguide.

2.4 TE and TM Modes (H and E Modes)

2.4.1 Mode Orthogonality

Mode Orthogonality for Axial Fields

As shown previously, for TE and TM waves in a cylindrical waveguide, we characterize the waves by H_z and E_z respectively. The equations governing these two scalar field components are given by (2.1.11) and (2.1.12). In order to obtain a unique solution to (2.1.11) and (2.1.12), we have to specify the boundary conditions for H_z and E_z. For a metallic waveguide, we need to impose the boundary condition that $\hat{n} \times \mathbf{E} = 0$ on the metallic surface. This requires $E_z = 0$ on the metallic surface. From Equation (2.1.8), we see that if $E_z = 0$, and furthermore, if $\hat{n} \cdot \nabla_s H_z = \frac{\partial}{\partial n} H_z = 0$, then $\hat{n} \times \mathbf{E} = 0$ on the metal surface. Hence, the equations governing the TE and TM modes are

$$(\nabla_s^2 + k_s^2) H_z = 0, \qquad \frac{\partial}{\partial n} H_z = 0 \quad \text{on} \quad S, \quad \text{TE}, \tag{2.4.1}$$

$$(\nabla_s^2 + k_s^2) E_z = 0, \qquad E_z = 0 \quad \text{on} \quad S, \quad \text{TM}, \tag{2.4.2}$$

where S is the metallic surface. The homogeneous boundary conditions in Equation (2.4.1) is the homogeneous Neumann boundary condition, while that for Equation (2.4.2) is the homogeneous Dirichlet boundary condition. They are sufficient to uniquely determine the solutions to the partial differential equations.

For a closed waveguide, solutions exist for (2.4.1) and (2.4.2) at only discrete values of k_s^2. Furthermore, k_s^2 is purely real because (2.4.1) and (2.4.2) are self-adjoint problems (see Problem 2.6). This property is independent of the homogeneous material filling the waveguide, and hence is true even for a lossy isotropic material. It can be shown easily that different solutions of (2.4.1) or (2.4.2) corresponding to different k_{is}^2 are orthogonal, i.e.,

$$\int_S H_{iz} H_{jz} dS = C_{ih} \delta_{ij}, \qquad \text{TE}, \tag{2.4.3}$$

$$\int_S E_{iz} E_{jz} dS = C_{ie} \delta_{ij}, \qquad \text{TM}, \tag{2.4.4}$$

where the integration is over S, the cross-section of the waveguide. We can prove the above assertion quite easily, e.g., by writing down

$$(\nabla_s^2 + k_{is}^2) \psi_{iz} = 0, \tag{2.4.5}$$

$$(\nabla_s^2 + k_{js}^2)\psi_{jz} = 0, \tag{2.4.6}$$

where ψ_z in this case can be either H_z or E_z. Multiplying the first equation by ψ_{jz} and the second equation by ψ_{iz}, subtracting the two equations, and integrating over S, we have

$$(k_{is}^2 - k_{js}^2) \int_S \psi_{iz}\psi_{jz} dS = \int_S (\psi_{iz}\nabla_s^2\psi_{jz} - \psi_{jz}\nabla_s^2\psi_{iz}) dS,$$

$$\text{Gauss' theorem} \longrightarrow = \oint_C \hat{n} \cdot (\psi_{iz}\nabla_s\psi_{jz} - \psi_{jz}\nabla_s\psi_{iz}) dl. \tag{2.4.7}$$

The above usage of Gauss' theorem is also known as Green's theorem. With either a homogeneous Neumann, Dirichlet or mixed boundary condition (Neumann on one part of C, and Dirichlet on the other parts), the right-hand side of (2.4.7) vanishes, and we have

$$(k_{is}^2 - k_{js}^2) \int_S \psi_{iz}\psi_{jz} dS = 0. \tag{2.4.8}$$

For $i \neq j$, we have

$$\int_S \psi_{iz}\psi_{jz} dS = 0. \tag{2.4.9}$$

Since k_{js}^2 can be shown to be pure real [see Problem 2-6], ψ_{jz}^* is also a solution of (2.4.6). Hence, we can further say that $\int_S \psi_{iz}\psi_{jz}^* dS = 0, \quad i \neq j$, since ψ_{jz}^* is also a solution to (2.4.6). The property described by (2.4.3) and (2.4.4), is known as the mode orthogonality of the axial components of the field. The real-value of $k_j s^2$ is related to that the operator ∇_s^2 is a Hermitian operator under appropriate boundary condition of the field.

A point is in order regarding the orthogonality relations listed in (2.4.3) and (2.4.4). When $i = j$ in (2.4.3),

$$\int_S H_{iz}H_{iz} = C \tag{2.4.10}$$

It is possible that $C = 0$ if H_{iz} is a complex function. Hence, it is prudent to rewrite the orthogonality relations as

$$\int_S H_{iz}H_{jz}^* dS = C_{ih}\delta_{ij}, \qquad \text{TE}, \tag{2.4.11}$$

$$\int_S E_{iz}E_{jz}^* dS = C_{ie}\delta_{ij}, \qquad \text{TM}, \tag{2.4.12}$$

In this manner, C_{ih} and C_{ie} are ensured to be positive real values, and they can be used to normalize the modes yielding the orthonormal relations that

$$\int_S H_{iz}H_{jz}^* dS = \delta_{ij}, \qquad \text{TE}, \tag{2.4.13}$$

$$\int_S E_{iz}E_{jz}^* dS = \delta_{ij}, \qquad \text{TM}, \tag{2.4.14}$$

Mode Orthogonality for Transverse Fields

The transverse components of **E** or **H** fields are also orthogonal. This property can be proven a number of ways. One way is to relate their orthogonality to the orthogonality of the scalar wave functions mentioned before. However, to demonstrate that the orthogonality is also related to the symmetry of the differential equation, which will prove this property as follows: the **E**-field of either the TE or the TM mode of a waveguide satisfies

$$\nabla \times \nabla \times \mathbf{E}_i - k^2 \mathbf{E}_i = 0. \tag{2.4.15}$$

Assuming that the field has $e^{ik_z z}$ dependence, and extracting the transverse component of the above equation, we have, for any two distinct modes,

$$\nabla_s \times \nabla_s \times \mathbf{E}_{is} - \nabla_s \nabla_s \cdot \mathbf{E}_{is} - k_{is}^2 \mathbf{E}_{is} = 0, \tag{2.4.16}$$

$$\nabla_s \times \nabla_s \times \mathbf{E}_{js} - \nabla_s \nabla_s \cdot \mathbf{E}_{js} - k_{js}^2 \mathbf{E}_{js} = 0, \tag{2.4.17}$$

where $k_{is}^2 = k^2 - k_{iz}^2, \quad k_{js}^2 = k^2 - k_{jz}^2$. Dot multiplying (2.4.16) by $\mathbf{E}_{js}$, (2.4.17) by $\mathbf{E}_{is}$, and subtracting, we have, after integrating,

$$\begin{aligned}(k_{js}^2 - k_{is}^2) &\int_S \mathbf{E}_{is} \cdot \mathbf{E}_{js} dS \\ &= \int_S dS(\mathbf{E}_{js} \cdot \nabla_s \times \nabla_s \times \mathbf{E}_{is} - \mathbf{E}_{is} \cdot \nabla_s \times \nabla_s \times \mathbf{E}_{js}) \\ &- \int_S dS(\mathbf{E}_{js} \cdot \nabla_s \nabla_s \cdot \mathbf{E}_{is} - \mathbf{E}_{is} \cdot \nabla_s \nabla_s \cdot \mathbf{E}_{js}).\end{aligned} \tag{2.4.18}$$

By noting that $\nabla \cdot (\mathbf{A} \times \mathbf{B}) = \mathbf{B} \cdot \nabla \times \mathbf{A} - \mathbf{A} \cdot \nabla \times \mathbf{B}$, and hence,

$$\begin{aligned}-\nabla_s \cdot (\mathbf{E}_{js} \times \nabla_s \times \mathbf{E}_{is} - \mathbf{E}_{is} \times \nabla_s \times \mathbf{E}_{js}) \\ = \mathbf{E}_{js} \cdot \nabla_s \times \nabla_s \times \mathbf{E}_{is} - \mathbf{E}_{is} \cdot \nabla_s \times \nabla_s \times \mathbf{E}_{js},\end{aligned} \tag{2.4.19}$$

and that $\nabla \cdot (\mathbf{A}\phi) = \phi \nabla \cdot \mathbf{A} + \mathbf{A} \cdot \nabla \phi$, and hence,

$$\nabla_s \cdot (\mathbf{E}_{js} \nabla_s \cdot \mathbf{E}_{is} - \mathbf{E}_{is} \nabla_s \cdot \mathbf{E}_{js}) = \mathbf{E}_{js} \cdot \nabla_s \nabla_s \cdot \mathbf{E}_{is} - \mathbf{E}_{is} \cdot \nabla_s \nabla_s \cdot \mathbf{E}_{js}, \tag{2.4.20}$$

we can convert the right-hand side of (2.4.18) into line integrals using Gauss' theorem in two dimensions, giving

$$\begin{aligned}(k_{js}^2 - k_{is}^2) &\int_S \mathbf{E}_{is} \cdot \mathbf{E}_{js} dS \\ &= -\oint_C dl\ \hat{n} \cdot [\mathbf{E}_{js} \times (\nabla_s \times \mathbf{E}_{is}) - \mathbf{E}_{is} \times (\nabla_s \times \mathbf{E}_{js})] \\ &- \oint_C dl\ \hat{n} \cdot [\mathbf{E}_{js} \nabla_s \cdot \mathbf{E}_{is} - \mathbf{E}_{is} \nabla_s \cdot \mathbf{E}_{js}].\end{aligned} \tag{2.4.21}$$

The right-hand side of (2.4.21) vanishes by virtue of the boundary condition. Since $\hat{n} \cdot [\mathbf{E}_{js} \times (\nabla_s \times \mathbf{E}_{is})] = \hat{n} \times \mathbf{E}_{js} \cdot \nabla_s \times \mathbf{E}_{is}$,[1] $\hat{n} \times \mathbf{E}_s = 0$ implies the zero of the first integral on the right hand size of (2.4.21). Furthermore, $\nabla_s \cdot \mathbf{E}_{is} = -ik_z E_z = 0$ on the waveguide wall implies the zero of the second term in the right-hand side of (2.4.21). Therefore,

$$\int_S \mathbf{E}_{is} \cdot \mathbf{E}_{js} dS = 0, \qquad i \neq j, \tag{2.4.22}$$

for any two distinct modes with different k_s^2, irrespective of whether they are TE or TM modes. A similar proof follows for $\mathbf{H}_s$, i.e.,

$$\int_S \mathbf{H}_{is} \cdot \mathbf{H}_{js} dS = 0, \qquad i \neq j. \tag{2.4.23}$$

We can further show that

$$\int_S \mathbf{F}_{is} \cdot \mathbf{F}_{js}^* \, dS = 0, \quad i \neq j, \tag{2.4.24}$$

where $\mathbf{F}$ is either $\mathbf{E}$ or $\mathbf{H}$. This is because $\mathbf{E}^*$ or $\mathbf{H}^*$ is also a solution of (2.4.16). Moreover, it is prudent to write the orthogonality relations of these fields as

$$\int_S \mathbf{F}_{is} \cdot \mathbf{F}_{js}^* \, dS = C_j \delta_{ij}, \quad i \neq j, \tag{2.4.25}$$

so that C_j is guaranteed to be positive real and these functions can be orthonormalized.

The orthogonality of modes in (2.4.22) follows from the fact that the differential operator in (2.4.16) is symmetric with the defined boundary conditions, i.e.,

$$\langle \mathbf{E}_{js}, (\nabla_s \times \nabla_s \times \; - \; \nabla_s \nabla_s \cdot) \mathbf{E}_{is} \rangle = \langle \mathbf{E}_{is}, (\nabla_s \times \nabla_s \times \; - \; \nabla_s \nabla_s \cdot) \mathbf{E}_{js} \rangle. \tag{2.4.26}$$

It is analogous to the fact that eigenvectors of a symmetric matrix with distinct eigenvalues are orthogonal. Moreover, since these operators are real, they are also Hermitian operators with real eigenvalues k_{js}^2.

When the medium is inhomogeneous, (2.4.16) is not valid for describing the field, and the differential operators are not symmetric anymore. For an inhomogeneously filled waveguide, (2.4.22) and (2.4.23) are not true in general.

In the preceding proof, we can also decompose the transverse fields into their TE and TM components, and express them in terms of E_z and H_z. The orthogonality of the transverse fields can hence be related to the orthogonality of the axial fields.

Mode Orthogonality for Reaction

A more general orthogonality condition which we shall derive later, and is true even for inhomogeneously filled waveguides is

$$\int_S (\mathbf{E}_{is} \times \mathbf{H}_{js}) \cdot \hat{z} dS = 0, \qquad i \neq j. \tag{2.4.27}$$

[1] This follows from $\mathbf{a} \cdot \mathbf{b} \times \mathbf{c} = \mathbf{c} \cdot \mathbf{a} \times \mathbf{b} = \mathbf{b} \cdot \mathbf{c} \times \mathbf{a}$.

The above is the reaction as is used in the Lorentz reciprocity theorem. To prove (2.4.27) for hollow waveguides, we first write down the equations satisfied by $\mathbf{E}_{is}$ and $\mathbf{H}_{js}$, i.e.,

$$\nabla_s \times \nabla_s \times \mathbf{E}_{is} - \nabla_s \nabla_s \cdot \mathbf{E}_{is} - k_{is}^2 \mathbf{E}_{is} = 0, \tag{2.4.28}$$

$$\nabla_s \times \nabla_s \times \mathbf{H}_{js} - \nabla_s \nabla_s \cdot \mathbf{H}_{js} - k_{js}^2 \mathbf{H}_{js} = 0. \tag{2.4.29}$$

We cross multiply (2.4.28) by $\mathbf{H}_{js}$ and (2.4.29) by $\mathbf{E}_{is}$. Upon subtraction and integration, we have

$$\begin{aligned}(k_{js}^2 - k_{is}^2) &\int_S (\mathbf{E}_{is} \times \mathbf{H}_{js}) \cdot \hat{z} dS \\ &= \int_S [(\nabla_s \times \nabla_s \times \mathbf{E}_{is}) \times \mathbf{H}_{js} - \mathbf{E}_{is} \times (\nabla_s \times \nabla_s \times \mathbf{H}_{js})] \cdot \hat{z} dS \\ &- \int_S [(\nabla_s \nabla_s \cdot \mathbf{E}_{is}) \times \mathbf{H}_{js} - \mathbf{E}_{is} \times (\nabla_s \nabla_s \cdot \mathbf{H}_{js})] \cdot \hat{z} dS.\end{aligned} \tag{2.4.30}$$

Using the fact that

$$\begin{aligned}(\nabla_s \times \nabla_s \times \mathbf{E}_s) \times \mathbf{H}_s \cdot \hat{z} &= (\mathbf{H}_s \times \hat{z}) \cdot (\nabla_s \times \nabla_s \times \mathbf{E}_s) \\ &= \nabla_s \cdot [(\nabla_s \times \mathbf{E}_s) \times (\mathbf{H}_s \times \hat{z})] \\ &\quad + (\nabla_s \times \mathbf{E}_s) \cdot \nabla_s \times (\mathbf{H}_s \times \hat{z})\end{aligned} \tag{2.4.31}$$

and that

$$\begin{aligned}\hat{z} \cdot \mathbf{E}_s \times \nabla_s \nabla_s \cdot \mathbf{H}_s &= (\hat{z} \times \mathbf{E}_s) \cdot \nabla_s \nabla_s \cdot \mathbf{H}_s \\ &= \nabla_s \cdot (\nabla_s \cdot \mathbf{H}_s \hat{z} \times \mathbf{E}_s) - \nabla_s \cdot \mathbf{H}_s \nabla_s \cdot \hat{z} \times \mathbf{E}_s,\end{aligned} \tag{2.4.32}$$

plus the fact that $\nabla_s \times (\mathbf{H}_s \times \hat{z}) = -\hat{z} \nabla_s \cdot \mathbf{H}_s$, and that $\nabla_s \cdot \hat{z} \times \mathbf{E}_s = -\hat{z} \cdot \nabla_s \times \mathbf{E}_s$, it can be seen that the last terms in (2.4.31) and (2.4.32) are identical except for a sign difference. Therefore, after using Gauss' theorem,

$$\begin{aligned}\int_S dS[(\nabla_s \times \nabla_s \times \mathbf{E}_s) \times \mathbf{H}_s + \mathbf{E}_s \times \nabla_s \nabla_s \cdot \mathbf{H}_s] \cdot \hat{z} = \oint_C dl \hat{n} \cdot [(\nabla_s \times \mathbf{E}_s) \\ \times (\mathbf{H}_s \times \hat{z}) + (\nabla_s \cdot \mathbf{H}_s) \hat{z} \times \mathbf{E}_s].\end{aligned} \tag{2.4.33}$$

Since $\nabla_s \times \mathbf{E}_s = i\omega\mu \mathbf{H}_z$ and $\nabla_s \cdot \mathbf{H}_s = -ik_z H_z$, the right-hand side vanishes if we have either an electric wall or a magnetic wall or a mixture thereof. Similarly, the other terms on the right hand side of (2.4.30) vanish by the same argument. Therefore, in general,

$$\int_S (\mathbf{E}_{is} \times \mathbf{H}_{js}) \cdot \hat{z} dS = 0, \qquad i \neq j, \tag{2.4.34}$$

for any two distinct modes with different propagation constants k_z or k_s.

Also, the transverse fields can be related to the axial fields, and their orthogonality can also be related to the orthogonality of the axial fields. We shall show later a more general proof of the above using Lorentz reciprocity theorem. This proof is even valid for inhomogeneously-filled waveguides.

Power Orthogonality

Since, $\mathbf{H}_{js}^*$ is also a solution to (2.4.29), we have

$$\int_S (\mathbf{E}_{is} \times \mathbf{H}_{js}^*) \cdot \hat{z} dS = 0, \qquad i \neq j, \qquad \text{power orthogonality.} \tag{2.4.35}$$

Since $\mathbf{E} \times \mathbf{H}^*$ represents the complex Poynting vector, Equation (2.4.35) implies that the power flow in a waveguide is independently carried by each mode. Cross interactions between the **E** and **H** fields of two different modes do not result in power flow as testified by Equation (2.4.35).

The above orthogonality principles assume that the modes have distinct eigenvalues k_{is}^2 or distinct axial propagation constants k_{iz}^2. When k_{is}^2 for two different modes are the same, the modes are called degenerate. If there are N degenerate, independent modes, we can use the Gram-Schmidt orthogonalization procedure to obtain N orthogonal modes if we so desire.

2.5 Rectangular Waveguides

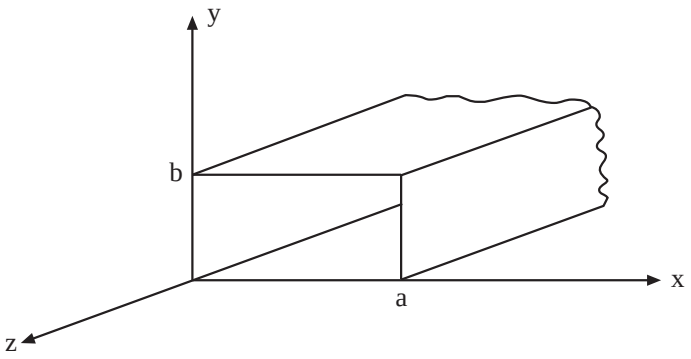


Figure 2.8: A rectangular waveguide.

The rectangular waveguide is the most commonly used hollow waveguide. By adjusting the aspect ratio, a to b, of the waveguide, one can obtain a good bandwidth for single mode propagation. Furthermore, the analysis of this waveguide is simple.

2.5.1 TE Modes (H Modes)

A TE mode in a rectangular waveguide is characterized by H_z satisfying Equation (2.4.1) with the requisite Neumann boundary condition. An H_z that will satisfy (2.4.1) with $\frac{\partial}{\partial n} H_z = 0$ on the waveguide wall is

$$H_z = H_0 \cos\left(\frac{m\pi}{a} x\right) \cos\left(\frac{n\pi}{b} y\right) e^{ik_z z}, \tag{2.5.1}$$

where $k_s^2 = (\frac{m\pi}{a})^2 + (\frac{n\pi}{b})^2$, $k_z = \sqrt{k^2 - k_s^2}$. The fact that k_s^2 has to satisfy the prescribed values is also known as the guidance condition. The transverse fields can be found using (2.1.7) and (2.1.8). The mode becomes evanescent or non-propagating when $k_s^2 > k^2$, i.e., when k_z becomes imaginary. Since $k^2 = \omega^2\mu\epsilon$, the cutoff frequency (the frequency below which the mode is evanescent) is given by

$$\omega_{mnc} = \frac{1}{\sqrt{\mu\epsilon}} \left[\left(\frac{m\pi}{a}\right)^2 + \left(\frac{n\pi}{b}\right)^2 \right]^{\frac{1}{2}}. \tag{2.5.2}$$

The corresponding mode is usually labeled as TE_{mn} (or H_{mn}) mode. The wavelength of a wave at ω_c in the medium denoted by μ, ϵ is the cutoff wavelength. It is

$$\lambda_{mnc} = \frac{2}{[(\frac{m}{a})^2 + (\frac{n}{b})^2]^{\frac{1}{2}}}. \tag{2.5.3}$$

In waveguide conventions, a is assumed larger than b. Then, the dominant mode (fundamental mode) with the lowest cutoff frequency is the $m = 1$, $n = 0$ mode, also known as the TE_{10} mode (or H_{10} mode). The TE_{00} mode does not exist because, in this case, $k_s = 0$, and $\mathbf{E}_s$ and $\mathbf{H}_s$ diverge from (2.1.7) and (2.1.8) when $k_s \to 0$.

2.5.2 TM Modes (E Modes)

A TM mode in a rectangular waveguide is characterized by E_z satisfying Equation (2.4.2) with the requisite Dirichlet boundary condition. An E_z that will satisfy (2.4.2) with $E_z = 0$ on the waveguide wall is

$$E_z = E_0 \sin\left(\frac{m\pi}{a}x\right) \sin\left(\frac{n\pi}{b}y\right) e^{ik_z z}, \tag{2.5.4}$$

where $k_s^2 = (\frac{m\pi}{a})^2 + (\frac{n\pi}{b})^2$, $k_z = \sqrt{k^2 - k_s^2}$. The TM_{mn} mode has the same cutoff frequency as the TE_{mn} mode. However, when either $m = 0$, or $n = 0$, the mode does not exist since $E_z = 0$ then. Therefore, the lowest TM mode is the TM_{11} mode with a cutoff frequency above that of the TE_{10} mode. Given the z components of the fields, all other field components of a waveguide can be derived. Figure 2.20 shows the field plots of some modes of a rectangular waveguide [8].[2]

2.6 Circular Waveguides

Certain modes of a circular waveguide have less attenuation from wall loss compared to a rectangular waveguide. Hence, it is sometimes preferred over a rectangular waveguide.

2.6.1 TE Modes (H Modes)

The H_z component of a TE mode satisfies Equation (2.4.1) in cylindrical coordinates, i.e.,

$$\left[\frac{1}{\rho}\frac{\partial}{\partial\rho}\rho\frac{\partial}{\partial\rho} + \frac{1}{\rho^2}\frac{\partial^2}{\partial\phi^2} + k_s^2\right] H_z = 0, \tag{2.6.1}$$

[2]The plots here are reproduced by A. Greenwood according to this reference.

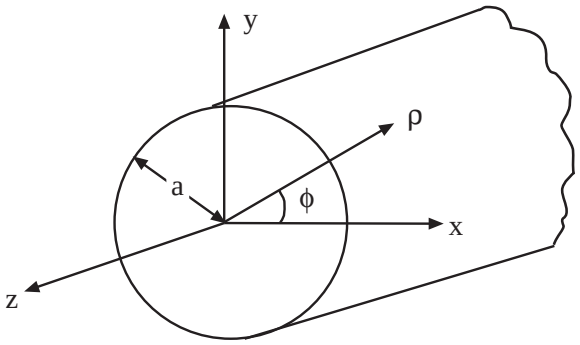


Figure 2.9: A circular waveguide.

with the boundary condition that $\frac{\partial}{\partial\rho}H_z = 0$, $\rho = a$, where a is the radius of the waveguide. If we assume that H_z has $e^{\pm in\phi}$, $\sin n\phi$ or $\cos n\phi$ dependence, where n is an integer, we can replace $\frac{\partial^2}{\partial\phi^2}$ by $-n^2$. Equation (2.6.1) then becomes

$$\left[\frac{1}{\rho}\frac{\partial}{\partial\rho}\rho\frac{\partial}{\partial\rho} - \frac{n^2}{\rho^2} + k_s^2\right] H_z = 0. \tag{2.6.2}$$

Equation (2.6.2) is the Bessel equation, whose solutions are either $J_n(k_s\rho)$, $Y_n(k_s\rho)$, $H_n^{(1)}(k_s\rho)$ or $H_n^{(2)}(k_s\rho)$. Of these four solutions, only two are independent because of the relations

$$H_n^{(1)}(k_s\rho) = J_n(k_s\rho) + iY_n(k_s\rho), \tag{2.6.3a}$$

$$H_n^{(2)}(k_s\rho) = J_n(k_s\rho) - iY_n(k_s\rho), \tag{2.6.3b}$$

$J_n(k_s\rho)$ is regular about the origin when $\rho \to 0$, but $Y_n(k_s\rho)$ is singular (so are $H_n^{(1)}(k_s\rho)$ and $H_n^{(2)}(k_s\rho)$). Since the field cannot be infinite at the center of the waveguide due to the absence of sources, the solution to (2.6.1) is of the form

$$H_z = H_0 J_n(k_s\rho)e^{\pm in\phi + ik_z z}. \tag{2.6.4}$$

We require that $\frac{\partial}{\partial\rho}H_z = 0$ at $\rho = a$, implying that

$$J_n'(k_s a) = 0, \tag{2.6.5}$$

with $k_z = \sqrt{k^2 - k_s^2}$. If the m-th zero of $J_n'(x)$ is defined to be β_{nm} such that $J_n'(\beta_{nm}) = 0$, the values of possible k_s are

$$k_s = \frac{\beta_{nm}}{a}. \tag{2.6.6}$$

The above is also the guidance condition for the waveguide mode. The subscript n denotes the orders of the Bessel function $J_n(x)$ and the circular harmonic $e^{\pm in\phi}$. The subscript m denotes the m-th zero of $J_n'(x)$ discounting the zero at the origin. The corresponding mode

is usually denoted as the TE_{nm} mode. Cutoff occurs when $k = \omega\sqrt{\mu\epsilon} < k_s$. From Figure 2.10, we see that the TE_{11} mode corresponding to the first zero of $J_1(x)$ has the lowest cutoff frequency. The cutoff frequency for the TE_{nm} mode is given by

$$\omega_{nmc} = \frac{1}{\sqrt{\mu\epsilon}}\frac{\beta_{nm}}{a}, \tag{2.6.7}$$

and the corresponding cutoff wavelength is

$$\lambda_{nmc} = \frac{2\pi}{\beta_{nm}}a. \tag{2.6.8}$$

2.6.2 TM Modes (E Modes)

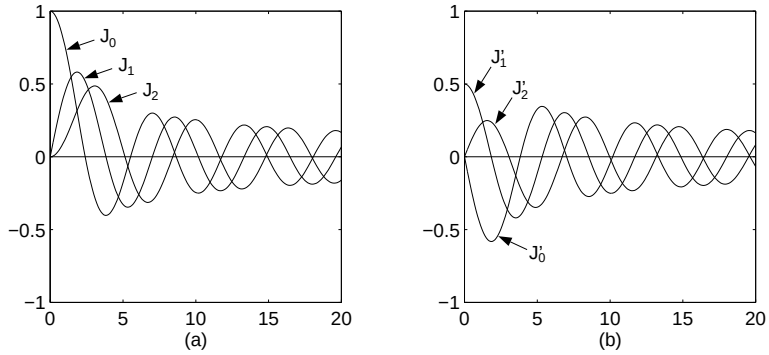


Figure 2.10: Plots of Bessel functions and their derivatives.

Similar to a TE mode, the E_z component of a TM mode has $e^{\pm in\phi}$ dependence, satisfying the equation

$$\left[\frac{1}{\rho}\frac{\partial}{\partial\rho}\rho\frac{\partial}{\partial\rho} - \frac{n^2}{\rho^2} + k_s^2\right]E_z = 0, \tag{2.6.9}$$

with the boundary condition $E_z(\rho = a) = 0$. Hence,

$$E_z = E_0 J_n(k_s\rho)e^{\pm in\phi + ik_z z}, \tag{2.6.10}$$

with $J_n(k_s a) = 0$. If we denote the m-th zero of $J_n(x)$ by α_{nm}, then k_s has possible values of

$$k_s = \frac{\alpha_{nm}}{a}, \tag{2.6.11}$$

where the subscript n denotes the order of the Bessel function, $J_n(x)$, and the subscript m denotes the m-th zero of $J_n(x)$, discounting the zero at the origin. It is also the guidance condition, and the corresponding mode is known as the TM_{nm} mode. The cutoff frequency of the TM_{nm} mode is given by

$$\omega_{nmc} = \frac{1}{\sqrt{\mu\epsilon}}\frac{\alpha_{nm}}{a}, \tag{2.6.12}$$

and the corresponding cutoff wavelength is

$$\lambda_{nmc} = \frac{2\pi}{\alpha_{nm}} a. \tag{2.6.13}$$

Looking at Figure 2.10, we see that the lowest TM mode is the TM_{01} mode, but it has a higher cutoff frequency compared to the TE_{11} mode.

Table 2.3.1. Roots of $J_n'(x) = 0$.

n	β_{n1}	β_{n2}	β_{n3}	β_{n4}
0	3.832	7.016	10.174	13.324
1	1.841	5.331	8.536	11.706
2	3.054	6.706	9.970	13.170
3	4.201	8.015	11.346	14.586
4	5.318	9.282	12.682	15.964
5	6.416	10.520	13.987	17.313

Table 2.3.2. Roots of $J_n(x) = 0$.

n	α_{n1}	α_{n2}	α_{n3}	α_{n4}
0	2.405	5.520	8.654	11.792
1	3.832	7.016	10.174	13.324
2	5.135	8.417	11.620	14.796
3	6.380	9.761	13.015	16.223
4	7.588	11.065	14.373	17.616
5	8.771	12.339	15.700	18.980

It can be shown that the TE_{01} mode has the lowest loss at high frequencies. The TE_{01} mode is axially symmetric with $\mathbf{E} = \hat{\phi} E_\phi$. Therefore, it can be enhanced by various means. One way is to use a mode filter as shown in Figure 2.11(a) [9]. The radial conducting wire will short out modes with radial components of the electric field. However, the TE_{01} mode is oblivious to the presence of the radial conducting strips, and is little affected. Another way to enhance the TE_{01} mode in a circular waveguide is to use a ribbed waveguide. For the TE_{01} mode, the current is purely circumferential and is oblivious to the presence of the ribs. However, the TM modes, which have axial components of the current, will be affected by the ribbed wall of the waveguide. In other words, the ribbed waveguide wall does not support the axial current flow effectively. If the periodicity of the waveguide corrugation is small compared to the wavelength, the TE_{01} mode will not be affected much.

The transverse field components of a circular waveguide are easily obtained given the axial components. Figure 2.21 shows the field plots of some modes in a circular waveguide [8].

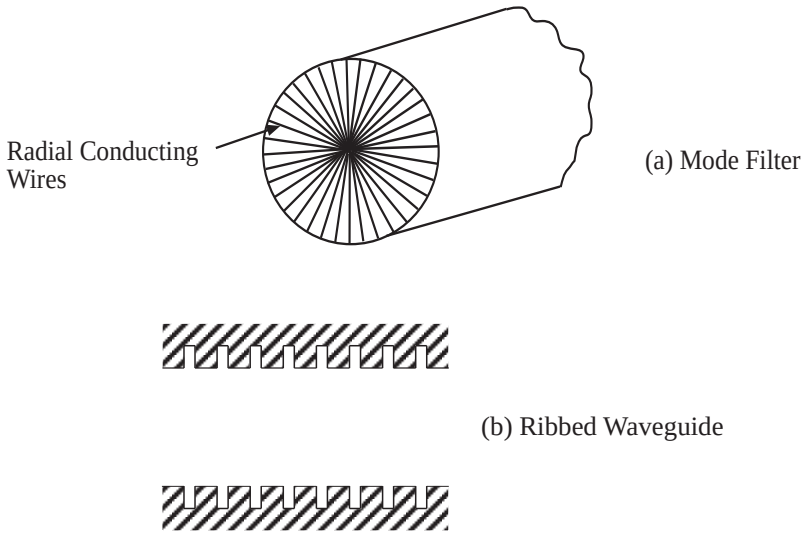


Figure 2.11: (a) A mode filter used to filter out most other modes, but it allows the TE_{01} modes to pass through. (b) A ribbed waveguide prevents axial current flow, hence attenuates the TM modes, which has $E_z \neq 0$.

2.7 Power Flow in a Waveguide

Because of the power orthogonality of the distinct modes in a waveguide with perfectly impenetrable walls (perfect electric conductors or perfect magnetic conductors), we can study the power flow due to each mode and the total power flow is the sum of the power flow from each mode.

2.7.1 Power Flow and Group Velocity

The time average power flow in a waveguide is given by

$$P_f = \frac{1}{2}\Re e \int_S (\mathbf{E}_s \times \mathbf{H}_s^*) \cdot \hat{z} dS. \tag{2.7.1}$$

For a TE mode with $e^{ik_z z}$ dependence, from (2.1.8),

$$\mathbf{E}_s = \frac{i\omega\mu}{k_s^2} \nabla_s \times \hat{z} H_z, \tag{2.7.2}$$

$$\mathbf{H}_s = \frac{ik_z}{k_s^2} \nabla_s H_z. \tag{2.7.3}$$

Substituting into (2.7.1), we can show that

$$P_f = \frac{1}{2}\Re e\left\{\frac{\omega\mu k_z^*}{k_s^4}\hat{z}\int_S (\nabla_s H_z)\cdot(\nabla_s H_z^*)dS\right\}. \tag{2.7.4}$$

(k_s^2 is always real in a homogeneously-filled waveguide). Using the fact that $\nabla_s\cdot(H_z\nabla_s H_z^*) = (\nabla_s H_z)\cdot(\nabla_s H_z^*) + H_z\nabla_s^2 H_z^*$, the above becomes

$$P_f = \frac{1}{2}\Re e\left\{\frac{\omega\mu k_z^*}{k_s^4}\hat{z}\left[\int_C dl\hat{n}\cdot H_z\nabla_s H_z^* - \int_S dS H_z\nabla_s^2 H_z^*\right]\right\}. \tag{2.7.5}$$

The first integral vanishes by virtue of the boundary condition. While using $\nabla_s^2 H_z^* = -k_s^2 H_z^*$, we have

$$P_f = \frac{\hat{z}}{2}\Re e\left\{\frac{\omega\mu k_z^*}{k_s^2}\int_S dS|H_z|^2\right\}. \tag{2.7.6}$$

For a TM mode, we can similarly show that

$$P_f = \frac{\hat{z}}{2}\Re e\left\{\frac{\omega\epsilon^* k_z}{k_s^2}\int_S dS|E_z|^2\right\}. \tag{2.7.7}$$

When the mode is cutoff, i.e., when $k^2 < k_s^2$, so that $k_z = \sqrt{k^2 - k_s^2}$ is pure imaginary, and the waveguide is lossless, there is no time average power flow down the waveguide. In such a case, the mode is evanescent.

The time average energy stored per unit length in the electric field for the TE mode is given by

$$\begin{aligned}\langle W_e\rangle &= \frac{1}{4}\epsilon\int_S \mathbf{E}_s\cdot\mathbf{E}_s^* dS \\ &= \frac{\omega^2|\mu|^2\epsilon}{4k_s^4}\int_S (\nabla_s H_z)\cdot(\nabla_s H_z^*)dS.\end{aligned} \tag{2.7.8}$$

In the same manner as (2.7.4), we can show that

$$\langle W_e\rangle = \frac{\omega^2|\mu|^2\epsilon}{4k_s^2}\int_S |H_z|^2 dS. \tag{2.7.9}$$

It can be shown that $\langle W_e\rangle = \langle W_m\rangle$ for a lossless waveguide where $\langle W_m\rangle$ is the time average energy stored in the magnetic field. Therefore, the total time average energy stored per unit length is

$$\langle W_T\rangle = \langle W_e\rangle + \langle W_m\rangle = \frac{\omega^2\mu^2\epsilon}{2k_s^2}\int_S |H_z|^2 dS. \tag{2.7.10}$$

For a lossless waveguide with a propagating mode, k_z is pure real and (2.7.6) for the TE mode becomes

$$P_f = \frac{\hat{z}}{2}\frac{\omega\mu k_z}{k_s^2}\int_S dS|H_z|^2. \tag{2.7.11}$$

Comparing (2.7.10) and (2.7.11), we note that for a lossless waveguide,

$$P_f = \frac{k_z}{\omega\mu\epsilon}\langle W_T\rangle, \tag{2.7.12}$$

where $k_z/\omega\mu\epsilon$ has the dimension of velocity. In a waveguide,

$$k_z^2 = \omega^2\mu\epsilon - k_s^2. \tag{2.7.13}$$

The group velocity in a waveguide is derived to be

$$v_g = \frac{d\omega}{dk_z} = \frac{k_z}{\omega\mu\epsilon}. \tag{2.7.14}$$

Therefore, (2.7.12) is just

$$P_f = v_g\langle W_T\rangle. \tag{2.7.15}$$

In other words, in a lossless waveguide, the time average energy stored per unit length, moving at the group velocity v_g contributes to the power flow. The group velocity in a waveguide is the velocity of energy propagation, and it is also the signal velocity.

The phase velocity of a wave in a waveguide is defined to be

$$v_{ph} = \frac{\omega}{k_z} = \frac{\omega}{\sqrt{k^2 - k_s^2}}. \tag{2.7.16}$$

It is the velocity of the phase of the wave. Since a signal does not travel at the phase velocity, it could be larger than the speed of light. This happens near cutoff when $k^2 \to k_s^2$ for a mode. Group velocity or signal velocity cannot be larger than the speed of light, a limit dictated by Einstein's theory of special relativity. Note that $v_{ph}v_g = c^2$ where c is the velocity of light in the medium.

2.7.2 Pulse Propagation in a Waveguide

Since the phase and the group velocities inside a hollow waveguide are frequency dispersive, a pulse propagating inside a hollow waveguide will be distorted due to frequency dispersion. Different frequency components will travel with different velocities. Hence, after a certain distance of propagation, different Fourier components lose their phase coherence, causing pulse distortion. Therefore, to minimize pulse distortion, the bandwidth of the pulse should be narrow. It can be shown that for a narrow-band pulse, the envelope of the pulse propagates with the group velocity while the carrier signal propagates with the phase velocity.

A narrow-band pulse can be written as:

$$\begin{aligned} p(\mathbf{r},t) &= \frac{1}{2\pi}\int_{-\infty}^{\infty} d\omega\tilde{p}(\mathbf{r},\omega)e^{-i\omega t} \\ &= \frac{1}{\pi}\Re e\int_0^{\infty} d\omega\tilde{p}(\mathbf{r},\omega)e^{-i\omega t} \end{aligned} \tag{2.7.17}$$

where $p(\mathbf{r},t)$ may represent a component (e.g., the z component) of the electric field or magnetic field inside a waveguide. Since

$$\left(\nabla^2 - \frac{1}{c^2}\frac{\partial^2}{\partial t^2}\right) p(\mathbf{r},t) = 0, \tag{2.7.18}$$

by substituting (2.7.17) into (2.7.18), we require that

$$\left(\nabla^2 + \frac{\omega^2}{c^2}\right) \tilde{p}(\mathbf{r},\omega) = 0, \tag{2.7.19}$$

If $p(\mathbf{r},t)$ corresponds to a particular mode, for example, in the case of a single mode propagation, then

$$\left(\nabla_s^2 + k_s^2\right) \tilde{p}(\mathbf{r},\omega) = 0, \tag{2.7.20}$$

and subtracting (2.7.19) from (2.7.20), we have

$$\left(\frac{\partial^2}{\partial z^2} + k_z^2\right) \tilde{p}(\mathbf{r},\omega) = 0, \tag{2.7.21}$$

where $k_z^2 = \frac{\omega^2}{c^2} - k_s^2$. Hence

$$\tilde{p}(\mathbf{r},\omega) = \tilde{p}_o(\mathbf{r}_s,\omega) e^{ik_z z}. \tag{2.7.22}$$

where $\mathbf{r}_s = \hat{x}x + \hat{y}y$. Consequently, we can rewrite (2.7.17) as

$$p(\mathbf{r},t) = \frac{1}{\pi}\Re e \int_0^\infty d\omega \tilde{p_o}(\mathbf{r}_s,\omega) e^{ik_z z - i\omega t}. \tag{2.7.23}$$

If $p(\mathbf{r},t)$ is a narrow-band pulse with a carrier frequency at ω_o, the above can be approximated by an integral

$$p(\mathbf{r},t) \simeq \frac{1}{\pi}\Re e \int_{\omega_0-\Delta}^{\omega_0+\Delta} d\omega \tilde{p}_o(\mathbf{r}_s,\omega) e^{ik_z z - i\omega t}. \tag{2.7.24}$$

In the above, we can approximate $k_z(\omega)$ as

$$k_z(\omega) \simeq k_z(\omega_0) + (\omega-\omega_0)\frac{dk_z(\omega_0)}{d\omega} = k_z(\omega_0) + \frac{(\omega-\omega_0)}{v_g} \tag{2.7.25}$$

and obtain

$$\begin{aligned} p(\mathbf{r},t) &\simeq \frac{1}{\pi}\Re e \left\{ e^{i[k_z(\omega_0)z - \omega_0 t]} \int_{\omega_0-\Delta}^{\omega_0+\Delta} d\omega\, \tilde{p}_o(\mathbf{r}_s,\omega) e^{\frac{i(\omega-\omega_0)(z-v_g t)}{v_g}} \right\} \\ &\simeq 2\Re e \left\{ e^{i[k_z(\omega_0)z - \omega_0 t]} F(\mathbf{r}_s, z - v_g t) \right\} \end{aligned} \tag{2.7.26}$$

where

$$F(\mathbf{r}_s,z) = \frac{1}{2\pi} \int_{\omega_0-\Delta}^{\omega_0+\Delta} d\omega \tilde{p}_o(\mathbf{r}_s,\omega) e^{\frac{i(\omega-\omega_0)z}{v_g}} \tag{2.7.27}$$

Since $\omega - \omega_0$ is small, $F(\mathbf{r}_s, z)$ is a slowly varying function of z. It represents an envelope function, which in (2.7.26), propagates at the group velocity v_g. The envelope function modulates a rapidly varying function

$$e^{i(k_z z - \omega_0 t)} \tag{2.7.28}$$

which travels at the phase velocity $v_{ph} = \frac{\omega_0}{k_z}$. Hence for a narrow-band signal, a shape-retaining envelope pulse can propagate in a dispersive waveguide.

2.7.3 Attenuation in a Waveguide

When we have a lossy dielectric medium inside a perfectly conducting waveguide, the attenuation due to the lossy dielectric can be ascertained from the formula

$$k_z = \sqrt{\omega^2 \mu\epsilon - k_s^2} = {k_z}' + i{k_z}'', \tag{2.7.29}$$

where $\epsilon = \epsilon' + i\epsilon''$ is complex. This is because the mathematical boundary value problem has not changed when ϵ becomes complex. Hence, (2.4.1) and (2.4.2) hold true even for lossy dielectric.

When the attenuation is due to wall losses because of the finite conductivity of the metallic wall, the calculation is more involved. In this case, we can use the following formula derived from energy conservation

$${k_z}'' = \frac{P_d}{2P_f}, \tag{2.7.30}$$

where P_d is the time average power dissipated per unit length while P_f is the total time average power flow in the waveguide. Since $\hat{n} \times \mathbf{E}$ is not identically zero on the waveguide wall now, $\hat{n} \cdot (\mathbf{E} \times \mathbf{H}^*)$ is not zero on the waveguide wall. We can calculate the time average power dissipated per unit length by integrating $\hat{n} \cdot (\mathbf{E} \times \mathbf{H}^*)$ over the circumference of the waveguide wall, i.e.,

$$P_d = \frac{1}{2} \Re e \oint_C dl \hat{n} \cdot (\mathbf{E} \times \mathbf{H}^*), \tag{2.7.31}$$

where C is a contour defining the cross-section of the waveguide. Since $\hat{n} \times \mathbf{E}$ is not zero, it can be approximated by

$$\hat{n} \times \mathbf{E} = \sqrt{\frac{\mu\omega}{i\sigma}} \mathbf{H} = (1 - i) \frac{1}{\sigma\delta} \mathbf{H}, \tag{2.7.32}$$

where $\delta = \sqrt{\frac{2}{\omega\mu\sigma}}$ is the skin depth in a metallic conductor. The above follows from that in the metal of the waveguide, the $\mathbf{E}$ and $\mathbf{H}$ are related by the intrinsic impedance of metal which is $\sqrt{\mu\omega/i\sigma}$ [see Problem 2-2]. Therefore, (2.7.31) becomes

$$P_d = \frac{1}{2\sigma\delta} \oint_C dl |\mathbf{H}|^2, \tag{2.7.33}$$

where $\mathbf{H}$ is purely tangential on the waveguide wall, if the wall is a perfect conductor. For a perturbation calculation, we can use the $\mathbf{H}$ field of a perfectly conducting waveguide to estimate (2.7.33). Note that $\sigma\delta$ is also the surface conductance of the waveguide.

For a TE mode,

$$\mathbf{H} = \hat{z}H_z + \frac{ik_z}{k_s^2}\nabla_s H_z. \tag{2.7.34}$$

Therefore, from (2.7.33), P_d becomes

$$P_d = \frac{1}{2\sigma\delta}\oint_C dl\left\{|H_z|^2 + \frac{|k_z|^2}{k_s^4}|\nabla_s H_z|^2\right\}. \tag{2.7.35}$$

Using (2.7.11), (2.7.30), and (2.7.35), we obtain that

$$k_z{}'' = \frac{k_s^2\delta}{4k_z}\frac{\oint_C dl\left\{|H_z|^2 + \frac{|k_z|^2}{k_s^4}|\nabla_s H_z|^2\right\}}{\int\limits_S dS|H_z|^2}. \tag{2.7.36}$$

For a TM mode,

$$\mathbf{H} = -\frac{i\omega\epsilon}{k_s^2}\nabla_s \times \hat{z}E_z. \tag{2.7.37}$$

Then, P_d becomes

$$P_d = \frac{\omega^2\epsilon^2}{2\sigma\delta k_s^4}\oint_C dl|\nabla_s E_z|^2, \tag{2.7.38}$$

and

$$k_z{}'' = \frac{k^2\delta}{4k_z k_s^2}\frac{\oint_C dl|\nabla_s E_z|^2}{\int\limits_S dS|E_z|^2}. \tag{2.7.39}$$

The above method of computing the attenuation of a waveguide is also known as the power-loss method. It is inadequate when the modes of the waveguides are degenerate.

For the TE case, $k_z \to k$ when $\omega \to \infty$, while k_s^2 remains a constant independent of frequencies. Therefore, from (2.7.36), we have

$$k_z{}'' \sim \frac{\delta k}{4k_s^2}\frac{\oint_C dl|\nabla_s H_z|^2}{\int\limits_S dS|H_z|^2}, \tag{2.7.40}$$

which increases as the frequency increases since $\delta k = \sqrt{2\omega\epsilon/\sigma}$.

For the TM case, when $\omega \to \infty$, from (2.7.39)

$$k_z{}'' \sim \frac{\delta k}{4k_s^2}\frac{\oint_C dl|\nabla_s E_z|^2}{\int\limits_S dS|E_z|^2}, \tag{2.7.41}$$

which also increases as the frequency increases. Therefore, a metallic waveguide becomes more inefficient at high frequencies.

From (2.7.36) and (2.7.39), we see that $k_z{}''$ diverges when $k_z \to 0$, i.e., when the wave tends to cutoff. This is because there is no real power flow P_f at cutoff, and hence $P_f \to 0$ in (2.7.30). Since Equation (2.7.30) embodies a perturbation concept, it is only valid when $k_z{}'' \ll k_z{}'$. Therefore, Equations (2.7.36) and (2.7.39) are invalid near the cutoff of the wave. But still, the trend is that k_z'' becomes larger close to cutoff.

The Magic Modes

For some special modes of a waveguide, due to symmetry, the second term in (2.7.34) vanishes. This can happen to some modes of a highly symmetrical waveguide such as a parallel plate waveguide or a circular waveguide. In this case, (2.7.36) becomes

$$k_z'' = \frac{k_s^2 \delta}{4k_z} \frac{\oint_C dl \left\{|H_z|^2\right\}}{\int\limits_S dS|H_z|^2}. \tag{2.7.42}$$

In this case, k_z'' becomes smaller as the frequency increases. For these modes, the electric field is tangential to the waveguide wall, with magnetic field normal to the k vector and the electric field. As the frequency increases, the k vector becomes almost parallel to the z axis. The magnetic field becomes almost vertical to the waveguide wall with a small tangential component. Hence, the induced surface current on the waveguide wall actually becomes smaller. Consequently, the attenuation of the waveguide mode actually decreases with increasing frequency. Some TE modes of the circular waveguide are such a mode, and they are known as "magic modes" (see Problems 2.9 and 2.12).

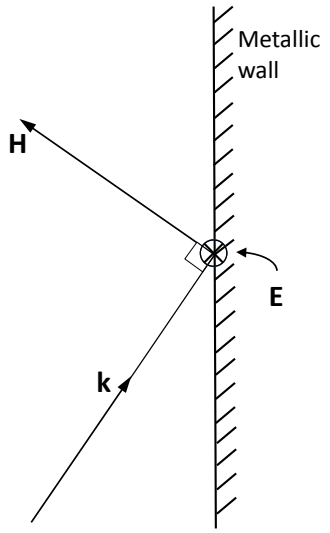


Figure 2.12: The field configuration of a magic mode on the wall of a waveguide. As the frequency increases, the k vector becomes more parallel to the waveguide wall, and tangential **H** becomes smaller, since **H** is orthogonal to both **E** and **k**. This reduces the surface current, and hence, the wall loss.

Figure 2.13 shows some typical losses of different modes in a rectangular and a circular waveguide including a "magic mode". Such low loss is desirable in radio astronomy where the frequency is high and the signal low. Hence, circular waveguides with a corrugated wall to "discourage" other modes, but promote the propagation of this magic mode, is actually used in the design of the VLA (very large array) of NRAO (National Radio Astronomy Observatory) in New Mexico.

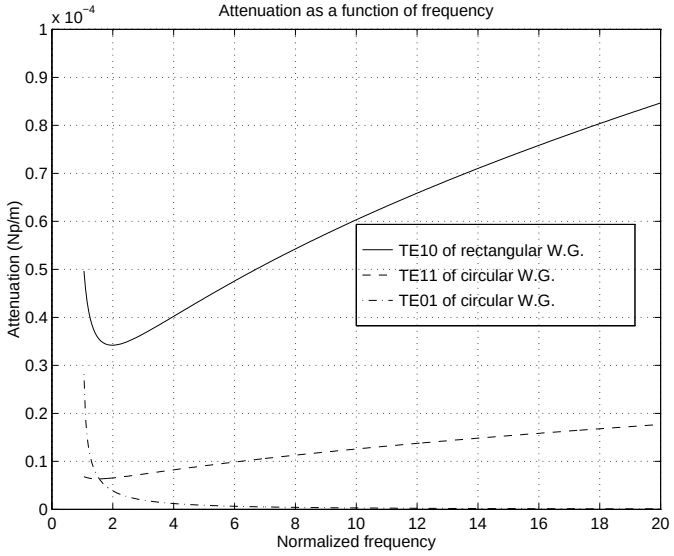


Figure 2.13: Loss as a function of frequency for different waveguide modes.

2.8 Excitation of Modes in a Waveguide

The modes of a waveguide are excited by putting sources inside a waveguide. This is usually in terms of waveguide probe carrying a current or a charge. So the probe can be dipole-like, producing mainly an electric field, or it can be loop-like producing mainly a magnetic field. The location of the probe is imperative if certain desirable modes are to be excited. We will study the relation between the probe location, the source type, and the modes

Periodic Boundary Condition

The excitation of modes in a waveguide can be made very similar to the excitation of modes in a cavity by the use of periodic boundary condition. Take the example of a rectangular cavity. The modes are countably infinite in all three directions and we have $k^2 = \left(\frac{m\pi}{a}\right)^2 + \left(\frac{n\pi}{b}\right)^2 + \left(\frac{p\pi}{d}\right)^2$ where m, n, and p are the indices for counting the modes in the x, y, and z directions respectively. In a rectangular waveguide of infinite length, then $k^2 = \left(\frac{m\pi}{a}\right)^2 + \left(\frac{n\pi}{b}\right)^2 + k_z^2$

where k_z^2 now becomes uncountably infinite as it becomes a continuum variable rather than a discrete variable in the cavity case.

However, we can use periodic boundary condition to discretize the waveguide wavenumber associated with the z axis. If we have a traveling wave in the z direction indicated by $\exp(ik_z z)$ with the requirement that this function repeats itself after distance d, then

$$e^{ik_z z}|_{z=0} = e^{ik_z z}|_{z=d} \tag{2.8.1}$$

The above implies that

$$e^{ik_z d} = 1 \tag{2.8.2}$$

implying that

$$k_z = \frac{2p\pi}{d}, \quad \forall \quad \text{integer} \quad p \tag{2.8.3}$$

The traveling wave then becomes a Fourier mode

$$e^{ik_z z} = e^{i\frac{2p\pi}{d}z} \tag{2.8.4}$$

When $d \to \infty$, k_z assumes a continuum of modes as in Fourier transform. When d is finite, the modes in the z direction is countably infinite just as the modes in a cavity.

Generalized Eigenfunction Expansion for Vector Wave Equation

Given an electric field that satisfies

$$\nabla \times \nabla \times \mathbf{E}(\mathbf{r}) - k_0^2 \mathbf{E}(\mathbf{r}) = i\omega\mu \mathbf{J}(\mathbf{r}) \tag{2.8.5}$$

we can expand the field in terms of the eigenfunctions of the following equation

$$\nabla \times \nabla \times \mathbf{F}_m(\mathbf{r}) - k_m^2 \mathbf{F}_m(\mathbf{r}) = 0 \tag{2.8.6}$$

assuming that $\mathbf{F}_m$ satisfies the same boundary condition as the electric field $\mathbf{E}$. First, we let

$$\mathbf{E}(\mathbf{r}) = \sum_m a_m \mathbf{F}_m(\mathbf{r}) \tag{2.8.7}$$

On substituting (2.8.7) into (2.8.5), we have

$$\sum_m a_m (k_m^2 - k_0^2)\mathbf{F}_m(\mathbf{r}) = i\omega\mu \mathbf{J}(\mathbf{r}) \tag{2.8.8}$$

Assuming an orthonormal relationship for the eigenfunctions such that[3]

$$\int_V d\mathbf{r} \mathbf{F}_{m'}^*(\mathbf{r}) \mathbf{F}_m(\mathbf{r}) = \delta_{m'm} \tag{2.8.9}$$

[3]Such an orthogonality relation with a conjugation is needed since the operator $\nabla \times \nabla \times$ is a Hermitian (self-adjoint) operator.

we can deduce that

$$a_m = i\omega\mu \frac{\langle \mathbf{F}_m^*, \mathbf{J} \rangle}{k_m^2 - k_0^2} \tag{2.8.10}$$

Consequently, we get

$$\mathbf{E}(\mathbf{r}) = i\omega\mu \sum_m \mathbf{F}_m(\mathbf{r}) \frac{\langle \mathbf{F}_m^*, \mathbf{J} \rangle}{k_m^2 - k_0^2} \tag{2.8.11}$$

The above is a general eigenfunction expansion formula if we orthonormalize the eigenfunctions. It says physically that the field due to a source in a waveguide can be expanded in terms of the eigenfunctions of the waveguide. The same expression also holds for cavity mode expansion. The excitation coefficients of the eigenfunctions are given by (2.8.10). These coefficients are proportional to $\langle \mathbf{F}_m^*, \mathbf{J} \rangle$ which is the inner product between the eigenfunction and the source. Hence, it is important that we learn how to find these eigenfunctions.

Modes are often excited by current probes in the waveguide or cavity. The above expression tells us if we want a certain mode $\mathbf{F}_m$ to be strongly excited, we need the inner product $\langle \mathbf{F}_m^*, \mathbf{J} \rangle$ to be large. Hence, the current on the probe should be located at where the field of the mode $\mathbf{F}_m$ is strong. If the probe is a short wire, it can be approximated by an electric dipole with strong charge accumulation that produces a strong electric field. This electric dipole should be placed close to the maxima of the mode in order to excite it.

On the other hand, if the current source $\mathbf{J}$ consists of a current loop with constant current,[4] then the inner product

$$\langle \mathbf{F}_m^*, \mathbf{J} \rangle = \int_V d\mathbf{r} \mathbf{F}_m^*(\mathbf{r}) \cdot \mathbf{J}(\mathbf{r}) = \oint_C d\mathbf{l} \cdot \mathbf{F}_m^*(\mathbf{r}) = \int_A dS \hat{n} \cdot \nabla \times \mathbf{F}_m^*(\mathbf{r}) \tag{2.8.12}$$

where C is the contour of the loop, and A is the cross section area of the loop. If $\mathbf{F}_m$ represents electric field, then $\nabla \times \mathbf{F}_m$ represents the magnetic field. Hence, the current loop has to be placed in location where the magnetic field of the mode is strong in order to excite it. A small current loop behaves like a magnetic dipole and hence, it needs to be placed near strong magnetic field in order to excite the mode strongly.

Also, due to the $1/(k_m^2 - k_0^2)$ dependence of the excitation coefficient, if the operating frequency of the source is closed to the resonant frequency of the mode, that mode will be strongly excited. This is the phenomenon of resonance coupling. Energy can be coupled to a mode if we operate close to the resonant frequency of the mode.

2.8.1 Vector Wave Functions in a Waveguide

In scalar function theory, we know that an arbitrary function can be expanded as $f(x) = \sum_{n=0}^{\infty} a_n \cos\left(\frac{n\pi x}{a}\right)$ by the completeness of the Fourier cosine basis for $0 < x < a$. It turns out that the waveguide modes that we have previously studied do not constitute a complete set. First, they satisfy source-free Maxwell's equations and hence, they are divergence free. Hence, they cannot constitute a non-divergence free field in a waveguide. Second, they propagate in

[4] This is possible if the loop size is much smaller than the wavelength.

either plus or minus z directions. If a source is present in a waveguide, it is not clear what direction the field is propagating in the source region. To obtain a complete set, we need to derive the vector wave functions.

An arbitrary vector function in a waveguide can be expanded in terms of vector wave functions of the waveguide which is the analogue of Fourier basis. However, drawing such an analogy has its pitfall. The expansion of modes in a cavity excited by a current source has a colorful history. There are basically two kinds of vector wave functions in a waveguide or a cavity: the divergence-free type (solenoidal) and the curl-free type (irrotational). It was believed by some for a while that the usual divergence free functions are complete. But in fact, both kinds of functions are needed for completeness [4, 10, 11]. More references on this topic can be found in [13].

We shall discuss the derivation of such vector wave functions for a uniform hollow waveguide. Both the electric field and the magnetic field in a waveguide satisfy the equation

$$\nabla \times \nabla \times \mathbf{F} - k^2 \mathbf{F} = 0. \tag{2.8.13}$$

It is to be noted that the above is an eigenvalue problem where k^2 is the eigenvalue, and $\mathbf{F}$ is the eigenfunction. It can be shown that with the appropriate boundary conditions, the $\nabla \times \nabla \times$ operator is Hermitian, and that the eigenvalues k^2 are always real.

The above is not actually the equation satisfied by the fields of the waveguide since k^2 is arbitrary and not fixed to be $\omega^2 \mu \epsilon$. However, solving the above equation for all possible k^2 or eigenvalues generates enough eigenfunctions that form a complete set. This includes the cases where $k^2 = 0$, implying the functions that belong to the null space of the $\nabla \times \nabla \times$ operator. Since k^2 is equivalent to frequency, we can think of the above as solving for all the resonance solutions (eigensolutions) of a very long cavity. As we shall see next, if the cavity is of finite length, then k^2 will assume discrete values. But if the cavity is of infinite length, then k^2 assumes a continuum of values.

Divergence-Free Eigenfunctions

We can convert the (2.8.13) into a scalar wave equation via the transformation

$$\mathbf{F} = \nabla \times \mathbf{c}\psi. \tag{2.8.14}$$

The above vector function is clearly divergence free. Substituting (2.8.14) into (2.8.13), and if ψ satisfies

$$(\nabla^2 + k^2)\psi = 0, \tag{2.8.15}$$

then $\mathbf{F}$ in (2.8.14) will satisfy (2.8.13). It is clear that $\mathbf{F}$ is transverse to $\mathbf{c}$. Since in a waveguide, as well as an infinitely long cavity, we can decompose the field into TE and TM types with respect to z, it is natural to choose $\mathbf{c} = \hat{z}$. For example, if ψ_h satisfies (2.8.15) with the Neumann boundary condition, i.e.,

$$(\nabla^2 + k^2)\psi_h = 0, \qquad \frac{\partial \psi_h}{\partial n} = 0, \qquad \text{on} \qquad C \tag{2.8.16}$$

then the vector wave function

$$\mathbf{M}_e = \nabla \times (\hat{z}\psi_h), \tag{2.8.17}$$

is analogous to the **E**-field of a TE mode. Again, it is to be emphasized that (2.8.16) defines an eigenvalue problem.

Since $\nabla \times \mathbf{M}_e$ is also a solution to (2.8.13), we can define another vector wave function

$$\mathbf{N}_h = \frac{1}{k}\nabla \times \mathbf{M}_e = \frac{1}{k}\nabla \times \nabla \times \hat{z}\psi_h. \tag{2.8.18}$$

$\mathbf{N}_h$ is a vector wave function analogous to the magnetic field of a TE mode.

Similarly, we define a ψ_e satisfying

$$(\nabla^2 + k^2)\psi_e = 0, \qquad \psi_e = 0, \qquad \text{on} \quad C. \tag{2.8.19}$$

Then,

$$\mathbf{M}_h = \nabla \times \hat{z}\psi_e, \tag{2.8.20}$$

is analogous to the **H**-field of the TM mode in a waveguide. Similarly, we have

$$\mathbf{N}_e = \frac{1}{k}\nabla \times \mathbf{M}_h = \frac{1}{k}\nabla \times \nabla \times \hat{z}\psi_e, \tag{2.8.21}$$

which is analogous to the **E** field of a TM mode. The eigenfunctions obtained from solving (2.8.16) and (2.8.19) are complete, and hence they can be used to generate all possible **M** and **N** functions above.

In the previous discussion, the subscript e denotes a quantity that is related to the **E** field while the subscript h denotes a quantity that is related to the **H**-field in terms of their boundary conditions. ψ_h and ψ_e are sometimes known as the magnetic Hertzian potential and the electric Hertzian potential respectively. By looking back at Equations (2.1.7) and (2.1.8), one notes that they are analogous to the H_z and E_z components of the field respectively.

As an example, for a rectangular waveguide,

$$\psi_{hmn}(k_z, \mathbf{r}) = \cos\left(\frac{m\pi x}{a}\right)\cos\left(\frac{n\pi y}{b}\right)e^{ik_z z}, \tag{2.8.22a}$$

$$\psi_{emn}(k_z, \mathbf{r}) = \sin\left(\frac{m\pi x}{a}\right)\sin\left(\frac{n\pi y}{b}\right)e^{ik_z z}. \tag{2.8.22b}$$

We can generate the vector wave function from the above by

$$\mathbf{M}_{emn}(k_z, \mathbf{r}) = \nabla \times \hat{z}\cos\left(\frac{m\pi x}{a}\right)\cos\left(\frac{n\pi y}{b}\right)e^{ik_z z}, \tag{2.8.23a}$$

$$\mathbf{N}_{hmn}(k_z, \mathbf{r}) = \frac{1}{k}\nabla \times \nabla \times \hat{z}\cos\left(\frac{m\pi x}{a}\right)\cos\left(\frac{n\pi y}{b}\right)e^{ik_z z}, \tag{2.8.23b}$$

and

$$\mathbf{M}_{hmn}(k_z, \mathbf{r}) = \nabla \times \hat{z}\sin\left(\frac{m\pi x}{a}\right)\sin\left(\frac{n\pi y}{b}\right)e^{ik_z z}, \tag{2.8.24a}$$

$$\mathbf{N}_{emn}(k_z, \mathbf{r}) = \frac{1}{k}\nabla \times \nabla \times \hat{z}\sin\left(\frac{m\pi x}{a}\right)\sin\left(\frac{n\pi y}{b}\right)e^{ik_z z}. \tag{2.8.24b}$$

We allow k_z to be a free variable (arbitrary variable) in Equations (2.8.22a) and (2.8.22b). This would render ψ_{hmn} and ψ_{emn} to be complete scalar functions in a hollow waveguide. At least, we know that an arbitrary function of z can be Fourier expanded in terms of $e^{ik_z z}$.

The above is analogous to finding the eigenvalues of a rectangular cavity of dimension $a \times b \times d$ with periodic boundary condition in the z direction. In this case, the eigenvalue is given by $k^2 = \left(\frac{m\pi}{a}\right)^2 + \left(\frac{n\pi}{b}\right)^2 + \left(\frac{2p\pi}{d}\right)^2$, where m, n, and p are all integers. We can identify $k_z = \frac{2p\pi}{d}$. As we let $d \to \infty$, k_z which previously takes on discrete values, becomes a continuum variable. Since k_z can be any continuum real variable, the eigenvalue $k^2 = k_s^2 + k_z^2$ can also take on any continuum real variable. In the above k_s is the transverse eigenvalue of the waveguide problem similar to (2.4.1) and (2.4.1). We shall denote it as k_{is} subsequently, where the index i implies an ordered pair (m, n) in the case of a rectangular waveguide.

As of this point, these vector wave functions are not physical modes of a waveguide. In order for them to be physical modes, k_z^2 has to satisfy the dispersion relation $k_0^2 = k_s^2 + k_z^2$ where k_0 is the wavenumber inside the waveguide. Since the dispersion relation describes an equation of a sphere in the k space, this sphere is known as the Ewald sphere or the energy shell in physics. The eigenfunctions derived so far do not satisfy the dispersion relation, and they are said to be off the energy shell. When they are forced to satisfy the dispersion relation, they are said to be on the energy shell.

Given the above information, we can easily show that

$$\int dV \psi_{hmn}(k_z, \mathbf{r}) \psi^*_{hm'n'}(k_z{}', \mathbf{r}) = (1+\delta_{0m})(1+\delta_{0n}) \frac{\pi ab}{2} \delta(k_z - k_z{}') \delta_{mn,m'n'}, \tag{2.8.25}$$

where $\delta(x)$ is the Dirac delta function and δ_{ij}, the Kronecker delta function. In general, for a waveguide of arbitrary cross-section, orthogonality relations exist for the ψ_h's and ψ_e's as [analogous to (2.4.3) and (2.4.4)] [5]

$$\int dV \psi_{hi}(k_z, \mathbf{r}) \psi^*_{hj}(k_z{}', \mathbf{r}) = A_{hi} \delta(k_z - k_z{}') \delta_{ij}, \tag{2.8.26a}$$

$$\int dV \psi_{ei}(k_z, \mathbf{r}) \psi^*_{ej}(k_z{}', \mathbf{r}) = A_{ei} \delta(k_z - k_z{}') \delta_{ij}, \tag{2.8.26b}$$

where A_{hi} and A_{ei} are the appropriate normalization constants.[6]

For the vector wave functions, the orthogonality relationships are

$$\int dV \mathbf{M}^{\ e}_{h\ i}(k_z, \mathbf{r}) \cdot \mathbf{M}^*_{e\ j}(k_z{}', \mathbf{r}) = k^2_{is}{}^{\ h}_{e} A^{\ h}_{e\ i} \delta(k_z - k_z{}') \delta_{ij}, \tag{2.8.27a}$$

$$\int dV \mathbf{N}^{\ e}_{h\ i}(k_z, \mathbf{r}) \cdot \mathbf{N}^*_{e\ j}(k_z{}', \mathbf{r}) = k^2_{is}{}^{\ e}_{h} A^{\ e}_{h\ i} \delta(k_z - k_z{}') \delta_{ij}, \tag{2.8.27b}$$

$$\int dV \mathbf{M}^{\ e}_{h\ i}(k_z, \mathbf{r}) \cdot \mathbf{N}^*_{e\ j}(k_z{}', \mathbf{r}) = 0, \qquad \text{for all} \quad i,\ j,\ k_z,\ k_z{}'. \tag{2.8.27c}$$

Because the vector functions $\mathbf{M}$ and $\mathbf{N}$ are divergence free, and that their curls are not zero, these functions are also termed the solenoidal vector wave functions. They can be used to expand divergence-free fields.

[5] See [13] p. 388.

[6] Notice that when k_z is real, $\psi_{ej}(-k_z) = \psi^*_{ej}(k_z)$, but since k_z will become complex in Cauchy integration technique applied on the complex plane, a technique that we will use later, we will retain this notation in the above.

Curl-Free Eigenfunctions

The expressions (2.8.20) and (2.8.22a) and (2.8.22b) are clearly divergence free. However, **M** and **N** vector wave functions are not, in general, complete. They cannot be used to represent fields whose divergence is non zero. To remedy this, the functions in the null-space of the $\nabla \times \nabla \times$ operator need to be considered. We need the **L** functions, which are defined as[7]

$$\mathbf{L}_h = \nabla \psi_h, \qquad \mathbf{L}_e = \nabla \psi_e, \tag{2.8.28}$$

where ψ_h and ψ_e are as defined in (2.8.16) and (2.8.19). $\mathbf{L}_h$ satisfies the magnetic field boundary condition while $\mathbf{L}_e$ satisfies the electric field boundary condition on the waveguide wall. The vector wave functions have zero curl and non-zero divergence. Hence, they are also known as the irrotational vector wave functions. They are solutions to Equation (2.8.13) corresponding to when $k^2 = 0$. Hence, they are the null-space solution of the $\nabla \times \nabla \times$ operator.

Their divergence is

$$\nabla \cdot \mathbf{L}_h = -k^2 \psi_h, \qquad \nabla \cdot \mathbf{L}_e = -k^2 \psi_e. \tag{2.8.29}$$

Since the divergence of field is proportional to charge, the right hand sides of the above represent charges. But since ψ_h and ψ_e form complete sets, they can be used to expand fields due to arbitrary sources inside the waveguide.

It can be shown that

$$\int dV \mathbf{L}_{\substack{e \\ h}\, i}(k_z, \mathbf{r}) \cdot \mathbf{L}^*_{\substack{e \\ h}\, j}(k_z{}', \mathbf{r}) = k^2 A_{\substack{e \\ h}\, i} \delta(k_z - k_z') \delta_{ij}, \tag{2.8.30a}$$

$$\int dV \mathbf{M}_{\substack{e \\ h}\, i}(k_z, \mathbf{r}) \cdot \mathbf{L}^*_{\substack{e \\ h}\, j}(k_z{}', \mathbf{r}) = 0, \qquad \text{all} \quad i, \quad j, \quad k_z, \quad k_z{}', \tag{2.8.30b}$$

$$\int dV \mathbf{N}_{\substack{e \\ h}\, i}(k_z, \mathbf{r}) \cdot \mathbf{L}^*_{\substack{e \\ h}\, j}(k_z{}', \mathbf{r}) = 0, \qquad \text{all} \quad i, \quad j, \quad k_z, \quad k_z{}', \tag{2.8.30c}$$

Eigenfunction Expansion of Arbitrary Fields

An arbitrary field in a waveguide can, in general, be expanded as

$$\mathbf{E}(\mathbf{r}) = \int_{-\infty}^{\infty} dk_z \sum_i [a_{ei}(k_z)\mathbf{M}_{ei}(k_z, \mathbf{r}) + b_{ei}(k_z)\mathbf{N}_{ei}(k_z, \mathbf{r}) + c_{ei}(k_z)\mathbf{L}_{ei}(k_z, \mathbf{r})], \tag{2.8.31a}$$

$$\mathbf{H}(\mathbf{r}) = \int_{-\infty}^{\infty} dk_z \sum_i [a_{hi}(k_z, \mathbf{r})\mathbf{M}_{hi}(k_z, \mathbf{r}) + b_{hi}(k_z)\mathbf{N}_{hi}(k_z, \mathbf{r}) + c_{hi}(k_z)\mathbf{L}_{hi}(k_z, \mathbf{r})]. \tag{2.8.31b}$$

The coefficients can be found from the orthogonality relationships.

[7] It is to be noted that Helmholtz theorem says that an arbitrary vector field can be decomposed into the sum of divergence-free field and curl-free field. In other words, $\mathbf{F} = \nabla \times A + \nabla \psi$. We expect to see such decomposition here.

2.8.2 Dyadic Green's Function

The dyadic Green's function in a waveguide is the solution to the equation

$$\nabla \times \nabla \times \overline{\mathbf{G}}(\mathbf{r}, \mathbf{r}') - k_0^2 \overline{\mathbf{G}}(\mathbf{r}, \mathbf{r}') = \overline{\mathbf{I}} \delta(\mathbf{r} - \mathbf{r}'), \tag{2.8.32}$$

satisfying the boundary condition $\hat{n} \times \overline{\mathbf{G}}(\mathbf{r}, \mathbf{r}') = 0$, for $\mathbf{r}$ on the waveguide wall. Once this Green's function is known, the field due to an arbitrary distributed source in a waveguide can be written as

$$\mathbf{E} = i\omega\mu \int \overline{\mathbf{G}}(\mathbf{r}, \mathbf{r}') \cdot \mathbf{J}(\mathbf{r}') d\mathbf{r}'. \tag{2.8.33}$$

A General Dyadic Green's Function

From the definition of the dyadic Green's function as given by (2.8.33), and from the generalized formula for mode expansion due to a current source, as given in (2.8.11), we can deduce that the general dyadic Green's function is of the form

$$\overline{\mathbf{G}}(\mathbf{r}, \mathbf{r}') = \sum_m \frac{\mathbf{F}_m(\mathbf{r}) \mathbf{F}_m^*(\mathbf{r}')}{k_m^2 - k_0^2} \tag{2.8.34}$$

The above is a succinct way to express the dyadic Green's function in terms of the modes of the cavity, where the modes are assumed to be orthonormal.

Dyadic Green's Function for Hollow Waveguide

The derivation of the dyadic Green's function has a colorful history as the mode expansion in a cavity [4, 14, 15, 17]. The controversy comes from the incompleteness of the divergence-free modes in a waveguide or a cavity. More references can be found in [13].

To solve Equation (2.8.32), we expand $\overline{\mathbf{G}}(\mathbf{r}, \mathbf{r}')$ in terms of the vector wave functions $\mathbf{M}_e$, $\mathbf{N}_e$, and $\mathbf{L}_e$. In other words,

$$\begin{aligned} \overline{\mathbf{G}}(\mathbf{r}, \mathbf{r}') = \int_{-\infty}^{\infty} dk_z \sum_i [&\mathbf{M}_{ei}(k_z, \mathbf{r}) \mathbf{a}_{ei}(k_z, \mathbf{r}') + \mathbf{N}_{ei}(k_z, \mathbf{r}) \mathbf{b}_{ei}(k_z, \mathbf{r}') \\ &+ \mathbf{L}_{ei}(k_z, \mathbf{r}) \mathbf{c}_{ei}(k_z, \mathbf{r}')]. \end{aligned} \tag{2.8.35}$$

Substituting $\overline{\mathbf{G}}(\mathbf{r}, \mathbf{r}')$ into (2.8.32), and using Equation (2.8.49), we have

$$\begin{aligned} \int_{-\infty}^{\infty} dk_z \sum_i \{&(k^2 - k_0^2)[\mathbf{M}_{ei}(k_z, \mathbf{r}) \mathbf{a}_{ei}(k_z, \mathbf{r}') + \mathbf{N}_{ei}(k_z, \mathbf{r}) \mathbf{b}_{ei}(k_z, \mathbf{r}')] \\ &- k_0^2 \mathbf{L}_{ei}(k_z, \mathbf{r}) \mathbf{c}_{ei}(k_z, \mathbf{r}')\} = \overline{\mathbf{I}} \delta(\mathbf{r} - \mathbf{r}'). \end{aligned} \tag{2.8.36}$$

From the orthogonality relations for $\mathbf{M}$, $\mathbf{N}$, and $\mathbf{L}$ functions, we have

$$\mathbf{a}_{ei}(k_z, \mathbf{r}') = \frac{1}{k_{ish}^2 A_{hi} (k_z^2 - k_{izh}^2)} \mathbf{M}_{ei}^*(k_z, \mathbf{r}'), \tag{2.8.37a}$$

$$\mathbf{b}_{ei}(k_z,\mathbf{r}') = \frac{1}{k_{ise}^2 A_{ei}(k_z^2-k_{ize}^2)}\mathbf{N}_{ei}^*(k_z,\mathbf{r}'), \tag{2.8.37b}$$

$$\mathbf{c}_{ei}(k_z,\mathbf{r}') = \frac{-1}{k^2 A_{ei}k_0^2}\mathbf{L}_{ei}^*(k_z,\mathbf{r}'). \tag{2.8.37c}$$

In the above, we have replaced $k^2-k_0^2$ with $k_z^2-k_{iz\,h}^{2\,e}$, because $k^2 = k_{is\,h}^{2\,e}+k_z^2$, $k_0^2 = k_{is\,h}^{2\,e}+k_{iz\,h}^{2\,e}$. Here, k_{iz} denotes values on the energy shell, while k_z denotes values off the energy shell. The subscripts e and h on the transverse eigenvalue k_{is} and longitudinal wavenumber k_{iz} denote the association of these values with either the Dirichlet or the Neumann problem respectively.

Therefore, the dyadic Green's function is

$$\begin{aligned}\overline{\mathbf{G}}(\mathbf{r},\mathbf{r}') = \int_{-\infty}^{\infty} dk_z \sum_i \Bigg[&\frac{\mathbf{M}_{ei}(k_z,\mathbf{r})\mathbf{M}_{ei}^*(k_z,\mathbf{r}')}{k_{ish}^2 A_{hi}(k_z^2-k_{izh}^2)} \\ &+\frac{\mathbf{N}_{ei}(k_z,\mathbf{r})\mathbf{N}_{ei}^*(k_z,\mathbf{r}')}{k_{ise}^2 A_{ei}(k_z^2-k_{ize}^2)} - \frac{\mathbf{L}_{ei}(k_z,\mathbf{r})\mathbf{L}_{ei}^*(k_z,\mathbf{r}')}{k^2 A_{ei}k_0^2}\Bigg].\end{aligned} \tag{2.8.38}$$

Cauchy Integration Technique

The first two integrals are of the form

$$I = \int_{-\infty}^{\infty} dk_z \frac{e^{ik_z(z-z')}f(k_z)}{k_z^2-k_{iz}^2}. \tag{2.8.39}$$

There are poles at $k_z = \pm k_{iz}$. If we assume a small loss in the medium, then the poles are off the real axis and the integral (2.8.39) is well-defined. If $f(k_z)/k_z^2 \to 0$ when $k_z \to \infty$, for $z > z'$, we can deform the contour of integration from the real axis to the contour C. By virtue of Jordan's lemma, the integral over C vanishes and the integral (2.8.39) is then equal to the residue of the pole at $k_z = k_{iz}$. When $z < z'$, we can deform the path of integration downward, and equate the integral (2.8.39) to the residue of the pole at $k_z = -k_{iz}$. Therefore, it follows that

$$I = \pi i \frac{e^{\pm ik_{iz}(z-z')}}{k_{iz}} f(\pm k_{iz}), \qquad \begin{matrix} z>0 \\ z<0. \end{matrix} \tag{2.8.40}$$

Note that the process of Cauchy integration technique forces k_z to be on the energy shell or on the Ewald sphere.

We can apply Cauchy integration technique to the first two integrals since $\mathbf{M}_{ei}$ and $\mathbf{N}_{ei}$ tend to be constants when $k_z \to \infty$. However, because

$$\mathbf{N}_{ei} = \frac{1}{k}\nabla\times\mathbf{M}_{hi} = \frac{1}{k}\nabla\times\nabla\times\hat{z}\psi_{ei}, \tag{2.8.41}$$

$\mathbf{N}_{ei}(\mathbf{k}_z,\mathbf{r})\mathbf{N}_{ei}(-k_z,\mathbf{r}')$ is proportional to $1/k^2$. But $k^2 = k_{ise}^2 + k_z^2$, implying that there are additional poles at $k_z = \pm ik_{ise}$ for the second term in (2.8.38). Similarly, the third term in (2.8.38) also has a $1/k^2$ dependence with poles at $k_z = \pm ik_{ise}$. It can be shown that the pole contributions at $k_z = \pm ik_{ise}$ for the second and the third terms cancel each other.

Since

$$\mathbf{L}_{ei}(k_z,\mathbf{r}) \sim \hat{z}ik_z\psi_{ei}, \qquad k_z \to \infty, \tag{2.8.42}$$

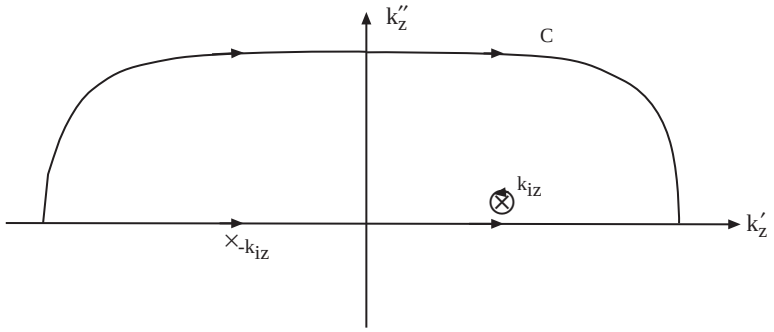


Figure 2.14: Contour integration on the complex k_z plane.

the third term tends to be a constant when $k_z \to \infty$. Therefore, contour integration cannot be applied to the third term, since Jordan's lemma is not satisfied. To remedy this, we write (2.8.38) as

$$\begin{aligned}\overline{\mathbf{G}}(\mathbf{r},\mathbf{r}') = \int_{-\infty}^{\infty} dk_z \sum_i \Bigg[& \frac{\mathbf{M}_{ei}(k_z,\mathbf{r})\mathbf{M}^*_{ei}(k_z,\mathbf{r}')}{k_{ish}^2 A_{hi}(k_z^2 - k_{izh}^2)} \\ & + \frac{\mathbf{N}_{ei}(k_z,\mathbf{r})\mathbf{N}^*_{ei}(k_z,\mathbf{r}')}{k_{ise}^2 A_{ei}(k_z^2 - k_{ize}^2)} \\ & - \left(\frac{\mathbf{L}_{ei}(k_z,\mathbf{r})\mathbf{L}^*_{ei}(k_z,\mathbf{r}')}{k^2 A_{ei} k_0^2} - \frac{\hat{z}\hat{z}\psi_{ei}(k_z,\mathbf{r})\psi^*_{ei}(k_z,\mathbf{r}')}{k_0^2 A_{ei}} \right) \Bigg] \\ & - \int_{-\infty}^{\infty} dk_z \sum_i \frac{\hat{z}\hat{z}\psi_{ei}(k_z,\mathbf{r})\psi^*_{ei}(k_z,\mathbf{r}')}{k_0^2 A_{ei}}.\end{aligned} \tag{2.8.43}$$

Contour integrations can now be performed on the third term, giving rise to a pole contribution that cancels a similar contribution from the second term. The second term contains a pole at $k = 0$ becuse the $\mathbf{N}$ function is proportional to $1/k$ as evident from (2.8.18). Hence, the second term is proportional to $1/k^2$ just as the third term. Since $k = 0$ for this pole contribution, for the sake of discussion, we will call this the static pole.

It is to be noted that if the $\mathbf{L}$ functions are not used (which is erroneous) in the expansion, a modal contribution from the second term due to this static pole will exist. This mode will satisfy the dispersion relation $k_x^2 + k_y^2 + k_z^2 = 0$, making it an nonphysical mode in the presence of a time-harmonic excitation. Fortunately, it is cancelled by a similar contribution from the third term, the curl-free term. On first sight, it may seem strange that a contribution from the divergence-free term should cancel one from the curl-free term. However, a closer examination shows that this static pole contributes both to a divergence-free and curl-free field. The role of the curl-free modes outside the source region is also discussed in [16].

Consequently, we have

$$\overline{\mathbf{G}}(\mathbf{r},\mathbf{r}') = \sum_i \pi i \left[\frac{\mathbf{M}_{ei}(\pm k_{izh},\mathbf{r})\mathbf{M}^*_{ei}(\pm k_{izh},\mathbf{r}')}{k_{izh}k^2_{ish}A_{hi}} + \frac{\mathbf{N}_{ei}(\pm k_{ize},\mathbf{r})\mathbf{N}^*_{ei}(\pm k_{ize},\mathbf{r}')}{k_{ize}k^2_{ise}A_{ei}} \right] - \frac{\hat{z}\hat{z}}{k_0^2}\delta(\mathbf{r}-\mathbf{r}'), \quad \begin{matrix} z > z' \\ z < z'. \end{matrix} \tag{2.8.44}$$

The upper sign is chosen when $z > z'$ and the lower sign is chosen when $z < z'$. The identity

$$\delta(\mathbf{r}-\mathbf{r}') = \int_{-\infty}^{\infty} dk_z \sum_i \frac{\psi_{ei}(k_z,\mathbf{r})\psi^*_{ei}(k_z,\mathbf{r}')}{A_{ei}}, \tag{2.8.45}$$

has been used to simplify the last integral. All the vector wave functions in (2.8.44) evaluated on the Ewald sphere. Hence, they are now physical wave functions which are solutions to Maxwell's equations.

The Dirac delta function part of the Green's function in (2.8.44) has a $\hat{z}\hat{z}$ component. This is because we have performed the dk_z integration first, letting $k_z \to \infty$, before letting the index i goes to infinity. When $k_z \to \infty$, it also implies that we are looking at length scales in the $\hat{z}$ direction with infinite resolution before the length scales in the transverse direction. Hence, the singularity in (2.8.44) is exactly the one obtained if one were to use a disk-shaped pill box in performing the principal volume integral. Therefore, in order to obtain a unique, correct solution when applying (2.8.44), one has to use a disk-shaped principal volume integral.[8]

Once the dyadic Green's function of a hollow waveguide is known, the excitation of modes due to an arbitrary current source in a waveguide can be found using (2.8.33). By substituting Equation (2.8.44) into (2.8.33), we have

$$\begin{aligned} \mathbf{E}(\mathbf{r}) = & \\ & -\omega\mu\pi \sum_i \left[\frac{1}{k_{izh}k^2_{ish}A_{hi}} \int \mathbf{M}_{ei}(\pm k_{izh},\mathbf{r})\mathbf{M}^*_{ei}(\pm k_{izh},\mathbf{r}')\cdot\mathbf{J}(\mathbf{r}')d\mathbf{r}' \right. \\ & \left. + \frac{1}{k_{ize}k^2_{ise}A_{ei}} \int \mathbf{N}_{ei}(\pm k_{ize},\mathbf{r})\mathbf{N}^*_{ei}(\pm k_{ize},\mathbf{r}')\cdot\mathbf{J}(\mathbf{r}')d\mathbf{r}' \right] \\ & - \frac{i}{\omega\epsilon}\hat{z}J_z(\mathbf{r}). \end{aligned} \tag{2.8.46}$$

The above integral is over the support of the current $\mathbf{J}(\mathbf{r}')$. If $z > z'_{max}$ or $z < z'_{min}$, where z'_{max} and z'_{min} define the range of the support of $\mathbf{J}(\mathbf{r}')$ in the z-direction, i.e., $\mathbf{J}(\mathbf{r}') = 0$ for $z > z'_{max}$ and $z < z'_{min}$, then the above can be written as

$$\begin{aligned} \mathbf{E}(\mathbf{r}) = & \\ & -\omega\mu\pi \sum_i \left[\frac{1}{k_{izh}k^2_{ish}A_{hi}} \mathbf{M}_{ei}(\pm k_{izh},\mathbf{r}) \int \mathbf{M}^*_{ei}(\pm k_{izh},\mathbf{r}')\cdot\mathbf{J}(\mathbf{r}')d\mathbf{r}' \right. \\ & \left. + \frac{1}{k_{ize}k_{ise}A_{ei}} \mathbf{N}_{ei}(\pm k_{ize},\mathbf{r}) \int \mathbf{N}^*_{ei}(\pm k_{ize},\mathbf{r}')\cdot\mathbf{J}(\mathbf{r}')d\mathbf{r}' \right], \quad \begin{matrix} z > z'_{max} \\ z < z'_{nmn}. \end{matrix} \end{aligned} \tag{2.8.47}$$

[8] See [13], Chapter 7.

Notice that $\mathbf{M}_{ei}$ and $\mathbf{N}_{ei}$ denote the TE and TM modes of a waveguide, respectively. The integrals are proportional to the excitation coefficients of the waveguide modes. From the above, we can see that the excitation coefficient of a waveguide mode $\mathbf{E}_j$ is proportional to

$$\text{Excitation Coefficient} \sim \int \mathbf{E}_j(\mathbf{r}') \cdot \mathbf{J}(\mathbf{r}') d\mathbf{r}'. \tag{2.8.48}$$

In other words, to excite a certain mode $\mathbf{E}_j$ in a waveguide maximally, there should be as much projection of $\mathbf{J}$ onto $\mathbf{E}_j$. To avoid the excitation of a mode $\mathbf{E}_j$, the current $\mathbf{J}$ should be chosen to be orthogonal to the mode $\mathbf{E}_j$.

2.8.3 Excitation of Modes by a Filamental Current

Consider a probe in a waveguide as shown in Figure 2.15 [6]. A current in the probe will produce an electromagnetic field that couples to the modes of the waveguide. We shall discuss how to calculate the amplitudes of the excited waveguide modes. Let us assume that the current on the probe is described by a current sheet

$$\mathbf{J}_s = \hat{y} I_0 \delta(x - d). \tag{2.8.49}$$

The above is a current sheet in the xy plane, and has a dimension of amperes per meter. In this simplified case, it is a current flowing in the y direction, and is a function of x only. This probe current does not produce an E_z component of the electric field. Hence, we do not expect the TM modes to be excited. However, the TE modes will be excited because the probe current will produce an H_z component of the magnetic field. If, on the other hand, the current on the probe is not a constant, there will be charge build up on the probe from $\nabla \cdot \mathbf{J} - i\omega\rho = 0$. This charge will induce an E_z component of the electric field, coupling to the TM modes [see Problem 2-11].

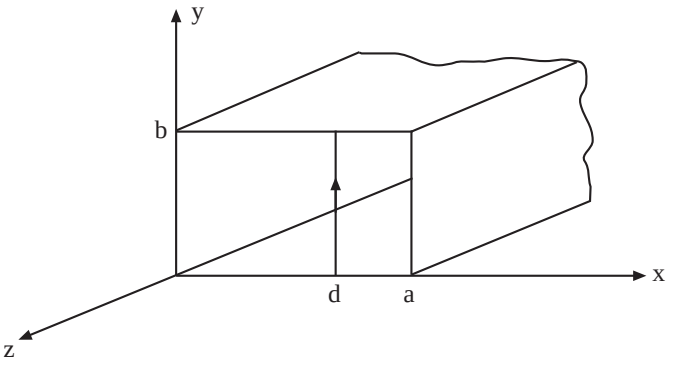


Figure 2.15: Excitation of a rectangular waveguide by a current probe.

For the current probe described by (2.8.49), only TE modes will be excited and we can

write the field as

$$H_z = \begin{cases} \displaystyle\sum_{m,n} H^+_{mn} \cos(\frac{m\pi x}{a}) \cos(\frac{n\pi y}{b}) e^{ik_z z}, & z > 0 \\ \displaystyle\sum_{m,n} H^-_{mn} \cos(\frac{m\pi x}{a}) \cos(\frac{n\pi y}{b}) e^{-ik_z z}, & z < 0. \end{cases} \tag{2.8.50}$$

where $k_z = \sqrt{k_0^2 - \left(\frac{m\pi}{a}\right)^2 - \left(\frac{n\pi}{b}\right)^2}$. The boundary condition across a current sheet is that

$$\hat{z} \times [\mathbf{H}(z+) - \mathbf{H}(z-)] = \mathbf{J}_s, \qquad \hat{z} \times [\mathbf{E}(z+) - \mathbf{E}(z-)] = 0, \tag{2.8.51}$$

where $\hat{z} \times \mathbf{H}$ gives the transverse field component of the magnetic field. Since $\mathbf{J}_s$ has no y variation, we expect $\hat{z} \times \mathbf{H}$ not to have any y variation. Hence, we can assume that $H^{\pm}_{mn} = 0$, $n \neq 0$. Assuming this, we can derive the transverse field components from (2.8.50) using (2.1.7) and (2.1.8), giving

$$\mathbf{H}_s = -\sum_m H^+_{m0} \frac{ik_z}{k_s^2} \hat{x} \frac{m\pi}{a} \sin\left(\frac{m\pi x}{a}\right) e^{ik_z z}, \qquad z > 0, \tag{2.8.52a}$$

$$\mathbf{H}_s = \sum_m H^-_{m0} \frac{ik_z}{k_s^2} \hat{x} \frac{m\pi}{a} \sin\left(\frac{m\pi x}{a}\right) e^{-ik_z z}, \qquad z < 0, \tag{2.8.52b}$$

where $k_s^2 = (\frac{m\pi}{a})^2 + (\frac{n\pi}{b})^2$. Applying the first boundary condition, we get

$$-\sum_m \frac{ik_z}{k_s^2} \hat{y} \frac{m\pi}{a} \sin\left(\frac{m\pi x}{a}\right) (H^+_{m0} + H^-_{m0}) = \hat{y} I_0 \delta(x - d). \tag{2.8.53}$$

Therefore, from the above boundary condition, only the TE_{m0} modes will be excited. Equation (2.8.53) alone is not sufficient to determine the unknowns H^+_{m0} and H^-_{m0}. Therefore, we need to apply the second boundary condition in (2.8.51). However, if one were to note that $H_z(z+) = H_z(z-)$ (a boundary condition which is the subset of the second boundary condition in (2.8.51), we note that

$$H^+_{m0} = H^-_{m0}. \tag{2.8.54}$$

Using (2.8.54) in (2.8.53) will uniquely determine the solution of (2.8.53). Consequently, (2.8.53) becomes

$$\sum_{m=1}^{\infty} -\frac{2ik_z}{k_s^2} H^+_{m0} \frac{m\pi}{a} \sin\left(\frac{m\pi x}{a}\right) = I_0 \delta(x - d). \tag{2.8.55}$$

A simple Fourier series analysis shows that

$$H^+_{m0} = \frac{-k_s^2 I_0}{ik_z m\pi} \sin\left(\frac{m\pi d}{a}\right). \tag{2.8.56}$$

where $k_s^2 = (\frac{m\pi}{a})^2$, and $k_z = \sqrt{k_0^2 - k_s^2}$. This analysis shows that only TE modes with no $\hat{y}$-variation of the field are excited. This is because the probe current in (2.8.49) generates a field with no y-variation. H^+_{m0} is sometimes known as the excitation coefficient of the mode. For a particular TE_{m0} mode, we can change d to alter the amplitude of the excitation coefficient. For example, letting $d/a = 1/2$ will maximize the excitation coefficient of the TE_{10} mode, while the TE_{m0} mode with m even will not be excited.

2.9 Modes of a Hollow Waveguide of Arbitrary Cross-Section

If a metallic waveguide has an arbitrary cross-section, whose shape does not fall on any of the curvilinear coordinates, we will have to find the modes numerically. We have learned that the TE and TM modes are characterized by solutions of the following equations:

$$\text{TE} \qquad (\nabla_s^2 + k_s^2)H_z(\mathbf{r}_s) = 0, \qquad \frac{\partial H_z}{\partial n} = 0 \quad \text{on} \quad C, \tag{2.9.1}$$

$$\text{TM} \qquad (\nabla_s^2 + k_s^2)E_z(\mathbf{r}_s) = 0, \qquad E_z = 0 \quad \text{on} \quad C. \tag{2.9.2}$$

2.9.1 Differential Equation Method

The above are eigenvalue problems. Eigenvalue problems can be converted into variational problems by defining Rayleigh quotient

$$k_s^2 = -\frac{\langle \phi, \nabla_s^2 \phi \rangle}{\langle \phi, \phi \rangle} \tag{2.9.3}$$

The numerator can be simplified by integration by parts.

$$\begin{aligned} \langle \phi, \nabla_s^2 \phi \rangle &= \int_S dS \phi \nabla_s^2 \phi \\ &= -\int_S dS \nabla_s \phi \cdot \nabla_s \phi + \oint_C dl \hat{n} \cdot (\phi \nabla_s \phi) \end{aligned} \tag{2.9.4}$$

With the choice of appropriate boundary conditions, the last term can be made to vanish. Hence, the Rayleigh quotient becomes

$$k_s^2 = \frac{\langle \nabla_s \phi, \nabla_s \phi \rangle}{\langle \phi, \phi \rangle} \tag{2.9.5}$$

The above can be shown to be variational, meaning that a first order error in ϕ gives rise to a second-order error in k_s^2. By letting

$$\phi = \phi_e + \delta\phi \tag{2.9.6}$$

$$k_s^2 = k_{se}^2 + \delta k_s^2 \tag{2.9.7}$$

where ϕ_e and k_{se}^2 are exact values for the function and the eigenvalue, respectively. Then, after cross-multiplying, and taking the first variation, we have

$$k_{se}^2 2\langle \phi_e, \delta\phi \rangle + \langle \phi_e, \phi_e \rangle \delta k_s^2 = -2\langle \nabla_s \phi_e, \nabla_s \delta\phi \rangle \tag{2.9.8}$$

After doing integration by parts on the right-hand side, the term involving $\delta\phi$ cancel each other, and hence $\delta k_s^2 = 0$. In other words, the exact eigenfunctions and eigenvalues of the problems (2.9.1) and (2.9.2) are at the stationary values or stationary points of the Rayleigh quotient (2.9.5).

When complex function ϕ is allowed, a Rayleigh quotient

$$k_s^2 = \frac{\langle \nabla_s \phi^*, \nabla_s \phi \rangle}{\langle \phi^*, \phi \rangle} \tag{2.9.9}$$

The above ensures that k_s^2 is always real for all ϕ.

Rayleign-Ritz Method

In this method, we let

$$\phi = \sum_{n=1}^{N} a_n \phi_n \tag{2.9.10}$$

and pick a_n to make (2.9.5) stationary. By using (2.9.10) in (2.9.5), we arrive at

$$k_s^2 = \frac{\sum_{n,n'} a_n a_{n'} M_{nn'}}{\sum_{n,n'} a_n a_{n'} B_{nn'}} = \frac{\mathbf{a}^t \cdot \overline{\mathbf{M}} \cdot \mathbf{a}}{\mathbf{a}^t \cdot \overline{\mathbf{B}} \cdot \mathbf{a}} \tag{2.9.11}$$

where $\overline{\mathbf{M}}$ and $\overline{\mathbf{B}}$ are symmetric matrices, $M_{nn'} = \langle \nabla_s \phi_n, \nabla_s \phi_{n'} \rangle, B_{nn'} = \langle \phi_n, \phi_{n'} \rangle$. Here, $B_{nn'}$ is also called the Gram matrix. In the equation above, (2.9.11), has stationary points. When we increase the number of unknowns in (2.9.10), the stationary points of (2.9.11) will approach the exact stationary points and hence, the exact answers. Meanwhile, (2.9.10) will approach the exact eigenfunction. We assume that

$$\mathbf{a} = \mathbf{a}_0 + \delta \mathbf{a}, \qquad k_s^2 = k_{s0}^2 + \delta k_s^2 \tag{2.9.12}$$

where $\mathbf{a_0}$ is the value that will optimize (2.9.11), and ${k_{s0}}^2$ is the optimal value of ${k_s}^2$. In other words, we want $\mathbf{a_0}$ to be at the stationary point of (2.9.11). In this case, δk_s^2 will be zero. Subsequently, after cross-multiplying, and taking the first variation, (2.9.11) becomes

$${k_{s0}}^2 2\delta \mathbf{a^t} \cdot \overline{\mathbf{B}} \cdot \mathbf{a_0} + \delta {k_s}^2 \mathbf{a_0^t} \cdot \overline{\mathbf{B}} \cdot \mathbf{a_0} = 2\delta \mathbf{a^t} \cdot \overline{\mathbf{M}} \cdot \mathbf{a_0} \tag{2.9.13}$$

In order for the $\delta {k_s}^2$ to be zero, so that $\mathbf{a_0}$ represents the optimal solution, we require that the $\delta \mathbf{a^t}$ terms cancel each other. Then it is necessary that

$$\overline{\mathbf{M}} \cdot \mathbf{a_0} = {k_{s0}}^2 \overline{\mathbf{B}} \cdot \mathbf{a_0} \tag{2.9.14}$$

The above is the matrix eigenvalue form which we can solve for $\mathbf{a_0}$ and ${k_{s0}}^2$. Once $\mathbf{a_0}$ is found, the eigenfunction ϕ is found via (2.9.10).

Myriads of methods can be used to choose ϕ_n in (2.9.10). If the cross section of the waveguide is arbitrary, it is more practical to triangulate the cross section and pick subdomain basis functons such as pyrimidal functions. Such a method of solution is known as the finite element method. Finite element method (FEM) is vastly popular in solving many differential equation problems [18–21].

In the choice of basis functions in (2.9.10), we can pick the functions to satisfy the homogeneous Dirichlet or Neumann boundary condition to make the last term in (2.9.4) vanish. If we pick basis functions whose values float at the contour C, the solutions will satisfy the homogeneous Neumann boundary condition. Hence, the homogeneous boundary condition is also known as the natural boundary condition [13, p. 308].

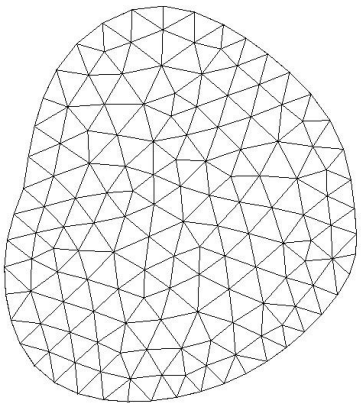


Figure 2.16: Two-dimensional FEM mesh for the waveguide.

2.9.2 Integral Equation Method

A more rigorous way of finding the modes of an arbitrarily shaped waveguide is to use the integral equation approach [22–24]. For this, we need a Green's function which is the solution to the equation

$$(\nabla_s^2 + k_s^2)g(\mathbf{r}_s, \mathbf{r}_s') = -\delta(\mathbf{r}_s - \mathbf{r}_s'), \quad \mathbf{r}_s \in S_1, \quad \mathbf{r}_s' \in S. \tag{2.9.15}$$

Multiplying (2.9.1) by $g(\mathbf{r}_s, \mathbf{r}_s')$ and (2.9.15) by $H_z(\mathbf{r}_s)$, subtracting the two resultant equations, and integrating over S, we obtain

$$\int_S [g(\mathbf{r}_s, \mathbf{r}_s')\nabla_s^2 H_z(\mathbf{r}_s) - H_z(\mathbf{r}_s)\nabla_s^2 g(\mathbf{r}_s, \mathbf{r}_s')]dS = H_z(\mathbf{r}_s'), \quad \mathbf{r}_s' \in S \tag{2.9.16}$$

But since $g\nabla_s^2 H_z - H_z\nabla_s^2 g = \nabla_s \cdot (g\nabla_s H_z - H_z\nabla_s g)$, Gauss' theorem can be used to convert (2.9.16) into

$$\oint_C dl\hat{n} \cdot [g(\mathbf{r}_s, \mathbf{r}_s')\nabla_s H_z(\mathbf{r}_s) - H_z(\mathbf{r}_s)\nabla_s g(\mathbf{r}_s, \mathbf{r}_s')] = H_z(\mathbf{r}_s'), \quad \mathbf{r}_s' \in S. \tag{2.9.17}$$

Since $\hat{n} \cdot \nabla_s H_z(\mathbf{r}_s) = \frac{\partial}{\partial n} H_z(\mathbf{r}_s) = 0$ for $\mathbf{r}_s \in C$ from Equation (2.9.1) we have

$$-\oint dl H_z(\mathbf{r}_s)\hat{n} \cdot \nabla_s g(\mathbf{r}_s, \mathbf{r}_s') = H_z(\mathbf{r}_s'), \quad \mathbf{r}_s' \in C. \tag{2.9.18}$$

We have imposed $\mathbf{r}_s' \in C$ in the above so that it is now an integral equation with the unknown $H_z(\mathbf{r}_s')$, $\mathbf{r}_s' \in C$. Since $\mathbf{r}_s$ can be equal to $\mathbf{r}_s'$ in the integral, the singularity of the Green's function $\hat{n} \cdot \nabla g(\mathbf{r}_s, \mathbf{r}_s')$ makes a straight forward evaluation of the above integral divergent. A principal value integral has to be taken to obtain a convergent integral [13, p. 455].

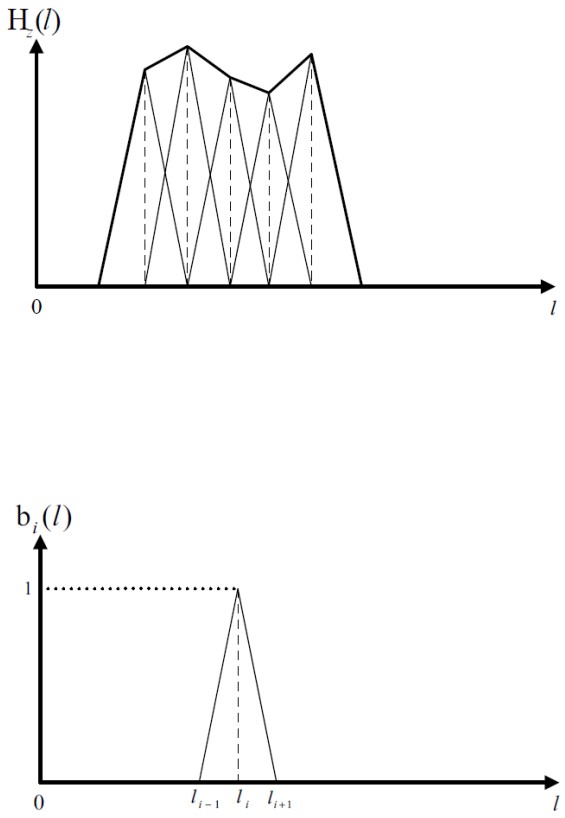


Figure 2.17: Choice of basis functions for $H_z(l)$. $b_i(l)$ is a triangular (hat or chapeau) basis function. It can approximate an arbitrary function as a piecewise linear function.

In the above, $H_z(\mathbf{r}_s)$ can be parameterized as a function of l where l is a variable defining the contour C of the waveguide. For simplicity, we can expand $H_z(l)$ in terms of triangular (also known as hat or chapeau) basis functions, i.e., $H_z(l) = \sum_{i=1}^{N} a_i b_i(l)$. Then, (2.9.18) becomes

$$\sum_{i=1}^{N} a_i \int_{l_{i-1}}^{l_{i+1}} dl \hat{n} \cdot \nabla_s g(l, l') p_i(l) = -\sum_{i=1}^{N} a_i p_i(l'). \tag{2.9.19}$$

After the dl integration, the summand in the above equation is a function of l'. A matrix equation can be obtained by point matching and fixing the above equation at $l' = l_j$, $j =$

$1, \ldots, N$, yielding

$$\sum_{i=1}^{N} A_{ji} a_j = 0, \quad j = 1, \ldots, N, \tag{2.9.20}$$

where

$$A_{ji} = \int_{l_{i-1}}^{l_{i+1}} dl \hat{n} \cdot \nabla_s g(l, l_j) p_i(l) + p_i(l_j). \tag{2.9.21}$$

Equation (2.9.20), hence, reduces to a matrix equation

$$\overline{\mathbf{A}}(k_s) \cdot \mathbf{a} = 0. \tag{2.9.22}$$

A nontrivial solution exists for $\mathbf{a}$ only if

$$\det(\mathbf{A}(k_s)) = 0. \tag{2.9.23}$$

Here, $\mathbf{A}$ is a function of k_s because the Green's function in Equation (2.9.15) is a function of k_s. The above can be satisfied only at certain values of k_s. At these values, (2.9.22) has a non-trivial null space solution, and hence $\mathbf{a}$ is nonzero, and can be found.

The derivation so far requires $g(\mathbf{r}_s, \mathbf{r}'_s)$ to be a solution of (2.9.15) for $\mathbf{r}_s \in S$, and $\mathbf{r}'_s \in S$. A simple Green's function that satisfies this requirement is

$$g(\mathbf{r}_s, \mathbf{r}'_s) = g(\mathbf{r}_s - \mathbf{r}'_s) = \frac{i}{4} H_0^{(1)}(k_s |\mathbf{r}_s - \mathbf{r}'_s|). \tag{2.9.24}$$

A similar integral equation can be derived for TM polarization which is

$$\oint_C dl g(\mathbf{r}_s, \mathbf{r}'_s) \hat{n} \cdot \nabla_s E_z(\mathbf{r}_s) = 0 \quad \mathbf{r}'_s \in C. \tag{2.9.25}$$

This can also be used to find the waveguide modes of an arbitrarily shaped waveguide. In the above, there is no derivative on the Green's function. Hence, this integral equation is less singular compared to the one for TE polarization.

2.9.3 Ad Hoc Method

We describe an ad hoc method for solving for the waveguide modes, but this method is not rigorous. It can only yield the modes satisfactorily if the shape of the waveguide is not too oblong, or if the wall of the waveguide does not have sharp corners.

One way to solve (2.9.1) and (2.9.2) is to expand the scalar field H_z or E_z in terms of functions which are known to be solutions of (2.9.1) or (2.9.2), but do meet the specified boundary conditions. For example, we can let

$$H_z(\mathbf{r}_s) = \sum_n a_n \psi_n(\mathbf{r}_s), \tag{2.9.26}$$

where $\psi_n(\mathbf{r}_s)$ could be [9]

$$\psi_n(\mathbf{r}_s) = J_n(k_s\rho)e^{in\phi}. \tag{2.9.27}$$

$\psi_n(\mathbf{r}_s)$ is clearly a solution to (2.9.1), but the boundary condition is not met. In order to satisfy the boundary condition, we require that the normal derivative of $H_z = 0$, or

$$\frac{\partial H_z}{\partial n} = \sum_n a_n \frac{\partial \psi_n(\mathbf{r}_s)}{\partial n} = 0 \quad \text{on} \quad C. \tag{2.9.28}$$

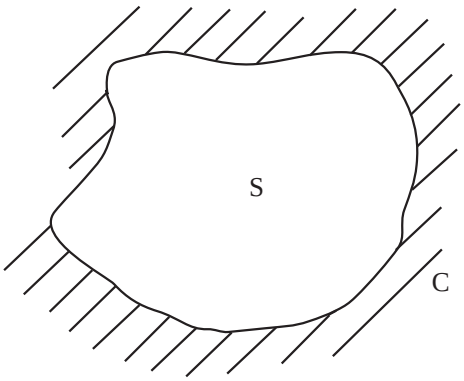


Figure 2.18: A waveguide of arbitrary cross-section.

The general method to solve (2.9.28) is to weight it with a test function and integrate over C, e.g.,

$$\sum_n a_n \left\langle \phi_m, \frac{\partial \psi_n}{\partial n} \right\rangle = 0, \tag{2.9.29}$$

where

$$\left\langle \phi_m, \frac{\partial \psi_n}{\partial n} \right\rangle = \int_C dl \phi_m(\mathbf{r}_s) \frac{\partial \psi_n}{\partial n} = A_{mn}(k_s) \tag{2.9.30}$$

is only a number dependent on k_s and the indices m and n. Equation (2.9.29), hence, reduces to a matrix equation

$$\overline{\mathbf{A}}(k_s) \cdot \mathbf{a} = 0. \tag{2.9.31}$$

A nontrivial solution exists for $\mathbf{a}$ only if

$$\det\left(\overline{\mathbf{A}}(k_s)\right) = 0. \tag{2.9.32}$$

We can search Equation (2.9.32) numerically for the values of k_s that satisfy the equation. These are the eigenvalues of the problem for the TE modes. The eigenvector $\mathbf{a}$ can be found from (2.9.31) at these values of k_s. With the knowledge of $\mathbf{a}$, we can construct H_z, the eigenfunction.

[9] The basis given by (2.9.27) is actually not complete in an arbitrarily shaped waveguide. This method is hence akin to the method of Rayleigh's hypothesis [13].

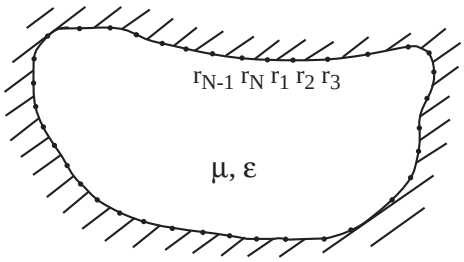


Figure 2.19: The solution of a waveguide problem via the method of point-matching.

For the TM problem, similar analysis gives rise to an **A** matrix given by

$$A_{mn}(k_s) = \int_C dl \phi_m(\mathbf{r}_s)\psi_n(\mathbf{r}_s). \tag{2.9.33}$$

Another way of solving (2.9.28) is to first assign N points on the perimeter of the wave guide C and evaluate the Equation (2.9.29) at these location. Equation (2.9.28) then becomes

$$\sum_n a_n \frac{\partial \psi_n(\mathbf{r}_m)}{\partial n} = 0, \quad m = 1, \ldots, N. \tag{2.9.34}$$

The above is a matrix equation similar to (2.9.31) where

$$A_{mn}(k_s) = \frac{\partial \psi_n(\mathbf{r}_m)}{\partial n}. \tag{2.9.35}$$

Similar analysis for the TE modes gives

$$A_{mn}(k_s) = \psi_n(\mathbf{r}_m). \tag{2.9.36}$$

This particularly simple way of solving Equation (2.9.28) is known as the point matching technique.

Exercises for Chapter 2

Problem 2-1: Many transmission line systems have more than two conductors, like your home telephone line, computer bus etc. If you have an N conductors (a general cylindrical waveguide with translational invariance in z direction) embedded in a homogeneous isotropic medium, how many *independent* TEM (Transverse Electromagnetic) modes can propagate down this line.

Problem 2-2: The reflection of a plane wave by a half space has a closed form solution. Write down the solutions of the reflection of plane TE and TM waves from a half space. Now

make the lower half space a metallic conductor whose conductivity is high but not infinite. Deduce a relationship between the tangential component of the electric field and the magnetic field at the metallic surface, i.e.,

$$E_t = Z_m H_t.$$

Z_m is also known as the surface impedance of the metallic conductor. It is useful for calculating the power absorbed by the metallic conductor.

Problem 2-3: For a coaxial transmission line with an inner conductor of radius a and an outer conductor of radius b, find the electric potential ϕ_s and the magnetic potential ψ_s. Is ψ_s a single value or a multi-value function for each point in space? Comment on this.

There are two ways to calculate the attenuation in a transmission line when the conductor is not perfect. One way is to first calculate the series resistance per unit length, R, and then find the attenuation constant from the formula $(k' + ik'')^2 = -YZ$ where Y is the shunt admittance per unit length, and Z is the series impedance per unit length, and k'' is the attenuation constant. Another way is to find P_d, the power dissipated per unit length by integrating the time average Poynting vector $\langle \mathbf{E} \times \mathbf{H} \rangle$ over the surface of the conductors. Since the tangential electric field is not zero anymore on a non-perfectly conducting metallic surface, the time average Poynting vector is not zero on the metallic surfaces. After finding the P_d, one can calculate the attenuation constant from

$$k'' = \frac{P_d}{2P_0}$$

where P_0 is the time average total power flow down the transmission line. You can use the surface impedance found from Problem 2 to calculate the power absorbed by the metallic conductor.

(a) Find the attenuation constant for a coaxial line using these two methods, and show that they are the same.

(b) For a 50 ohm coaxial transmission line with an outer radius of .5 cm, filled with teflon with $\epsilon = 2\epsilon_0$, calculate the loss due to attenuation in terms of dB/Km at 20 KHz and 1 GHz, if we assume that the conductor is made out of copper with $\sigma = 5.8 \times 10^7$ mho/m.

Problem 2-4: Can a single piece of conductor embedded in a homogeneous medium support a TEM mode? Discuss why and why not.

Problem 2-5:

(a) Explain why the line capacitance (farad/m) of the TEM mode in a transmission line can be found by solving an electrostatic problem.

(b) Given the knowledge of the line capacitance, show that it has to be of the form $C = \epsilon K$ where K is a dimensionless, geometry dependent factor.

(c) Show that now if there is dielectric loss such that $\epsilon = \epsilon' + i\sigma/\omega$, then the admittance per unit length to be used in the telegraphists equation (for lossy line) is $Y = -i\omega\epsilon' K + \sigma K$. Hence, $G = \sigma K$.

Problem 2-6: The eigenvalue k_{is}^2 of the partial differential equation $(\nabla_s^2 + k_{is}^2)\psi_{iz} = 0$ is always real, and can be proven as follows:

(a) Assume that k_{is}^2 is complex so that ψ_{iz}^* is a solution of $(\nabla_s^2 + {k_{is}^*}^2)\psi_{iz}^* = 0$. Show that

$$\psi_{iz}^* \nabla_s^2 \psi_{iz} - \psi_{iz} \nabla_s^2 \psi_{iz}^* = ({k_{is}^*}^2 - k_{is}^2)|\psi_{iz}|^2.$$

(b) Assume that ψ_{iz} is either E_{iz} or H_{iz}, integrate the equation in part (a) over the cross-section of a metallic waveguide and show that

$$({k_{is}^*}^2 - k_{is}^2) \int_S |\psi_{iz}|^2 dS = 0.$$

From the above, explain why k_{is}^2 has to be real.

Problem 2-7: Prove that

$$\int_S dS \mathbf{H}_{is} \cdot \mathbf{H}_{js} = 0, \qquad i \neq j$$

for any two distinct modes for a homogeneously filled closed waveguide. S in the above refers to the cross-sectional area of the waveguide that is permeated by the field.

Problem 2-8:

(a) Prove that $\mathbf{E}_{is}$ and $\mathbf{E}_{js}$ are always orthogonal if one field is coming from a TM (TE) mode while the other field is coming from a TE (TM) mode. Hint: Express these fields in terms of the longitudinal components of the fields.

(b) Prove that

$$\int_S (\mathbf{E}_{is} \times \mathbf{H}_{js}^*) \cdot \hat{z} dS = 0$$

for $i \neq j$.

(c) From the above, explain how the concept of energy orthogonality and power orthogonality follow.

Problem 2-9: For the TE_{10} mode of a rectangular waveguide, find its attenuation constant due to wall loss. Do the same for the TE_{11} and the TE_{01} modes of a circular waveguide. Sketch the attenuation constants as functions of frequency. Which mode has the lowest loss at high frequencies?

Problem 2-10: Derive the complete expression for the fields of the TE_{11} mode of a rectangular waveguide and sketch the field patterns for both the **E** and **H** fields in the cross-section of the waveguide. Do the same for the TM_{11} mode of the rectangular waveguide, and the TE_{12} mode of the circular waveguide. (A good feel for the field pattern of the waveguide mode is essential in determining how to excite it with a probe, see Equation (2.8.48).)

Problem 2-11:

A probe with a current distribution represented by a current sheet given by

$$\mathbf{J}_s(\mathbf{r}) = \hat{y}\delta(x - a/2)\sin(\frac{\pi y}{b})$$

is placed in a waveguide. Calculate the excitation coefficients of all the modes excited by this probe.

Problem 2-12: Certain modes of metallic waveguides have decreasing loss for increasing frequency.

(a) For the TE_{10} mode of a rectangular waveguide, find its attenuation constant due to wall loss. Do the same for the TE_{11} and the TE_{01} modes of a circular waveguide. Hint: Integral of product of Bessel functions is needed in the solution. Equation (11.4.5) of Abramowitz and Stegun comes in useful here.

(b) Sketch the attenuation constants as functions of frequency. Which mode has the lowest loss at high frequencies?

(c) For the TE_1 mode of a parallel plate waveguide, does its attenuation constant increase or decrease with frequency?

(d) From the above, explain the feature of the "magic" mode that has decreasing loss for increasing frequency.

Problem 2-13:

(a) Give an intuitive explanation as to why the TE_{11} mode of a circular waveguide has a lower cutoff frequency than the TE_{01} mode.

(b) Explain why the TE_{00} mode of a rectangular waveguide cannot exist.

Problem 2-14: Prove the mutual orthogonality of the vector wave functions in a hollow waveguide, that is

$$\int dV \mathbf{M}_{\substack{e\\h}\, i}(k_z, \mathbf{r}) \cdot \mathbf{N}_{\substack{e\\h}\, j}(-k_z', \mathbf{r}) = 0,$$

$$\int dV \mathbf{L}_{\substack{e\\h}\, i}(k_z, \mathbf{r}) \cdot \mathbf{N}_{\substack{e\\h}\, j}(-k_z', \mathbf{r}) = 0,$$

$$\int dV \mathbf{L}_{\substack{e\\h}\, i}(k_z, \mathbf{r}) \cdot \mathbf{M}_{\substack{e\\h}\, j}(-k_z', \mathbf{r}) = 0,$$

for all i, j, k_z and k_z'.

Problem 2-15:

For the coaxial waveguide shown:

(a) Write down the guidance conditions for all the modes in the waveguide using Bessel and Neumann functions. Is the TEM mode a special case of these guidance conditions?

(b) Derive the **M**, **N**, and **L** vector wave functions for the electric and magnetic fields of this waveguide. (Hint: Use sine and cosine functions for the ϕ dependence.)

(c) Write down the orthogonality relationships for the vector wave functions with the A_{ei} and A_{hi} (defined in the text) derived explicitly.

(d) For a source in the circular waveguide described by

$$\mathbf{J}(\mathbf{r}) = \hat{z} I \ell \delta \left(\rho - \frac{a+b}{2} \right) \delta(\phi) \delta(z)/\rho,$$

find the **E** field in terms of the vector wave function.

Problem 2-16: Go through the exercise of deriving the dyadic Green's function of a waveguide. Most of the derivation is already outlined in the text. Fill in the details.

Problem 2-17: For a rectangular waveguide:

(a) Write down the guidance conditions for all the modes in the waveguide.

(b) Derive the **M**, **N** and **L** vector wave functions for the electric and magnetic fields of this waveguide.

(c) You are given three Hertzian electric dipoles. Describe how you would place these dipoles in a rectangular waveguide and weight their amplitudes so that the TE_{10} is excited but the TE_{20} and TE_{30} modes would not be excited (note: you may excite all the other modes).

Problem 2-18:

(a) Derive the integral equation for finding the modes of the TM polarization of a hollow waveguide:

$$\oint_C dl g(\mathbf{r}_s, \mathbf{r}'_s) \hat{n} \cdot \nabla_s E_z(\mathbf{r}_s) = 0, \quad \mathbf{r}'_s \in C.$$

(b) Describe how you would solve the above equation numerically.

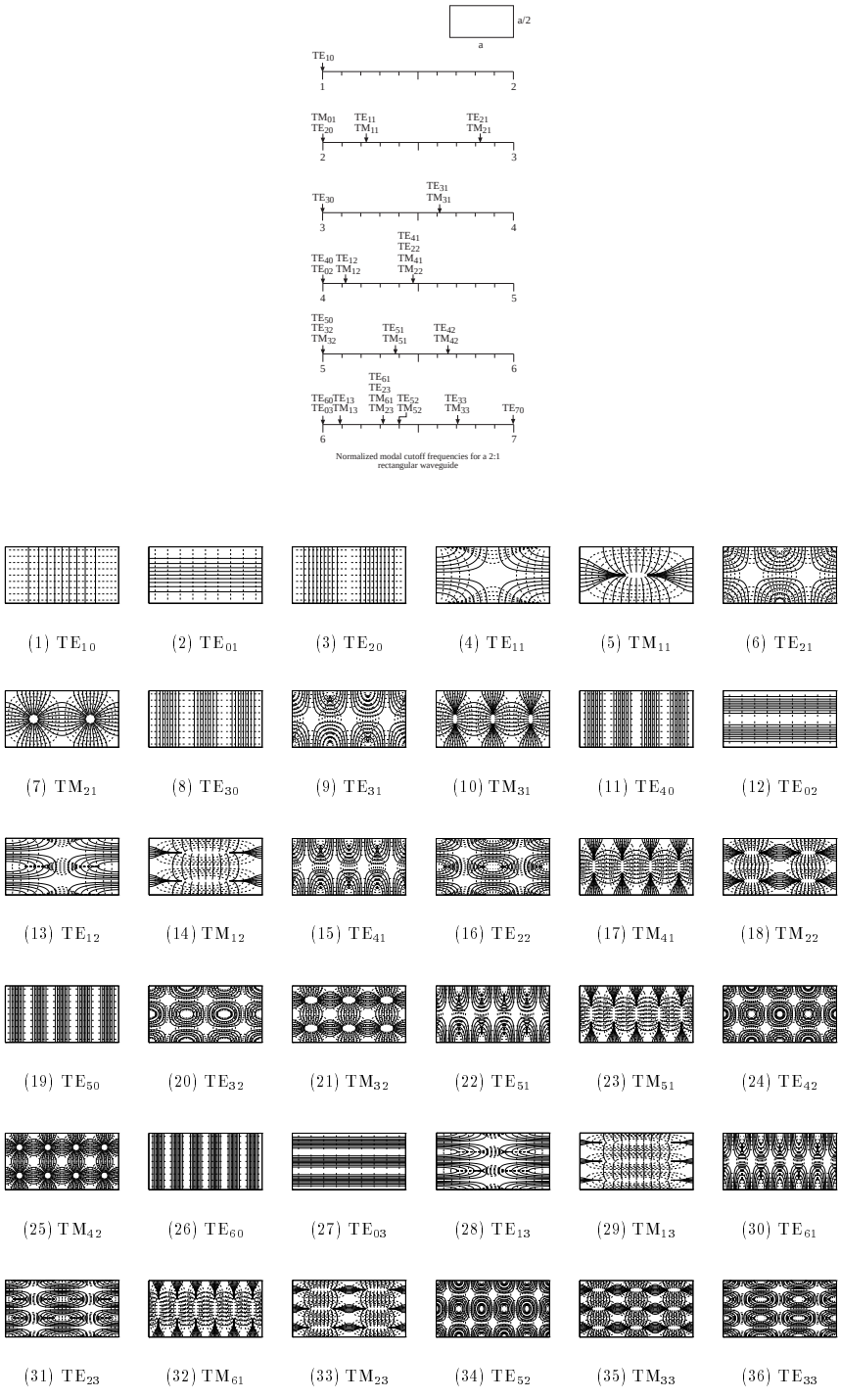
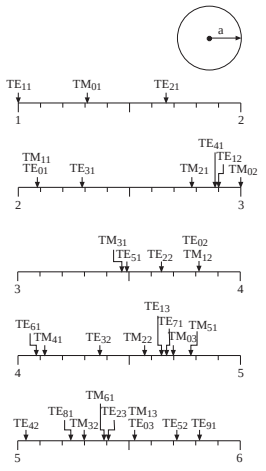


Figure 2.20: Cutoff frequencies (normalized with $\omega_{10c} = 1$) and field plots of the first 36 modes of a rectangular waveguide (courtesy of A. Greenwood).



Normalized model cutoff frequencies for a circular waveguide.

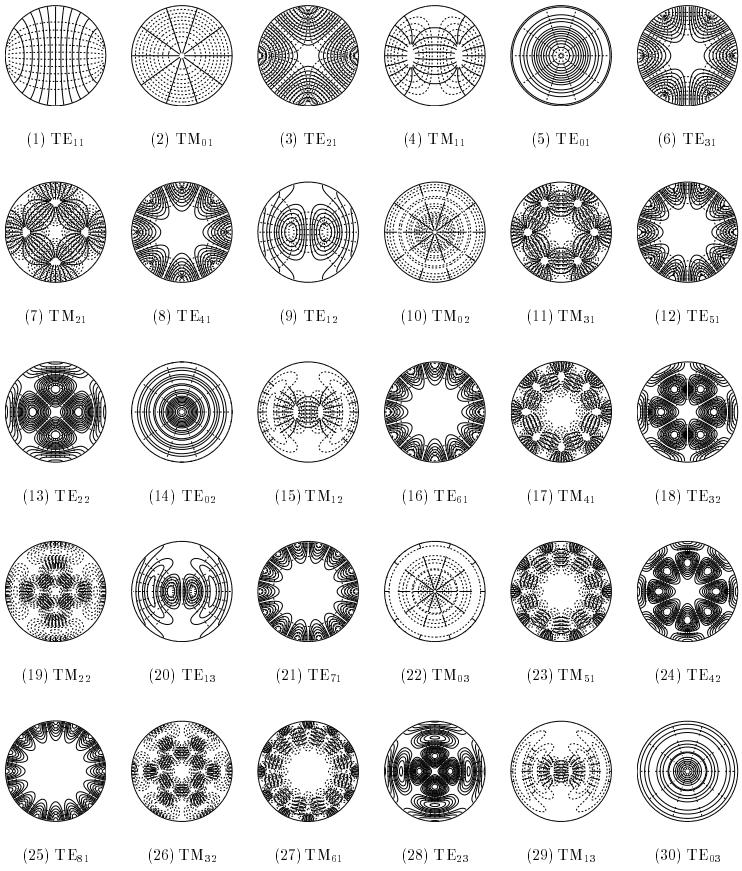


Figure 2.21: Cutoff frequencies (normalized with $\omega_{11c} = 1$) and field plots of the first 30 modes of a circular waveguide (courtesy of A. Greenwood).

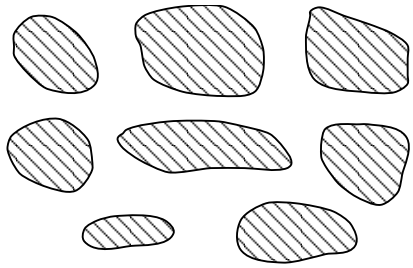


Figure 2.22: Problem 2-1

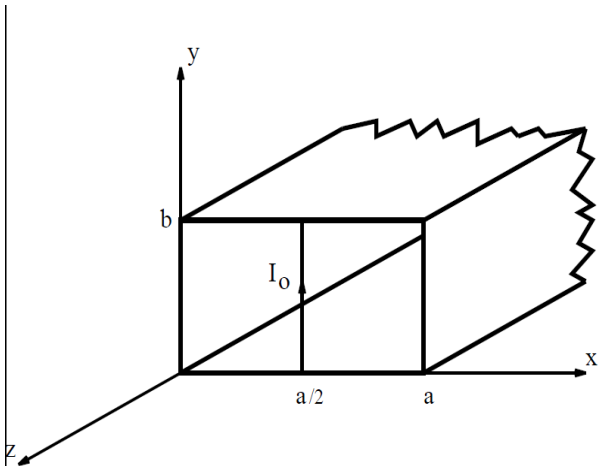


Figure 2.23: Problem 2-10

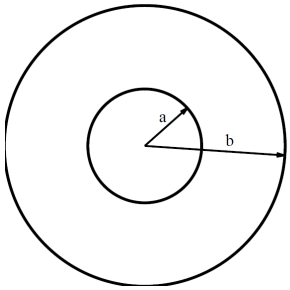


Figure 2.24: Problem 2-12

Bibliography

[1] H.R.L. Lamont, *Wave Guides,* third ed., Methuen, London, 1950.

[2] N. Marcuvitz, ed., *Waveguide Handbook,* MIT Radiation Laboratory Series, vol, 10, McGraw-Hill, New York, 1951.

[3] K.G. Budden, *The Wave-Guide Mode Theory of Wave Propagation,* Englewood Cliffs, NJ, Prentice Hall, 1961.

[4] R.E. Collin, *Field Theory of Guided Waves,* IEEE Press, Piscataway, NJ, 1991.

[5] R.E. Collin, *Foundation for Microwave Engineering,* IEEE Press, Piscataway, NJ, 2001.

[6] J.A. Kong, *Electromagnetic Wave Theory,* EMW Publishing, Cambridge, MA, 2000.

[7] D.M. Pozar, *Microwave Engineering,*, John Wiley & Sons, 2000.

[8] C.S. Lee, S.W. Lee, and S.L. Chuang, "Plot of modal field distribution in rectangular and circular waveguides," *IEEE Trans. Micro. Theory Tech.,* vol. MTT-33, no. 3, pp. 271-274, March 1985.

[9] J.W. Archer, "TE0n-mode filter for VLA circular waveguide system," *Electronic Letters*, vol. 15, pp. 343-345, June 7, 1979.

[10] J.A. Stratton, *Electromagnetic Theory,* McGraw-Hill, NY, 1941.

[11] W.W. Hansen, "A new type of expansion in radiation problems," *Phys. Rev.*, 47, 139143, 1935).

[12] R.E. Collin, "On the incompleteness of E and H modes in waveguides," Can. J. Phys., 51, 1135-1140, 1973.

[13] W.C. Chew, *Waves and Fields in Inhomogeneous Media,* Van Nostrand Reinhold, New York, 1990, reprinted, IEEE Press, Piscataway, NJ, 1995.

[14] C.T. Tai, *Dyadic Green's Functions in Electromagnetic Theory,*, Intext Pub., NY, 1971.

[15] C.T. Tai, "On the eigenfunction expansion of dyadic Green's functions," *Proc. IEEE*, 61, 480-481, 1973.

[16] W.A. Johnson, A.Q. Howard, and D.G. Dudley, "On the irrotational component of the electric Green's dyadic," *Radio Sci.*, 14, 961-967, 1979.

[17] A.D. Yaghjian, "Electric dyadic Green's functions in the source region," *Proc. IEEE*, 68, 248-263, 1980.

[18] O.C. Zienkiewicz, *The finite element method in engineering science,* McGraw-Hill, 1971.

[19] P.P. Silvester and R.L. Ferrari, *Finite Elements for Electrical Engineers*, first edition, Cambridge University Press, 1983.

[20] J.M. Jin, *The Finite Element Method in Electromagnetics*, Wiley, 1993.

[21] J.L. Volakis, A. Chatterjee, L.C. Kempel, *Finite Element Method Electromagnetics: Antennas, Microwave Circuits, and Scattering Applications*, Wiley, 1998.

[22] R.F. Harrington, *Field computation by moment methods*, Macmillan, 1968.

[23] W.C. Chew, J.M. Jin, E. Michielssen, and J.M. Song, (editors), *Fast and Efficient Algorithms in Computational Electromagnetics*, Artech House, Boston, MA, 2001.

[24] W.C. Chew, M.S. Tong, and B. Hu, *Integral Equations Methods for Electromagnetic and Elastic Waves*, Morgan & Claypool, 2008.

Chapter 3

Inhomogeneously Filled Waveguides

The theory of inhomogeneously filled waveguides covers a large class of waveguides [1, 3–6]. Waveguides are filled with inhomogeneous material to give the waveguide a certain property. For instance, phase-shifters, polarizers, and attenuators are made out of inhomogeneously filled waveguides. Often, waveguides are filled inhomogenously with ferrite material to make non-reciprocal waveguides. To make the fabrication of waveguides simpler, many waveguides are also filled with inhomogeneous materials. Optical waveguides almost exclusively comprise inhomogeneously filled material.

3.1 The Need for Hybrid modes

When a hollow waveguide is filled with inhomogeneity, the most general case is the existence of the hybrid modes. In these modes, the TE_z and TM_z modes are coupled. Only for cases with special symmetry are they decoupled. The reason is that both TE_z and TM_z fields are needed to match the boundary condition at the dielectric interface.

Assume that we only have TE_z field inside the waveguide. Then,

$$\mathbf{H}_s = \frac{1}{k_s^2}\left[\frac{\partial}{\partial z}\nabla_s H_z\right] = \frac{1}{k_s^2} ik_z \nabla_s H_z \tag{3.1.1}$$

The boundary condition for H_z is that it is continuous, or $H_{1z} = H_{2z}$ at the interface between two dielectric regions. Furthermore, we require that,

$$\hat{n} \times \mathbf{H}_{1s} = \hat{n} \times \mathbf{H}_{2s} \tag{3.1.2}$$

implying that at the dielectric interface,

$$\frac{1}{k_{1s}^2} ik_z \hat{n} \times \nabla_s H_{1z} = \frac{1}{k_{2s}^2} ik_z \hat{n} \times \nabla_s H_{2z} \tag{3.1.3}$$

Due to phase matching, k_z is the same in all regions. If H_{1z} is continuous at a dielectric interface, then $\hat{n}\times\nabla_s H_{1z} = \hat{n}\times\nabla_s H_{2z}$ also, since these are tangential derivatives. Therefore, (3.1.3) cannot be satisfied since $k_{1s} \neq k_{2s}$ in general. In order to satisfy the boundary condition, (3.1.1) has to be augmented with the contribution from TM_z field.

However, under special circumstances, Equation (3.1.3) can be satisfied if:

(1) $k_z = 0$, implying that, $\mathbf{H}_s = 0$ and $\mathbf{H} = \hat{z}H_z$ only. In this case, it reduces to a two-dimensional problem;

(2) $\hat{n}\times\nabla_s H_{iz} = 0$ at the interface, implying symmetry that exists for certain modes, for instance, in a dielectric slab or in an axi-symmetric geometry such as a circular optical fiber;

(3) The surface is a PMC surface so that $H_z = 0$, $\hat{n}\times\mathbf{H}_s = 0$. This boundary condition is sufficient to guarantee the uniqueness of the TM_z mode alone in the waveguide;

(4) The surface is a PEC surface so that the boundary condition is for $\hat{n}\times\mathbf{E} = 0$, giving rise to $\hat{n}\cdot\nabla H_z = 0$ on the surface. Again, this boundary condition is sufficient to guarantee the uniqueness of the solution, needless for the coupling of the TE_z and TM_z modes.

Similarly, arguments above apply to the TM_z field. Under these special circumstances, the field is not depolarized at the interface. Namely, if the field is TE_z or TM_z before impinging on the interface, the scattered field off the interface remains the same as the original polarization.

3.2 Derivation of Pertinent Equation

When a uniform waveguide is filled with inhomogeneous materials, the guided modes of the structure cannot be decomposed into TE and TM modes, except for some very special cases. In other words, the E_z and H_z components of the fields are always coupled together. Such modes are also called the hybrid modes. This coupling can be shown from Maxwell's equations, which imply

$$\nabla\times\mu^{-1}\nabla\times\mathbf{E} - \omega^2\epsilon\mathbf{E} = 0. \tag{3.2.1}$$

The above governs the **E**-field propagating in an inhomogeneously filled waveguide. Assuming $e^{ik_z z}$ dependence, we can decompose

$$\mathbf{E} = \mathbf{E}_s + \mathbf{E}_z, \qquad \nabla = \nabla_s + \hat{z}ik_z, \tag{3.2.2}$$

for a guided mode in the waveguide. With the use of the above,

$$\begin{aligned}\nabla\times\mu^{-1}\nabla\times\mathbf{E}&\\ &= \nabla_s\times\mu^{-1}\nabla_s\times\mathbf{E}_s + ik_z\hat{z}\times\mu^{-1}\nabla_s\times\mathbf{E}_z - k_z^2\mu^{-1}\hat{z}\times\hat{z}\times\mathbf{E}_s\\ &+ \nabla_s\times\mu^{-1}\nabla_s\times\mathbf{E}_z + ik_z\nabla_s\times\mu^{-1}\hat{z}\times\mathbf{E}_s.\end{aligned} \tag{3.2.3}$$

Because μ is a function of $\mathbf{r}_s$, it does not commute with ∇_s. The last two terms in (3.2.3) are directed in the z direction. Therefore, by equating the z components in (3.2.1), we have

$$\nabla_s\times\mu^{-1}\nabla_s\times\mathbf{E}_z + ik_z\nabla_s\times\mu^{-1}\hat{z}\times\mathbf{E}_s - \omega^2\epsilon\mathbf{E}_z = 0. \tag{3.2.4}$$

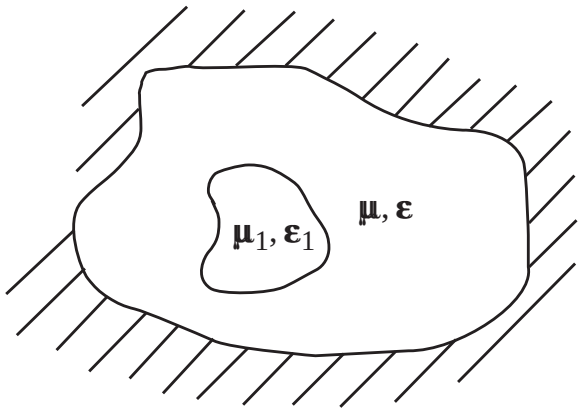


Figure 3.1: An inhomogeneously filled waveguide.

3.2.1 E_z-H_z Formulation

It can be shown that

$$\nabla_s \times \mu^{-1} \nabla_s \times \mathbf{E}_z = -\nabla_s \cdot \mu^{-1} \nabla_s \mathbf{E}_z, \tag{3.2.5a}$$

$$\nabla_s \times \mu^{-1} \hat{z} \times \mathbf{E}_s = \hat{z} \nabla_s \cdot \mu^{-1} \mathbf{E}_s. \tag{3.2.5b}$$

In (3.2.5a), $\nabla_s \cdot \mu^{-1} \nabla_s$ is a scalar operator, and hence, the right-hand side is still a vector pointing in the $\hat{z}$ direction. Using[1]

$$\mathbf{E}_s = \frac{i}{k_s^2} [k_z \nabla_s E_z + \omega \mu \nabla_s \times \mathbf{H}_z], \tag{3.2.6}$$

where now $k_s^2 = \sqrt{\omega^2 \mu \epsilon - k_z^2}$ is a function of $\mathbf{r}$, we can rewrite (3.2.4) as

$$\nabla_s \cdot \mu^{-1} \nabla_s E_z - i k_z \nabla_s \cdot \left(\frac{i k_z}{\mu k_s^2} \nabla_s E_z + \frac{i \omega}{k_s^2} \nabla_s \times \mathbf{H}_z \right) + \omega^2 \epsilon E_z = 0. \tag{3.2.7}$$

In the above, $k_s = \sqrt{k^2 - k_z^2}$. Since k is a function of $\mathbf{r}_s$, so is k_s. Therefore, in general, the equation governing E_z is coupled to H_z as well. By duality principle, the equation governing H_z, which is also coupled to E_z, is

$$\nabla_s \cdot \epsilon^{-1} \nabla_s H_z - i k_z \nabla_s \cdot \left(\frac{i k_z}{\epsilon k_s^2} \nabla_s H_z - \frac{i \omega}{k_s^2} \nabla_s \times \mathbf{E}_z \right) + \omega^2 \mu H_z = 0. \tag{3.2.8}$$

However, if μ and ϵ are constants, then k_s^2 is a constant, and $\nabla_s \cdot \nabla_s \times \mathbf{E}_z = \nabla_s \cdot \nabla_s \times \mathbf{H}_z = 0$, the E_z and H_z equations are decoupled again. Therefore, for general μ and ϵ which are inhomogeneous, the TE and the TM fields in a waveguide are coupled. Exceptions are sometimes found, e.g., in the axial symmetric modes of a circular optical fiber.

[1]This equation is valid in an inhomogeneously filled waveguide, because the only assumption made in deriving it is that the field varies as $\exp(ik_z z)$. However, k_s is not constant anymore.

The above equations show that the TE and TM fields have to co-exist in the waveguide, hence giving rise to hybrid modes. However, they are not cast in terms of eigenvalue problems. We will derive equations from which the eigenvalues and eigenfunctions of the waveguide can be derived.

3.2.2 Transverse Field Formulation

The transverse components of the field in Equation (3.2.1) can be isolated to obtain

$$\nabla_s \times \mu^{-1}\nabla_s \times \mathbf{E}_s + ik_z\mu^{-1}\hat{z} \times \nabla_s \times \mathbf{E}_z + k_z^2\mu^{-1}\mathbf{E}_s - \omega^2\epsilon\mathbf{E}_s = 0. \tag{3.2.9}$$

Using $\nabla \cdot \epsilon\mathbf{E} = 0$, we have

$$ik_z E_z = -\epsilon^{-1}\nabla_s \cdot \epsilon\mathbf{E}_s. \tag{3.2.10}$$

Therefore

$$ik_z\hat{z} \times \nabla_s \times \mathbf{E}_z = ik_z\nabla_s E_z = -\nabla_s\epsilon^{-1}\nabla_s \cdot \epsilon\mathbf{E}_s. \tag{3.2.11}$$

Consequently, an equation that governs the transverse electric field for the i-th mode is

$$\mu\nabla_s \times \mu^{-1}\nabla_s \times \mathbf{E}_{is} - \nabla_s\epsilon^{-1}\nabla_s \cdot \epsilon\mathbf{E}_{is} - k^2\mathbf{E}_{is} + k_{iz}^2\mathbf{E}_{is} = 0, \tag{3.2.12}$$

where $k^2 = \omega^2\mu\epsilon$. For reason to be explained later, we can multiply the above by $z\times$ to get

$$z \times \mu\nabla_s \times \mu^{-1}\nabla_s \times \mathbf{E}_{is} - z \times \nabla_s\epsilon^{-1}\nabla_s \cdot \epsilon\mathbf{E}_{is} - k^2 z \times \mathbf{E}_{is} + k_{iz}^2 z \times \mathbf{E}_{is} = 0, \tag{3.2.13}$$

In addition, an equation for the transverse magnetic field for the j-th mode is easily obtained by invoking duality, namely,

$$z \times \epsilon\nabla_s \times \epsilon^{-1}\nabla_s \times \mathbf{H}_{js} - z \times \nabla_s\mu^{-1}\nabla_s \cdot \mu\mathbf{H}_{js} - k^2 z \times \mathbf{H}_{js} + k_{jz}^2 z \times \mathbf{H}_{js} = 0. \tag{3.2.14}$$

In the above, k_z^2 are the eigenvalues since it is constant throughout the uniform waveguide. Due to the phase matching condition, k_z is a constant everywhere in an inhomogeneous waveguide. We need to solve either Equation (3.2.13) or (3.2.14) since they can be shown to be transpose of each other.

Notice that whether if we describe the modes in a waveguide with Equation (3.2.7) and (3.2.8), or Equations (3.2.13) and (3.2.14), only two components of the electric field or the magnetic field are required to describe the modes in an inhomogeneously filled waveguide. This is because $\nabla \cdot \mathbf{B} = 0$ and $\nabla \cdot \mathbf{D} = 0$ and hence, not all three components of the $\mathbf{B}$ field or the $\mathbf{D}$ field are independent of each other.

As a last note, as shall be shown later, (3.2.13) and (3.2.14) are transpose of each other. Hence, eigenfunctions from these equations are mutually orthogonal.

3.2.3 Physical Interpretation of the Depolarization Effect

The fact that the E_z and H_z waves are in general coupled at a dielectric interface is also known as depolarization effect. An H_z (TE to z) wave incident at a dielectric rod, in general, generates both H_z and E_z (TM to z) waves, and hence, causes the depolarization of the wave. This happens conversely for E_z wave incident on a dielectric rod. This depolarization effect

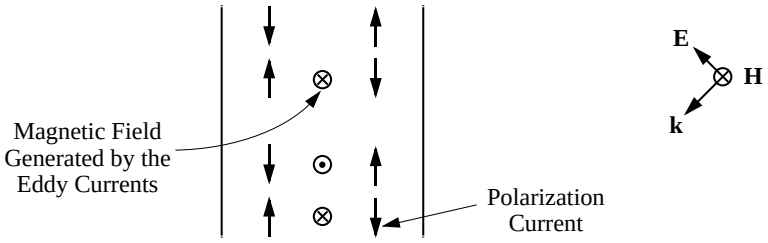


Figure 3.2: Polarization current and magnetic field generated by the induced current on a dielectric rod illuminated by an E_z wave.

occurs only for E_z or H_z waves that vary as a function of z ($k_z \neq 0$). The depolarization effect disappears when the waves do not vary as a function of z ($k_z = 0$), or when the scatterer is a cylindrical perfect electric conductor (PEC) or a perfect magnetic conductor.

A guided mode can be thought of as a wave bouncing around in a waveguide such that phase coherence or constructive interference occurs. The condition for phase coherence or constructive interference is precisely the guidance condition of the waveguide modes. In the case of a hollow waveguide, the waves are bouncing off a PEC cylindrical surface and hence, the polarization purity can be maintained. Therefore, the mode can be either purely E_z or H_z type. When a waveguide is inhomogeneously filled, the waves have to bounce off a dielectric rod, and in general, E_z or H_z mode purity cannot be maintained. These modes are termed the hybrid modes.

We can argue by contradiction, except for special symmetric cases, that only a hybrid mode consisting of TE and TM waves is possible. When an E_z polarized (TM to z) wave is obliquely incident on a dielectric slab as shown in Figure 3.2, polarization purity can be kept if the slab is infinitely wide coming out of the paper. The magnetic field will be horizontal in the slab with the electric polarization current flowing around it. The magnetic field alternates in its polarity as one moves in the z direction. The k vector lies in the plane of the paper both in the air and in the slab. Now assume that the slab is truncated so that it extends in finitely into the paper as well as out of the paper. Assume that polarization purity is still preserved. Then at the interface at the truncated surfaces, the k vectors are both parallel to the surface inside the dielectric as well as in the air. This is an impossibility since the phase velocity of the waves on two sides of the interface are different and the boundary condition can never be met since phase matching is violated. Hence, the k vector has to "tilt" in order to satisfy the boundary condition, introducing the z component of the magnetic field.

Alternatively, we can consider an H_z incident wave (TE to z). Let us assume that the field remains TE inside the dielectric rod and see that it will lead to a contradiction. If this is the case, the transverse electric field will at least induce polarization currents flowing in the xy directions. When these currents meet a dielectric interface, polarization charges are induced at the interface. Because of the z-variation of the incident field, these charges must be sign-changing in the z-direction. The electric field has to turn around due to the different phase velocity it has in the air compared to the dielectric. Therefore, an E_z field must exist due to these charges.

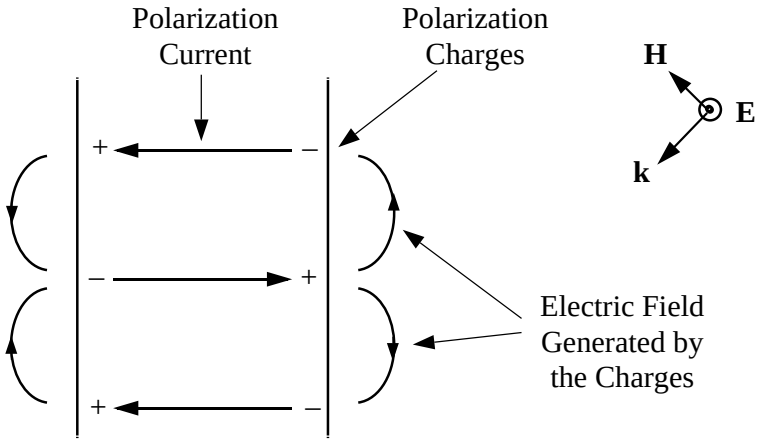


Figure 3.3: Induced polarization currents and charges in a dielectric rod if the field remains TE to z when the rod is illuminated by an H_z wave.

The above explains the general depolarization of the wave. However, certain symmetrical rods may not depolarize certain incident field. For instance, an infinite dielectric slab will not depolarize a TE (or TM) wave where the **E** (or **H**) field is aligned parallel to the dielectric interface. A circular dielectric rod will not depolarize an axially symmetric E_ϕ or H_ϕ polarized TE or TM wave incident on it. In the above discussion, we are referring to TE to z or TM to z waves.

3.2.4 Mode Orthogonality

In an inhomogeneous waveguide, the differential operators in (3.2.13) and (3.2.14) are not self-transpose (or symmetric) which is unlike a homogeneously filled cylindrical waveguide. Therefore, general orthogonality between $\mathbf{E}_{is}$ and $\mathbf{E}_{js}$, or $\mathbf{H}_{is}$ and $\mathbf{H}_{js}$ does not exist. However, mode orthogonality exists between $\mathbf{E}_{is}$ and $\mathbf{H}_{js}$. We can prove the orthogonality of $\mathbf{E}_{is}$ and $\mathbf{H}_{js}$ for a general, inhomogeneous, anisotropic waveguide using the Lorentz reciprocity theorem.

Consider two waveguides with identical walls but different anisotropic, inhomogeneous media: one waveguide is filled with $\overline{\boldsymbol{\mu}}$, $\overline{\boldsymbol{\epsilon}}$, while the other one is filled with $\overline{\boldsymbol{\mu}}^t(\mathbf{r}_s)$, $\overline{\boldsymbol{\epsilon}}^t(\mathbf{r}_s)$ where the superscript t stands for transpose (see Figure 3.4, the second problem is the auxiliary of the first one). It can be shown that

$$\int_S (\mathbf{E}_{is} \times \mathbf{H}_{js}) \cdot \hat{z} dS = 0, \qquad i \neq j, \tag{3.2.15}$$

where $\mathbf{E}_i$ is the **E**-field inside the first waveguide, while $\mathbf{H}_j$ is the **H**-field inside the second waveguide. In the first waveguide,

$$\nabla \times \mathbf{E}_i = i\omega \overline{\boldsymbol{\mu}} \cdot \mathbf{H}_i, \qquad \nabla \times \mathbf{H}_i = -i\omega \overline{\boldsymbol{\epsilon}} \cdot \mathbf{E}_i, \tag{3.2.16}$$

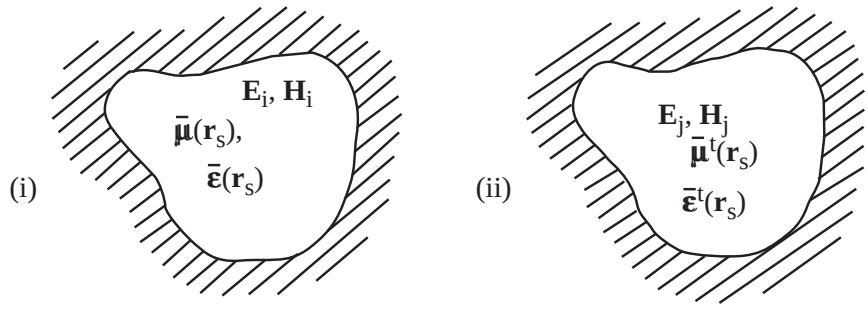


Figure 3.4: Proof of mode orthogonality. The second problem is the auxiliary of the first where the waveguide is filled with transposed material tensors.

while in the second waveguide,

$$\nabla \times \mathbf{E}_j = i\omega \overline{\boldsymbol{\mu}}^t \cdot \mathbf{H}_j, \qquad \nabla \times \mathbf{H}_j = -i\omega \overline{\boldsymbol{\epsilon}}^t \cdot \mathbf{E}_j. \tag{3.2.17}$$

We can show that

$$\begin{aligned} \nabla \cdot (\mathbf{E}_i \times \mathbf{H}_j - \mathbf{E}_j \times \mathbf{H}_i) = {} & i\omega(\mathbf{H}_j \cdot \overline{\boldsymbol{\mu}} \cdot \mathbf{H}_i - \mathbf{H}_i \cdot \overline{\boldsymbol{\mu}}^t \cdot \mathbf{H}_j) \\ & + i\omega(\mathbf{E}_i \cdot \overline{\boldsymbol{\epsilon}}^t \cdot \mathbf{E}_j - \mathbf{E}_j \cdot \overline{\boldsymbol{\epsilon}} \cdot \mathbf{E}_i). \end{aligned} \tag{3.2.18}$$

Since $\mathbf{A} \cdot \overline{\mathbf{B}}^t \cdot \mathbf{C} = \mathbf{C} \cdot \overline{\mathbf{B}} \cdot \mathbf{A}$,

$$\nabla \cdot (\mathbf{E}_i \times \mathbf{H}_j - \mathbf{E}_j \times \mathbf{H}_i) = 0. \tag{3.2.19}$$

The above is the generalized Lorentz reciprocity theorem. If $\mathbf{E}_i$, $\mathbf{H}_i \sim e^{ik_{iz}z}$ while $\mathbf{E}_j$, $\mathbf{H}_j \sim e^{ik_{jz}z}$ in their z-dependence, the above becomes

$$\nabla_s \cdot (\mathbf{E}_i \times \mathbf{H}_j - \mathbf{E}_j \times \mathbf{H}_i) = -i(k_{iz} + k_{jz})\hat{z} \cdot (\mathbf{E}_{is} \times \mathbf{H}_{js} - \mathbf{E}_{js} \times \mathbf{H}_{is}) \tag{3.2.20}$$

Integrating (3.2.20) over the cross-section of the waveguide, we have

$$\oint_C dl\hat{n} \cdot (\mathbf{E}_i \times \mathbf{H}_j - \mathbf{E}_j \times \mathbf{H}_i) = -i(k_{iz} + k_{jz}) \int_S dS\hat{z} \cdot (\mathbf{E}_{is} \times \mathbf{H}_{js} - \mathbf{E}_{js} \times \mathbf{H}_{is}). \tag{3.2.21}$$

The line integral vanishes by virtue of the boundary conditions. Hence,

$$(k_{iz} + k_{jz}) \int dS\hat{z} \cdot (\mathbf{E}_{is} \times \mathbf{H}_{js} - \mathbf{E}_{js} \times \mathbf{H}_{is}) = 0. \tag{3.2.22}$$

If the waveguide has reflection symmetry, a mode propagating in the $+z$ direction is almost the same as a mode propagating in the $-z$ direction, except for a change of the sign of the field. If $\mathbf{E}_{js}$, $\mathbf{H}_{js}$ corresponding to a mode with k_{jz} propagating in $+z$ direction, then by

Poynting theorem, $\mathbf{E}_{js}$, $-\mathbf{H}_{js}$ are the transverse fields of a mode with $-k_{jz}$ propagating in the $-z$ direction. Then, Equation (3.2.22) can be rewritten as

$$(k_{iz} - k_{jz}) \int dS\hat{z} \cdot (\mathbf{E}_{is} \times \mathbf{H}_{js} + \mathbf{E}_{js} \times \mathbf{H}_{is}) = 0. \tag{3.2.23}$$

If k_{iz} and k_{jz} are both non-zero, then the integral in (3.2.22) must be zero. If $k_{iz} \neq k_{jz}$, then the integral in (3.2.23) must be zero. The combination of (3.2.22) and (3.2.23) implies that

$$\int_S dS\hat{z} \cdot (\mathbf{E}_{is} \times \mathbf{H}_{js}) = 0, \qquad i \neq j. \tag{3.2.24}$$

If the medium is reciprocal, then $\overline{\boldsymbol{\mu}} = \overline{\boldsymbol{\mu}}^t$, $\overline{\boldsymbol{\epsilon}} = \overline{\boldsymbol{\epsilon}}^t$ and waveguides (i) and (ii) in Figure 3.1 are the same waveguide, and $\mathbf{E}_{is}$ and $\mathbf{H}_{js}$ are the fields from the same waveguide.

Furthermore, if $\overline{\boldsymbol{\mu}}$ and $\overline{\boldsymbol{\epsilon}}$ are Hermitian corresponding to a lossless medium with reflection symmetry, similar proof shows that

$$\int_S dS\hat{z} \cdot (\mathbf{E}_{is} \times \mathbf{H}_{js}^*) = 0, \qquad i \neq j. \tag{3.2.25}$$

The above is the power orthogonality condition for two different modes in an inhomogeneous, anisotropic, lossless, waveguide with reflection symmetry. Otherwise, the mode of the original waveguide is orthogonal to another mode of another waveguide where the medium is filled will a conjugate transpose medium.

3.2.5 Reflection Symmetry and Conservation of Parity

A commonly accepted law of physics is that the classical laws of physics hold true in the mirrored world (the reflected world) [8]. For electromagnetics, this means that a right-hand rule becomes a left-hand rule. This is also known as the conservation of parity, and is found to be violated in modern physics by some weak interactions.

If a waveguide has reflection symmetry, we say that it appears to be the same waveguide in the mirrored world as it is in the real world. This is certainly true of all uniform hollow waveguide. If a mode propagates in the $+z$ direction in the real world, the corresponding mode propagates in the mirrored $+z$ direction, plus that all the field components are mirrored. We can rotate the mirrored waveguide by 180^o about an axis perpendicular to the z axis, and we recover the original waveguide. The field of the mirrored mode satisfies the left-hand rule rather than the right-hand rule. However, we can convert the field of the mirrored mode into a real-world mode by switching $\mathbf{H}$ to $-\mathbf{H}$, and now the field will satisfy the right-hand rule as before.

When the waveguide is filled with an anisotropic material, the problem is more tricky, because in the mirrored world, the anisotropic material may not reflect to be the same material in the real world. However, if the permeability and permittivity tensors are of the form

$$\overline{\boldsymbol{\mu}} = \begin{pmatrix} \overline{\boldsymbol{\mu}}_s & 0 \\ 0 & \mu_{zz} \end{pmatrix}, \qquad \overline{\boldsymbol{\epsilon}} = \begin{pmatrix} \overline{\boldsymbol{\epsilon}}_s & 0 \\ 0 & \epsilon_{zz} \end{pmatrix}, \tag{3.2.26}$$

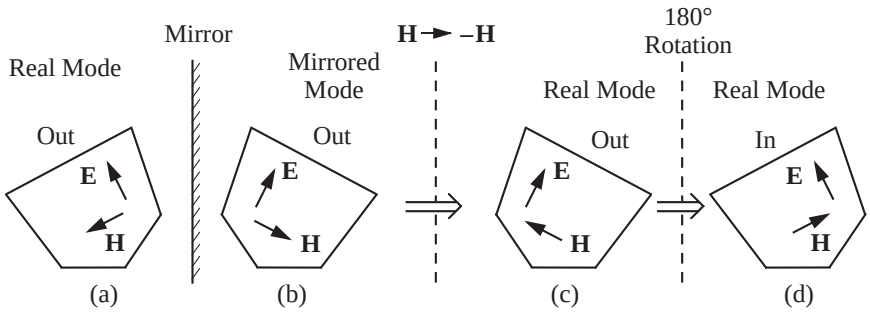


Figure 3.5: Reflection and the conservation of parity of a waveguide.

then the waveguide has reflection symmetry, i.e., the waveguide in the mirrored world is the same as the original waveguide. Again, if we switch $\mathbf{H}$ to $-\mathbf{H}$, we can obtain a real world mode.

We can convince ourselves more by looking at Maxwell's equations. By separating it into axial and transverse components, Maxwell's equations for a guided mode for the above tensors become

$$\nabla_s \times \mathbf{H}_s = -i\omega\epsilon_{zz}\mathbf{E}_z, \tag{3.2.27a}$$

$$ik_z\hat{z} \times \mathbf{H}_s + \nabla_s \times \mathbf{H}_z = -i\omega\overline{\boldsymbol{\epsilon}}_s \cdot \mathbf{E}_s, \tag{3.2.27b}$$

$$\nabla_s \times \mathbf{E}_s = i\omega\mu_{zz}\mathbf{H}_z, \tag{3.2.27c}$$

$$ik_z\hat{z} \times \mathbf{E}_s + \nabla_s \times \mathbf{E}_z = i\omega\overline{\boldsymbol{\mu}}_s \cdot \mathbf{H}_s. \tag{3.2.27d}$$

If the direction of propagation changes, which is what a mirror reflection does, k_z changes to $-k_z$. Notice that now if we let

$$\mathbf{H}_s \to -\mathbf{H}_s, \quad \mathbf{H}_z \to \mathbf{H}_z, \quad \mathbf{E}_s \to \mathbf{E}_s, \quad \mathbf{E}_z \to -\mathbf{E}_z, \tag{3.2.28}$$

the above equations remain invariant. Therefore if the waveguide mode changes direction, the above transformation in the field is needed to obtain another solution to Maxwell's equations.

To see how this is related to the conservation of parity, we imagine a waveguide in (a) above with the $\mathbf{E}$ and $\mathbf{H}$ fields as shown. They could have a $\hat{z}$ component perpendicular to the paper which is not shown. We assume that the mode is propagating out of the paper. In the mirrored world, as shown in (b), we have a mirrored mode and mirrored waveguide. The electromagnetic field does not satisfy Maxwell's equations according to the right-hand rule. To make the mode in (b) into a real mode, we let $\mathbf{H} \to -\mathbf{H}$ as shown in (c). However, compared to (a), the waveguide is not the original waveguide. If the constitutive parameters are those in (3.2.26), then (c) is just a 180^o rotation of the waveguide in (a) about an axis perpendicular to the z axis. By rotating the waveguide by 180^o, we obtain (d). However, $\mathbf{H}_s$ has changed sign while $\mathbf{H}_z$ remains unchanged compared to (a). This is precisely the field transformation prescribed by (3.2.28). A closer examination also indicates that $\mathbf{E}_s$ does not change sign, while E_z changes sign.

3.3 General Anisotropic Waveguide

The vector wave equation governing the field inside an anisotropic waveguide is

$$\nabla \times \overline{\boldsymbol{\mu}}^{-1} \cdot \nabla \times \mathbf{E} - \omega^2 \overline{\boldsymbol{\epsilon}} \cdot \mathbf{E} = 0. \tag{3.3.1}$$

In general, the modes inside such a waveguide cannot be decomposed into TE and TM modes. Hence, the problem is again vector, requiring at least two components of the field. Assuming the waveguide to have reflection symmetry in the z direction, then

$$\overline{\boldsymbol{\epsilon}} = \begin{bmatrix} \overline{\boldsymbol{\epsilon}}_s & 0 \\ 0 & \epsilon_{zz} \end{bmatrix}, \qquad \overline{\boldsymbol{\mu}} = \begin{bmatrix} \overline{\boldsymbol{\mu}}_s & 0 \\ 0 & \mu_{zz} \end{bmatrix}, \tag{3.3.2}$$

where $\overline{\boldsymbol{\epsilon}}_s$ and $\overline{\boldsymbol{\mu}}_s$ are 2×2 tensors with components in the transverse direction. The transverse component of (3.3.1) can be extracted to obtain

$$\nabla_s \times \mu_{zz}^{-1} \nabla_s \times \mathbf{E}_s - ik_z \hat{z} \times \overline{\boldsymbol{\mu}}_s^{-1} \cdot \hat{z} \times \nabla_s E_z - k_z^2 \hat{z} \times \overline{\boldsymbol{\mu}}_s^{-1} \cdot \hat{z} \times \mathbf{E}_s - \omega^2 \overline{\boldsymbol{\epsilon}}_s \cdot \mathbf{E}_s = 0. \tag{3.3.3}$$

With the use of the divergence condition,

$$ik_z E_z = -\epsilon_{zz}^{-1} \nabla_s \cdot \overline{\boldsymbol{\epsilon}}_s \cdot \mathbf{E}_s. \tag{3.3.4}$$

Hence, the z component of the field in (3.3.3) can be replaced to yield

$$\begin{aligned} \nabla_s \times \mu_{zz}^{-1} \nabla_s \times \mathbf{E}_s + \hat{z} \times \overline{\boldsymbol{\mu}}_s^{-1} \cdot \hat{z} \times \nabla_s \boldsymbol{\epsilon}_{zz}^{-1} \nabla_s \cdot \overline{\boldsymbol{\epsilon}}_s \mathbf{E}_s \\ - k_z^2 \hat{z} \times \overline{\boldsymbol{\mu}}_s^{-1} \cdot \hat{z} \times \mathbf{E}_s - \omega^2 \overline{\boldsymbol{\epsilon}}_s \cdot \mathbf{E}_s = 0. \end{aligned} \tag{3.3.5}$$

Consequently, the field in an anisotropic waveguide can be characterized by $\mathbf{E}_s$ alone. We can multiply the above by $\overline{\boldsymbol{\mu}}_s \cdot \hat{z} \times$ to get

$$\begin{aligned} \overline{\boldsymbol{\mu}}_s \cdot \hat{z} \times \nabla_s \times \mu_{zz}^{-1} \nabla_s \times \mathbf{E}_s - \hat{z} \times \nabla_s \epsilon_{zz}^{-1} \nabla_s \cdot \overline{\boldsymbol{\epsilon}}_s \cdot \mathbf{E}_s \\ - \omega^2 \overline{\boldsymbol{\mu}}_s \cdot \hat{z} \times \overline{\boldsymbol{\epsilon}}_s \cdot \mathbf{E}_s + k_z^2 \hat{z} \times \mathbf{E}_s = 0. \end{aligned} \tag{3.3.6}$$

Note that the above equation is a function of k_z^2, implying that if $\mathbf{E}_s$ with $e^{ik_z z}$ dependence is a solution to (3.3.6), an $\mathbf{E}_s$ with $e^{-ik_z z}$ dependence is also a solution. This is a consequence of reflection symmetry, or the assumptions about $\overline{\boldsymbol{\epsilon}}$ and $\overline{\boldsymbol{\mu}}$ in (3.3.2). The corresponding equation for the transverse magnetic field is obtained by duality, yielding

$$\begin{aligned} \overline{\boldsymbol{\epsilon}}_s \cdot \hat{z} \times \nabla_s \times \epsilon_{zz}^{-1} \nabla_s \times \mathbf{H}_s - \hat{z} \times \nabla_s \boldsymbol{\mu}_{zz}^{-1} \nabla_s \cdot \overline{\boldsymbol{\mu}}_s \cdot \mathbf{H}_s \\ - \omega^2 \overline{\boldsymbol{\epsilon}}_s \cdot \hat{z} \times \overline{\boldsymbol{\mu}}_s \cdot \mathbf{H}_s + k_z^2 \hat{z} \times \mathbf{H}_s = 0. \end{aligned} \tag{3.3.7}$$

Since the solution to (3.3.6) is orthogonal to the solution to (3.3.7), for two different modes, Equation (3.3.7) is also the transpose equation of (3.3.6) [9, 15]. This shall be elaborated in the next section.

3.4 Proof of Transpose of Operators

For an anisotropic waveguide with reflection symmetry,

$$\begin{aligned}\overline{\boldsymbol{\mu}}_s \cdot \hat{z} \times \nabla_s \times \mu_{zz}^{-1} \nabla_s \times \mathbf{E}_s - \hat{z} \times \nabla_s \epsilon_{zz}^{-1} \nabla_s \cdot \overline{\boldsymbol{\epsilon}}_s \cdot \mathbf{E}_s \\ -\omega^2 \overline{\boldsymbol{\mu}}_s \cdot \hat{z} \times \overline{\boldsymbol{\epsilon}}_s \cdot \mathbf{E}_s + k_z^2 \hat{z} \times \mathbf{E}_s = 0\end{aligned} \tag{3.4.1}$$

We shall show that the above equation is transpose to the equation

$$\begin{aligned}\overline{\boldsymbol{\epsilon}}_s^t \cdot \hat{z} \times \nabla_s \times \epsilon_{zz}^{-1} \nabla_s \times \mathbf{H}_s - \hat{z} \times \nabla_s \mu_{zz}^{-1} \nabla_s \cdot \overline{\boldsymbol{\mu}}_s^t \cdot \mathbf{H}_s \\ -\omega^2 \overline{\boldsymbol{\epsilon}}_s^t \cdot \hat{z} \times \overline{\boldsymbol{\mu}}_s^t \cdot \mathbf{H}_s + k_z^2 \hat{z} \times \mathbf{H}_s = 0\end{aligned} \tag{3.4.2}$$

The definition of the transpose operator is[2]

$$\langle \mathbf{u}, \mathcal{L}\mathbf{v} \rangle = \langle \mathbf{v}, \mathcal{L}^t \mathbf{u} \rangle \tag{3.4.3}$$

where inner product between two vector fields in the infinite dimensional space (also called the Hilbert space) is $\langle \mathbf{f}, \mathbf{g} \rangle = \int dS \mathbf{f} \cdot \mathbf{g}$. The integration in this case is taken over the cross-sectional area of the waveguide. Furthermore, the $\mathbf{H}_s$ field in (3.4.2) is the field of the auxiliary problem as indicated in Figure 3.4.

To prove this, we start with the first term in (3.4.1), and calculate the expression

$$I = \langle \mathbf{H}_s, \overline{\boldsymbol{\mu}}_s \cdot \hat{z} \times \nabla_s \times \mu_{zz}^{-1} \nabla_s \times \mathbf{E}_s \rangle \tag{3.4.4}$$

It can be easily shown that

$$I = \langle \hat{z} \times \overline{\boldsymbol{\mu}}_s^t \cdot \mathbf{H}_s, \nabla_s \times \mu_{zz}^{-1} \nabla_s \times \mathbf{E}_s \rangle \tag{3.4.5}$$

Using Gauss theorem, or integration by parts, and that $\nabla_s \cdot (\mathbf{A} \times \mathbf{B}) = \mathbf{B} \cdot \nabla_s \times \mathbf{A} - \mathbf{A} \cdot \nabla_s \times \mathbf{B}$ we have

$$I = -\int dS \nabla_s \cdot [(\hat{z} \times \overline{\boldsymbol{\mu}}_s^t \cdot \mathbf{H}_s) \times (\bar{\mu}_{zz}^{-1} \nabla_s \times \mathbf{E}_s)] + \int dS (\nabla_s \times (\hat{z} \times \overline{\boldsymbol{\mu}}_s^t \cdot \mathbf{H}_s)) \cdot \mu_{zz}^{-1} \nabla_s \times \mathbf{E}_s) \tag{3.4.6}$$

It can be shown that $\nabla_s \times (\hat{z} \times \overline{\boldsymbol{\mu}}_s^t \cdot \mathbf{H}_s) = \hat{z} \nabla_s \cdot (\overline{\boldsymbol{\mu}}_s^t \cdot \mathbf{H}_s)$, and that $\hat{z} \cdot \nabla_s \times \mathbf{E}_s = -\nabla_s \cdot (\hat{z} \times \mathbf{E}_s)$. The first term above can be converted to a boundary integral over the waveguide wall using Gauss' divergence theorem. It vanishes by virtue of the boundary condition on the waveguide wall that $\hat{n} \cdot \mathbf{B} = 0$. Hence, only the second term remains. Consequently,

$$I = -\int dS \nabla_s \cdot (\overline{\boldsymbol{\mu}}_s^t \cdot \mathbf{H}_s) \mu_{zz}^{-1} \nabla_s \cdot (\hat{z} \times \mathbf{E}_s) \tag{3.4.7}$$

Using integration by parts one more time, and then the vector identity, that $\nabla_s \cdot (\phi \mathbf{A}) = \phi \nabla_s \cdot \mathbf{A} + \nabla_s \phi \cdot \mathbf{A}$, where $\phi = \nabla_s \cdot (\overline{\boldsymbol{\mu}}_s^t \cdot \mathbf{H}_s) \mu_{zz}^{-1}$ and $\mathbf{A} = \hat{z} \times \mathbf{E}_s$, yields

$$I = -\int dS \nabla_s \cdot \left[\nabla_s \cdot (\overline{\boldsymbol{\mu}}_s^t \cdot \mathbf{H}_s) \mu_{zz}^{-1} (\hat{z} \times \mathbf{E}_s) \right] + \int dS \hat{z} \times \nabla_s \mu_{zz}^{-1} \nabla_s \cdot (\overline{\boldsymbol{\mu}}_s^t \cdot \mathbf{H}_s) \cdot \mathbf{E}_s \tag{3.4.8}$$

[2]When applied to a matrix operator, this becomes $\mathbf{u}^t \cdot \overline{\mathbf{L}} \cdot \mathbf{v} = \left(\mathbf{v}^t \cdot \overline{\mathbf{L}}^t \cdot \mathbf{u} \right)$. Similar formula can be derived for defining conjugate transpose or adjoint of an operator [15].

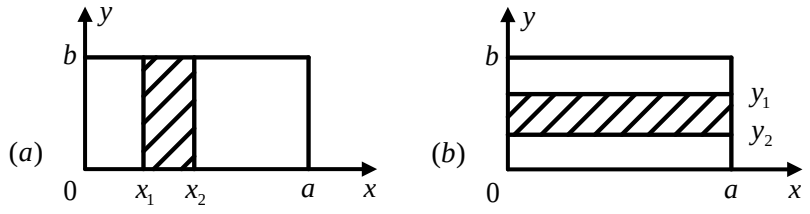


Figure 3.6: Dielectric-slab-loaded waveguide.

where the cyclic relation of cross-dot product has been used. The first integral can be converted to a boundary integral and vanishes by virtue of the boundary condition and that $\hat{\times}\mathbf{E}_s = 0$ on the waveguide wall. Hence,

$$I = \int dS\hat{z} \times \nabla_s \mu_{zz}^{-1} \nabla_s \cdot (\overline{\boldsymbol{\mu}}_s^t \cdot \mathbf{H}_s) \cdot \mathbf{E}_s = -\langle \mathbf{E}_s, \hat{z} \times \nabla_s \mu_{zz}^{-1} \nabla_s \cdot (\overline{\boldsymbol{\mu}}_s^t \cdot \mathbf{H}_s) \rangle \tag{3.4.9}$$

Hence, the first term in (3.4.1) is the negative transpose of the second term in (3.4.2). Similarly, the second term in (3.4.1) is the transpose of the first term in (3.4.2).

Furthermore, it can be shown that $(\hat{z}\times)^t = -\hat{z}\times$, $(\overline{\boldsymbol{\epsilon}} \cdot \hat{z} \times \overline{\boldsymbol{\mu}}_s)^t = -\overline{\boldsymbol{\mu}}_s^t \cdot \hat{z} \times \overline{\boldsymbol{\epsilon}}_s^t$. Hence if,

$$\mathcal{L}(*) = \overline{\boldsymbol{\mu}}_s \cdot \hat{z} \times \nabla_s \times \mu_{zz}^{-1} \nabla_s \times (*) - \hat{z} \times \nabla_s \epsilon_{zz}^{-1} \nabla_s \cdot \overline{\boldsymbol{\epsilon}}_s(*) - \omega^2 \overline{\boldsymbol{\mu}}_s \cdot \hat{z} \times \overline{\boldsymbol{\epsilon}}_s(*) + k_z^2 \hat{z} \times (*) \tag{3.4.10}$$

Then

$$-\mathcal{L}^t(*) = \overline{\boldsymbol{\epsilon}}_s^t \cdot \hat{z} \times \nabla_s \times \epsilon_{zz}^{-1} \nabla_s \times (*) - \hat{z} \times \nabla_s \mu_{zz}^{-1} \nabla_s \cdot \overline{\boldsymbol{\mu}}_s^t(*) - \omega^2 \overline{\boldsymbol{\epsilon}}_s^t \cdot \hat{z} \times \overline{\boldsymbol{\mu}}_s^t(*) + k_z^2 \hat{z} \times (*) \tag{3.4.11}$$

where $(*)$ represents a vector function that these operators act on. In general, (3.4.2) is the field equation for a waveguide filled with transpose medium compared to the original waveguide equation (3.4.1), and (3.4.2) is the negative transpose of (3.4.1). However, if the waveguide is filled with reciprocal medium, then (3.4.1) and (3.4.2) are field equations for the same waveguide.

3.5 Dielectric-Slab-Loaded Rectangular Waveguides

For the analysis of a general inhomogeneously filled waveguide, a numerical method has to be sought [9–11]. However, when the waveguide has certain symmetry such as a slab loaded rectangular waveguide, analytic method for its analysis is possible.

Dielectric-slab-loaded waveguides find applications in a number of microwave components, because the phase velocity of a mode can be altered with dielectric loading [3,12]. A waveguide can also be loaded with a ferrite slab [13, 14]. When a ferrite slab is biased with a magnetic field, it becomes an anisotropic, gyrotropic medium. Such a medium is non-reciprocal. Hence, ferrite slabs can be used to design non-reciprocal devices such as isolators.

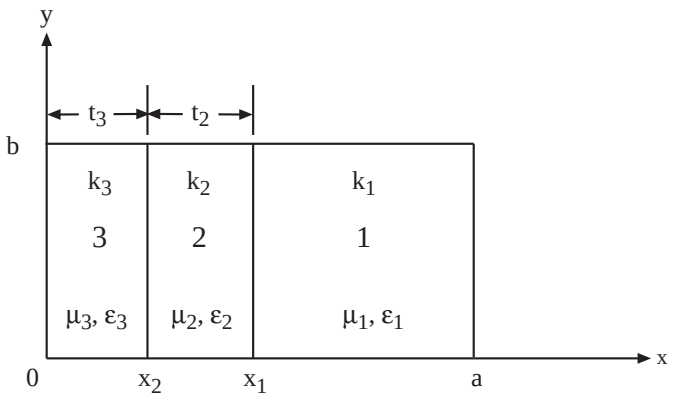


Figure 3.7: A special case of a slab loaded waveguide.

The analysis of a ferrite-slab-loaded waveguide is complicated. We will focuss our analysis on a dielectric-slab-loaded waveguide. In such a waveguide, with the exception of special cases, the modes cannot be decomposed into TE and TM modes with respect to the z direction. Hence, a general mode is hybrid. However, for the slab-loaded waveguide as shown in Figure 3.6(a), the modes in the waveguide can be decomposed into modes with **E** field transverse to x or modes with **H** field transverse to x. Similar decomposition exists for Figure 3.6(b) since it is just a 90^o rotation of Figure 3.6(a). The modes with **E** field transverse to x are known as the LSE (Longitudinal Section Electric) modes, while the modes with **H** field transverse to x are known as the LSM (Longitudinal Section Magnetic) modes. There exist closed form expressions for the guidance condition of these modes, but the exact wave number k_z has to be found numerically.

Since the LSE mode is transverse to x, it can be characterized with H_x. Similarly, the LSM mode can be characterized with E_x. The field has to have $e^{ik_z z}$ dependence everywhere inside the waveguide due to the phase matching condition. Consequently, the equations satisfied by H_x and E_x are

$$(\nabla_s^2 + k_i^2 - k_z^2)H_{ix} = 0, \qquad \text{LSE modes,} \tag{3.5.1}$$

$$(\nabla_s^2 + k_i^2 - k_z^2)E_{ix} = 0, \qquad \text{LSM modes.} \tag{3.5.2}$$

where subscript i denotes the region i and subscript s denotes transverse to x. The solutions to (3.5.1) and (3.5.2) are of the form

$$H_{ix} = H_0\{e^{\pm ik_{ix}x}\}\{e^{\pm ik_y y}\}e^{ik_z z}, \tag{3.5.3}$$

$$E_{ix} = E_0\{e^{\pm ik_{ix}x}\}\{e^{\pm ik_y y}\}e^{ik_z z}, \tag{3.5.4}$$

where the braces imply linear superpositions. Here, k_y in each region must be the same due to the phase matching condition. The fields transverse to x can be found from H_{ix} and E_{ix} in each region, i.e.,

$$\mathbf{E}_{is} = \frac{1}{k_y^2 + k_z^2}\left[\frac{\partial}{\partial x}\nabla_s E_{ix} - i\omega\mu_i \hat{x} \times \nabla_s H_{ix}\right], \tag{3.5.5a}$$

$$\mathbf{H}_{is} = \frac{1}{k_y^2 + k_z^2}\left[\frac{\partial}{\partial x}\nabla_s H_{ix} + i\omega\epsilon_i \hat{x} \times \nabla_s E_{ix}\right], \tag{3.5.5b}$$

where the subscript s implies transverse to x. It can be seen from the above that the boundary conditions for H_{ix} and E_{ix} are that

$$\frac{\partial H_{ix}}{\partial y} = 0, \qquad \text{at} \quad y = 0 \quad \text{and} \quad y = b, \tag{3.5.6a}$$

$$E_{ix} = 0, \qquad \text{at} \quad y = 0 \quad \text{and} \quad y = b. \tag{3.5.6b}$$

Therefore, forms for H_{ix} and E_{ix} that satisfy the above boundary conditions are

$$H_{ix} = H_0\{e^{\pm ik_{ix}x}\}\cos\left(\frac{n\pi y}{b}\right)e^{ik_z z}, \tag{3.5.7a}$$

$$E_{ix} = E_0\{e^{\pm ik_{ix}x}\}\sin\left(\frac{n\pi y}{b}\right)e^{ik_z z}, \tag{3.5.7b}$$

where we have let $k_y = \frac{n\pi}{b}$, and n is an integer. In the above, $k_{ix} = \sqrt{k_i^2 - k_z^2 - (\frac{n\pi}{b})^2}$. It is now clear that H_{ix} represents bouncing waves that are TE to x while E_{ix} represents bouncing waves that are TM to x. Hence, in region 1, we can write the solution as

$$H_{1x} = H_0[e^{-ik_{1x}(x-x_1)} + \tilde{R}_{12}^{TE}e^{ik_{1x}(x-x_1)}]\cos\left(\frac{n\pi y}{b}\right)e^{ik_z z}, \tag{3.5.8a}$$

$$E_{1x} = E_0[e^{-ik_{1x}(x-x_1)} + \tilde{R}_{12}^{TM}e^{ik_{1x}(x-x_1)}]\sin\left(\frac{n\pi y}{b}\right)e^{ik_z z}, \tag{3.5.8b}$$

where $\tilde{R}_{12}^{TE,TM}$ is a generalized Fresnel reflection coefficient for a TE or a TM wave incident from the left at $x = x_1$. It includes subsurface reflections. However, we need to impose the boundary conditions that

$$H_{ix}(x = a) = 0, \qquad \frac{\partial E_{ix}}{\partial x}(x = a) = 0. \tag{3.5.9}$$

The above boundary conditions can be derived from (3.5.5a) and (3.5.5b) as the most general case. They can also be derived from $\hat{n} \cdot \mathbf{H} = 0$ and $\nabla \cdot \mathbf{E} = 0$, respectively. They could only be satisfied if

$$1 + \tilde{R}_{12}^{TE}e^{2ik_{1x}(a-x_1)} = 0, \qquad \text{LSE modes}, \tag{3.5.10a}$$

$$1 - \tilde{R}_{12}^{TM}e^{2ik_{1x}(a-x_1)} = 0, \qquad \text{LSM modes}. \tag{3.5.10b}$$

The above are the guidance conditions for the LSE modes and the LSM modes in a slab loaded rectangular waveguide.

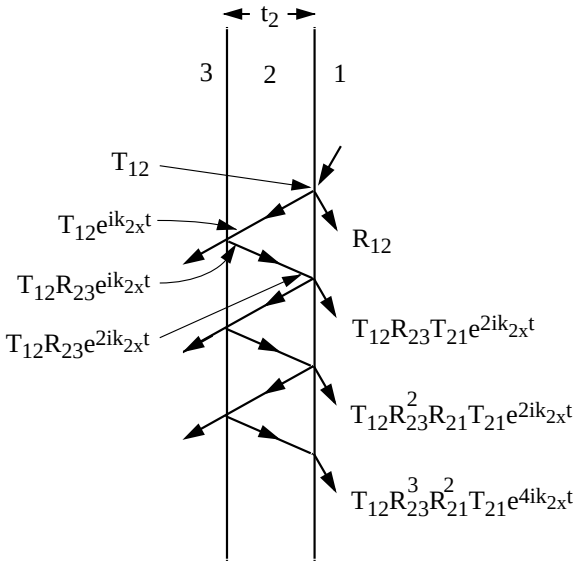


Figure 3.8: Multiple reflections and transmissions through a dielectric slab.

To find $\tilde{R}_{12}$, we note that a wave impinging on a slab with one subsurface interface will consist of a surface reflection from the top interface plus a sequence of subsurface reflections. The single interface reflections are governed by the Fresnel reflection coefficients. Hence [15]

$$\begin{aligned}\tilde{R}_{12} = R_{12} + T_{12}R_{23}T_{21}e^{2ik_{2x}t_2} + T_{12}R_{23}^2R_{21}T_{21}e^{4ik_{2x}t_2} \\ + T_{12}R_{23}^3R_{21}^2T_{21}e^{6ik_{2x}t_2} + \ldots,\end{aligned} \tag{3.5.11}$$

where $T_{ij} = 1 + R_{ij}$ is the Fresnel transmission coefficient at the ij interface. The above could be summed to yield

$$\tilde{R}_{12} = R_{12} + \frac{T_{12}R_{23}T_{21}e^{2ik_{2x}t_2}}{1 - R_{23}R_{21}e^{2ik_{2x}t_2}} = \frac{R_{12} + R_{23}e^{2ik_{2x}t_2}}{1 - R_{23}R_{21}e^{2ik_{2x}t_2}}. \tag{3.5.12}$$

In the above,

$$R_{ij}^{TE} = \frac{\mu_j k_{ix} - \mu_i k_{jx}}{\mu_j k_{ix} + \mu_i k_{jx}}, \qquad R_{ij}^{TM} = \frac{\epsilon_j k_{ix} - \epsilon_i k_{jx}}{\epsilon_j k_{ix} + \epsilon_i k_{jx}}, \tag{3.5.13}$$

depending on whether we are calculating $\tilde{R}_{12}$ for a TE wave or a TM wave.

If there are subsurface layers below region 3, R_{23} in (3.5.12) can be replaced with $\tilde{R}_{23}$, or

$$\tilde{R}_{12} = \frac{R_{12} + \tilde{R}_{23}e^{2ik_{2x}t_2}}{1 - R_{21}\tilde{R}_{23}e^{2ik_{2x}t_2}}. \tag{3.5.14}$$

The above is a recursive relation from which one can calculate $\tilde{R}_{ij}$ for any number of layers. For example, if there is a metallic wall at $x = 0$, as in Figure 3.7, $R_{34}^{TE} = -1$, $R_{34}^{TM} = 1$, then

$$\tilde{R}_{23}^{TE} = \frac{R_{23}^{TE} - e^{2ik_{3x}t_3}}{1 + R_{32}^{TE}e^{2ik_{3x}t_3}}, \qquad \tilde{R}_{23}^{TM} = \frac{R_{23}^{TM} + e^{2ik_{3x}t_3}}{1 - R_{32}^{TM}e^{2ik_{3x}t_3}}. \tag{3.5.15}$$

With $\tilde{R}_{12}$ defined by (3.5.14) and (3.5.15), Equations (3.5.10a) and (3.5.10b), in general, have to be solved numerically with a root solver like the Newton-Raphson method, or the Muller's method. All the R_{ij}'s are defined in terms of k_{ix} in Equation (3.5.13), where

$$k_{ix} = \sqrt{k_i^2 - k_z^2 - \left(\frac{n\pi}{b}\right)^2}. \tag{3.5.16}$$

Hence, Equations (3.5.10a) and (3.5.10b) can be solved for values of k_z. For simple geometry, graphical solutions to Equations (3.5.10a) and (3.5.10b) may be found.

The guidance condition given by Equations (3.5.10a) and (3.5.10b) are also known as the transverse resonance condition which will be discussed in greater detail in the next section. The guidance conditions in (3.5.10a) and (3.5.10b) are obtained by considering waves bouncing in the air region. If a guided mode is trapped in the dielectric slab region, the wave becomes evanescent in the air region and Equations (3.5.10a) and (3.5.10b) can become ill-posed. To remedy this, it is better to write down the guidance condition in the slab region using the transverse resonance condition described in the next section.

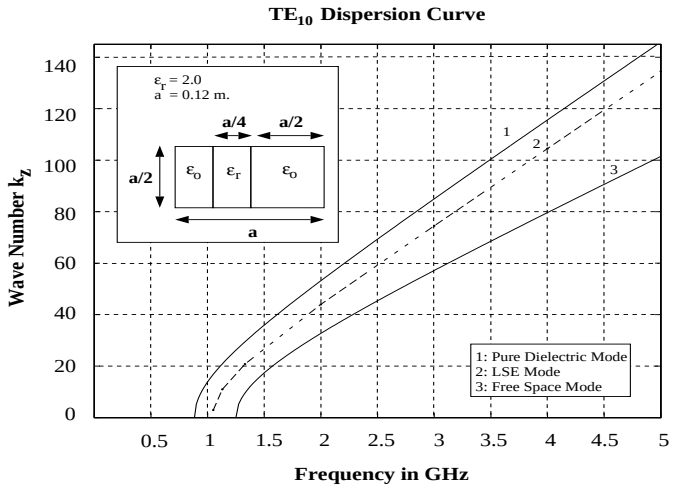
Figure 3.9 shows the dispersion curves of the TE_{10} and TE_{01} (with respect to z) modes of a slab loaded dielectric waveguide. Case 1 is when the whole waveguide is filled with dielectric material, and case 3 is when the waveguide is empty. Notice that for the perturbed modes (case 2), the dispersion curve is parallel to case 1, The reason is that the mode is entrapped in the dielectric slab for high frequencies, and the group velocity of the mode $(d\omega/dk_z)$ approaches that of the dielectric slab.

Also, notice that at lower frequencies, the TE_{10} mode is perturbed more by the slab than the TE_{01} mode, because for the TE_{01} mode, the electric field is normal to the slab while for the TE_{10} mode, the electric field is parallel to the slab. This fact can be used to create anisotropy in a symmetric waveguide like a circular waveguide or a square waveguide. Figure 3.10 shows the dispersion curves for the LSE mode, which is the perturbed TE_{10} mode, and the LSM mode, which is the perturbed TE_{01} mode. Notice that the LSE mode is affected more by the dielectric slab than the LSM mode. This anisotropic effect can be used to make a quarter-wave plate out of a dielectric slab loaded waveguide. This will be discussed in greater detail later.

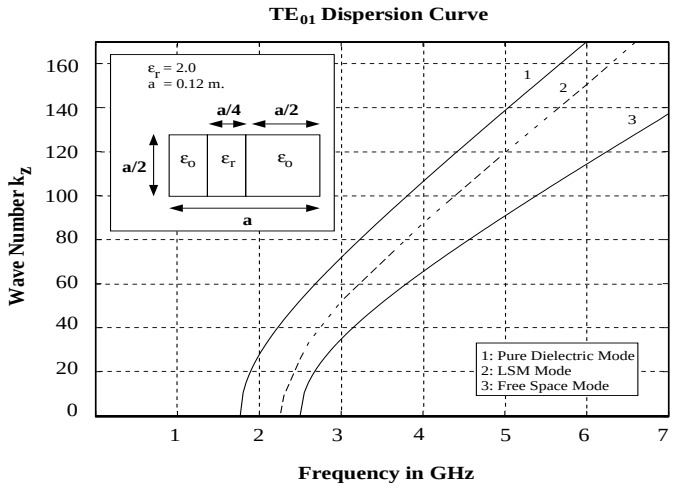
3.6 Transverse Resonance Condition

The guidance conditions given by (3.5.10a) and (3.5.10b) can also be derived by the transverse resonance condition. The transverse resonance condition is a powerful condition that can be used to derive the guidance condition of a mode in a layered medium.

To derive this condition, we first have to realize that a guided mode in a waveguide is due to the coherent or constructive interference of the waves. This implies that if a plane wave



(a)



(b)

Figure 3.9: Dispersion curves for various dielectric loading for (a) TE_{10} mode, and (b) TE_{01} mode.

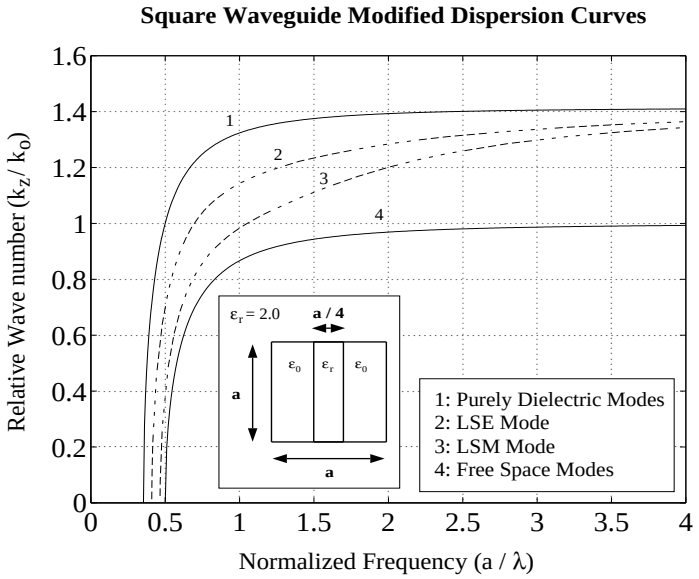


Figure 3.10: The dispersion curves for a slab-loaded square waveguide showing anisotropy.

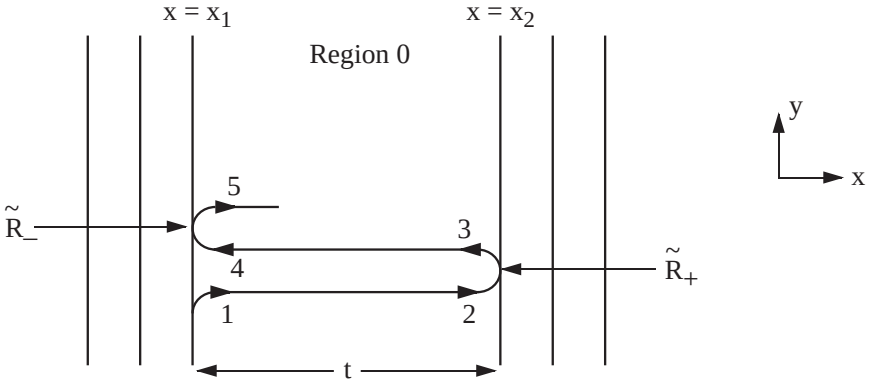


Figure 3.11: The transverse resonance condition for a layered medium. The phase of the wave at position 5 should be equal to the phase at position 1.

starts at position 1 and is multiply reflected as shown, it will regain its original phase in the x direction at position 5. Since this mode progresses in the y direction. it will gain a phase in the y direction. But, for it to coherently interfere in the x direction, the transverse phase at 5 must be the same as 1.

Assuming that the wave starts with amplitude 1 at position 1, it will gain a transverse phase of $e^{ik_{0x}t}$ when it reaches position 2. Upon reflection at $x = x_2$, at position 3, the wave becomes $\tilde{R}_+ e^{ik_{0x}t}$. Finally, at position 5, it becomes $\tilde{R}_- \tilde{R}_+ e^{2ik_{0x}t}$. For constructive interference to occur or for the mode to exist, we require that

$$\tilde{R}_- \tilde{R}_+ e^{2ik_{0x}t} = 1. \tag{3.6.1}$$

The above is the transverse resonance condition. It is also the guidance condition for a mode travelling in a layered medium.

In (3.5.10a), a metallic wall has a reflection coefficient of 1 for a TM wave, hence if $\tilde{R}_+$ is 1, Equation (3.6.1) becomes

$$1 - \tilde{R}_- e^{2ik_{0x}t} = 0. \tag{3.6.2}$$

In (3.5.10b), a metallic wall has a reflection coefficient of -1, and Equation (3.6.1) becomes

$$1 + \tilde{R}_- e^{2ik_{0x}t} = 0. \tag{3.6.3}$$

3.7 Fabry-Perot Etalon

Since we have the machinery in place, it is convenient to study the Fabry-Perot etalon which is often used as an optical filter [5]. By tracing the plane wave or ray through the slabs, we can show that the generalized transmission coefficient from region 1 to region 3 is

$$\tilde{T}_{13} = \frac{T_{12}T_{23}e^{ik_{2x}d}}{1 - R_{23}R_{21}e^{2ik_{2x}d}} \tag{3.7.1}$$

where d is the thickness of the slab. In etalon application, regions 1 and 3 are free space while region 2 is dielectric. Specializing to the case when regions 1 and 3 have the same parameters, we have

$$\tilde{T} = \frac{T_{12}T_{21}e^{ik_{2x}d}}{1 - R_{21}^2 e^{2ik_{2x}d}} \tag{3.7.2}$$

Since $T_{12} = 1 + R_{12}$, $T_{21} = 1 + R_{21} = 1 - R_{12}$, we have

$$\tilde{T} = \frac{(1 - R_{21}^2)e^{ik_{2x}d}}{1 - R_{21}^2 e^{2ik_{2x}d}} \tag{3.7.3}$$

For normally incident wave, $k_{2x} = k_2$, and the above becomes

$$\tilde{T} = \frac{(1 - R_{21}^2)e^{ik_2 d}}{1 - R_{21}^2 e^{2ik_2 d}} \tag{3.7.4}$$

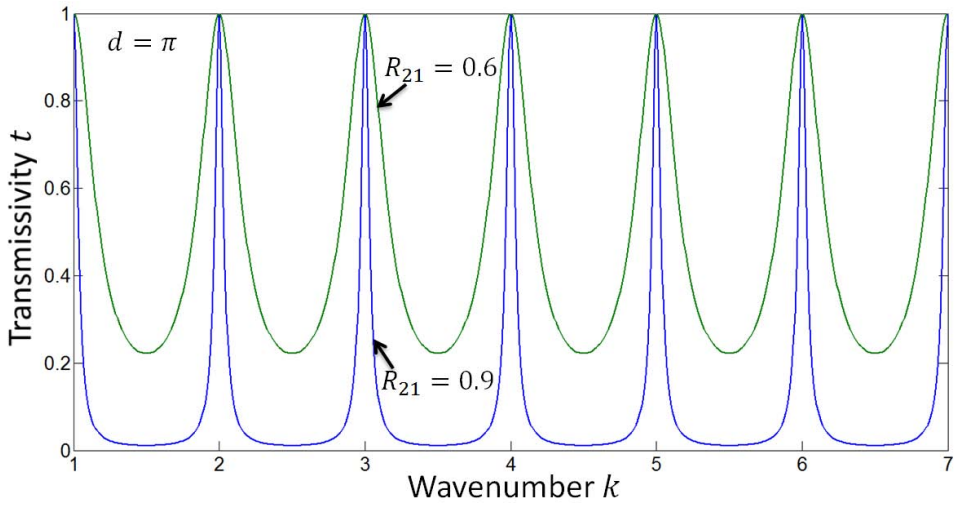


Figure 3.12: The transmissivity of the etalon as a function of frequency.

where

$$R_{21} = \frac{k_2 - k_1}{k_2 + k_1} = \frac{\sqrt{\epsilon_2} - \sqrt{\epsilon_1}}{\sqrt{\epsilon_2} + \sqrt{\epsilon_1}} \tag{3.7.5}$$

assuming that $\mu_1 = \mu_2 = \mu$. The transmissivity is

$$t = |\tilde{T}|^2 = \frac{|1 - R_{21}^2|^2}{|1 - R_{21}^2 e^{2ik_2 t}|^2} \tag{3.7.6}$$

The above is maximum with $t = 1$ when $e^{2ik_2 d} = 1$ or $2k_2 d = 2m\pi$ with integer m. In other words, at such a frequency, the etalon is transparent with no reflected wave. Specializing (3.5.14) to this case, and using the fact that $R_{12} = -R_{21}$,

$$\tilde{R} = \frac{R_{12}\left(1 - e^{2ik_2 d}\right)}{1 - R_{21}^2 e^{2ik_2 d}}. \tag{3.7.7}$$

It is seen that the above can be zero if $2k_2 d = 2m\pi$. The zero comes about because of the destructive interference of the reflected waves. The transmissivity of the etalon as a function of frequency is shown in Figure 3.12.

Next, we can perform the pole analysis of the etalon. In the vicinity of the maxima, $k_2 d - m\pi$ is small, and

$$e^{2ik_2 d} = e^{2ik_2 d - 2im\pi} \approx 1 + \frac{2id}{c_2}(\omega - \omega_m) = 1 + \frac{2id}{c_2}\Delta\omega \tag{3.7.8}$$

where $\Delta\omega = \omega - \omega_m$,

$$\omega_m = \frac{m\pi c_2}{d} \tag{3.7.9}$$

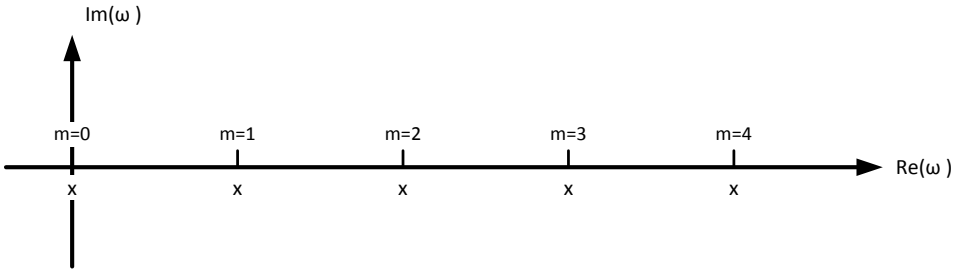


Figure 3.13: The locations of the poles in the complex ω plane for the Fabry-Perot modes.

and

$$c_2 = \frac{1}{\sqrt{\mu\epsilon_2}} \tag{3.7.10}$$

Hence,

$$\tilde{T} \approx \frac{1 - R_{21}^2}{1 - R_{21}^2 - R_{21}^2 2id\frac{\Delta\omega}{c_2}} \tag{3.7.11}$$

near the maxima. The pole locations can be found to be when

$$\Delta\omega \cong \frac{1 - R_{21}^2}{R_{21}^2}\left(\frac{-ic_2}{2d}\right) \tag{3.7.12}$$

Hence, the pole location for the m-th resonance mode in the complex plane is

$$\omega_{pm} \simeq \omega_m - \frac{ic_2}{2d}\frac{1 - R_{21}^2}{R_{21}^2} = \frac{m\pi c_2}{d} - \frac{ic_2}{2d}\frac{1 - R_{21}^2}{R_{21}^2} \tag{3.7.13}$$

The above is accurate when $1 - R_{21}^2$ is small. It is slightly below the real ω axis which is in agreement with a lossy resonant mode with $e^{-i\omega t}$ time convention. Notice that the imaginary part of the pole location is independent of frequency (See Figure 3.13).

The mechanism for transmission here is resonance tunneling. The Fabry-Perot etalon is used as an optical filter. To obtain a narrow band filter, the Q's of these modes have to be high. The Q of a resonator is

$$Q = \frac{\omega_r W_T}{P_d} \tag{3.7.14}$$

where

$$\frac{W_T}{P_d} = T$$

is the time constant for which the energy store W_T will diminish to e^{-1} of its original value. So $Q = \omega_r T$ is also the number of cycles in radian for this decay to happen. Q is an

asymptotic concept that has meaning only when Q is large. When the pole location is known, $\omega_p = \omega_p^{'} + i\omega_p^{''}$, then

$$Q = -\frac{\omega_p^{'}}{2\omega_p^{''}} \tag{3.7.15}$$

Applying the above to the m-th mode in (11), we have

$$Q_m = \frac{m\pi R_{21}^2}{(1 - R_{21}^2)} \tag{3.7.16}$$

3.8 Rod-Loaded Circular Waveguide

When a circular waveguide is loaded with a concentric circular dielectric rod, the equation for the guidance condition can be found in closed form. Ferrite rods are usually used to load a circular waveguide to engender Faraday rotation. Since ferrites are anisotropic, the analysis of a ferrite loaded circular waveguide is in general very complex. Hence, we will analyze the case of a circular waveguide loaded with a circular dielectric rod. The general case of a waveguide periodically loaded with dielectric rods can be used to make microwave filters [16].

3.8.1 Reflection off a Dielectric Rod

Since E_z and H_z waves are in general coupled in uniform dielectric rod, we have to consider both polarizations together in this reflection problem. We will follow an analysis presented in [15]. The z-components of the fields satisfy

$$\left(\nabla^2 + k^2\right) \begin{bmatrix} E_z \\ H_z \end{bmatrix} = 0. \tag{3.8.1}$$

Assuming $e^{ik_z z + in\phi}$ dependence in the wave, the equation becomes

$$\left(\frac{1}{\rho}\frac{\partial}{\partial\rho}\rho\frac{\partial}{\partial\rho} + \frac{n^2}{\rho^2} + k^2 - k_z^2\right) \begin{bmatrix} E_z \\ H_z \end{bmatrix} = 0. \tag{3.8.2}$$

The general solution to the above is of the form

$$\begin{bmatrix} E_z \\ H_z \end{bmatrix} = [\mathbf{a}_n J_n(k_\rho \rho) + \mathbf{b}_n H_n^{(1)}(k_\rho \rho)] e^{ik_z z + in\phi}, \tag{3.8.3}$$

where $k_\rho = \sqrt{k^2 - k_z^2}$, and $H_n^{(1)}(x)$ is the Hankel function of the first kind.

We can assume an incident field on a dielectric rod in region 0 as

$$\mathbf{f}_{0z}^i = \begin{bmatrix} E_{0z}^i \\ H_{0z}^i \end{bmatrix} = \mathbf{a}_{0n} J_n(k_{0\rho}\rho) \tag{3.8.4}$$

where $k_{0\rho} = \sqrt{k_0^2 - k_z^2}$, and the $e^{ik_z z + in\phi}$ dependence is implied. The above is the incident wave in the absence of the dielectric rod. Hence, it cannot have a Hankel wave for it would

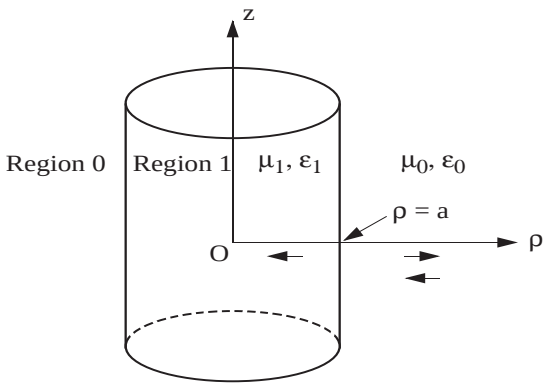


Figure 3.14: Geometry for defining the reflection of a wave off a circular dielectric rod.

be singular at the origin. When a circular dielectric rod is put at the origin, it will reflect the incident wave generating an outgoing wave satisfying the radiation condition at infinity. Therefore, The total solution in region 0 must be of the form

$$\mathbf{f}_{0z} = \begin{bmatrix} E_{oz} \\ H_{0z} \end{bmatrix} = \mathbf{a}_{0n} J_n(k_{0\rho}\rho) + \mathbf{b}_{0n} H_n^{(1)}(k_{0\rho}\rho). \tag{3.8.5}$$

Since $\mathbf{b}_n$ must be linearly dependent on $\mathbf{a}_n$, we express

$$\mathbf{b}_{0n} = \overline{\mathbf{R}}_{01} \cdot \mathbf{a}_{0n} \tag{3.8.6}$$

In region 1, the solution can only admit Bessel waves since a Hankel wave is singular at the origin. Therefore, the general solution is of the form

$$\mathbf{f}_{1z} = \begin{bmatrix} E_{1z} \\ H_{1z} \end{bmatrix} = \mathbf{a}_{1n} J_n(k_{1\rho}\rho). \tag{3.8.7}$$

Since $\mathbf{a}_n$ is linearly related to $\mathbf{a}_{0n}$, we let

$$\mathbf{a}_{1n} = \overline{\mathbf{T}}_{01} \cdot \mathbf{a}_{0n}. \tag{3.8.8}$$

The transverse to z components of the field can be obtained from the following equations in the j-th region

$$\mathbf{E}_{js} = \frac{1}{k_{j\rho}^2}[ik_z \nabla_s E_{jz} - i\omega\mu\hat{z} \times \nabla_s H_{jz}], \tag{3.8.9a}$$

$$\mathbf{H}_{js} = \frac{1}{k_{j\rho}^2}[ik_z \nabla_s H_{jz} + i\omega\epsilon\hat{z} \times \nabla_s E_{jz}], \tag{3.8.9b}$$

where $\nabla_s = \hat{\rho}\frac{\partial}{\partial\rho} + \hat{\phi}\frac{1}{\rho}\frac{\partial}{\partial\phi} = \hat{\rho}\frac{\partial}{\partial\rho} + \hat{\phi}\frac{in}{\rho}$, and $k_{j\rho}^2 = k_j^2 - k_z^2$. From the phase-matching condition, k_z is the same in all regions.

In order to find $\overline{\mathbf{R}}_{01}$ and $\overline{\mathbf{T}}_{01}$ in (3.8.6) and (3.8.8), we need to match the continuity of the tangential components of $\mathbf{E}$ and $\mathbf{H}$ across the interface. These tangential components are the z and the ϕ components. The z components are already given in (3.8.5) and (3.8.7). The corresponding ϕ components can be derived using (3.8.9a) and (3.8.9b). From (3.8.5) and (3.8.9a) and (3.8.9b), we deduce that

$$\mathbf{f}_{0\phi} = \begin{bmatrix} H_{0\phi} \\ E_{0\phi} \end{bmatrix} = \overline{\mathbf{J}}_n(k_{0\rho}\rho) \cdot \mathbf{a}_{0n} + \overline{\mathbf{H}}_n^{(1)}(k_{0\rho}\rho) \cdot \overline{\mathbf{R}}_{01} \cdot \mathbf{a}_{0n} \tag{3.8.10}$$

$$\mathbf{f}_{1\phi} = \begin{bmatrix} H_{1\phi} \\ E_{1\phi} \end{bmatrix} = \overline{\mathbf{J}}_n(k_{0\rho}\rho) \cdot \overline{\mathbf{T}}_{01} \cdot \mathbf{a}_{0n}, \tag{3.8.11}$$

where

$$\overline{\mathbf{B}}_n(k_{j\rho}\rho) = \frac{1}{k_{j\rho}^2\rho} \begin{bmatrix} i\omega\epsilon_j k_{j\rho}\rho B_n'(k_{j\rho}\rho) & -nk_z B_n(k_{j\rho}\rho) \\ -nk_z B_n(k_{j\rho}\rho) & -i\omega\mu_j k_{j\rho}\rho B_n'(k_{j\rho}\rho) \end{bmatrix}. \tag{3.8.12}$$

In the above, B_n is either $H_n^{(1)}$ or J_n depending on if we are defining $\overline{\mathbf{H}}_n^{(1)}$ or $\overline{\mathbf{J}}_n$ matrix. Note that $\overline{\mathbf{B}}_n$ is diagonal when $n = 0$. This also implies the decoupling of E_z and H_z waves when $n = 0$.

Matching the boundary condition at the boundary where $\rho = a$, we have

$$[J_n(k_{0\rho}a) + H_n^{(1)}(k_{0\rho}a)\overline{\mathbf{R}}_{01}] \cdot \mathbf{a}_{0n} = J_n(k_{1\rho}a)\overline{\mathbf{T}}_{01} \cdot \mathbf{a}_{0n}, \tag{3.8.13a}$$

$$[\overline{\mathbf{J}}_n(k_{0\rho}a) + \overline{\mathbf{H}}_n^{(1)}(k_{0\rho}a) \cdot \overline{\mathbf{R}}_{01}] \cdot \mathbf{a}_{0n} = \overline{\mathbf{J}}_n(k_{1\rho}a) \cdot \overline{\mathbf{T}}_{01} \cdot \mathbf{a}_{0n}, \tag{3.8.13b}$$

The above can be solved to yield

$$\overline{\mathbf{R}}_{01} = \overline{\mathbf{D}}^{-1} \cdot [J_n(k_{1\rho}a)\overline{\mathbf{J}}_n(k_{0\rho}a) - J_n(k_{0\rho}a)\overline{\mathbf{J}}_n(k_{1\rho}a)], \tag{3.8.14a}$$

$$\overline{\mathbf{T}}_{01} = \frac{2\omega}{\pi k_{0\rho}^2 a}\overline{\mathbf{D}}^{-1} \cdot \begin{bmatrix} \epsilon_0 & 0 \\ 0 & -\mu_0 \end{bmatrix}, \tag{3.8.14b}$$

where

$$\overline{\mathbf{D}} = \left[\overline{\mathbf{J}}_n(k_{1\rho}a)H_n^{(1)}(k_{0\rho}a) - \overline{\mathbf{H}}_n^{(1)}(k_{0\rho}a)J_n(k_{1\rho}a)\right]. \tag{3.8.14c}$$

The Wronskian for Hankel function, which is

$$H_n^{(1)}(x)J_n'(x) - J_n(x)H_n^{(1)'}(x) = -\frac{2i}{\pi x} \tag{3.8.15}$$

has been used to simplify the above.

In general $\overline{\mathbf{R}}_{01}$ and $\overline{\mathbf{T}}_{01}$ are non-diagonal implying the coupling of the E_z and H_z waves by the dielectric rod.

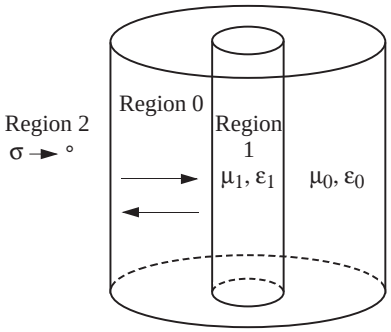


Figure 3.15: Geometry of a dielectric-rod-loaded circular waveguide.

3.8.2 Reflection off a PEC Waveguide Wall

We have asserted that a PEC or a PMC cylindrical surface does not depolarize an E_z or H_z wave, We can further confirm this assertion by looking at Equation (3.8.9a). For a PEC, we require that $E_z = 0$ and that $\hat{n} \times \mathbf{E}_s = 0$ on the cylindrical surface. It is seen that if we set $E_z = 0$ on a surface, then $\hat{n} \times \nabla E_z = 0$ or the first term in (3.8.9a) is zero without any help from the second term in (3.8.9a). Therefore, the E_z wave alone can satisfy the boundary condition on a PEC independently of the H_z wave. In order for $\hat{n} \times \mathbf{E}_s = 0$ for the H_z wave, we require that

$$\hat{n} \times \hat{z} \times \nabla_s H_z = 0, \tag{3.8.16a}$$

or that

$$\hat{n} \cdot \nabla_s H_z = 0. \tag{3.8.16b}$$

In other words, if we impose the homogeneous Neumann boundary condition (3.8.16b) on a PEC cylindrical surface, the tangential electric field that arises from the H_z wave will satisfy the requisite boundary condition independently of the E_z wave.

As a result of the above discussion, when an outgoing Hankel Wave impinges on a PEC waveguide wall which is circular, it reflects back into a Bessel wave. Hence, the wave in region 0 can be written as

$$\mathbf{f}_{0z} = \begin{bmatrix} E_{0z} \\ H_{0z} \end{bmatrix} = H_n^{(1)}(k_{0\rho}\rho)\mathbf{b}_{0n} + J_n(k_{0\rho}\rho)\overline{\mathbf{R}}_{02} \cdot \mathbf{b}_{0n} \tag{3.8.17}$$

where $\overline{\mathbf{R}}_{02}$ is a diagonal 2×2 matrix due to the decoupling of the E_z and H_z waves. Matching the requisite boundary condition on the waveguide wall at $\rho = b$, we obtain that

$$\overline{\mathbf{R}}_{02} = \begin{bmatrix} -H_n^{(1)}(k_{0\rho}b)/J_n(k_{0\rho}b) & 0 \\ 0 & -H_n^{(1)'}(k_{0\rho}b)/J_n'(k_{0\rho}b) \end{bmatrix}. \tag{3.8.18}$$

Note that if region 2 is a dielectric region, the above will be a non-diagonal matrix as shown in the next subsection.

3.8.3 Reflection off an Outer Dielectric Wall

If region 2 is a dielectric region rather than a perfectly conducting region, a reflection matrix can be similarly derived as in Subsection 3.5.1. In this case, we can show that

$$\overline{\mathbf{R}}_{02} = \overline{\mathbf{D}}^{-1} \cdot [H_n^{(1)}(k_{0\rho}a)\overline{\mathbf{H}}_n^{(1)}(k_{2\rho}a) - H_n^{(1)}(k_{2\rho}a)\overline{\mathbf{H}}_n^{(1)}(k_{0\rho}a)] \tag{3.8.19a}$$

$$\overline{\mathbf{T}}_{02} = \frac{2\omega}{\pi k_{0\rho}^2 a}\overline{\mathbf{D}}^{-1} \cdot \begin{bmatrix} \epsilon_0 & 0 \\ 0 & -\mu_0 \end{bmatrix}. \tag{3.8.19b}$$

where

$$\overline{\mathbf{D}} = [\overline{\mathbf{J}}_n(k_{0\rho}a)H_n^{(1)}(k_{2\rho}a) - \overline{\mathbf{H}}_n^{(1)}(k_{2\rho}a)J_n(k_{0\rho}a)]. \tag{3.8.19c}$$

3.8.4 The Guidance Condition

The guidance condition in a dielectric-rod-loaded circular waveguide can be obtained by considering the solution in region 0 and matching boundary condition on the waveguide wall at $\rho = b$. The solution in region 0 is given by (3.8.5) is rewritten here as

$$\mathbf{f}_{0z} = [J_n(k_{0\rho}\rho)\bar{\mathbf{I}} + H_n^{(1)}(k_{0p}\rho)\overline{\mathbf{R}}_{01}] \cdot \mathbf{a}_{0n}. \tag{3.8.20}$$

However, according to (3.8.17), the field in region 0 can also be written as

$$\mathbf{f}_{0z} = [J_n(k_{0\rho}\rho)\overline{\mathbf{R}}_{02} + H_n^{(1)}(k_{0\rho}\rho)\bar{\mathbf{I}}] \cdot \mathbf{b}_{0n}. \tag{3.8.21}$$

Hence, we conclude that

$$\overline{\mathbf{R}}_{02} \cdot \mathbf{b}_{0n} = \mathbf{a}_{0n}, \tag{3.8.22a}$$

$$\overline{\mathbf{R}}_{01} \cdot \mathbf{a}_{0n} = \mathbf{b}_{0n}, \tag{3.8.22b}$$

or that

$$(\overline{\mathbf{R}}_{02} \cdot \overline{\mathbf{R}}_{01} - \bar{\mathbf{I}}) \cdot \mathbf{a}_{0n} = 0. \tag{3.8.23}$$

In order for $\mathbf{a}_{0n} \neq 0$, we require that

$$\det(\overline{\mathbf{R}}_{02} \cdot \overline{\mathbf{R}}_{01} - \bar{\mathbf{I}}) = 0. \tag{3.8.24}$$

The above is the guidance condition or the transverse resonance condition for a cylindrically layered circular waveguide.

3.9 Applications of Inhomogeneously Filled Waveguides

Inhomogeneous filled waveguides can be used to make variable phase shifters and attenuators [3, 4]. When the inhomogeneity filling the waveguide is nonreciprocal like ferrite, isolators, gyrators, and attenuators can be made.

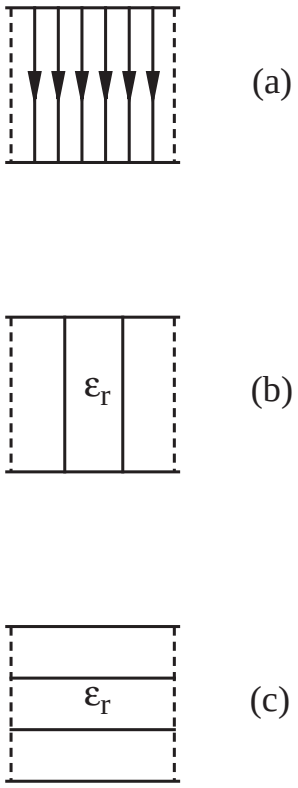


Figure 3.16: Different dielectric-slab loading of a parallel plate waveguide.

3.9.1 The Effect of Inhomogeneous Fillings on the Phase Velocity

An intuitive understanding of how inhomogeneous fillings affect the phase velocity of a guided mode can be acquired by studying a parallel plate waveguide. Considering a square region of a parallel plate waveguide away from the edges so that fringing field effect can be ignored. Hence, we assume that the field lines for the TEM mode is perfectly vertical.

The capacitance of (a) can be thought of as three capacitances in parallel or as three capacitances in series. When the waveguide is filled as in (b), it affects one of the three capacitances in parallel, yielding a resultant capacitance given by

$$C_b = \frac{2}{3}C_a + \frac{\epsilon_r}{3}C_{a,} = \frac{2+\epsilon_r}{3}C_a \tag{3.9.1}$$

where C_a is the capacitance of (a). When the waveguide is filled as in (c), it affects one of the three capacitances in series, yielding a resultant capacitance given by

$$C_c^{-1} = 2(3C_a)^{-1} + (3\epsilon_r C_a)^{-1} \tag{3.9.2}$$

or

$$C_c = \frac{3\epsilon_r C_a}{2\epsilon_r + 1}. \tag{3.9.3}$$

It is seen that

$$\frac{3\epsilon_r}{2\epsilon_r+1} < \frac{2+\epsilon_r}{3}, \quad \epsilon_r > 1. \tag{3.9.4}$$

Therefore $C_b > C_c$ always for $\epsilon_r > 1$. The reason is that enhancing a capacitor in parallel has more effect on the total capacitance than enhancing a capacitor in parallel. When $\epsilon_r \to \infty, C_c \to \frac{3}{2}C_a$ or saturates while $C_b \to \infty$. Since the phase velocity of a TEM mode in a parallel plate waveguide is given by

$$\upsilon = \frac{1}{\sqrt{LC}} \tag{3.9.5}$$

where L and C are line inductance and line capacitance respectively, a dielectric loading in case (b) slows down the wave more than case (c).

When the waveguide is a square waveguide, the TE_{10} mode is affected even more by a symmetrically located dielectric slab because the electric field has a maximum at the center of the waveguide. Hence, by dielectric-slab loading as in case (b), the TE_{10} mode propagates with a slower phase velocity than the TE_{01} mode. This gives rise to anisotropy in a waveguide.

3.9.2 Quarter-Wave Plate

By slab loading a circular waveguide, one can make a quarter wave plate. If a dielectric slab is oriented at 45^o with respect to the TE_{11} mode of a circular waveguide, the mode can be decomposed into two orthogonal modes, one with **E** field perpendicular to the dielectric slab, and another parallel to the slab. The one with electric field parallel to the slab is going to be slowed down more than the one with electric field perpendicular to the slab. Hence after a certain distance, the phases of these two modes are going to be out of phase. If the length of the dielectric slab is chosen judiciously such that these two modes are 90^o out of phase,

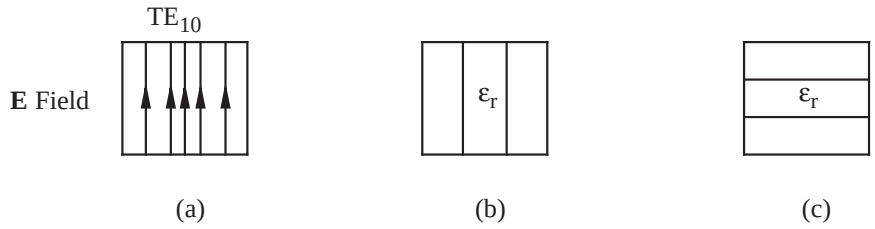


Figure 3.17: Dielectric-slab loading of a square waveguide. The TE_{10} mode is affected more by case (b) than case (c).

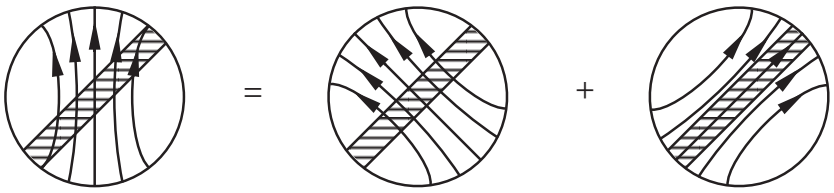


Figure 3.18: Decomposition of a TE_{11} mode of a circular waveguide into two orthogonal modes. One perpendicular to the slab and another parallel to the slab.

then one obtains a circularly polarized mode at the other end of the waveguide. Such an application of a dielectric slab loading is similar to a quarter-wave plate in optics and hence its name. A half-wave plate will shift one component of the wave by 180^o compared to the other orthogonal component.

3.9.3 Variable Phase Shifter

Dielectric loading can be used to make variable phase shifters. One way of achieving this is to vary the position of the slab position in a rectangular waveguide. Since the TE_{10} mode of a rectangular waveguide has a weaker field near the side walls of the waveguide, the phase velocity will be slowed down less when the slab is near to the waveguide wall compared to near the center of the waveguide (See Figure 3.20).

A linear phase changer can also be constructed by dielectric slab loading as shown below. When the center slab is moved by a distance Δz, line 1 and line 3 are increased by a length Δz while line 2 and line 4 are decreased by a length Δz. Therefore, the total phase shift is

$$\Delta\phi = (k_{z1} + k_{z3} - k_{z2} - k_{z4})\Delta z. \tag{3.9.6}$$

Because of the sinusoidal distribution of the electric field of a TE_{10} mode of a rectangular

waveguide, it is clear that

$$k_{z4} < k_{z1} < k_{z3} < k_{z2}. \tag{3.9.7}$$

However, if the center section is chosen to be about $0.3a$, we can have

$$k_{z1} + k_{z3} > k_{z2} + k_{z4}, \tag{3.9.8}$$

and a net positive phase shift linearly proportional to Δz becomes possible.

A rotary phase shifter can be made by sandwiching a section of half-wave plate between two sections of quarter-wave plates as shown in Figure 3.21. The quarter-wave plate and half-wave plate sections are made by dielectric slab loading as described previously. By rotating the middle section, one can vary the phase of the wave going from port A to port B.

To explain the operating principle of this device, we will use linearly polarized waves in free space. The first quarter-wave plate converts a linearly polarized wave into a circularly polarized wave where the $\hat{x}$ and $\hat{y}$ components are 90^o out of phase, viz.,

$$\mathbf{E} = (i\hat{x} + \hat{y})E_0, \tag{3.9.9}$$

where $\hat{x}$ direction is parallel to the dielectric slab in the quarter-wave plate section, and E_0 is a complex number. The above represents a left-hand circularly polarized wave. When this wave impinges on the half-wave plate section, the dielectric slab can be oriented at any angles. In Figure 3.22, we assume it to be at an angle θ with respect to the coordinates of the dielectric slab in the first section. In this case, we have

$$\hat{x} = \hat{x}'\cos\theta - \hat{y}'\sin\theta, \quad \hat{y} = \hat{x}'\sin\theta + \hat{y}'\cos\theta. \tag{3.9.10}$$

Using (3.9.10) in (3.9.9), we have

$$\mathbf{E} = (i\hat{x}' + \hat{y}')e^{-i\theta}E_0 \tag{3.9.11}$$

The above is still a left-hand circularly polarized wave, and $e^{-i\theta}$ comes about because we are rotating the coordinates counterclockwise while the polarization is rotating clockwise. After this wave has gone through the half-wave section, the $\hat{y}'$ component will be 180^o out of phase with respect to the $\hat{x}'$ component and we have

$$\mathbf{E} = (-i\hat{x}' + \hat{y}')e^{-i\theta}E_0' \tag{3.9.12}$$

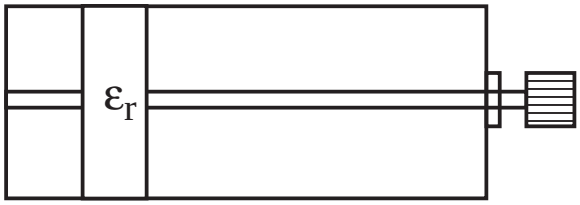
which is a right-hand circularly polarized wave, and E_0' is a new complex number accounting for the additional phase delay the wave has acquired in propagating through the middle section. Projecting this back to the original coordinates by using

$$\hat{x}' = \hat{x}\cos\theta + \hat{y}\sin\theta, \quad \hat{y}' = -\hat{x}\sin\theta + \hat{y}\cos\theta, \tag{3.9.13}$$

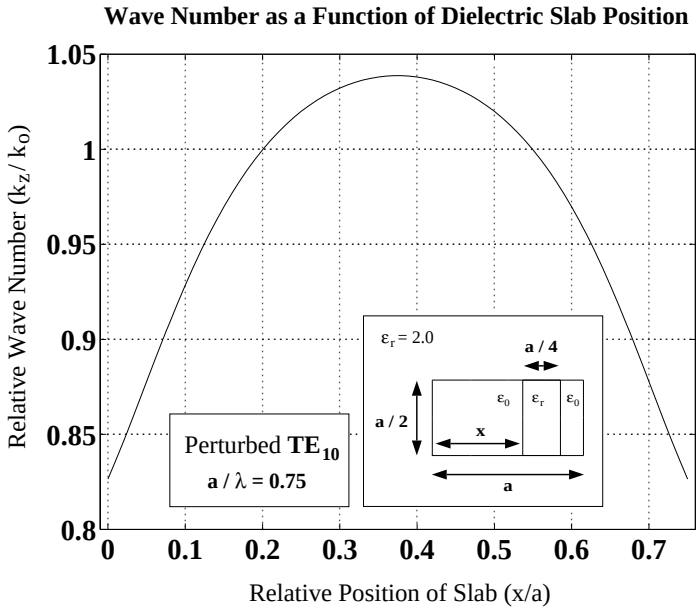
we have

$$\mathbf{E} = (-i\hat{x} + \hat{y})e^{-2i\theta}E_0'. \tag{3.9.14}$$

If the wave had remained left-hand circularly polarized, the projection back would have annulled the phase $e^{-i\theta}$. But because it becomes a right-hand circularly polarized, it introduces an additive phase instead.



(a)



(b)

Figure 3.19: (a) Variable dielectric slab locations can be used as a variable phase shifter. (b) Variable dielectric slab locations affects the phase velocity of the TE_{10} mode.

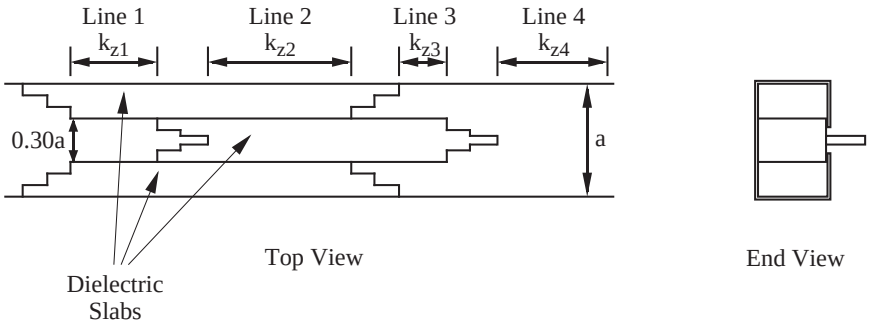


Figure 3.20: A linear phase shifter using four different sections of dielectric loading.

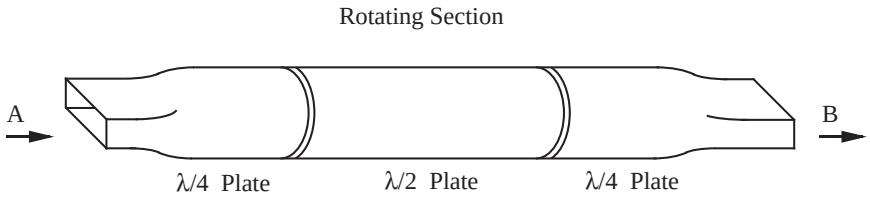


Figure 3.21: A rotary phase shifter.

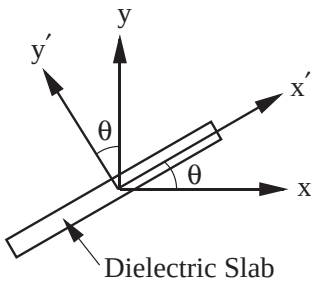


Figure 3.22: Orientation of the dielectric slab in the half-wave plate section compared to the original coordinates.

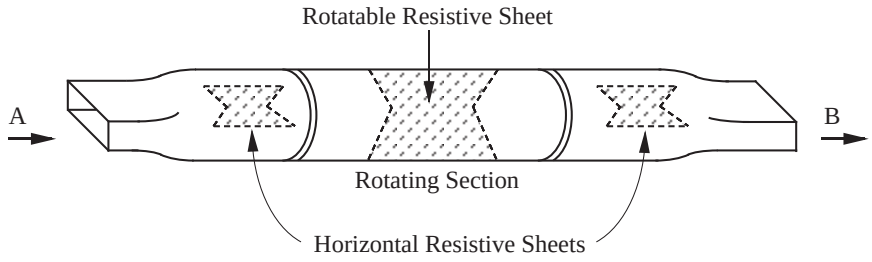


Figure 3.23: A rotary variable attenuator.

When this wave in (3.9.14) passes through the last section of the quarter-wave plate, the $\hat{x}$ component gains another 90^o of phase with respect to the $\hat{y}$ component, and it becomes

$$\mathbf{E} = (\hat{x} + \hat{y})e^{-2i\theta}E_0'' \tag{3.9.15}$$

where E_0'' is a new complex number. This wave is again linearly polarized with respect to the rectangular waveguide in the output port. Note that there is a phase shift of -2θ which is dependent on the orientation of the center rotating section .

3.9.4 Variable Attenuator

For the same reason that a dielectric slab parallel to the electric field affect a mode more than a slab perpendicular to the electric field, a resistive (lossy) dielectric sheet parallel to the electric field will incur more loss on a mode than a resistive sheet perpendicular to the electric field. Hence, a variable attenuator can be made similar to a variable phase shifter, except that in the first and the last section, the resistive sheets are loaded horizontally, and the middle section the resistive sheet can be rotated.

The horizontal resistive sheet in the first section ensures that the mode is predominantly vertically polarization. The resistive sheet in the middle section will attenuate the mode proportional to the component of the electric field parallel to the resistive sheet. Hence, the attenuation of the mode by the middle section can be varied by rotating it. The last section ensures that any horizontally polarized modes due to mode conversion be removed, and only the vertically polarized component will exit from port B.

3.10 Spin Dynamics and Ferrite Materials

The understanding of the interaction of particle spins with electromagnetic field is commonly encountered in the study of nuclear magnetic resonance, and ferrite materials. The first is extremely useful in magnetic resonance imaging (MRI) which is an important medical imaging modality. A particle spin has angular momentum as well as a magnetic moment. The magnetic moment will interact with an ambient magnetic field giving rise to the precession of the spins. The precession of the spins, in the case of MRI, yields spin echoes that can

be measured for imaging and spectrocopic purposes. In the case of ferrites, it gives rise to anisotropic, gyrotropic materials that exhibit Faraday rotation. Such effect can be used to design non-reciprocal microwave devices such as isolators.

As mentioned, ferrite is commonly used as an anisotropic material in a waveguide to make nonreciprocal waveguides [3, 4, 18–20]. Ferrite obtains its anisotropy by having its electron spins interact with a static magnetic field. An electron spin has a magnetic dipole moment as well as angular momentum. In the presence of a static magnetic field, this dipole moment aligns itself with the magnetic field. If a transverse force, via an electromagnetic field, is applied to tilt the direction of the dipole moment, and hence, the direction of angular momentum, the spin precesses about the static magnetic field just as a spinning top precesses about a gravitational field.

The correct description of the motion of an electron in the presence of a magnetic field requires quantum mechanics. However, when a large number of electrons are considered, their average motion can be described by a classical equation of motion similar to the equation governing the motion of a spinning top. From this equation, we can understand the anisotropic nature of ferrites being biased by a magnetic field.

For an electron with a magnetic dipole moment $\mathbf{m}$ in the presence of a magnetic field $\mathbf{B}_0$, the torque exerted on the electron is given by

$$\mathbf{T} = \mathbf{m} \times \mathbf{B}_0. \tag{3.10.1}$$

This torque is exerted on the angular momentum $\mathbf{P}$ of the electron, causing it to change. Therefore, we have

$$\frac{d\mathbf{P}}{dt} = \mathbf{T} = \mathbf{m} \times \mathbf{B}_0. \tag{3.10.2}$$

But the magnetic dipole moment of an electron is antiparallel to its angular momentum, i.e.,

$$\mathbf{m} = -\gamma \mathbf{P} \tag{3.10.3}$$

where γ is known as the gyromagnetic ratio. Consequently, the equation of motion for a spinning electron in a magnetic field is

$$\frac{d\mathbf{P}}{dt} = -\gamma \mathbf{P} \times \mathbf{B}_0 = +\boldsymbol{\omega}_0 \times \mathbf{P} \tag{3.10.4}$$

where $\boldsymbol{\omega}_0 = \gamma \mathbf{B}_0$, and $\omega_0 = |\boldsymbol{\omega}_0| = \gamma |\mathbf{B}_0|$ is also known as the Larmor frequency. For electron spins in a reasonably strong magnetic field, the Larmor frequency can be in the microwave regime.

Often time, the equation of motion is written as

$$\frac{d\mathbf{P}}{dt} = -\gamma \mathbf{P} \times \mathbf{B}_0 - \lambda \mathbf{P} \times (\mathbf{P} \times \mathbf{B}_0) \tag{3.10.5}$$

The last term above accounts for damping. The above is known as the Landau-Lifshitz equation. Later, in 1955, Gilbert, starting from first principles, replaced the last term with a different expression yielding

$$\frac{d\mathbf{P}}{dt} = -\gamma \left[\mathbf{P} \times \mathbf{B}_0 - \eta \mathbf{P} \times \left(\mathbf{P} \times \frac{d\mathbf{P}}{dt} \right) \right] \tag{3.10.6}$$

The above is known as the Landau-Lifshitz-Gilbert equation. It can be shown that the above reduces to the same form as the Landau-Lifshitz equation

$$\frac{d\mathbf{P}}{dt} = -\gamma'\mathbf{P} \times \mathbf{B}_0 - \lambda'\mathbf{P} \times (\mathbf{P} \times \mathbf{B}_0) \tag{3.10.7}$$

but with different γ' and λ'.

Even though we say that (3.10.4) to (3.10.7) are the equations of motion for an electron, it actually governs the average motion over an ensemble of a large number of electrons. Hence, a classical picture applies here. Other subatomic particles like protons also possess a spin, but because of their larger mass, the Larmor frequency is much lower. Proton spins, prevalent in MRI, has a Larmour frequency of 42.6 MHz per Tesla,[3] while electron spins have Larmor frequencies in the GHz regime.

The Larmor frequency is the free precession frequency of a spin when it is tipped from the vertical position in the presence of the static biasing magnetic field. Equation (3.10.4) is sometimes known as the Bloch equation.

Assuming that the field $\mathbf{B}_0$ is static, then it is easy to show that the natural solution to (3.10.4) consists of a $\mathbf{P}$ with a $\hat{z}$ component and a circulating $\hat{x}$ and $\hat{y}$ component. Let us assume that

$$\mathbf{P} = \frac{1}{\sqrt{2}}(\hat{x} + i\hat{y})P_s e^{-i\omega t} + \hat{z}P_z. \tag{3.10.8}$$

The above is not a pure time-harmonic signal as the second term can have a different frequency from the first term. Note that $|\mathbf{P}| = \sqrt{P_s^2 + P_z^2} = P_0 = \textit{constant}$ even though it is time varying as angular momentum has to be conserved. Also, this angular momentum comes from the intrinsic spin of the particle, which is a constant. Also, $\mathbf{P}(t)$ can be obtained in the real world by taking the real part of Equation (3.10.8), or by adding a complex conjugate term to the above. For convenience, we will leave (3.10.8) with its phasors. Then, using (3.10.8) in (3.10.4) yields

$$-i\omega(\hat{x} + i\hat{y})P_s\frac{e^{-i\omega t}}{\sqrt{2}} + \hat{z}\frac{dP_z}{dt} = (\omega_0\hat{y} - i\omega_0\hat{x})P_s\frac{e^{-i\omega t}}{\sqrt{2}}. \tag{3.10.9}$$

It is seen that the above is satisfied when we have

$$\omega = \omega_0, \quad \frac{dP_z}{dt} = 0. \tag{3.10.10}$$

In other words, the spin precesses at the Larmor frequency ω_0 while the $\hat{z}$-component of its angular momentum remains unchanged since the system is non-dissipative. In a dissipative system, the kinetic energy in the angular momentum will be lost to the environment. The precession of the spin will slow down, and eventually, the spin will be completely aligned with the background static magnetic field.

Now if we include an additional RF field $\mathbf{B}_1$, in the transverse direction which is circularly polarized such that

$$\mathbf{B}_{1+} = \frac{1}{\sqrt{2}}(\hat{x} + i\hat{y})B_{1+}e^{-i\omega t}, \tag{3.10.11}$$

[3] 1 T=10^4 gauss=1 weber/m^2=1 Volt Second/m^2.

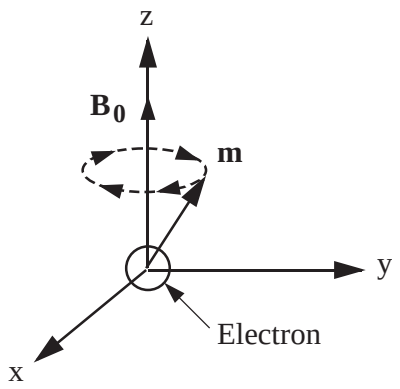


Figure 3.24: The magnetic dipole moment of an electron precesses just like a spinning top when an RF transverse to z magnetic field is applied.

then the equation of motion becomes

$$\frac{d\mathbf{P}}{dt} = \boldsymbol{\omega}_0 \times \mathbf{P} + \gamma \mathbf{B}_{1+} \times \mathbf{P}. \tag{3.10.12}$$

The transverse RF field, which is the driving field, will tilt the spin and force it to precess at the same frequency. Therefore, we let

$$\mathbf{P} = \frac{1}{\sqrt{2}}(\hat{x} + i\hat{y})P_s e^{-i\omega t} + \hat{z}P_z. \tag{3.10.13}$$

Using (3.10.13) in (3.10.12), we have

$$\begin{aligned} -i\omega(\hat{x} + i\hat{y})P_s \frac{e^{-i\omega t}}{\sqrt{2}} + \hat{z}\frac{dP_z}{dt} &= -i\omega_0(\hat{x} + i\hat{y})P_s \frac{e^{-i\omega t}}{\sqrt{2}} \\ &\quad + i\gamma(\hat{x} + i\hat{y})B_{1+}P_z \frac{e^{-i\omega t}}{\sqrt{2}} \end{aligned} \tag{3.10.14}$$

Again we have $dP_z/dt = 0$ and

$$P_s = \frac{\gamma B_{1+}}{\omega_0 - \omega} P_z. \tag{3.10.15}$$

The precession now is at the driving frequency ω. This equation so far, has been derived with no approximation. Note that if the frequency $\omega = \omega_0$, a resonance occur, but P_s does not go to infinity as $|\mathbf{P}| = \sqrt{P_s^2 + P_z^2} = P_0 = constant$ and $P_z \to 0$ when $P_s \to P_0$. In other words, as one drives the spin system closer to the resonance frequency, $P_s \gg P_z$, but $P_s^2 + P_z^2 = constant$. Therefore, the vector $\mathbf{P}$ has to tilt, and at resonance, the vector $\mathbf{P}$ is precessing horizontally.

Next, we assume that the applied RF field has a magnetic field component given by a circularly polarized field of opposite polarity such that

$$\mathbf{B}_{1-} = \frac{1}{\sqrt{2}}(\hat{x} - i\hat{y})B_{1-}e^{-i\omega t} \tag{3.10.16}$$

and assuming then the spins are tipped from the vertical and precess at the same frequency as the driving RF field, we let

$$\mathbf{P} = \frac{1}{\sqrt{2}}(\hat{x} - i\hat{y})P_s e^{-i\omega t} + \hat{z}P_z. \tag{3.10.17}$$

Using (3.10.17) in (3.10.12) with $\mathbf{B}_{1+}$ replaced by $\mathbf{B}_{1-}$, we have

$$-i\omega(\hat{x} - i\hat{y})P_s\frac{e^{-i\omega t}}{\sqrt{2}} + \hat{z}\frac{dP_z}{dt} = i\omega_0(\hat{x} - i\hat{y})P_s\frac{e^{-i\omega t}}{\sqrt{2}} - i\gamma(\hat{x} - i\hat{y})B_{1-}P_z\frac{e^{-i\omega t}}{\sqrt{2}}. \tag{3.10.18}$$

The above implies that $dP_z/dt = 0$, and that

$$P_s = \frac{\gamma B_{1-}}{\omega_0 + \omega}P_z. \tag{3.10.19}$$

Assuming that the RF field is much smaller than the static field so that $B_1 \ll B_0$, then we can assume that P_s is small or that $P_z \simeq P_0$. In this case, (3.10.15) and (3.10.19) together become

$$P_{s\pm} \approx \frac{\gamma B_{1\pm}}{\omega_0 \mp \omega}P_0, \tag{3.10.20}$$

and the transverse part of the spin angular momentum becomes

$$\mathbf{P}_{s\pm} \approx \frac{1}{\sqrt{2}}(\hat{x} \pm i\hat{y})\frac{\gamma B_{1\pm}}{\omega_0 \mp \omega}P_0 e^{-i\omega t}. \tag{3.10.21}$$

By the above approximation, we have simplified the relationship between $\mathbf{P}_{s\pm}$ and the driving field $B_{1\pm}$.

The magnetic moment of a spin is given by $\mathbf{m} = -\gamma\mathbf{P}$. Hence, the transverse part of the magnetic moment is

$$\mathbf{m}_{s\pm} = \frac{1}{\sqrt{2}}(\hat{x} \pm i\hat{y})\frac{\gamma B_{1\pm}m_0}{\omega_0 \mp \omega}e^{-i\omega t}. \tag{3.10.22}$$

where $m_0 = -\gamma P_0$ is the component of the magnetic moment in the $\hat{z}$-direction, where in the approximation here it is the magnitude of the magnetic moment.

The magnetic dipole moment density or magnetization density is given by $\mathbf{M} = N\mathbf{m}$ and we have

$$\mathbf{M}_{s\pm} = \frac{1}{\sqrt{2}}(\hat{x} \pm i\hat{y})\frac{\gamma B_{1\pm}M_0}{\omega_0 \mp \omega}e^{-i\omega t}. \tag{3.10.23}$$

where $\mathbf{M}_s = N\mathbf{m}_s$ and $M_0 = Nm_0$ are the dipole moment density in the transverse and axial directions respectively.

The above is the RF response of the transverse magnetization density in response to an applied RF magnetic field given by

$$\mathbf{B}_{1\pm} = \frac{1}{\sqrt{2}}(\hat{x} \pm i\hat{y})B_{1\pm}e^{-i\omega t}. \tag{3.10.24}$$

In other words, we have

$$\mathbf{M}_{s\pm} = \frac{\gamma M_0}{\omega_0 \mp \omega}\mathbf{B}_{1\pm}. \tag{3.10.25}$$

One can always make a linear polarization out of a linear superposition of two circular polarization via

$$\mathbf{B} = \hat{x}B_x e^{-i\omega t} = \frac{1}{2}\left[(\hat{x} + i\hat{y})B_x + (\hat{x} - i\hat{y})B_x\right]e^{-i\omega t}. \tag{3.10.26}$$

When this RF field is used to excite the spins, the RF magnetization response is given by

$$\begin{aligned}\mathbf{M}_s &= \left(\frac{\gamma M_0}{\omega_0 - \omega}\right)\frac{1}{2}(\hat{x} + i\hat{y})B_x e^{-i\omega t} + \left(\frac{\gamma M_0}{\omega_0 + \omega}\right)\frac{1}{2}(\hat{x} - i\hat{y})B_x e^{-i\omega t} \\ &= \hat{x}\gamma M_0 B_x\left(\frac{\omega_0}{\omega_0^2 - \omega^2}\right)e^{-i\omega t} + i\hat{y}\gamma M_0 B_x\left(\frac{\omega}{\omega_0^2 - \omega^2}\right)e^{-i\omega t}.\end{aligned} \tag{3.10.27}$$

Similarly, when we make the RF field $\hat{y}$ polarized such that

$$\mathbf{B} = \hat{y}B_y e^{-i\omega t} = \frac{1}{2i}[(\hat{x} + i\hat{y})B_y - (\hat{x} - i\hat{y})B_y]e^{-i\omega t}, \tag{3.10.28}$$

the RF magnetization response is

$$\begin{aligned}\mathbf{M}_s &= \left(\frac{\gamma M_0}{\omega_0 - \omega}\right)\frac{1}{2i}(\hat{x} + i\hat{y})B_y e^{-i\omega t} - \left(\frac{\gamma M_0}{\omega_0 + \omega}\right)\frac{1}{2i}(\hat{x} - i\hat{y})B_y e^{-i\omega t} \\ &= -i\hat{x}\gamma M_0 B_y\left(\frac{\omega}{\omega_0^2 - \omega^2}\right)e^{-i\omega t} + \hat{y}\gamma M_0 B_y\left(\frac{\omega_0}{\omega_0^2 - \omega^2}\right)e^{-i\omega t}.\end{aligned} \tag{3.10.29}$$

Consequently, $B_x = \mu_0 H_x$, and $B_y = \mu_0 H_y$, we have in matrix form

$$\begin{bmatrix} M_x \\ M_y \end{bmatrix} = \begin{bmatrix} \chi_{xx} & \chi_{xy} \\ \chi_{yx} & \chi_{yy} \end{bmatrix}\begin{bmatrix} H_x \\ H_y \end{bmatrix}, \quad \text{or} \quad \mathbf{M} = \overline{\chi}\cdot\mathbf{H} \tag{3.10.30}$$

where

$$\chi_{xx} = \chi_{yy} = \frac{\mu_0\gamma M_0\omega_0}{\omega_0^2 - \omega^2}, \tag{3.10.31a}$$

$$\chi_{xy} = -\chi_{yx} = \frac{-i\mu_0\gamma M_0\omega}{\omega_0^2 - \omega^2}. \tag{3.10.31b}$$

The above is the RF magnetization response to an RF magnetic field excitation in a ferrite medium when the RF signal is assumed small, and we assume that $\hat{z}$ component of the

magnetization response is small and is of higher order. The reason being that the z component of the RF field has little effect on the spin momentum when it is pointed primarily in the z direction. In general,

$$\mathbf{B} = \mu_0(\mathbf{H} + \mathbf{M}) = \mu_0(\bar{\mathbf{I}} + \overline{\chi}) \cdot \mathbf{H} = \overline{\mu} \cdot \mathbf{H} \tag{3.10.32}$$

where

$$\overline{\mu} = \mu_0 \begin{bmatrix} 1 + \chi_{xx} & \chi_{xy} & 0 \\ \chi_{yx} & 1 + \chi_{yy} & 0 \\ 0 & 0 & 1 \end{bmatrix}. \tag{3.10.33}$$

This is an example of an anisotropic magnetic medium, or a gyrotropic magnetic medium. The above permeability tensor is also Hermitian implying that it represents a lossless medium.

3.10.1 Natural Plane Wave Solutions in an Infinite Homogeneous Ferrite Medium

A ferrite medium,which is a gyrotropic material, admits circularly polarized plane waves as the natural plane waves propagating in a homogeneous anisotropic ferrites [21]. We can see this by looking at the solution of the vector wave equation. When ϵ is homogeneous and isotropic, and $\overline{\mu}$ is anisotropic, the magnetic field satisfies the following vector wave equation:

$$\nabla \times \nabla \times \mathbf{H} - \omega^2 \epsilon \overline{\boldsymbol{\mu}} \cdot \mathbf{H} = 0. \tag{3.10.34}$$

Assuming that a plane wave solution propagating in the z direction exists such that

$$\mathbf{H} = \mathbf{H}_0 e^{ikz}, \tag{3.10.35}$$

then

$$-k^2 \hat{z}(\hat{z} \cdot \mathbf{H}) + k^2 \mathbf{H} - \omega^2 \epsilon \overline{\boldsymbol{\mu}} \cdot \mathbf{H} = 0. \tag{3.10.36}$$

If $\overline{\boldsymbol{\mu}}$ is of the form

$$\overline{\boldsymbol{\mu}} = \begin{bmatrix} \overline{\boldsymbol{\mu}}_s & 0 \\ 0 & \mu \end{bmatrix} \tag{3.10.37}$$

where $\overline{\boldsymbol{\mu}}_s$ is a 2×2 tensor, then $H_z = 0$ by equating the z component of (3.10.36). Therefore $\mathbf{H}$ has only transverse to z components, and Equation (3.10.36) becomes

$$k^2 \mathbf{H}_s - \omega^2 \epsilon \overline{\boldsymbol{\mu}}_s \cdot \mathbf{H}_s = 0 \tag{3.10.38}$$

where $\mathbf{H}_s$ is a vector in the xy plane. From Equation (3.10.33), it is clear that

$$\mathbf{B}_{s\pm} = \overline{\boldsymbol{\mu}}_s \cdot \mathbf{H}_{s\pm} = \mu_{\pm} \mathbf{H}_{s\pm} \tag{3.10.39}$$

for a ferrite medium where $\chi_{xx} = \chi_{yy}$, and

$$\mathbf{H}_{s\pm} = H_{0\pm}(\hat{x} \pm i\hat{y}). \tag{3.10.40}$$

Consequently, (3.10.38) becomes

$$k^2 \mathbf{H}_{s\pm} - \omega^2 \epsilon \mu_{\pm} \mathbf{H}_{s\pm} = 0 \tag{3.10.41}$$

where k can have two possible values given by

$$k_{\pm} = \omega\sqrt{\mu_{\pm}\epsilon}. \tag{3.10.42}$$

Therefore, the natural solution in a ferrite medium is of the form

$$\mathbf{H}_{\pm} = H_{0\pm}(\hat{x} \pm i\hat{y})e^{ik_{\pm}z}. \tag{3.10.43}$$

In essence, in a ferrite medium, a right-hand circularly polarized wave "feels" a different permeability compared to a left-hand circularly polarized wave. Therefore, the two polarizations propagate with different velocities. When the wave is not propagating in the z direction, the propagation of the wave is more complicated, and will not be discussed here.

3.10.2 Faraday Rotation

Faraday rotation occurs in a ferrite. It can also occur in the earth ionosphere where the electron spins are biased by the earth magnetic field. To understand Faraday rotation, we decompose a linearly polarized wave into two circularly polarized waves, viz.,

$$\mathbf{E} = \hat{x}E_0 = \frac{1}{2}(\hat{x} + i\hat{y})E_0 + \frac{1}{2}(\hat{x} - i\hat{y})E_0, \tag{3.10.44}$$

where the first term is left-hand circularly polarized for a wave propagating in the $\hat{z}$ direction, and likewise, the second term is right-hand circularly polarized. In a ferrite medium, these two polarizations will propagate with different phase velocities. After a certain distance, there could be a phase difference between them, and we have

$$\mathbf{E} = \frac{1}{2}(\hat{x} + i\hat{y})E_0'e^{i\theta} + \frac{1}{2}(\hat{x} - i\hat{y})E_0'. \tag{3.10.45}$$

Combining the terms in the above, we have

$$\begin{aligned}\mathbf{E} &= \hat{x}e^{i\frac{\theta}{2}}\cos\left(\frac{\theta}{2}\right)E_0' - \hat{y}e^{i\frac{\theta}{2}}\sin\left(\frac{\theta}{2}\right)E_0' \\ &= \left[\hat{x}\cos\left(\frac{\theta}{2}\right) - \hat{y}\sin\left(\frac{\theta}{2}\right)\right]E_0'e^{i\theta/2}.\end{aligned} \tag{3.10.46}$$

Hence, the wave vector is now tilted by an angle $-\frac{\theta}{2}$ or has rotated clockwise.

For a wave propagating in the negative z direction, the polarizations in the first and second term reverse roles. Now, the second term in (3.10.45) will have a phase gain over the first term, and the wave vector is rotated counterclockwise after propagating through a certain distance. Hence, the phenomenon is nonreciprocal meaning that the waves propagating in the $\pm z$ directions are quite different in behavior.

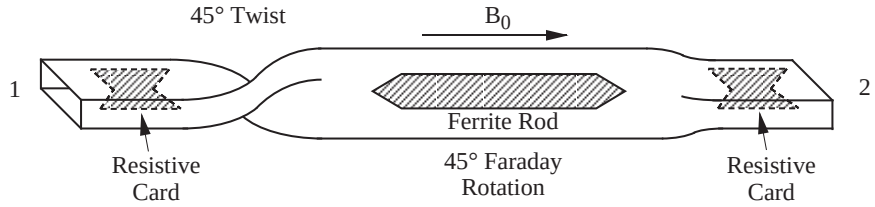


Figure 3.25: A Faraday-rotation isolator.

3.10.3 Applications of Faraday Rotation

Faraday rotation can be used to make a number of nonreciprocal microwave devices. One simple example is a gyrator, which is defined as a device whose transmission from port 1 to port 2 has a 180^o phase shift compared to its transmission from port 2 to port 1. The polarization rotation and the nonreciprocal natures of ferrites can be used to make such a device.

Another device that can be made from ferrite loading is an isolator. An isolator consists of first a 45^o mechanical twist which is a reciprocal section. Then it is followed by another section of 45^o Faraday rotation which is the nonreciprocal section. Resistive cards are added to filter undesired modes other than the TE_{10} mode.

For a mode transmitting from port 1 to port 2, the mechanical twist section rotates the polarization counterclockwise by 45^o while the ferrite section undoes it by rotating the polarization clockwise by 45^o. Hence the transmission from port 1 to port 2 is little affected.

For transmission from port 2 to port 1, the rotation in the ferrite section is counterclockwise, and hence is additive with respect to the rotation in the twisted section. Hence, when the wave arrives at port 1, it has been rotated by 90^o, and will not be able to transmit as the TE_{10} mode.

Another way to use ferrite as an isolator is to notice the resonance behavior of ferrites. The precessing ferrite spins acquire a larger component transverse to the static magnetic fields when it is excited by one circular polarization compared to the other. When there is a loss mechanism to dissipate the energy of the spins, the polarization that excites a larger spin amplitude will dissipate more energy into the ferrite than the other polarization.

If we take the top view of the magnetic field distribution of the TE_{10} mode of a rectangular waveguide, the magnetic field is actually circularly polarized away from the midsection of the waveguide. The polarization changes sign for a $+z$ or a $-z$ propagating wave. Therefore, if a waveguide is loaded with ferrites away from the middle of the waveguide as shown, it will attenuate a TE_{10} mode by a different amount depending on the direction of propagation of the mode. This can be used as an isolator, and it is called the resonance isolator.

3.10.4 Spintronics

A fervent area of research related to spin dynamics is spintronics, the art of making electronic devices by exploiting spin physics. Some materials are highly magnetic because there are

Figure 3.26: Ferrite loading for resonance isolators.

many unpaired electrons in these materials. These unpaired electrons give these materials magnetic dipole moments, such as ferromagnetic materials. In ferromagnets, the magnetic dipoles of the same orientation cluster together to form microscopic domains that are random on a macroscopic scale. These domains can be aligned macroscopically by magnetization, making these materials into magnets. For instance, Fe, Co, Ni and their alloys have this property.

The conduction property of these materials can also be affected by remnant magnetic field in the domain, or externally applied magnetic field. In the presence of an external magnetic field, the energy levels of the down spins are much higher than the energy levels of the up spins. Hence, the down spins can be pushed into their conduction band while the up spins are relegated to the valence band. An flux of down-spin electrons can pass through such a medium, but not the up-spin electrons. This phenomenon can be used to generate giant magneto-resistance (GMR), and it has been used to generate magnetic storage devices.

Exercises for Chapter 3

Problem 3-1: Prove the identities in Equations (3.2.5a) (3.2.5b) of Section 3.1.

Problem 3-2: Consider the scattering problem involving a circular dielectric rod. The incident wave is described by a TM wave given as

$$E_z = E_0 J_0(k_\rho \rho) e^{ik_z z}.$$

Match boundary condition at the surface of the dielectric rod, and find the scattered field. Does the scattered field involve both TE and TM wave? Explain why.

Problem 3-3: If this world is dominated by left-handed people, we may have used left-hand rule in cross products rather than right-hand rule. How should Maxwell's equations be

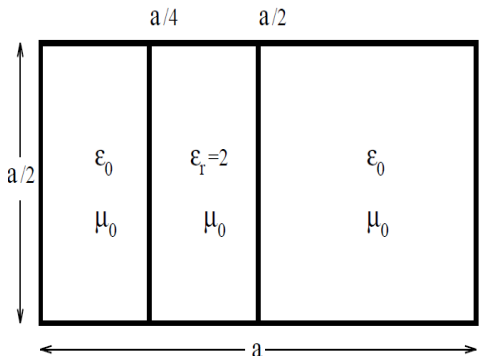


Figure 3.27: Problem 3-5

rewritten if left-hand rule is used instead? Would the law of electromagnetics be affected if left-hand rule is used instead?

Problem 3-4: Explain why if a general uniform waveguide is filled with an anisotropic material such that $\overline{\boldsymbol{\epsilon}}$ and $\overline{\boldsymbol{\mu}}$ are 3×3 tensors, and that ϵ_{xz}, ϵ_{yz}, ϵ_{zx}, ϵ_{zy}, μ_{xz}, μ_{yz}, μ_{zx}, and μ_{zy} are all not zero, then the waveguide does not have reflection symmetry. That is the waveguide in the mirrored world is not the original waveguide anymore even after a 180^o rotation.

Problem 3-5:

(a) Prove that if two matrices are the transpose of each other, they share the same eigenvalues. Do they share the same eigenvectors also? Proof that their eigenvectors are orthogonal to each other.

(b) Prove that Equation (3.3.7) is negative-transpose to Equation (3.3.6) in Section 3.3.

Problem 3-6: For the dielectric-slab-loaded rectangular waveguide shown:

(a) Assume that the dielectric slab is absent, plot the dispersion curve for the TE_{10} and TE_{01} modes above cut-off. That is plot k_z as a function of frequency.

(b) Now with the dielectric slab in place, the guidance properties of the aforementioned two modes will be perturbed. Write down the guidance conditions from which you can find the dispersion curves of the perturbed TE_{10} and TE_{01} modes. Define all the variables in the guidance conditions so that if you need to calculate these guidance conditions, you know how to.

(c) Write a computer program to solve for the roots of the guidance conditions and plot the dispersion curves for the perturbed modes. (Note: You can use a Muller root solver which is available in IMSL.)

Problem 3-7: Simplify Equation (3.8.24) of Section 3.5 for the $n = 0$ mode, or the axially symmetric mode.

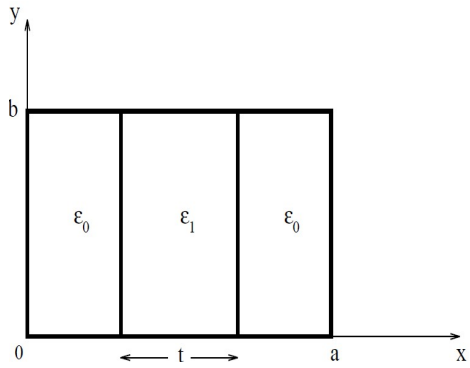


Figure 3.28: Problem 3-9

Problem 3-8: For an inhomogeneously filled waveguide, the equations governing the E_z and H_z components of the fields are

$$\mu\nabla_s \cdot \mu^{-1}\nabla_s E_z - ik_z\mu\nabla_s \cdot \left[\frac{ik_z}{\mu k_s^2}\nabla_s E_z + \frac{i\omega}{k_s^2}\nabla_s \times \mathbf{H}_z\right] + \omega^2\mu\epsilon E_z = 0,$$

$$\epsilon\nabla_s \cdot \epsilon^{-1}\nabla_s H_z - ik_z\epsilon\nabla_s \cdot \left[\frac{ik_z}{\epsilon k_s^2}\nabla_s H_z - \frac{i\omega}{k_s^2}\nabla_s \times \mathbf{E}_z\right] + \omega^2\mu\epsilon H_z = 0.$$

(a) For inhomogeneities which are piecewise constant, show that these two equations are coupled only at the discontinuities of the piecewise constant inhomogeneity.

(b) For a homogeneously filled waveguide, show that the two equations are decoupled from each other. What are the equations in this case?

Problem 3-9:
Find the guidance conditions for the symmetrically placed slab in the rectangular waveguide for the LSM and LSE modes. Simplify the expressions for the guidance conditions as much as possible.

Problem 3-10: If the equation for spin precession now has small loss terms, so that $P_z \to P_{z0}$, and $P_s \to 0$, when $t \to \infty$, namely,

$$\frac{d\mathbf{P}}{dt} = \boldsymbol{\omega}_0 \times \mathbf{P} - \mathbf{P}_s\frac{1}{T_2} - \hat{z}\,(P_z - P_{z0})\,\frac{1}{T_1},$$

where T_1 is the relaxation time for P_z, and T_2 is the relaxation for $\mathbf{P}_s$. Find the solution to the above equation. What is the steady solution if a time-harmonic RF $\mathbf{B}_1$ field as in (3.10.11) is applied to excite the system?

Problem 3-11: Analyze the problem in Subsection 3.10.1 when the wave is propagating in the negative z direction for Faraday rotation.

Bibliography

[1] N. Marcuvitz, ed., *Waveguide Handbook,* MIT Radiation Laboratory Series, vol, 10, McGraw-Hill, New York, 1951.

[2] L.G. Chambers, "Compilation of the propagation constants of an inhomogeneously-filled waveguide," *Br. J. Appl. Phys.*, 3, 19-21, 1952.

[3] R.E. Collin, *Field Theory of Guided Waves,* IEEE Press, Piscataway, NJ, 1991.

[4] R.E. Collin, *Foundation for Microwave Engineering,* IEEE Press, Piscataway, NJ, 2001.

[5] A. Yariv, *Optical Electronics,* Holt, Rinehart, and Winston, New York, 1985.

[6] T. Okoshi, *Opitcal Fibers,* Academic Press, New York, 1992.

[7] R.H. Sheikh and M.W. Gunn, "Wave Propagation in a Rectangular Waveguide Inhomogeneously Filled with Semiconductors (Correspondence)," *IEEE Trans. Micro. Theory and Techniques*, vol. MTT-19, no. 2, pp. 117-121, 1968.

[8] R. Feynman, R.B. Leighton, and M.L. Sands, *The Feynman Lectures on Physics*, vol. I, Chapter 52, Addison-Wesley Publishing Co., 1965.

[9] W. C. Chew and M. Nasir, "A variational analysis of anisotropic, inhomogeneous dielectric waveguides," *IEEE Trans. Microwave Theory Techniques*, vol. 37, no. 4, pp. 661-668, Apr. 1989.

[10] J.F. Lee, D.K. Sun, and Z.J. Cendes, "Full wave analysis of dielectric waveguides using tangential vector finite elements," *IEEE Trans. Microwave Theory and Techniques,* vol. 39, no. 8, pp.1262-1271, August 1991.

[11] J. Jin, *The finite element method in electromagnetics*, John Wiley & Sons, Inc., New York, 1993.

[12] P.H. Vartanian, W.P. Ayres, and A.L. Helgesson, "Propagation in Dielectric Slab Loaded Rectangular Waveguide," *IRE Trans. Micro. Theory Tech.*, vol. 6, no. 4, pp. 215-222, April 1958.

[13] B. Lax, K.J.B Utkm, and L.M. Roth, "Ferrite phase shifters in rectangular waveguide," *J. Appl. Phys.*, vol. 25, pp. 1413-1421, November, 1954.

[14] A.D. Bresler, "On the TE_{n0} modes of a ferrite slab loaded rectangular waveguide and the associated thermodynamic paradox," *IEEE Trans. Microwave Theory and Tech.* vol. 8, no. 1, pp. 81-95, Jan 1960.

[15] W.C. Chew, *Waves and Fields in Inhomogeneous Media,* Van Nostrand Reinhold, New York, 1990, reprinted, IEEE Press, Piscataway, NJ, 1995.

[16] S. Amari, R. Vahldieck, J. Bornemann and P. Leuchtmann, "Propagation in a circular waveguide periodically loaded with thick dielectric disks," *IEEE MTT-S Int. Microwave Symp. Dig.*, pp. 1535-1538, Baltimore, USA, June 1998.

[17] W. Li, M.L. Gong, Y.Y. Wei, H.Q. Xie, "The dispersive properties of a dielectric- rod loaded waveguide immersed in a magnetized annular plasma," *Chinese Phys.*, 13, 54-59, 2004.

[18] B. Lax, "The Status of Microwave Applications of Ferrites and Semiconductors," *IRE Trans. Micro. Theory Tech.*, vol. 6, no. 1, pp. 5-18, Jan 1958.

[19] L. Zhou and L. E. Davis, "Finite element method with edge elements for waveguides loaded with ferrite magnetized in arbitrary direction," *IEEE Trans. Microwave Theory and Techniques,* vol. 44, no. 6, pp. 809-815, June 1996.

[20] B. Lax and K.J. Button, *Microwave Ferrites and Ferrimagnetics,* McGraw-Hill, NY, 1962.

[21] K.G. Budden, *Radio Waves in the Ionosphere,* Cambridge University Press, Cambridge, UK, 1961.

Chapter 4

Coupling of Waveguides and Cavities

Once we have a waveguide or a cavity, it is important to know how to couple energy into it. Energy can be coupled into a waveguide by use of a probe, or an aperture, or simply just by connecting one waveguide to another. We will first study the excitation of modes in a waveguide by the use of a probe. We will develop the integral equation from which such a problem can be solved exactly, as well as calculating the input impedance by a variational formula. Such method can also be applied as well to cavity coupling, and coupling of electromagnetic energy into free space as in antennas.

To study aperture coupling, we will also discuss the pertinent equivalent principle needed. Such problems have also been addressed in [1, 2].

4.1 Excitation of Waveguides by a Probe

We will study the coupling of modes from a coaxial line to a waveguide as shown in Figure 4.1. Coaxial line is quite prevalent, and its characteristics can be easily understood via transmission line theory. This coupling problem is important in understanding the transfer of power from a coaxial cable to a waveguide system. By the proper adjustment of d and l, we can cause almost all the power from the coaxial cable to be transferred to the waveguide. From a transmission line theory viewpoint, the coupling to the waveguide is reflected in the transmission line being terminated with a load. The load can be changed by the proper adjustment of the dimension and location of the probe until a matched load is arrived at.

4.1.1 Derivation of the Equivalent Problem and the Integral Equation

In the waveguide volume enclosed by the surface S, the electric field satisfies the following vector wave equation

$$\nabla \times \nabla \times \mathbf{E}(\mathbf{r}) - k^2 \mathbf{E}(\mathbf{r}) = 0. \tag{4.1.1}$$

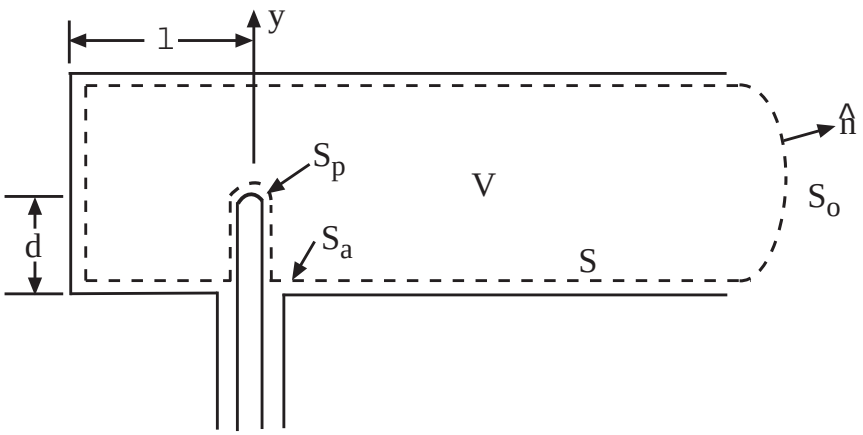


Figure 4.1: Excitation of a cylindrical, hollow waveguide by a probe.

A dyadic Green's function is defined to be a solution to the following equation, i.e.,

$$\nabla \times \nabla \times \overline{\mathbf{G}}(\mathbf{r}, \mathbf{r}') - k^2 \overline{\mathbf{G}}(\mathbf{r}, \mathbf{r}') = \overline{\mathbf{I}} \delta(\mathbf{r} - \mathbf{r}'). \tag{4.1.2}$$

Dot-multiplying (4.1.1) by $\overline{\mathbf{G}}(\mathbf{r}, \mathbf{r}')$ and (4.1.2) by $\mathbf{E}(\mathbf{r})$, upon subtraction and integration over V, we have

$$\int_V d\mathbf{r} [\mathbf{E}(\mathbf{r}) \cdot \nabla \times \nabla \times \overline{\mathbf{G}}(\mathbf{r}, \mathbf{r}') - \nabla \times \nabla \times \mathbf{E}(\mathbf{r}) \cdot \overline{\mathbf{G}}(\mathbf{r}, \mathbf{r}')] = \mathbf{E}(\mathbf{r}'). \tag{4.1.3}$$

Using

$$\begin{aligned} &\nabla \cdot [-\mathbf{E}(\mathbf{r}) \times \nabla \times \overline{\mathbf{G}}(\mathbf{r}, \mathbf{r}') - (\nabla \times \mathbf{E}(\mathbf{r})) \times \overline{\mathbf{G}}(\mathbf{r}, \mathbf{r}')] \\ &= \mathbf{E}(\mathbf{r}) \cdot \nabla \times \nabla \times \overline{\mathbf{G}}(\mathbf{r}.\mathbf{r}') - \nabla \times \nabla \times \mathbf{E}(\mathbf{r}) \cdot \overline{\mathbf{G}}(\mathbf{r}, \mathbf{r}'), \end{aligned} \tag{4.1.4}$$

and the divergence theorem, then

$$\mathbf{E}(\mathbf{r}') = -\int_S dS \hat{n} \cdot [\mathbf{E}(\mathbf{r}) \times \nabla \times \overline{\mathbf{G}}(\mathbf{r}, \mathbf{r}') + (\nabla \times \mathbf{E}(\mathbf{r})) \times \overline{\mathbf{G}}(\mathbf{r}, \mathbf{r}')], \quad \mathbf{r}' \in V. \tag{4.1.5}$$

The above can also be written as

$$\mathbf{E}(\mathbf{r}) = -\int_S dS' [\hat{n}' \times \mathbf{E}(\mathbf{r}') \cdot \nabla' \times \overline{\mathbf{G}}(\mathbf{r}', \mathbf{r}) + i\omega\mu \hat{n}' \times \mathbf{H}(\mathbf{r}') \cdot \overline{\mathbf{G}}(\mathbf{r}', \mathbf{r})], \quad \mathbf{r} \in V. \tag{4.1.6}$$

The above is actually a statement of Huygens' principle for vector electromagnetic field: Given the knowledge of tangential $\mathbf{E}$ and $\mathbf{H}$ fields on the closed surface of a volume V, the field is known everywhere inside V.[1]

[1] In (4.1.3), if $\mathbf{r}'$ is outside V, the right-hand side will evaluate to zero, leading to the left-hand side of (4.1.6) to be zero. This identity is known as the extinction theorem.

Furthermore, we can define the electric dyadic Green's function that will generate an electric field from an electric point current source. We will label such a Green's function $\overline{\mathbf{G}}_e(\mathbf{r},\mathbf{r}')$ satisfying the requisite boundary condition on the waveguide wall as shown in Figure 4.2. Similarly, we can define a magnetic dyadic Green's function that generates a magnetic field from a magnetic point current source, and label it $\overline{\mathbf{G}}_m(\mathbf{r},\mathbf{r}')$.

Moreover, using the reciprocity relations, it can be shown that [3, p. 32] [see also Problem 4-1]

$$[\overline{\mathbf{G}}_e(\mathbf{r}',\mathbf{r})]^t = \overline{\mathbf{G}}_e(\mathbf{r},\mathbf{r}'), \tag{4.1.7a}$$

$$[\nabla' \times \overline{\mathbf{G}}_e(\mathbf{r}',\mathbf{r})]^t = \nabla \times \overline{\mathbf{G}}_m(\mathbf{r},\mathbf{r}') \tag{4.1.7b}$$

Hence, we can rewrite the above as

$$\mathbf{E}(\mathbf{r}) = -\int_S dS'[\nabla \times \overline{\mathbf{G}}_m(\mathbf{r},\mathbf{r}') \cdot \hat{n}' \times \mathbf{E}(\mathbf{r}') + i\omega\mu\overline{\mathbf{G}}_e(\mathbf{r},\mathbf{r}') \cdot \hat{n}' \times \mathbf{H}(\mathbf{r}')], \quad \mathbf{r} \in V. \tag{4.1.8}$$

At this point, we have not specified the boundary conditions to be satisfied by the above dyadic Green's functions except that they are solutions of (4.1.2). A convenient choice for the electric dyadic Green's function for the waveguide is that it satisfies the boundary condition on the waveguide wall that

$$\hat{n} \times \overline{\mathbf{G}}_e(\mathbf{r},\mathbf{r}') \cdot \mathbf{a} = 0, \qquad \mathbf{r} \in \text{wall}, \tag{4.1.9}$$

where $\mathbf{a}$ is an arbitrary vector. In other words,

$$\hat{n} \times \overline{\mathbf{G}}_e(\mathbf{r},\mathbf{r}') = 0, \qquad \mathbf{r} \in \text{wall}, \tag{4.1.10}$$

Also, $\overline{\mathbf{G}}_m(\mathbf{r},\mathbf{r}')$ generates the magnetic field inside a waveguide due to a magnetic current source. Its curl produces the electric field with zero tangential component on the waveguide wall, or

$$\hat{n} \times \nabla \times \overline{\mathbf{G}}_m(\mathbf{r},\mathbf{r}') = 0, \qquad \mathbf{r} \in \text{wall}, \tag{4.1.11}$$

Consequently, it is seen that the electric field thus generated in (4.1.8) has zero tangential component on the waveguide wall. Hence, $\hat{n}' \times \mathbf{E}(\mathbf{r}') = 0$ on the waveguide wall except for S_a. Also, $\overline{\mathbf{G}}_e(\mathbf{r},\mathbf{r}') \cdot \hat{n}' \times \mathbf{H}(\mathbf{r}')$ is zero on the waveguide wall by reciprocity or by taking the transpose of this expression and applying (4.1.9). In other words,

$$\overline{\mathbf{G}}_e(\mathbf{r},\mathbf{r}') \cdot \hat{n}' \times \mathbf{H}(\mathbf{r}') = \hat{n}' \times \mathbf{H}(\mathbf{r}') \cdot \overline{\mathbf{G}}_e(\mathbf{r}',\mathbf{r}) = -\mathbf{H}(\mathbf{r}') \cdot \hat{n}' \times \overline{\mathbf{G}}_e(\mathbf{r}',\mathbf{r}) = 0, \qquad \mathbf{r}' \in \text{wall}. \tag{4.1.12}$$

Hence, the second integral in (4.1.8) is nonzero only on S_p. Therefore, (4.1.8) can be rewritten as

$$\mathbf{E}(\mathbf{r}) = -\int_{S_a} dS' \nabla \times \overline{\mathbf{G}}_m(\mathbf{r},\mathbf{r}') \cdot \hat{n}' \times \mathbf{E}(\mathbf{r}') - i\omega\mu \int_{S_p} dS' \overline{\mathbf{G}}_e(\mathbf{r},\mathbf{r}') \cdot \hat{n}' \times \mathbf{H}(\mathbf{r}'), \quad \mathbf{r} \in V. \tag{4.1.13}$$

The integral over S_0 in Figure 4.1 can be made to vanish by taking S_0 to infinity and introducing an infinitesimal amount of loss. Also, $-\hat{n}' \times \mathbf{H}(\mathbf{r}')$ can be identified as $\mathbf{J}_p(\mathbf{r}')$,

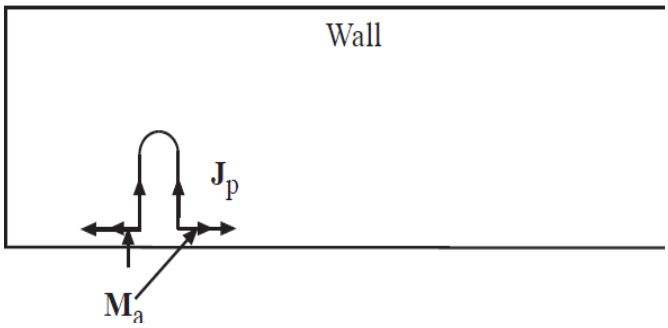


Figure 4.2: Equivalent sources for the probe excitation of a waveguide.

the surface current on the probe, and $\hat{n}' \times \mathbf{E}(\mathbf{r}')$ can be identified as $\mathbf{M}_a(\mathbf{r}')$, as equivalent magnetic current impressed on S_a. Rewriting (4.1.13), it becomes

$$\mathbf{E}(\mathbf{r}) = -\int_{S_a} dS' \nabla \times \overline{\mathbf{G}}_m(\mathbf{r}, \mathbf{r}') \cdot \mathbf{M}_a(\mathbf{r}') + i\omega\mu \int_{S_p} dS' \overline{\mathbf{G}}_e(\mathbf{r}, \mathbf{r}') \cdot \mathbf{J}_p(\mathbf{r}'). \tag{4.1.14}$$

If $\mathbf{M}_a(\mathbf{r}')$ is assumed known, then using $\hat{n} \times \mathbf{E}(\mathbf{r}) = 0$ on S_p as the additional boundary condition, an integral equation can be set up

$$i\omega\mu\hat{n} \times \int_{S_p} dS' \overline{\mathbf{G}}_e(\mathbf{r}, \mathbf{r}') \cdot \mathbf{J}_p(\mathbf{r}') = \hat{n} \times \int_{S_a} dS' \nabla \times \overline{\mathbf{G}}_m(\mathbf{r}, \mathbf{r}') \cdot \mathbf{M}_a(\mathbf{r}'), \tag{4.1.15}$$

from which $\mathbf{J}_p(\mathbf{r}')$ can be solved for. In Equation (4.1.14), the problem in Figure 4.1 is replaced with an equivalent problem consisting of impressed current $\mathbf{J}_p$ on S_p and $\mathbf{M}_a$ on S_a. These are impressed current because they are sources impressed in the space V radiating via the waveguide dyadic Green's function.

4.1.2 Generalization to Other Structures

The above theory is quite general. It obviously applies to arbitrarily shaped waveguides or cavities. When applied to an arbitrarily shaped structure as shown in Figure 4.3, the onus is on finding the dyadic Green's function. However, one can also use free-space dyadic Green's function, but at the expense of adding more unknowns to the integral equation.

The above formulation can also be applied to a monopole antenna mounted on a ground plane driven by a coaxial cable from below. In this case, we can use a dyadic Green's function that satisfies the boundary condition of a metallic half space. This Green's function can be found easily using image theorem.

It can also be applied to a metallic antenna driven by a magnetic current source in free space. We can apply the free-space dyadic Green's function in this case. One can assume that S_a and S_p form a closed surface. In this case, there are impressed electric current on the surface S_p. But on the surface S_a, there would be both impressed magnetic and electric current.

In the above cases, there is a bounding surface at infinity, S_{inf} that has to be included. By use of the radiation condition, the contribution from this bounding surface can be shown to vanish.

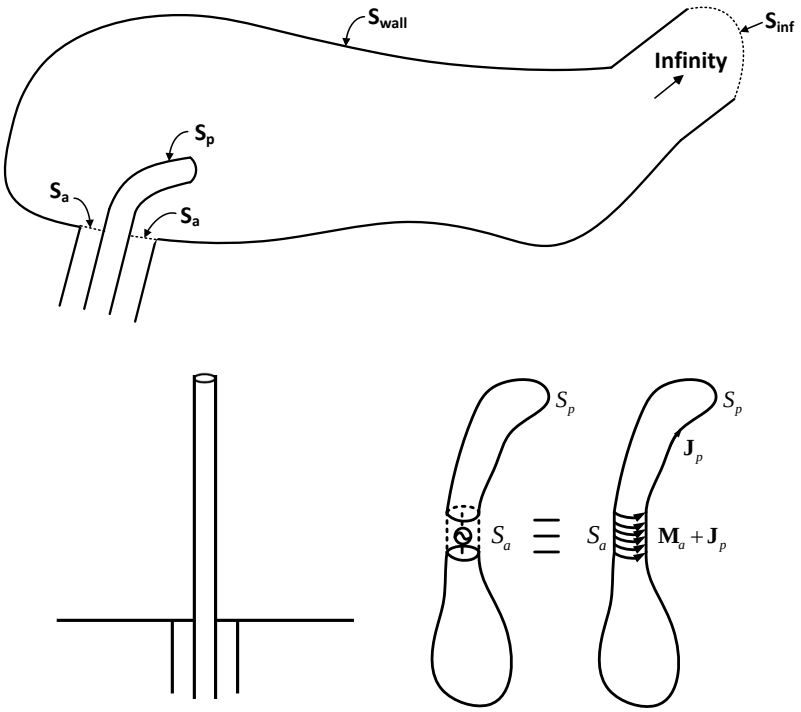


Figure 4.3: Excitation of a different structure with a magnetic current source. Top: A coaxial driven probe exciting an arbitrary waveguide. Bottom left: A monopole antenna driven by a coax via a ground plane. Bottom right: A general antenna driven by a voltage source represented by an equivalent magnetic current source.

4.2 Input Impedance of the Probe

The input impedance of a probe exciting a waveguide can be calculated. This yields information on how the position and length of the probe can be adjusted to arrive at the input

impedance we desire. For maximum transfer of the power, the input impedance should be matched to the characteristic impedance of the transmission line to minimize reflections.

4.2.1 Variational Expressions for Input Admittance

Variational expressions for the input impedance of an antenna has been discussed in [4–6]. Such expressions can also be used to find the input impedance of a probe inside the waveguide, as the two problems are very similar. The difference appears mainly in the Green's function. In an antenna, the source is radiating in free space, and hence, free-space Green's function usually suffices. However, in a waveguide, the free-space Green's function has to be replaced by the Green's function of the waveguide structure.

A variational expression for the input impedance or admittance of a source driven complex structure can be derived. The source can be either a voltage source or a current source. When it is a voltage source, it is a magnetic current ring, and when it is a current source, it is an electric dipole such as a Hertzian dipole or its equivalence. The case of a current source has been adequately discussed in [22]. The voltage source is usually modeled by a magnetic ring current, a ribbon current, or a magnetic frill. The case of a magnetic current exciting an antenna structure has been discussed in [6]. However, our discussion here is in accordance to [7], which is different from the previous treatment on the subject. However, the variational formula for a magnetic current source driven antenna seems to have been missed until presented in [7].

In any case, the input impedance of the structure is predominantly determined by the induced current on the structure. A variational expression has the advantage of yielding a second order error for the input impedance when the error of the current on the structure is first order. The current on the complex structure can be roughly estimated, or solved for from an integral equation.

If a probe current produces a magnetic field $\mathbf{H}_p$, and the aperture magnetic current produces an electric field $\mathbf{E}_a$ and a magnetic field $\mathbf{H}_a$, then

$$-\langle \mathbf{M}_a, \mathbf{H}_T \rangle = -\int_{S_a} dS' \mathbf{M}_a(\mathbf{r}') \cdot \mathbf{H}_T(\mathbf{r}') = -\int_{S_a} dS' \hat{n}' \cdot (\mathbf{E}_a \times \mathbf{H}_T) \tag{4.2.1}$$

where $\mathbf{H}_T = \mathbf{H}_p + \mathbf{H}_a$. On the aperture S_a, we can assume that only the TEM mode of the coax is important. Therefore,

$$\mathbf{E}_a = \hat{\rho} E_0, \quad \mathbf{H}_T = \hat{\phi} H_0, \tag{4.2.2}$$

Substituting (4.2.2) into (4.2.1), we have

$$-\int_{S_a} dS' \hat{n} \cdot (\mathbf{E}_a \times \mathbf{H}_T) = \int_a^b d\rho' E_0 \int_0^{2\pi} \rho' d\phi' H_0 = VI. \tag{4.2.3}$$

In the above, the fields are represented in a local coordinate system, and $\hat{n}' = -\hat{z}$. Also, even though $\mathbf{H}_T$ may depart from the TEM mode field of a waveguide, by the mode orthogonality

theorem, only the TEM-mode component of $\mathbf{H}_T$ will contribute to the integral in Equation (4.2.3).

Does this mean that the higher-order modes in $\mathbf{H}_T$ do not contribute to the current I? No, it does not. The higher-order modes do contribute to the current I near the aperture, but their effect diminishes rapidly away from the aperture. Therefore, the current I in (4.2.3) is only the TEM component of the current, which can be easily related to the current in the rest of the coaxial cable by transmission line theory.

Consequently,

$$VI = -\langle \mathbf{M}_a, \mathbf{H}_T \rangle. \tag{4.2.4}$$

By letting $I = Y_{in}V$, we deduce that

$$Y_{in} = -\frac{\langle \mathbf{M}_a, \mathbf{H}_T \rangle}{V^2}. \tag{4.2.5}$$

The above is an expression for the input admittance of the probe assuming only the TEM mode in the coax, but it is not variational. To derive a variational expression, we write [7]

$$Y_{in} = \frac{-\langle \mathbf{M}_a, \mathbf{H}_T \rangle + \langle \mathbf{J}_p, \mathbf{E}_T \rangle}{V^2} \tag{4.2.6}$$

where $\mathbf{E}_T = \mathbf{E}_a + \mathbf{E}_p$, the total electric field produced both by $\mathbf{M}_a$ and $\mathbf{J}_p$. The above also falls under the category of the reaction formula for the input impedance of antennas. Notice that if $\mathbf{E}_T$ is exact, then the tangential component of $\mathbf{E}_T$ is zero on the probe surface and $\langle \mathbf{J}_p, \mathbf{E}_T \rangle$ would be zero. However, the second term in the numerator of (4.2.6) is required to make it a variational expression. In other words, first order error $\mathbf{J}_p$ will result in a second order error in Y_{in}. To prove that (4.2.6) is variational, we let $\mathbf{M}_a$ and V be known and hence fixed, and let

$$\mathbf{J}_p = \mathbf{J}_{pe} + \delta\mathbf{J}, \quad Y_{in} = Y_{ine} + \delta Y \tag{4.2.7}$$

where the subscript e stands for "exact". Cross-multiplying (4.2.6), and taking the first variation, we have

$$\delta Y V^2 = -\langle \mathbf{M}_a, \delta\mathbf{H} \rangle + \langle \delta\mathbf{J}, \mathbf{E}_{Te} \rangle + \langle \mathbf{J}_{pe}, \delta\mathbf{E} \rangle. \tag{4.2.8}$$

From reciprocity,

$$\langle \delta\mathbf{J}, \mathbf{E}_{Te} \rangle = -\langle \mathbf{M}_a, \delta\mathbf{H} \rangle + \langle \mathbf{J}_{pe}, \delta\mathbf{E} \rangle. \tag{4.2.9}$$

Then

$$\delta Y V^2 = 2\langle \delta\mathbf{J}, \mathbf{E}_{Te} \rangle. \tag{4.2.10}$$

Since

$$\langle \delta\mathbf{J}, \mathbf{E}_{Te} \rangle = 0, \tag{4.2.11}$$

because $\mathbf{E}_{TE}$ has no tangential components on the probe surface and $\delta\mathbf{J}$ is purely tangential on the probe surface, (4.2.10) implies that

$$\delta Y = 0. \tag{4.2.12}$$

As a result, the first variation in the admittance about the exact admittance Y_{ine} is zero. Equation (4.2.6) is a variational expression for the input admittance. Given $\mathbf{M}_a$, $\mathbf{J}_p$ and V with first order errors, the errors incurred in Y_{in} is of second order.

The variational nature of Equation (4.2.6) can be better appreciated if its quadratic nature is written more explicitly. To this end, it can be written as

$$Y_{in} = -\frac{\langle \mathbf{M}_a, \mathbf{H}_a \rangle}{V^2} - \frac{\langle \mathbf{M}_a, \mathbf{H}_p \rangle}{V^2} + \frac{\langle \mathbf{J}_p, \mathbf{E}_p \rangle}{V^2} + \frac{\langle \mathbf{J}_p, \mathbf{E}_a \rangle}{V^2} \tag{4.2.13}$$

where we assume $\mathbf{M}_a$, and hence, V, $\mathbf{H}_a$, and $\mathbf{E}_a$ are fixed. When $\mathbf{J}_p$ is varied, only the last three terms would vary. By reciprocity, $-\langle \mathbf{M}_a, \mathbf{H}_p \rangle = \langle \mathbf{J}_p, \mathbf{E}_a \rangle$, and the above becomes

$$Y_{in} = -\frac{\langle \mathbf{M}_a, \mathbf{H}_a \rangle}{V^2} + 2\frac{\langle \mathbf{J}_p, \mathbf{E}_a \rangle}{V^2} + \frac{\langle \mathbf{J}_p, \mathbf{E}_p \rangle}{V^2}. \tag{4.2.14}$$

Furthermore,

$$\begin{aligned} \langle \mathbf{J}_p, \mathbf{E}_p \rangle &= i\omega\mu \langle \mathbf{J}_p, \overline{\mathbf{G}}_e, \mathbf{J}_p \rangle \\ &= i\omega\mu \int\limits_{S_p} dS \mathbf{J}_p(\mathbf{r}) \cdot \int\limits_{S_p} dS' \overline{\mathbf{G}}_e(\mathbf{r}, \mathbf{r}') \cdot \mathbf{J}_p(\mathbf{r}'). \end{aligned} \tag{4.2.15}$$

The double commas in the above implies that there is a double integration in the inner product. The above is analogous to $\mathbf{a}^t \cdot \overline{\mathbf{A}} \cdot \mathbf{a}$ in linear algebra. When it is used in (4.2.14), it becomes

$$Y_{in} = -\frac{\langle \mathbf{M}_a, \mathbf{H}_a \rangle}{V^2} + 2\frac{\langle \mathbf{J}_p, \mathbf{E}_a \rangle}{V^2} + i\omega\mu \frac{\langle \mathbf{J}_p, \overline{\mathbf{G}}_e, \mathbf{J}_p \rangle}{V^2}. \tag{4.2.16}$$

The above is clearly quadratic and has a stationary point about the exact solution.

For an exact $\mathbf{E}_T$, $\langle \mathbf{J}_p, \mathbf{E}_{Te} \rangle = 0$ and (4.2.6) reduces to (4.2.5) again. Since $\mathbf{H}_T = \mathbf{H}_a + \mathbf{H}_p$, we have from (4.2.5)

$$Y_{in} = -\frac{\langle \mathbf{M}_a, \mathbf{H}_a \rangle}{V^2} - \frac{\langle \mathbf{M}_a, \mathbf{H}_p \rangle}{V^2} \tag{4.2.17}$$

for exact solutions.

Furthermore, $-\langle \mathbf{M}_a, \mathbf{H}_p \rangle = \langle \mathbf{J}_p, \mathbf{E}_a \rangle = -\langle \mathbf{J}_p, \mathbf{E}_p \rangle$ from reciprocity and that $\langle \mathbf{J}_p, \mathbf{E}_T \rangle = 0$ for exact solutions. Hence, (4.2.17) becomes

$$Y_{in} = -\frac{\langle \mathbf{M}_a, \mathbf{H}_a \rangle}{V^2} - \frac{\langle \mathbf{J}_p, \mathbf{E}_p \rangle}{V^2}. \tag{4.2.18}$$

If $\mathbf{M}_a$ is assumed real, the first term is the complex conjugate of the complex power $\langle \mathbf{M}_a, \mathbf{H}_a^* \rangle$ due to $\mathbf{M}_a$ alone and can be related to the gap capacitance at the base of the probe. The second term in (4.2.18) is the complex conjugate of the complex power $\langle \mathbf{E}_p, \mathbf{J}_p^* \rangle$ due to $\mathbf{J}_p$ alone. Hence, it is due to the probe admittance. Since the terms in (4.2.18) are additive, the gap capacitor is in parallel connection with the probe admittance.

It is to be reminded that in the use of (4.2.6), $\mathbf{H}_T$ and $\mathbf{E}_T$ are to be calculated from $\mathbf{M}_a$ and $\mathbf{J}_p$. In particular,

$$\mathbf{H}_p(\mathbf{r}) = \int\limits_{S_p} dS' \nabla \times \overline{\mathbf{G}}_e(\mathbf{r}, \mathbf{r}') \cdot \mathbf{J}_p(\mathbf{r}'), \tag{4.2.19}$$

$$\mathbf{E}_a(\mathbf{r}) = -\int\limits_{S_a} dS' \nabla \times \overline{\mathbf{G}}_m(\mathbf{r}, \mathbf{r}') \cdot \mathbf{M}_a(\mathbf{r}'). \tag{4.2.20}$$

4.2.2 Rayleigh-Ritz Method

The Rayleigh-Ritz method is named after Lord Rayleigh [17], a prodigious English scientist, and Walter Ritz [18], a Swiss mathematician. It is extremely useful in solving complex physical problems. It seems that many physical phenomena are always described by an equation which corresponds to the minimization of a certain quantity. So instead of solving the equation direction, one can attempt to minimize the corresponding quantity instead.

This method is extremely useful if we have an expression for Y_{in} which will have Y_{ine} (exact Y_{in}) as the lower bound or the upper bound to all the approximate Y_{in}. (However, Y_{in} is actually a complex number, but for the sake of the ease for discussion, we will assume that Y_{in} is real.) Without loss of generality, let us discuss the lower bound case. For example, if

$$Y_{in} = f(\mathbf{J}_p) \geq Y_{ine}, \tag{4.2.21}$$

and the equality is satisfied only if $\mathbf{J}_p = \mathbf{J}_{pe}$, an optimal value of Y_{in} can be obtained even with an approximate $\mathbf{J}_p$. We can let

$$\mathbf{J}_p = \sum_{n=1}^{N} a_n \mathbf{J}_n, \tag{4.2.22}$$

where $\mathbf{J}_n$ is a set of basis functions with which an arbitrary $\mathbf{J}_p$ can be approximated fairly well. The coefficients a_n's are yet to be determined to give the best approximation to $\mathbf{J}_p$. The Rayleigh-Ritz procedure provides a systematic way to determine the optimal values of a_n's so as to best determine Y_{in} from (4.2.21). If we substitute (4.2.22) into (4.2.21), then

$$Y_{ina} = f\left(\sum_{n=1}^{N} a_n \mathbf{J}_n\right) > Y_{ine}. \tag{4.2.23}$$

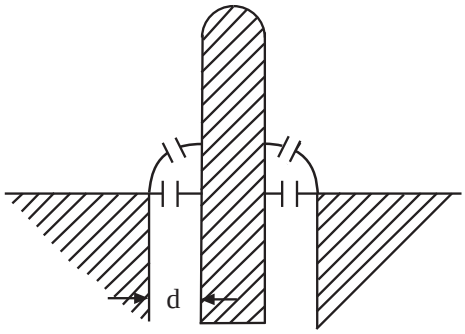


Figure 4.4: Gap capacitances can be important in the input impedance of a probe.

The best choice of a_n's will be those that minimizes the number Y_{ina}. In other words, from Figure 4.6. the optimal values of a_n's are those that would make Y_{ina} stationary. Hence,

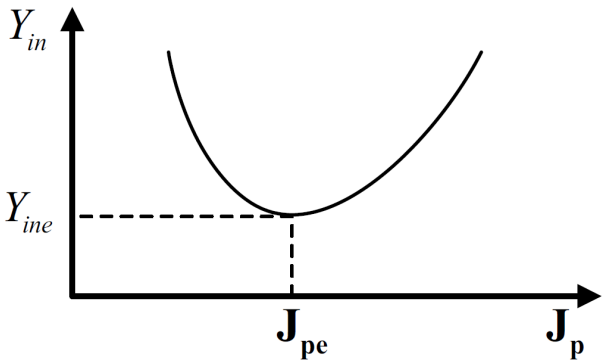


Figure 4.5: A pictorial representation of a stationary point in a multidimensional space.

there are N equations

$$\frac{\partial f\left(\sum_{n=1}^{N} a_n \mathbf{J}_n\right)}{\partial a_i} = 0, \qquad i = 1, \ldots, N, \tag{4.2.24}$$

from which we can solve for the optimal a_i's, the a_{io}'s.

This concept may not work as well when Y_{in} is a complex function, or when the stationary point is not a global minimum or maximum, but a saddle point instead. However, for these cases, the Rayleigh-Ritz procedure converges despite as we shall explain later.

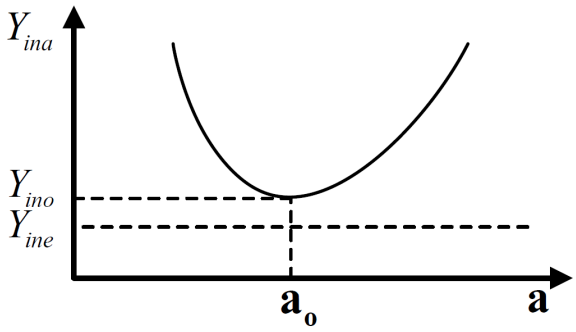


Figure 4.6: The minimum Y_{ino} achieved by the Rayleigh-Ritz procedure compared to the exact minimum, Y_{ine}.

Equation (4.2.14) is a variational expression when the solution is varied about $\mathbf{J}_{pe}$. Hence,

the Rayleigh-Ritz procedure can be used to find the optimal Y_{in}. To do so, we assume

$$\mathbf{J}_p = \sum_{n=1}^{N} a_n \mathbf{J}_n. \tag{4.2.25}$$

Then (4.2.16) becomes

$$Y_{in} = -\frac{\langle \mathbf{M}_a, \mathbf{H}_a \rangle}{V^2} + 2\frac{\sum\limits_n a_n \langle \mathbf{J}_n, \mathbf{E}_a \rangle}{V^2} + \frac{\sum\limits_n \sum\limits_{n'} a_n a_{n'} i\omega\mu \langle \mathbf{J}_n, \overline{\mathbf{G}}_e, \mathbf{J}_{n'} \rangle}{V^2}. \tag{4.2.26}$$

The above is of the form

$$Y_{in} = -\frac{\langle \mathbf{M}_a, \mathbf{H}_a \rangle}{V^2} + 2\frac{\mathbf{a}^t \cdot \mathbf{e}}{V^2} + \frac{\mathbf{a}^t \cdot \overline{\mathbf{A}} \cdot \mathbf{a}}{V^2} \tag{4.2.27}$$

where $[\mathbf{e}]_n = \langle \mathbf{J}_n, \mathbf{E}_a \rangle$, $\left[\overline{\mathbf{A}}\right]_{nn'} = i\omega\mu \langle \mathbf{J}_n, \overline{\mathbf{G}}_e, \mathbf{J}_{n'} \rangle$, and $[a]_n = a_n$. The matrix $\overline{\mathbf{A}}$ is the matrix representation of the dyadic Green's function $i\omega\mu\overline{\mathbf{G}}_e$, while the vector $\mathbf{e}$ is the vector representation of $\mathbf{E}$.

Taking the first variation of the above about $\mathbf{a}_o$, the optimal $\mathbf{a}$, we have

$$\delta Y_{in} = 2\frac{\delta\mathbf{a}^t \cdot \mathbf{e}}{V^2} + 2\frac{\delta\mathbf{a}^t \cdot \overline{\mathbf{A}} \cdot \mathbf{a}_o}{V^2}. \tag{4.2.28}$$

The first variation δY_{in} will vanish if

$$\overline{\mathbf{A}} \cdot \mathbf{a}_o = -\mathbf{e}. \tag{4.2.29}$$

The above could be solved to yield the optimal $\mathbf{a}_o$ which can in turn be used to find $\mathbf{J}_p$ in (4.2.25). Y_{in0} can also be found by the direct substitution of (4.2.29) into (4.2.27) yielding

$$Y_{in0} = -\frac{\langle \mathbf{M}_a, \mathbf{H}_a \rangle}{V^2} + \frac{\mathbf{a}_o^t \cdot \mathbf{e}}{V^2}. \tag{4.2.30}$$

Equation (4.2.29) is also more directly obtained by solving the following integral equation which follows from $-\hat{n} \times \mathbf{E}_a = \hat{n} \times \mathbf{E}_p$ on the probe surface

$$-\hat{n} \times \mathbf{E}_a = i\omega\mu\hat{n} \times \int\limits_{S_p} dS' \overline{\mathbf{G}}_e(\mathbf{r}, \mathbf{r}') \cdot \mathbf{J}_p(\mathbf{r}'). \tag{4.2.31}$$

The use of Galerkin's method in solving (4.2.31) yields (4.2.29).

The reason why the Rayleigh-Ritz procedure converges to the exact solution is because most of the variational expressions we work with are quadratic in nature. This quadratic nature is the generalization of the quadratic expression in one dimension such as $ax^2 + bx + c$ to higher dimensions. For example, (4.2.27) is a quadratic expression in the variable $\mathbf{a}$ while (4.2.13) is quadratic in the variable $\mathbf{J}_p$. The space for the inner product in (4.2.27) is an approximate finite dimensional space, while the space for the inner product in (4.2.13) is infinite dimensional. All quadratic expressions have only one stationary point. Consequently, as one increase the dimension of the approximate finite dimensional space, the stationary point will approach that of the infinite dimensional space.

4.2.3 Mode Matching Method—A Tour de Force Calculation

In the previous section, the equation for the input admittance, Equation (4.2.5) is valid only when the field at S_a can be approximated by only TEM modes. This is not true in general. A more accurate analysis of the input impedance of the probe requires the use of the mode-matching method [7]. This will allow us to use higher order modes in the coaxial waveguide. It also allows us to control the accuracy of the calculation as much as we want to. When very high accuracy solution is needed, we just need to add more modes in the solution procedure.

Inside the coaxial waveguide, the field is assumed to be

$$\mathbf{E}(\mathbf{r}) = \mathbf{E}_0(\mathbf{r}_s)e^{ikz} + \sum_{l=0}^{L-1} \Gamma_l \mathbf{E}_l(\mathbf{r}_s)e^{-ik_{lz}z}, \tag{4.2.32}$$

where $\mathbf{E}_0(\mathbf{r}_s)$ is the field distribution of the TEM mode, and $\mathbf{E}_m(\mathbf{r}_s)$, $m > 0$ are the fields of the higher order modes which are evanescent. The corresponding magnetic field is

$$\mathbf{H}(\mathbf{r}) = \mathbf{H}_0(\mathbf{r}_s)e^{ikz} - \sum_{l=0}^{L-1} \Gamma_l \mathbf{H}_l(\mathbf{r}_s)e^{-ik_{lz}z}. \tag{4.2.33}$$

Hence, the total electric field at the aperture in accordance with Equation (4.2.32) is

$$\mathbf{E}_a(\mathbf{r}_s) = \mathbf{E}_0(\mathbf{r}_s) + \sum_{l=0}^{L-1} \Gamma_l \mathbf{E}_l(\mathbf{r}_s). \tag{4.2.34}$$

The corresponding magnetic current is

$$\mathbf{M}_a(\mathbf{r}_s) = \mathbf{E}_a \times \hat{n} = \mathbf{M}_0 + \sum_{l=0}^{L-1} \Gamma_l \mathbf{M}_l \tag{4.2.35}$$

where $\mathbf{M}_l = \mathbf{E}_l \times \hat{n}$.

The electric field in the waveguide region in the volume V due to the magnetic current $\mathbf{M}_a$ is given by

$$\begin{aligned} \mathbf{E}_a(\mathbf{r}) &= \int_{S_a} dS' \nabla \times \overline{\mathbf{G}}_m(\mathbf{r}, \mathbf{r}') \cdot \mathbf{M}_a(\mathbf{r}') \\ &= \int_{S_a} dS' \nabla \times \overline{\mathbf{G}}_m(\mathbf{r}, \mathbf{r}') \cdot \mathbf{M}_0(\mathbf{r}') \\ &+ \sum_{l=0}^{L-1} \Gamma_l \int_{S_a} dS' \nabla \times \overline{\mathbf{G}}_m(\mathbf{r}, \mathbf{r}') \cdot \mathbf{M}_l(\mathbf{r}'), \quad \mathbf{r} \in V, \end{aligned} \tag{4.2.36}$$

where $\overline{\mathbf{G}}_m(\mathbf{r}, \mathbf{r}')$ is the magnetic-type dyadic Green's function.

$\mathbf{E}_a$ from (4.2.36) can be substituted into (4.2.31) to yield

$$\begin{aligned} -\hat{n} \times \int_{S_a} dS' \nabla \times \overline{\mathbf{G}}_m(\mathbf{r}, \mathbf{r}') \cdot \mathbf{M}_0(\mathbf{r}') \\ - \sum_{l=0}^{L-1} \Gamma_l \hat{n} \times \int_{S_a} dS' \nabla \times \overline{\mathbf{G}}_m(\mathbf{r}, \mathbf{r}') \cdot \mathbf{M}_l(\mathbf{r}') \\ = i\omega\mu \hat{n} \times \int_{S_p} dS' \overline{\mathbf{G}}_e(\mathbf{r}, \mathbf{r}') \cdot \mathbf{J}_p(\mathbf{r}'). \end{aligned} \tag{4.2.37}$$

Expanding $\mathbf{J}_p$ as in (4.2.25), we obtain

$$\begin{aligned} -\hat{n} \times \int_{S_a} dS' \nabla \times \overline{\mathbf{G}}_m(\mathbf{r}, \mathbf{r}') \cdot \mathbf{M}_0(\mathbf{r}') \\ - \sum_{l=0}^{L-1} \Gamma_l \hat{n} \times \int_{S_a} dS' \nabla \times \overline{\mathbf{G}}_m(\mathbf{r}, \mathbf{r}') \cdot \mathbf{M}_l(\mathbf{r}') \\ = i\omega\mu \sum_{n=1}^{N} a_n \hat{n} \times \int_{S_p} dS' \overline{\mathbf{G}}_e(\mathbf{r}, \mathbf{r}') \cdot \mathbf{J}_n(\mathbf{r}'). \end{aligned} \tag{4.2.38}$$

Testing the above equation with $\hat{n} \times \mathbf{J}_m(\mathbf{r})$, $m = 1, \ldots, N$, we have

$$\begin{aligned} -\langle \mathbf{J}_m(\mathbf{r}), \nabla \times \overline{\mathbf{G}}_m(\mathbf{r}, \mathbf{r}'), \mathbf{M}_0(\mathbf{r}') \rangle \\ - \sum_{l=0}^{L-1} \Gamma_l \langle \mathbf{J}_m(\mathbf{r}), \nabla \times \overline{\mathbf{G}}_m(\mathbf{r}, \mathbf{r}'), \mathbf{M}_l(\mathbf{r}') \rangle \\ = i\omega\mu \sum_{n=1}^{N} a_n \langle \mathbf{J}_m(\mathbf{r}), \overline{\mathbf{G}}_e(\mathbf{r}, \mathbf{r}'), \mathbf{J}_n(\mathbf{r}') \rangle, \quad m = 1, \ldots, N. \end{aligned} \tag{4.2.39}$$

In the above, the notation

$$\langle \mathbf{f}(\mathbf{r}), \overline{\mathbf{G}}(\mathbf{r}, \mathbf{r}'), \mathbf{g}(\mathbf{r}') \rangle = \int_{S_p} dS \mathbf{f}(\mathbf{r}) \cdot \int_{S_p} dS' \overline{\mathbf{G}}(\mathbf{r}, \mathbf{r}') \cdot \mathbf{g}(\mathbf{r}'). \tag{4.2.40}$$

Notice that the total field inside the waveguide is given by

$$\mathbf{E}_T = \mathbf{E}_a + \mathbf{E}_p \tag{4.2.41}$$

where $\mathbf{E}_p$ is the field produced by the probe current, i.e.,

$$\mathbf{E}_p = i\omega\mu \int_{S_p} dS' \overline{\mathbf{G}}_e(\mathbf{r}, \mathbf{r}') \cdot \mathbf{J}_p(\mathbf{r}'). \tag{4.2.42}$$

Because of the electric dyadic Green's function used here, $\hat{n} \times \mathbf{E}_p = 0$ on S_a. Therefore,

$$\hat{n} \times \mathbf{E}_T = \hat{n} \times \mathbf{E}_a \quad \text{on } S_a. \tag{4.2.43}$$

Consequently, the total field given by (4.2.41) calculated via the use of (4.2.36), and (4.2.43) satisfies the boundary condition that $\hat{n} \times \mathbf{E}_T$ is continuous at S_a from the coaxial waveguide to the main waveguide.

Next, we need to impose the boundary condition that the tangential component of the magnetic field is continuous across S_a. To this end, we find $\mathbf{H}_T = \mathbf{H}_a + \mathbf{H}_p$ where

$$\mathbf{H}_a(\mathbf{r}) = i\omega\epsilon \int_{S_a} dS' \overline{\mathbf{G}}_m(\mathbf{r}, \mathbf{r}') \cdot \mathbf{M}_a(\mathbf{r}'), \tag{4.2.44a}$$

$$\mathbf{H}_p(\mathbf{r}) = \int_{S_p} dS' \nabla \times \overline{\mathbf{G}}_e(\mathbf{r}, \mathbf{r}') \cdot \mathbf{J}_p(\mathbf{r}'). \tag{4.2.44b}$$

On expanding $\mathbf{J}_p$ as in (4.2.25), and $\mathbf{M}_a$ as in (4.2.35), we have

$$\begin{aligned} \mathbf{H}_T(\mathbf{r}) = {} & i\omega\epsilon \int_{S_a} dS' \overline{\mathbf{G}}_m(\mathbf{r}, \mathbf{r}') \cdot \mathbf{M}_0(\mathbf{r}') \\ & + i\omega\epsilon \sum_{l=0}^{L-1} \Gamma_l \int_{S_a} dS' \overline{\mathbf{G}}_m(\mathbf{r}, \mathbf{r}') \cdot \mathbf{M}_l(\mathbf{r}') \\ & + \sum_{n=1}^{N} a_n \int_{S_p} dS' \nabla \times \overline{\mathbf{G}}_e(\mathbf{r}, \mathbf{r}') \cdot \mathbf{J}_n(\mathbf{r}'). \end{aligned} \tag{4.2.45}$$

The boundary condition requires that the tangential components of $\mathbf{H}$ in (4.2.33) and (4.2.45) be continuous. Equating (4.2.33) and (4.2.45), and testing the result with $\mathbf{M}_{l'}(\mathbf{r})$, $l' = 0, \ldots, L-1$, we have

$$\begin{aligned} \delta_{0l'}\lambda_0 - \Gamma_{l'}\lambda_{l'} = {} & i\omega\epsilon \langle \mathbf{M}_{l'}(\mathbf{r}), \overline{\mathbf{G}}_m(\mathbf{r}, \mathbf{r}'), \mathbf{M}_0(\mathbf{r}') \rangle \\ & + i\omega\epsilon \sum_{l=0}^{L-1} \Gamma_l \langle \mathbf{M}_{l'}(\mathbf{r}), \overline{\mathbf{G}}_m(\mathbf{r}, \mathbf{r}'), \mathbf{M}_l(\mathbf{r}') \rangle \\ & + \sum_{n=1}^{N} a_n \langle \mathbf{M}_{l'}(\mathbf{r}), \nabla \times \overline{\mathbf{G}}_e(\mathbf{r}, \mathbf{r}'), \mathbf{J}_n(\mathbf{r}') \rangle, \quad l' = 0, \ldots, L-1. \end{aligned} \tag{4.2.46}$$

In the above, we have made use mode orthogonality to arrive at

$$\begin{aligned} \langle \mathbf{M}_{l'}, \mathbf{H}_l \rangle &= \int_{S_a} dS \mathbf{M}_{l'} \cdot \mathbf{H}_l = \int_{S_a} dS \mathbf{E}_{l'} \times \hat{n} \cdot \mathbf{H}_l \\ &= -\int_{S_a} dS \hat{n} \cdot \mathbf{E}_{l'} \times \mathbf{H}_l = -\delta_{l'l}\lambda_{l'} \end{aligned} \tag{4.2.47}$$

where

$$\lambda_{l'} = \int_{S_a} dS\hat{n} \cdot \mathbf{E}_{l'} \times \mathbf{H}_{l'} \tag{4.2.48}$$

Equations (4.2.39) and (4.2.46) constitute $N + L$ equations for $N + L$ unknowns, Γ_l, $l = 0, \ldots, L-1$, and a_n, $n = 1, \ldots, N$. They can be solved by using matrix inversion. Once Γ_0, the reflection coefficient of the TEM mode is found, and assuming that only the TEM mode propagates in the coaxial waveguide region, the input impedance of the probe is given by

$$Y_{in} = Y_0 \frac{1 - \Gamma_0}{1 + \Gamma_0}. \tag{4.2.49}$$

The advantage of this approach is that the input admittance can be found to any desired numerical accuracy by increasing the number of terms in (4.2.25) and (4.2.35). When only one mode is used in the coaxial region, it can be shown that this method, with the input impedance given by (4.2.49), yields the same answer as the method using the variational expression.

4.3 Excitation of a Microstrip Patch Antenna

A microstrip patch antenna is made by etching a patch on top of a dielectric substrate backed by a ground plane. It was first proposed by Deschamps in 1953 [8], and put into practice by Munson in 1972 [9]. Microstrip patch antenna is a very popular antenna because of its ease of fabrication, light weight, and conformal nature. Because of the proximity of the radiation source to a ground plane, and cancelation of the radiation field due to a negative image current on the ground plane, the current on a microstrip patch is a poor radiator ordinarily. However, it can be made to radiate well if resonant modes exist on the patch. At the resonant frequency of the patch, the current amplitude can be greatly enlarged, enhancing the radiation field despite negative image current cancelation. Hence, it radiates by resonance coupling [10–13].

A microstrip patch antenna can be thought of as a cavity-backed slot antenna. The radiation is actually from the side walls or slots of the antenna, and will not radiate well unless it is backed by a resonant structure. Because of the cavity nature of the antenna, it is generally narrow band, but much ingenious design has made these antennas operate with a broader bandwidth.

4.3.1 Magnetic Wall Model

A microstrip patch can be approximated by a magnetic wall model where the side walls of the patch are replaced with magnetic walls and the top and bottom patches remain metallic. The magnetic walls have $\hat{n} \times \mathbf{H} = 0$ boundary condition while that on the metallic walls is $\hat{n} \times \mathbf{E} = 0$. In short, they are perfectly conducting walls. We can assume that the substrate is thin so that $\frac{\partial}{\partial z} = 0$ for the field. Only very high order modes will have $\frac{\partial}{\partial z} \neq 0$. These modes will be far away from the operating frequency of the patch, so that they are weakly excited.

We assume that only TM_z modes are important since TE_z modes will be shorted out. So the field inside the cavity can be written as

$$\mathbf{E}_{mn} = \hat{z} E_{mn} \cos\left(\frac{m\pi x}{a}\right) \cos\left(\frac{n\pi y}{b}\right) \tag{4.3.1}$$

This is called the TM_{mn0} mode, but we will call this the TM_{mn} mode for short. Notice that we have chosen the solution to satisfy the Neumann boundary condition on the magnetic wall so that tangential magnetic field is zero there.[2]

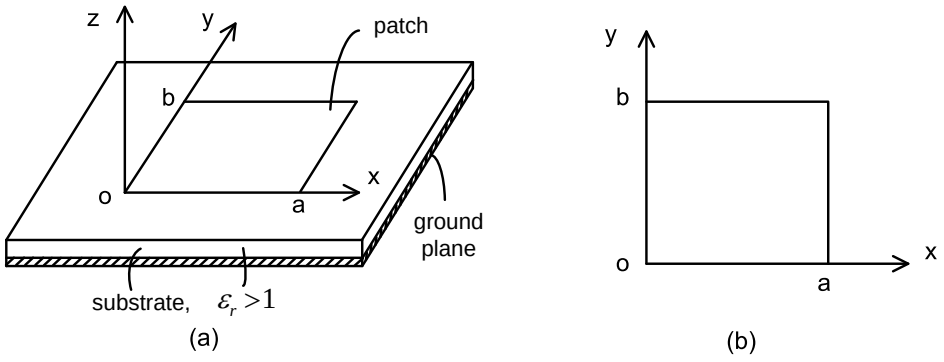


Figure 4.7: Microstrip patch antenna: (a) Side view. (b) Top view.

The resonant frequency of the TM_{mn} mode is given by

$$k_{mn}^2 = \left(\frac{m\pi}{a}\right)^2 + \left(\frac{n\pi}{b}\right)^2 \tag{4.3.2}$$

The normalization constant is

$$E_{mn} = \left[\frac{4}{ab(1+\delta_{0m})(1+\delta_{0n})}\right]^{\frac{1}{2}} \tag{4.3.3}$$

so that the modes are orthonormal. Using the fact that

$$\nabla \times \nabla \times \mathbf{E} - k^2 \mathbf{E} = i\omega\mu \mathbf{J} \tag{4.3.4}$$

and letting

$$\mathbf{E} = \sum_{m,n} a_{mn} \mathbf{E}_{mn}(x, y) \tag{4.3.5}$$

[2]We assume that the probe used is such that there is no charge accumulation on the probe and hence, only divergence-free modes need to be considered in the mode expansion. Such is the case if the current on the probe has constant current such that $\nabla \cdot \mathbf{J} = 0$.

and that

$$\nabla \times \nabla \times \mathbf{E}_{mn} - k_{mn}^2 \mathbf{E}_{mn} = 0 \tag{4.3.6}$$

we obtain that

$$a_{mn} = i\omega\mu \frac{\langle \mathbf{E}_{mn}^*, \mathbf{J} \rangle}{k_{mn}^2 - k^2} \tag{4.3.7}$$

If we assume that

$$\mathbf{J}(x, y) = \hat{z} I_0 \, \delta(y - y') B(x - x') \tag{4.3.8}$$

where

$$B(x) = \frac{1}{w} \begin{cases} 1, & |x| \le w/2 \\ 0, & |x| > w/2 \end{cases} \tag{4.3.9}$$

is a box function. Such a current is chosen so that it is not of zero thickness and width. Otherwise, it will have infinite inductance.

Then

$$\langle \mathbf{E}_{mn}^*, \mathbf{J} \rangle = E_{mn} \cos\left(\frac{n\pi y'}{b}\right) \frac{1}{w} \int_{x' - \frac{w}{2}}^{x' + \frac{w}{2}} \cos\left(\frac{m\pi x}{a}\right) dx \tag{4.3.10}$$

By letting

$$\begin{aligned} I = \frac{1}{w} \int_{x' - \frac{w}{2}}^{x' + \frac{w}{2}} dx \, \cos\left(\frac{m\pi x}{a}\right) &= \frac{1}{w} \left(\frac{m\pi}{a}\right)^{-1} \sin\left(\frac{m\pi x}{a}\right) \Bigg|_{x' - \frac{w}{2}}^{x' + \frac{w}{2}} \\ &= \frac{1}{w} \left(\frac{m\pi}{a}\right)^{-1} \left[\sin\left(\frac{m\pi (x' + \frac{w}{2})}{a}\right) - \sin\left(\frac{m\pi (x' - \frac{w}{2})}{a}\right) \right] \end{aligned} \tag{4.3.11}$$

and by using

$$\sin(A + B) - \sin(A - B) = 2 \cos A \sin B \tag{4.3.12}$$

then

$$I = \frac{2a}{m\pi w} \cos\left(\frac{m\pi x'}{a}\right) \sin\left(\frac{m\pi w}{2a}\right) = \cos\left(\frac{m\pi x'}{a}\right) \mathrm{sinc}\left(\frac{m\pi w}{2a}\right) \tag{4.3.13}$$

where $\mathrm{sinc}(x) = (\sin x)/x$.

Hence

$$\langle \mathbf{E}_{mn}^*, \mathbf{J} \rangle = E_{mn} \cos\left(\frac{m\pi x'}{a}\right) \cos\left(\frac{n\pi y'}{b}\right) \mathrm{sinc}\left(\frac{m\pi w}{2a}\right) \tag{4.3.14}$$

$$a_{mn} = i\omega\mu \frac{\phi_{mn}(x', y') \mathrm{sinc}\left(\frac{m\pi w}{2a}\right)}{k_{mn}^2 - k^2} \tag{4.3.15}$$

$$\mathbf{E} = i\omega\mu\hat{z} \sum_{m,n} \frac{1}{k_{mn}^2 - k^2} \phi_{mn}(x, y) \phi_{mn}(x', y') \mathrm{sinc}\left(\frac{m\pi w}{2a}\right) \tag{4.3.16}$$

where

$$\phi_{mn}(x,y) = E_{mn}\cos\left(\frac{m\pi x}{a}\right)\cos\left(\frac{m\pi y}{b}\right) \tag{4.3.17}$$

The above is the magnetic wall cavity model for the patch antenna. However, it has no loss and the resonant frequencies of the modes of the cavity are purely real. When the operating frequency coincides with the resonant frequency, from (4.3.7), it is seen that the excitation coefficient of the mode becomes infinite. This is unphysical, as the resonant frequency of the cavity is never real in practice: the resonant modes of the patch antenna are radiationally damped. Hence, its resonant frequencies are complex rather than real. In addition, there are material loss and copper loss of the antenna that causes the modes to have complex resonant frequencies, giving rise to damped resonances. The dielectric loss can be easily incorporated by using a complex dielectric. The radiation damping can be modeled by a lossy magnetic wall while the copper loss can be modeled by a lossy electric wall.[3]

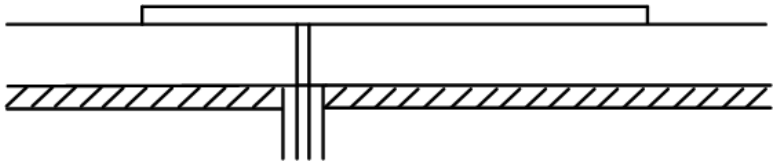


Figure 4.8: Excitation of the patch antenna with a probe (cross-section view).

For a real resonant frequency, the fields of the mode will have a $e^{-i\omega_{mn}t}$ time dependence. If the resonant frequency is complex with $\tilde{\omega}_{mn} = \omega_{mn} - i\alpha_{mn}$, then the time dependence of of the fields is $e^{-i\omega_{mn}t}e^{-\alpha_{mn}t}$. The energy density of the mode is proportional to $|\mathbf{E}|^2$ and $|\mathbf{H}|^2$, and hence the stored energy $W_{mn,T} \sim e^{-2\alpha_{mn}t}$. By energy conservation, the power radiated by the mode is proportional to the negative time rate of change of stored energy, or

$$P_{mn,rad} = -\frac{d}{dt}W_{mn,T} = 2\alpha_{mn}W_{mn,T}$$

Hence, if the stored energy and the power radiated is known, the imaginary part of the resonant frequency α_{mn} can be found. This is equivalent to finding the Q of the resonant modes of the antenna.

Actually, the reactive power leakage from the antenna also gives rise to a real resonant frequency shift. The estimate of this shift is more difficult and method of estimating this shift is given in [14–16] using perturbation approach and asymptotic approach. In retrospect, the fringing field at the open edge of the patch makes the patch effectively larger, lowering the real resonant frequency compared to that predicated by the magnetic wall model.

4.3.2 The Q of the Modes

Due to radiation damping, and other losses in the cavity, the resonant frequency of each mode is not purely real. We shall discuss how to estimate the damping rate due to radiation by

[3]The excitation of the cavity modes by a source is elaborately dealt with in [2, 20]. An application to microstrip antenna is given in [12].

using a perturbation approach. In this approach, we assume that the current distribution on the patch is not changed a lot when the magnetic walls are removed to allow for radiation of the patch current.

To estimate the power radiated by the patch, we need to find the current on the patch. To this end, we derive the magnetic field in the cavity:

$$\nabla \times \mathbf{E}_{mn} = i\omega\mu \mathbf{H}_{mn} \tag{4.3.18}$$

or that

$$\mathbf{H}_{mn} = -\frac{1}{E_{mn}\omega\mu}\left[\hat{x}\left(\frac{n\pi}{b}\right)\cos\left(\frac{m\pi x}{a}\right)\sin\left(\frac{n\pi y}{b}\right) - \hat{y}\left(\frac{m\pi}{a}\right)\sin\left(\frac{m\pi x}{a}\right)\cos\left(\frac{n\pi y}{b}\right)\right] \tag{4.3.19}$$

The corresponding current on the top patch is

$$\mathbf{J}_{mn} = -\hat{z}\times\mathbf{H}_{mn} = \frac{1}{i\omega\mu}\left[\hat{x}\left(\frac{m\pi}{a}\right)\sin\left(\frac{m\pi x}{a}\right)\cos\left(\frac{n\pi y}{b}\right) + \hat{y}\left(\frac{n\pi}{b}\right)\cos\left(\frac{m\pi x}{a}\right)\sin\left(\frac{n\pi y}{b}\right)\right] \tag{4.3.20}$$

If this current is radiating in free space, its field is given by the free space dyadic Green's function acting on the above current, namely,

$$\mathbf{E}_{mn,R} = i\omega\mu\int d\mathbf{r}'\overline{\mathbf{G}}\left(\mathbf{r},\mathbf{r}'\right)\cdot\mathbf{J}_{mn}\left(\mathbf{r}'\right) = i\omega\mu\left(\overline{\mathbf{I}} + \frac{\nabla\nabla}{k^2}\right)\cdot\int d\mathbf{r}'\frac{e^{ik|\mathbf{r}-\mathbf{r}'|}}{4\pi|\mathbf{r}-\mathbf{r}'|}\mathbf{J}_{mn}\left(\mathbf{r}'\right) \tag{4.3.21}$$

By letting letting $|\mathbf{r}| = r \to \infty$,

$$\frac{e^{ik|\mathbf{r}-\mathbf{r}'|}}{4\pi|\mathbf{r}-\mathbf{r}'|} \sim \frac{e^{ikr}}{4\pi r}e^{-ik\hat{r}\cdot\mathbf{r}'} \tag{4.3.22}$$

Hence,

$$\mathbf{E}_{mn,R} \approx i\omega\mu\left(\overline{\mathbf{I}} + \frac{\nabla\nabla}{k^2}\right)\cdot\frac{e^{ikr}}{4\pi r}\int d\mathbf{r}'e^{-ik\hat{r}\cdot\mathbf{r}'}\mathbf{J}_{mn}\left(\mathbf{r}'\right) \tag{4.3.23}$$

Furthermore, when $r \to \infty$, the spherical wave becomes like a plane wave. In otherwords, $e^{ikr} = e^{i\mathbf{k}\cdot\mathbf{r}}$, where $\mathbf{k} = k\hat{k}$ and $\mathbf{r} = r\hat{r}$. Clearly, $\hat{r} = \hat{k}$. Then, we let $\nabla \approx i\mathbf{k} = ik\hat{k}$, and we have

$$\left(\overline{\mathbf{I}} + \frac{\nabla\nabla}{k^2}\right)\frac{e^{ikr}}{r} \approx \left(\overline{\mathbf{I}} - \hat{k}\hat{k}\right)\frac{e^{ikr}}{r} = \left(\hat{\theta}\hat{\theta} + \hat{\phi}\hat{\phi}\right)\frac{e^{ikr}}{r} \tag{4.3.24}$$

In the above, we have made use of the fact that $\hat{k}\hat{k} = \hat{r}\hat{r}$, and that $\overline{\mathbf{I}} = \hat{r}\hat{r} + \hat{\theta}\hat{\theta} + \hat{\phi}\hat{\phi}$. Consequently,

$$\begin{aligned}\mathbf{E}_{mn,R} &\approx i\omega\mu\frac{e^{ikr}}{4\pi r}\left(\hat{\theta}\hat{\theta} + \hat{\phi}\hat{\phi}\right)\cdot\int d\mathbf{r}'e^{-ik\hat{r}\cdot\mathbf{r}'}\mathbf{J}(\mathbf{r}') \\ &= i\omega\mu\frac{e^{ikr}}{4\pi r}\left(\hat{\theta}\hat{\theta} + \hat{\phi}\hat{\phi}\right)\cdot\tilde{\mathbf{J}}(k\hat{r})\end{aligned} \tag{4.3.25}$$

The last integral corresponds to a Fourier integral with the Fourier spectral variable evaluated on the energy shell or Ewald sphere where $|\mathbf{k}| = k = \omega\sqrt{\mu\epsilon}$. Hence, $\tilde{\mathbf{J}}(k\hat{r})$ is the Fourier

transform of $\mathbf{J}(\mathbf{r}')$ with $\mathbf{k} = k\hat{r}$, or on the Ewald sphere. Since $\mathbf{J}(\mathbf{r})$ consists of sinusoidal functions, their Fourier transforms can be evaluated in closed form. Hence, $\tilde{\mathbf{J}}(k\hat{r})$ can be found. Also, the physical meaning is that only this Fourier component will radiate coherently in the $k\hat{r}$ direction.

The above is the electric field radiated via the free-space dyadic Green's function. In order to account for the fact that this current is radiating on a dielectric substance backed by a ground plane, we need only to add the reflected wave term. The reflected wave can be added using ray physics since the observation point is in the far field where ray physics applies [3]. Consequently,

$$\mathbf{E}_{mn,R} \approx i\omega\mu \frac{e^{ikr}}{4\pi r} \left\{ \hat{\theta}\tilde{J}_\theta\left(k\hat{r}\right)\left[1 - \tilde{R}^{TM}(\hat{r})\right] + \hat{\phi}\tilde{J}_\phi\left(k\hat{r}\right)\left[1 + \tilde{R}^{TE}\left(\hat{r}\right)\right]\right\} \tag{4.3.26}$$

where $\tilde{R}^{TM}$ and $\tilde{R}^{TE}$ are the generalized reflection coefficient for the layered medium representing the substrate with a ground plane. The minus sign in front of $\tilde{R}^{TM}$ is because J_θ produces TM fields of opposite polarities above and below the source which is assumed to be an infinitely thin sheet. The J_θ current resembles a Hertzian dipole radiating in endfire direction, and hence produces a TM field that is odd symmetric about $z = 0$ plane. But the J_ϕ current resembles a Hertzian dipole radiating in the broadside diection produces a TE field that is even symmetric about $z = 0$ plane.

The power density radiated by this mode is then

$$S_{mn}(r,\theta,\phi) = \frac{1}{2\eta}\left|\mathbf{E}_{mn,R}\right|^2 \tag{4.3.27}$$

$$= \frac{\omega^2\mu^2}{2\eta}\frac{1}{\left|4\pi r\right|^2}\left\{\left|\tilde{J}_\theta(k\hat{r})\left[1 - \tilde{R}^{TM}(\hat{r})\right]\right|^2 + \left|\tilde{J}_\phi(k\hat{r})\left[1 + \tilde{R}^{TE}(\hat{r})\right]\right|^2\right\} \tag{4.3.28}$$

where $\hat{r}$ is a function of (θ, ϕ).

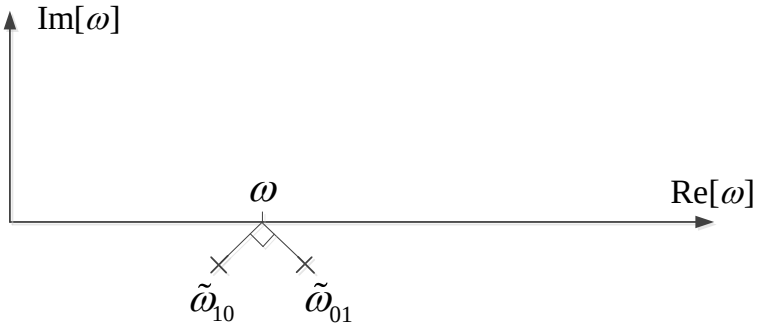


Figure 4.9: Splitting of the degenerate TM_{10} and TM_{01} modes.

The total power radiated can be found by performing the following integral over a hemisphere, namely,[4]

$$P_{mn,rad} = \int_0^{2\pi} \int_0^{\frac{\pi}{2}} S_{mn}(r,\theta,\phi) r^2 \sin\theta d\theta d\phi \tag{4.3.29}$$

The stored energy of the mode is

$$W_{mn,T} = \frac{1}{4}\int_V \left[\epsilon|\mathbf{E}_{mn}|^2 + \mu|\mathbf{H}_{mn}|^2\right] dV = \frac{1}{2}\int_V \epsilon|\mathbf{E}_{mn}|^2 dV \tag{4.3.30}$$

in the cavity between the patches. Therefore, for normalized modes,

$$W_{mn,T} = \frac{1}{2}\epsilon \int_V |\mathbf{E}_{mn}|^2 dV = \frac{1}{2}\epsilon \tag{4.3.31}$$

Consequently, the Q of the mn mode is

$$Q_{mn} = \frac{\omega_{mn} W_{mn,T}}{P_{mn,rad}} = \frac{\omega_{mn}}{2\alpha_{mn}} = \omega_{mn}\tau_{mn} \tag{4.3.32}$$

where τ_{mn} is the decay time constant of the mn modes, and the complex resonant frequency of the mode is

$$\tilde{\omega}_{mn} = \omega_{mn} - i\alpha_{mn} \tag{4.3.33}$$

From (4.3.32),

$$\alpha_{mn} = \frac{\omega_{mn}}{2Q_{mn}} \tag{4.3.34}$$

The Q can be estimated using the integral (4.3.29) and expression (4.3.32).

4.3.3 Circular Polarization Excitation

In a square microstrip patch, the TM_{10} and TM_{01} modes are degenerate. If the operating frequency is chosen close to that of these modes, they will be dominant. Hence, (4.3.5) can be approximated by only two modes, namely,

$$\mathbf{E} \cong a_{10}\mathbf{E}_{10}(x,y) + a_{01}\mathbf{E}_{01}(x,y) \tag{4.3.35}$$

where

$$a_{10} = i\omega\mu\frac{\langle \mathbf{E}_{10}^*, \mathbf{J}\rangle}{\tilde{k}_{10}^2 - k^2}, \qquad a_{01} = i\omega\mu\frac{\langle \mathbf{E}_{01}^*, \mathbf{J}\rangle}{\tilde{k}_{01}^2 - k^2} \tag{4.3.36}$$

where $\tilde{k}_{10}$ and $\tilde{k}_{01}$ are the complex resonant frequencies of the modes. If the probe is located such that $x' = y'$ or along diagonal of the square patch, and $\mathrm{sinc}(\frac{m\pi w}{2a}) \approx 1$ for both TM_{01} and TM_{10} modes, then $a_{10} \approx a_{01}$, as can be seen from (4.3.16). The mode currents of these modes are orthogonal to each other in space, but the field produced is not circularly polarization.

However, circular polarization can be obtained by making these modes non-degenerate by destroying the symmetry. We can let $a = b + \triangle$, so that the TM_{10} mode has a slightly lower

[4]In addition, the patch current can excite a surface wave mode in the dielectric substrate layer causing further loss, and damping of the mode. This is not included in this integral.

resonant frequency compared to the TM_{01} mode as shown in Figure 4.9. Consequently, we have

$$a_{10} \simeq i\omega\mu \frac{\langle \mathbf{E}_{10}^*, \mathbf{J} \rangle}{(\tilde{k}_{10} - k)2\tilde{k}_{10}}, \qquad a_{01} \simeq i\omega\mu \frac{\langle \mathbf{E}_{01}^*, \mathbf{J} \rangle}{(\tilde{k}_{01} - k)2\tilde{k}_{01}} \tag{4.3.37}$$

If we split the modes appropriately, and get $\tilde{k}_{10} - k$ to be 90^o out of phase with $\tilde{k}_{01} - k$. Then a_{10} and a_{01} will be 90^o out of phase, and $\mathbf{E}$ in (4.3.35) will become circularly polarized.

4.3.4 Perturbation Formula for Resonant Frequency Shift

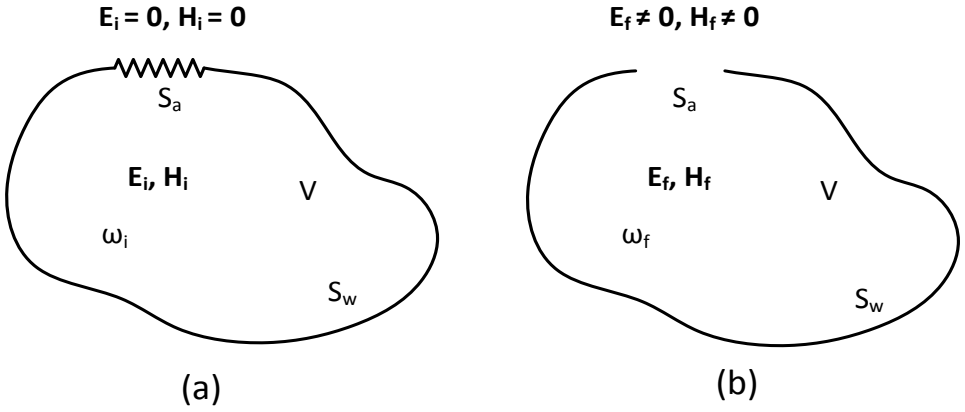


Figure 4.10: Derivation of the perturbation formula: (a) Geometry before perturbation. (b) Geometry after perturbation.

The above derivation for resonant frequency shift due to radiation damping is intuitive. But a more formal procedure for ascertaining the resonant frequency shift can be obtained by using perturbation concepts [4]. These concepts have been applied to derive the resonant frequency shift for microstrip antennas [14, 15]. To start, we take the divergence of the following quantity:

$$\begin{aligned} &\nabla \cdot \mathbf{E}_i^* \times \mathbf{H}_f + \mathbf{E}_f \times \mathbf{H}_i^*) \\ &\quad = \mathbf{H}_f \cdot \nabla \times \mathbf{E}_i^* - \mathbf{E}_i^* \cdot \nabla \times \mathbf{H}_f + \mathbf{H}_i^* \cdot \nabla \times \mathbf{E}_f - \mathbf{E}_f \cdot \nabla \times \mathbf{H}_i^* \\ &\quad = -i\omega_i \mathbf{H}_f \cdot \overline{\boldsymbol{\mu}}^* \cdot \mathbf{H}_i^* + i\omega_f \mathbf{E}_i^* \cdot \overline{\epsilon} \cdot \mathbf{E}_f + i\omega_f \mathbf{H}_i^* \cdot \overline{\mu} \cdot \mathbf{H}_f - i\omega_i \mathbf{E}_f \cdot \overline{\epsilon}^* \cdot \mathbf{E}_i^* \end{aligned} \tag{4.3.38}$$

Integrating the above over the original volume of the cavity as shown in Figure 4.10, we arrive at

$$\int_S dS\hat{\mathbf{n}} \cdot (\mathbf{E}_i^* \times \mathbf{H}_f + \mathbf{E}_f \times \mathbf{H}_i^*) = i(\omega_f - \omega_i) \int_V dV \, (\mathbf{H}_i^* \cdot \overline{\mu} \cdot \mathbf{H}_f + \mathbf{E}_i^* \cdot \overline{\epsilon} \cdot \mathbf{E}_f) \tag{4.3.39}$$

where we have assumed that the medium is lossless, and hence $\overline{\boldsymbol{\epsilon}}^{\dagger} = \overline{\boldsymbol{\epsilon}}$ and $\overline{\boldsymbol{\mu}}^{\dagger} = \overline{\boldsymbol{\mu}}$. Since this is a perturbation, the final fields and the initial fields are similar to each other. Hence, we can approximate the right-hand side with the initial field entirely. Consequently,

$$\int_S dS\hat{\mathbf{n}} \cdot (\mathbf{E}_i^* \times \mathbf{H}_f + \mathbf{E}_f \times \mathbf{H}_i^*) \cong i(\omega_f - \omega_i) \int_V dV[\mathbf{H}_i^* \cdot \overline{\mu} \cdot \mathbf{H}_i + \mathbf{E}_i^* \cdot \overline{\epsilon} \cdot \mathbf{E}_i] \tag{4.3.40}$$

The integral on the right-hand side is purely real now, and it can be written in terms of the initial time-average stored energy $\langle W_{T_i} \rangle$.

$$\int_{S_A} dS\hat{\mathbf{n}} \cdot (\mathbf{E}_i^* \times \mathbf{H}_f) + \int_{S_W} dS\hat{\mathbf{n}} \cdot (\mathbf{E}_f \times \mathbf{H}_i^*) \cong i(\omega_f - \omega_i) 4\langle W_{T_i} \rangle \tag{4.3.41}$$

In the above, the first integral on the left can be approximated as the complex power radiated by the hole.

$$P_{rad}^* \cong \int_{S_A} dS\hat{\mathbf{n}} \cdot (\mathbf{E}_f^* \times \mathbf{H}_f) = 2\langle P_{rad} \rangle - iR_{rad} \tag{4.3.42}$$

where R_{rad} is the reactive power leaked by the cavity to the outside. The second integral on the left of (4.3.41) can be related approximately to the wall loss on the interior of the cavity. Hence, it can be approximated by

$$\begin{aligned} P_{wall} &\cong \int_{S_W} dS\hat{\mathbf{n}} \cdot (\mathbf{E}_f \times \mathbf{H}_i^*) = \int_{S_W} dS(\hat{\mathbf{n}} \times \mathbf{E}_f) \cdot \mathbf{H}_i^* \\ &= \frac{1}{\eta} \int_{S_W} dS|\mathbf{H}_i^*|^2 \end{aligned} \tag{4.3.43}$$

where

$$\eta = \sqrt{\frac{\mu}{\epsilon}} = \sqrt{\frac{\omega\mu}{i\sigma}} = (1-i)\sqrt{\frac{\omega\mu}{2\sigma}}$$

Consequently, the wall loss becomes

$$P_{wall} = (1-i)\sqrt{\frac{\omega\mu}{2\sigma}} \int_{S_W} dS|\mathbf{H}_i|^2 = 2\langle P_{wall} \rangle - iR_{wall} \tag{4.3.44}$$

Hence, the resonant frequency shift is given by

$$\omega_f - \omega_i \approx -i\frac{2\langle P_{rad} \rangle + 2\langle P_{wall} \rangle - iR_{rad} - iR_{wall}}{4\langle W_{Ti} \rangle} \tag{4.3.45}$$

Therefore,

$$\Im m(\omega_f - \omega_i) \approx -i\frac{\langle P_{rad} + P_{wall} \rangle}{2\langle W_{Ti} \rangle} \tag{4.3.46}$$

$$\Re e(\omega_f - \omega_i) \approx -\frac{R_{rad} + R_{wall}}{4\langle W_{Ti} \rangle} \tag{4.3.47}$$

The dissipative loss causes the cavity to have a negative imaginary part of the resonant frequency giving rise to damping. The reactive power leakage and absorption by the wall make the cavity appear larger and lower the resonant frequency.

The fringing field effect at the edge of a microstrip patch can also be solved in closed form using Wiener-Hopf technique [23, 24]. It has been used to ascertain resonant frequency shift of microstrip antennas in [16].

4.3.5 Variational Impedance Formula for a Current Source

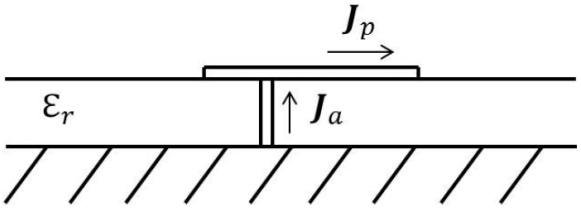


Figure 4.11: Excitation of a microstrip antenna by a current source.

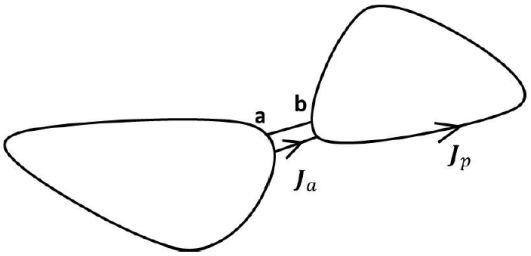


Figure 4.12: Excitation of a general antenna by a current source.

In the previous input impedance formula, the magnetic current $\mathbf{M}_a$ is assumed known and immutable by its environment. The magnetic current is related to the tangential electric field; hence, it is equivalent to a voltage source in circuit theory. Another formula that is of importance is when the structure is driven by an immutable applied (impressed) electric current $\mathbf{J}_a$. Hence, it is equivalent to a current source in circuit theory. The microstrip antenna driven by such a current source is shown in Figure 4.11 and a general structure case is shown in Figure 4.12. In this case, a variational formula for the input impedance based on

reaction is given by [4, 19][5]

$$Z_{in} = -\frac{\langle \mathbf{E}_T, \mathbf{J}_T \rangle}{I^2} \tag{4.3.48}$$

where $\mathbf{E}_T = \mathbf{E}_a + \mathbf{E}_p$, and $\mathbf{E}_a$ and $\mathbf{E}_p$ are the fields produced by applied (or impressed) current $\mathbf{J}_a$ and the induced current $\mathbf{J}_p$, respectively, and $\mathbf{J}_T = \mathbf{J}_a + \mathbf{J}_p$. As shall be shown later, the above formula can admit approximate solution for $\mathbf{J}_p$ with second order error in the input impedance Z_{in}.

In the exact limit, $\langle \mathbf{E}_T, \mathbf{J}_p \rangle = 0$ and the above reduces to

$$Z_{in} = -\frac{\langle \mathbf{E}_T, \mathbf{J}_a + \mathbf{J}_p \rangle}{I^2} = -\frac{\langle \mathbf{E}_T, \mathbf{J}_a \rangle}{I^2} \tag{4.3.49}$$

To prove the above impedance formula, we assume that the exciting source is small and hence, $\mathbf{J}_a$ is constant over space. Then

$$-\langle \mathbf{E}_T, \mathbf{J}_a \rangle = -\int_V dV \mathbf{E}_T \cdot \mathbf{J}_a = -I \int_a^b d\mathbf{l} \cdot \mathbf{E}_a = V_{ab} I = V I = Z_{in} I^2 \tag{4.3.50}$$

asserting the correctness of the impedance formula in the exact limit.

To prove the variational form of the above formula, we express the above in a quadratic form. To this end, we have

$$\begin{aligned} Z_{in} = -\frac{\langle \mathbf{E}_a + \mathbf{E}_p, \mathbf{J}_a + \mathbf{J}_p \rangle}{I^2} &= -\frac{\langle \mathbf{E}_a, \mathbf{J}_a \rangle + \langle \mathbf{E}_p, \mathbf{J}_a \rangle + \langle \mathbf{E}_a, \mathbf{J}_p \rangle + \langle \mathbf{E}_p, \mathbf{J}_p \rangle}{I^2} \\ &= -\frac{\langle \mathbf{E}_a, \mathbf{J}_a \rangle + 2\langle \mathbf{E}_a, \mathbf{J}_p \rangle + \langle \mathbf{E}_p, \mathbf{J}_p \rangle}{I^2} \\ &= -\frac{\langle \mathbf{E}_a, \mathbf{J}_a \rangle + 2\langle \mathbf{E}_a, \mathbf{J}_p \rangle + i\omega\mu \langle \mathbf{J}_p, \overline{\mathbf{G}}_e, \mathbf{J}_p \rangle}{I^2} \end{aligned} \tag{4.3.51}$$

It is straightforward to prove that the stationary point of the above expression is at the exact solution. In the above, the choice of $\overline{\mathbf{G}}_e(\mathbf{r}, \mathbf{r}')$ is important. It can be a free-space Green's function, or Green's function that satisfies specific boundary conditions. For instance, on a PEC surface where $\hat{n} \times \overline{\mathbf{G}}_e(\mathbf{r}, \mathbf{r}') = 0$, reciprocity implies immediately that an impressed electric current on such a surface does not radiate, and its contribution can be ignored in the above calculation.

Another popular formula for input impedance is the power formula where the input impedance is given by

$$Z_{in} = -\frac{\langle \mathbf{E}_T, \mathbf{J}_T^* \rangle}{|I|^2} \tag{4.3.52}$$

The above is based on power conservation, but its variational nature cannot be proved. However, in the limit when $\mathbf{E}_T$ is exact, the above reduces only to integration over the current $\mathbf{J}_a$, and it becomes

$$Z_{in} = -\frac{\langle \mathbf{E}_T, \mathbf{J}_a^* \rangle}{|I|^2} \tag{4.3.53}$$

[5]This formula was used for input impedance calculation before its variational nature was known.

If the current source $\mathbf{J}_a$ is electrically small and constant phase as is the case for a circuit component in circuit theory, the above formula reduces to the variational formula based on reaction.

The names for these formulas have been rather confusing in the literature. The above power formula has been called the induced EMF formula in [4], while the reaction formula has been called the induced EMF formula in [6]. Much of the controversy between the induced EMF formula and power formula has also been discussed in [6].

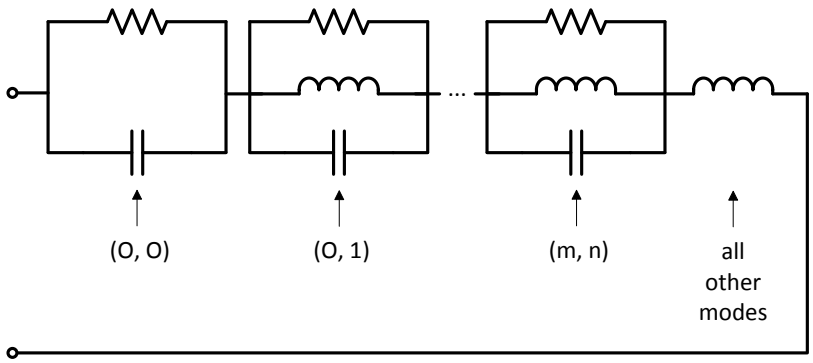


Figure 4.13: Circuit model for the input impedance of a microstrip patch antenna.

Application to Patch Antenna

We will next show how to apply the above result to a microstrip patch antenna. In the exact limit, as mentioned before,

$$Z_{in} = -\frac{\langle \mathbf{E}_T, \mathbf{J}_a \rangle}{I^2} \tag{4.3.54}$$

In the above, $\mathbf{E}_T$ is the total field in the structure produced by the applied current $\mathbf{J}_a$ and the induced current $\mathbf{J}_p$. For the microstrip patch antenna, we have learned how to find the total field in the cavity using a magnetic wall model when a current source is inserted into it. Hence, the total field $\mathbf{E}_T$ is approximately given by[6]

$$\mathbf{E}_T = i\omega\mu \sum_i \frac{\langle \mathbf{E}_i^*, \mathbf{J}_a \rangle}{\tilde{k}_i^2 - k^2} \mathbf{E}_i(\mathbf{r}) \tag{4.3.55}$$

In the above, we have assumed that the dyadic Green's function of the microstrip cavity can be approximated by

$$\overline{\mathbf{G}}(\mathbf{r}, \mathbf{r}') = \sum_i \frac{\mathbf{E}_i(\mathbf{r})\mathbf{E}_i^*(\mathbf{r}')}{\tilde{k}_i^2 - k^2} \tag{4.3.56}$$

[6]More elaborate model is given in [2, 12, 20].

Hence, this Green's function automatically accounts for the field generated by the induced current on the wall of the cavity. This is because we have chosen the eigenmodes to satisfy the requisite boundary conditions on the wall of the cavity.

Therefore,

$$\langle \mathbf{E}_T, \mathbf{J}_a \rangle = i\omega\mu \sum_i \frac{\langle \mathbf{E}_i^*, \mathbf{J}_a \rangle}{\tilde{k}_i^2 - k^2} \langle \mathbf{E}_i, \mathbf{J}_a \rangle \tag{4.3.57}$$

Hence, the input impedance, derivable from (4.3.54), is

$$Z_{in} = Z_a + \sum_i^M Z_i \tag{4.3.58}$$

where Z_a is higher-order mode contributions, and the second summation comes from the dominant mode contributions.

$$Z_i = -\frac{i\omega\mu}{I^2} \frac{\langle \mathbf{E}_i^*, \mathbf{J}_a \rangle \langle \mathbf{E}_i, \mathbf{J}_a \rangle}{\tilde{k}_i^2 - k^2} \tag{4.3.59}$$

is the contribution to the input impedance from individual modes of the cavity. Notice that the above frequency dependence can be fitted with a simple GLC model of a lossy tank circuit resonator. Hence, for the i-th mode, we can pick G_i, L_i, and C_i appropriately to fit the mathematical formula. A circuit approximation of the microstrip patch antenna hence can be expressed as in Figure 4.13 [11]. For the microstrip patch, there is a static mode TM_{00} mode with zero resonant frequency. This mode represents the static capacitor in the microstrip patch. It is denoted by the lossy capacitor model. Also, the probe produces a singular field, which can only be constituted by a linear superposition of many high order modes. Hence, the probe inductance comes from the higher order modes in the cavity. Notice that in this model, unlike the magnetic frill model, the gap capacitance at the base of the probe is ignored.

4.4 Aperture Coupling in Waveguide

In addition to using probes to couple energy into a waveguide, a simple way is to drill holes on the walls of the waveguide, and let energy flow naturally from one waveguide to another. The simple solution of aperture coupling was first derived by Hans Bethe [2,21,22], who eventually received a Nobel prize, not for one given contribution, but for his numerous contributions in physics.

4.4.1 Bethe Coupling

An **E** field in the vicinity of a waveguide wall, is predominantly normal to the waveguide wall. If now, an aperture is opened at the waveguide wall, the electric field in the vicinity of the waveguide wall will be as shown in Figure 4.14(b). The field looks like that of a vertical electric dipole in region B. It has been shown by Bethe [2, 21] that the dipole moment of the

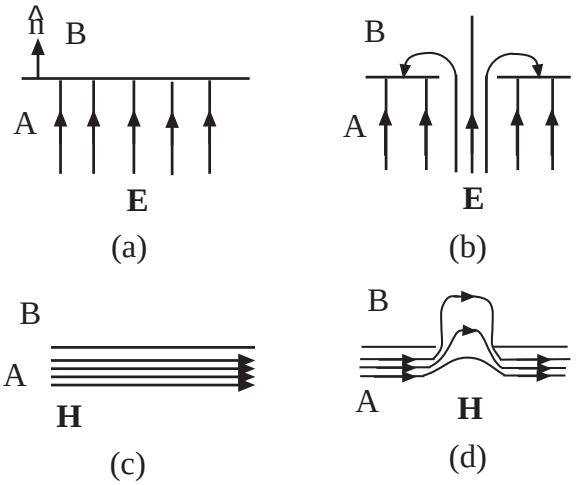


Figure 4.14: Fields in the vicinity of an aperture in a waveguide. Case (b) resembles an electric dipole, while case (d) resembles a magnetic dipole.

vertical electric dipole is proportional to the normal component of the electric field. For a circular aperture of radius a_0, it is

$$\mathbf{p} = \frac{2}{3} a_0^3 \hat{n}(\hat{n} \cdot \epsilon_0 \mathbf{E}) = \alpha_e \hat{n}(\hat{n} \cdot \epsilon_0 \mathbf{E}). \tag{4.4.1}$$

where $\alpha_e = \frac{2}{3} a_0^3$, and a_0 is the radius of the circular aperture.

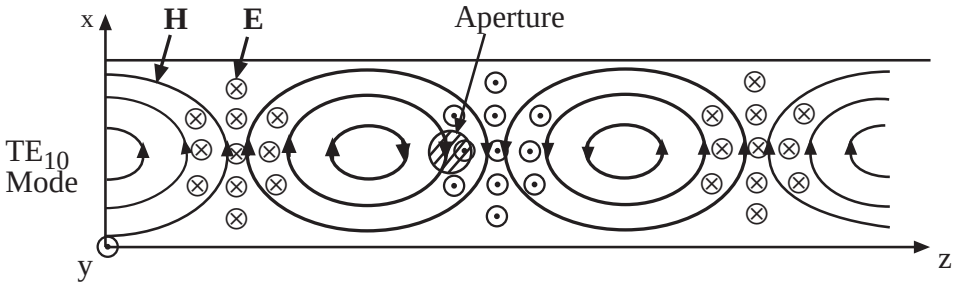


Figure 4.15: Top view of the field in the neighborhood of an aperture in a rectangular waveguide.

A magnetic field in the vicinity of the waveguide wall, is predominantly tangential. Now, if an aperture is present, the magnetic field will leak into region B as shown in Figure 4.14(d). It looks like the field due to a horizontal magnetic dipole in region B. Similarly, the dipole

moment of the horizontal magnetic dipole is [2, 21]

$$\mathbf{m} = -\frac{4}{3}a_0^3\mathbf{H}_t == -\alpha_m\mathbf{H}_t. \tag{4.4.2}$$

where $\alpha_m = \frac{4}{3}a_0^3$.

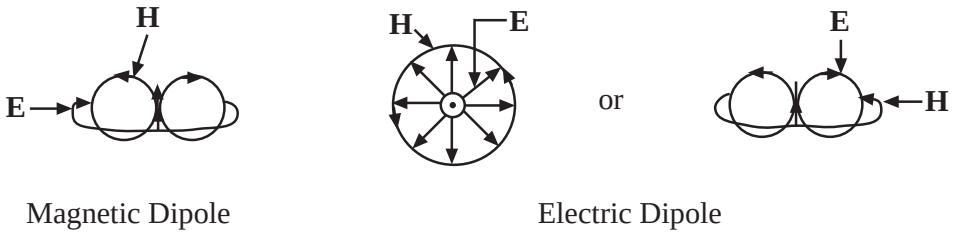


Figure 4.16: Equivalent sources at the aperture of a waveguide.

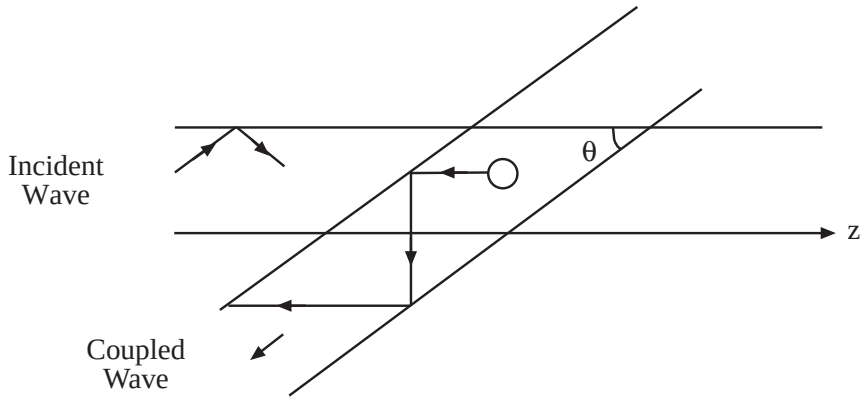


Figure 4.17: A directional coupler using two rectangular waveguides, one on top of another.

If a TE_{10} mode is propagating in a rectangular waveguide, and we have a small aperture on top of the waveguide, then a time harmonic electric and magnetic dipoles will be generated at the aperture. In this case, the electric dipole and magnetic dipole are in phase with respect to each other. For the case of Figure 4.15, the magnetic dipole is pointing in the x direction, and the electric dipole is pointing in the y direction. The superposition of the vertical electric dipole and horizontal magnetic dipole gives rise to the cancelation of fields in the $+z$ direction (see Figure 4.16). Hence, together, they radiate predominantly in the $-z$ direction.[7]

Now, if we lay another waveguide on top of the first waveguide, the radiating electric and magnetic dipoles couple most efficiently into the TE_{10} mode of the top waveguide if the top

[7]Antennas made by a superposition of an electric dipole and a loop to increase their directivity are known as Huygens antenna.

guide is oriented at an angle θ with respect to the bottom guide as shown in Figure 4.17. This is because a TE_{10} mode is actually a bouncing plane wave in a rectangular waveguide.

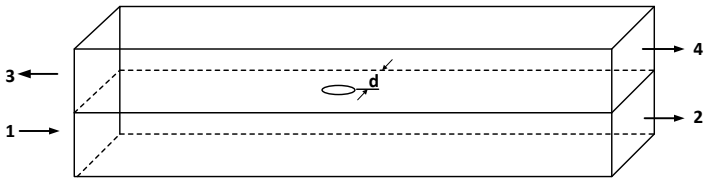


Figure 4.18: An asymmetrically located aperture can also be used to make a directional coupler where $d \neq a/2$.

When the aperture is located at the middle of the waveguide, the magnetic dipole is pointing in the x direction, requiring the top waveguide to be tilted. We can generate a dipole pointing away from the x direction by placing the aperture away from the center of the waveguide, or by using an elliptically shaped aperture. In this case, we can couple efficiently into the TE_{10} mode without having to tilt the top waveguide. Hence, a directional coupler can also be made with $\theta = 0$ if the aperture is not symmetrically located as shown in Figure 4.18 When an aperture is asymmetrically located, the magnetic field that excites it is elliptically polarized. It can be shown that the exciting field of this aperture due to an incident TE_{10} in port 1 of the bottom waveguide is given by

$$\mathbf{E} = \hat{y}E_y = \hat{y}E_0 \sin\left(\frac{\pi d}{a}\right), \tag{4.4.3}$$

$$\mathbf{H} = -E_0 \frac{k_z}{\omega\mu_0}\left[\hat{x}\sin\left(\frac{\pi d}{a}\right) + i\hat{z}\frac{\pi}{k_z a}\cos\left(\frac{\pi d}{a}\right)\right]. \tag{4.4.4}$$

The equivalent electric dipole moment for radiation into the upper guide is

$$\mathbf{p} = \hat{y}\epsilon_0\alpha_e E_0 \sin\left(\frac{\pi d}{a}\right), \tag{4.4.5}$$

The equivalent magnetic dipole moment is

$$\mathbf{m} = \alpha_m E_0 \frac{k_z}{\omega\mu_0}\left[\hat{x}\sin\left(\frac{\pi d}{a}\right) + i\hat{z}\frac{\pi}{k_z a}\cos\left(\frac{\pi d}{a}\right)\right], \tag{4.4.6}$$

Excitation of Modes by Electric and Magnetic Current Sources

Since there will be an electric dipole source and a magnetic dipole source induced in a waveguide due to coupling via a hole, it is prudent to study how they would couple to other modes

in a waveguide. We have learned previously that an electric current source excites modes in a waveguide or cavity as

$$\mathbf{E}(\mathbf{r}) = i\omega\mu \sum_i \frac{\langle \mathbf{E}_i, \mathbf{J}_e \rangle}{k_i^2 - k^2} \mathbf{E}_i(\mathbf{r}) \tag{4.4.7}$$

We define

$$\mathbf{H}_i = \frac{1}{k_i} \nabla \times \mathbf{E}_i \tag{4.4.8}$$

If $\mathbf{E}_i$ is normalized, $\mathbf{H}_i$ is also normalized. In fact one can easily show that

$$\begin{aligned} \langle \mathbf{H}_i^*, \mathbf{H}_i \rangle &= \frac{1}{k_i^2} \langle \nabla \times \mathbf{E}_i^*, \nabla \times \mathbf{E}_i \rangle \\ &= \frac{1}{k_i^2} \langle \mathbf{E}_i^*, \nabla \times \nabla \times \mathbf{E}_i \rangle \\ &= \langle \mathbf{E}_i^*, \mathbf{E}_i \rangle = 1 \end{aligned} \tag{4.4.9}$$

Also, it can be easily shown that

$$\nabla \times \mathbf{H}_i = \frac{1}{k_i^2} \nabla \times \nabla \times \mathbf{E}_i = k_i \mathbf{E}_i \tag{4.4.10}$$

Hence, if we have a vector wave equation given by[8]

$$\nabla \times \nabla \times \mathbf{H} - k^2 \mathbf{H} = i\omega\epsilon \mathbf{J}_m \tag{4.4.11}$$

the orthonormal eigenmode expansion gives

$$\mathbf{H} = i\omega\epsilon \sum_i \frac{\langle \mathbf{H}_i^*, \mathbf{J}_m \rangle}{(k_i^2 - k^2)} \mathbf{H}_i(\mathbf{r}) \tag{4.4.12}$$

where $\mathbf{H}_i$ is normalized. The corresponding $\mathbf{E}$ field, via the use of (4.4.10) and Maxwell's equations, is

$$\mathbf{E} = \sum \frac{\langle \mathbf{H}_i^*, \mathbf{J}_m \rangle}{(k_i^2 - k^2)} k_i \mathbf{E}_i \tag{4.4.13}$$

The polarization density can be expressed as

$$\mathbf{P} = \hat{y}\epsilon\alpha_e E_0 \sin\left(\frac{\pi d}{a}\right) \delta(\mathbf{r} - \hat{x}d) \tag{4.4.14}$$

to imply that the hole is located at $(x, y, z) = (d, 0, 0)$. The corresponding electric current density arising from time-varying polarization density is

$$\mathbf{J}_p = -i\omega\mathbf{P} \tag{4.4.15}$$

[8] We will use $\mathbf{J}_m$ to denote magnetic current and reserve $\mathbf{M}$ to denote magnetization density in this section.

Similarly, the magnetization density is

$$\mathbf{M} = \alpha_m \frac{k_z E_0}{\omega\mu}\left[\hat{x}\sin\left(\frac{\pi d}{a}\right) + i\hat{z}\frac{\pi}{k_z a}\cos\left(\pi d a\right)\right]\delta(\mathbf{r} - \hat{x}d) \tag{4.4.16}$$

The corresponding magnetic current density arising from a time-varying magnetization density is given as

$$\mathbf{J_m} = i\omega\mu\mathbf{M} \tag{4.4.17}$$

The corresponding TE_{10} mode in the upper waveguide is

$$\mathbf{E}_{10} = \hat{y}E_{10}\sin\left(\frac{\pi x}{a}\right)e^{\pm ik_z z} \tag{4.4.18}$$

where E_{10} is for normalization. The corresponding $\mathbf{H}_{10}$ field is

$$\mathbf{H}_{10} = \frac{E_{10}}{k_{10}}\left[\pm\hat{x}ik_z\sin\left(\frac{\pi x}{a}\right) + \hat{z}\frac{\pi}{a}\cos\left(\frac{\pi x}{a}\right)\right]e^{\pm ik_z z} \tag{4.4.19}$$

which is normalized. The $\pm$ sign implies $\pm z$ propagating waves.

To see if the mode excited by the electric dipole will cancel the one excited by the magnetic dipole, we need to compare (4.4.7) and (4.4.13). Hence, we need to sum the coefficients

$$k_{10}\langle\mathbf{H}_{10}^{\star}, \mathbf{J}_m\rangle$$

and

$$i\omega\mu\langle\mathbf{E}_{10}^{\star}, \mathbf{J}_p\rangle$$

In details,

$$k_{10}\langle\mathbf{H}_{10}^{\star}, \mathbf{J}_m\rangle = \alpha_m ik_z E_0 E_{10}\left[\pm ik_z\sin^2\left(\frac{\pi d}{a}\right) + i\frac{\pi^2}{k_z a^2}\cos^2\left(\frac{\pi d}{a}\right)\right] \tag{4.4.20}$$

Furthermore,

$$i\omega\mu\langle\mathbf{E}_{10}^{\star}, \mathbf{J}_p\rangle = \omega^2\mu\epsilon\alpha_e E_0 E_{10}\sin^2(\frac{\pi d}{a}) \tag{4.4.21}$$

Hence, the TE_{10} mode that is excited is proportional to

$$\begin{aligned} B_{\pm} \propto \quad & k_{10}\langle\mathbf{H}_{10}^{\star}, \mathbf{J}_m\rangle + i\omega\mu\langle\mathbf{E}_{10}^{\star}, \mathbf{J}_p\rangle \\ & = E_0 E_{10}\left[\pm k_z^2\alpha_m\sin^2\left(\frac{\pi d}{a}\right) + \left(\frac{\pi}{a}\right)^2\alpha_m\cos^2\left(\frac{\pi d}{a}\right) + k^2\alpha_e\sin^2\left(\frac{\pi d}{a}\right)\right] \\ & = 0, \quad z < 0. \end{aligned} \tag{4.4.22}$$

The solution of the above is

$$\frac{d}{a} = \frac{1}{\pi}\sin^{-1}\left(\frac{1}{\sqrt{6}}\frac{\lambda_0}{a}\right) \tag{4.4.23}$$

where λ_0 is the free-space wavelength. The above coupler using one hole to achieve directional coupling is known as the Bethe hole coupler.

When a hole is dug in a waveguide wall, the induced dipoles also give rise to back action in the original waveguide, altering the field. This in turn, gives rise to an alteration of the coupled field from the first waveguide to the second waveguide. This effect is taken into account by Collin [2]. The resulting hierarchy of equations are rather complicated but the above condition still holds true. This can be thought of as a multiple scattering or coupling effect, as in the Fabry-Perot etalon. It is the cancelation of the leading order term that is important, for the cancelation of the higher-order terms will follow suit.

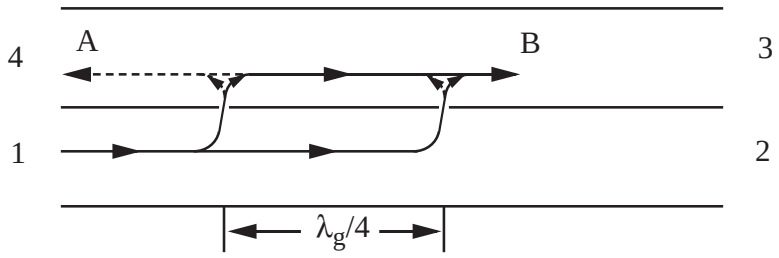


Figure 4.19: A two-hole directional coupler.

Other kinds of directional couplers are also possible. For instance, one can make a two-hole coupler with the holes spaced $\lambda_g/4$ apart as shown in Figure 4.19. The apertures need not have a preferred directional coupling. If the aperture coupling is weak, then the wave reaching the second hole is essentially the same as the wave that reaches the first hole. Hence, the two different waves that reach B via coupling through the two different holes are in phase and will interfere constructively. Because of the $\lambda_g/4$ separation of the two holes, the waves from the two different holes that reach A will be $\lambda_g/2$ or 180^o out of phase. Therefore, at A, the waves interfere destructively, and there is little energy coupled to port 4.

Since this coupler uses constructive and destructive interferences to enhance its directivity, the directivity is frequency sensitive. One remedy is to use more holes so as to broaden its bandwidth, or to use the Schwinger reversed-phase coupler. [2, 25]

4.4.2 Equivalence Principles in Aperture Coupling

Equivalence principles have been discussed in [2, 4]. In the actual calculation of the electric and magnetic dipole moments, certain equivalence principles have to be invoked. We shall discuss them as follows.

There will be two kinds of currents in the following discussion: induced currents and impressed currents. Induced currents are currents flowing in a conductor due to the presence of an incident or exciting field next to the conductor. A perfect conductor, for instance, cannot have a non-zero field in it, and hence, current flows on its surface to prevent the fields from penetrating it. On the other hand, impressed currents are currents we assume as sources in Maxwell's equations. They are currents assumed to exist in free space. They are

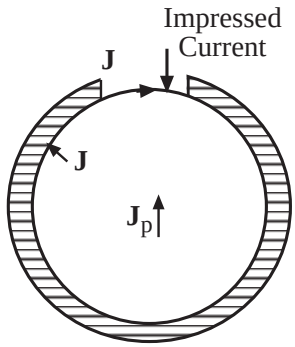


Figure 4.20: Equivalent problem of a waveguide with an aperture—nonradiating case.

the driving source terms in Maxwell's equations, that are immutable as we seek the solutions. On the other hand, induced currents follow from the solutions of Maxwell's equations. They are due to currents flowing in conductors as we seek solutions to Maxwell's equations.

Equivalence Principle I

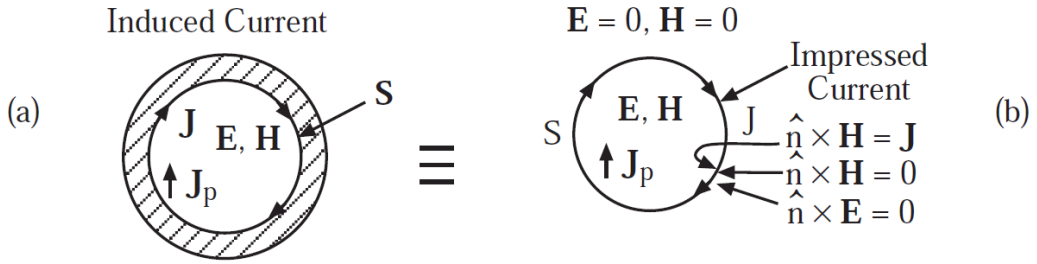


Figure 4.21: Equivalent problem of a hollow waveguide—equivalence principle I.

If we have a source in a metallic waveguide, the source will induce a current $\mathbf{J}$ on the inner surface of the waveguide. The current $\mathbf{J}$ is known as the induced current. On the surface of the waveguide, we have

$$\hat{n} \times \mathbf{E} = 0, \qquad \hat{n} \times \mathbf{H} = \mathbf{J}. \tag{4.4.24}$$

This current is responsible for expelling the electromagnetic field away from the perfect conductor. It is also responsible for the jump discontinuity for the magnetic field outside and inside the perfect conductor.

Now if we remove the metallic wall that supports the induced current and replace the induced current with an *impressed current* in vacuum, the field inside the surface S is identical to before. Furthermore, the field outside S is identically zero. This is because the impressed current $\mathbf{J}$ supports a discontinuity in the magnetic field. Hence, $\hat{n} \times \mathbf{H} = 0$ and $\hat{n} \times \mathbf{E} = 0$ just outside the impressed current $\mathbf{J}$. By Huygens' principle, the field must be zero everywhere outside S. We can check if the solution satisfies all the requisite boundary conditions. If it does, it is the only unique solution.

The equivalence principle can be proved by three means:

- By performing a Gedanken experiment whereby the conductive material with zero field inside is been chiseled away until the induced current is replaced by impressed current in free space. The induced current has held the internal field of the cavity in place, and the impressed current will still hold the internal fields in place.
- By using uniqueness principle argument, when the induced current is replaced by impressed current, the boundary conditions for the fields remain the same. By uniqueness principle, they must be the same;
- The equivalence principle can also be proved mathematically by the use of Huygens principle.

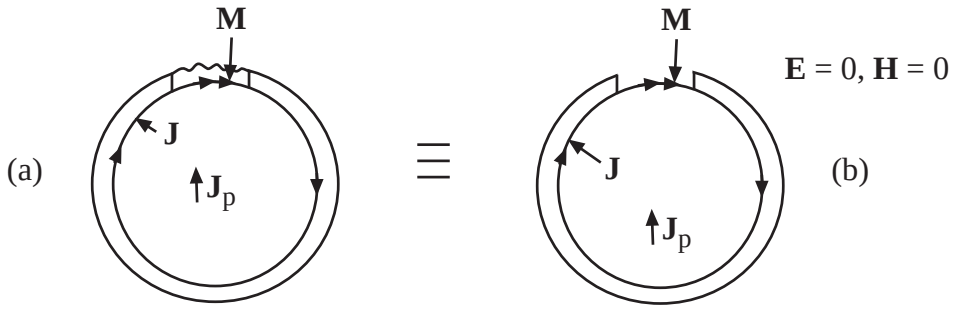


Figure 4.22: Equivalent problem of a waveguide which is partly covered with a magnetic wall-nonradiating case.

Equivalence Principle II

From equivalence principle I, if we have a small aperture in the cavity now, and at the aperture, the original induced current from Figure 4.21a is impressed, then the field will still be identically zero outside the waveguide. Hence, $\mathbf{J}_p$ and its own induced current on the PEC wall must have generated equal and opposite field to that produced by impressed $\mathbf{J}$ at the aperture and its own induced current on the PEC wall. Because of this, the two currents in Figure 4.20 generate zero field outside the waveguide. Also, because of this, the cases in Figure 4.23a and Figure 4.23b generate equivalent field outside the waveguide. Note that the current in Figure 4.23b is exactly opposite to that in Figure 4.20.

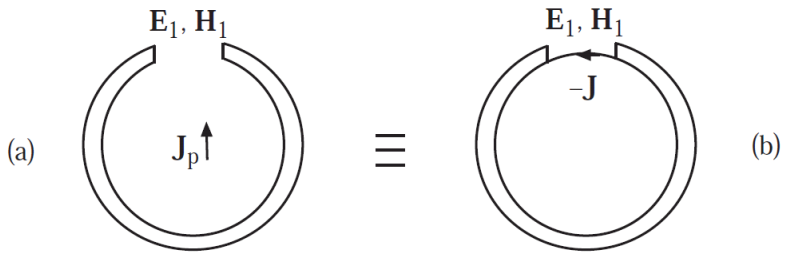


Figure 4.23: Equivalent problem of a waveguide with an aperture—radiating case—equivalence principle II.

Equivalence Principle III

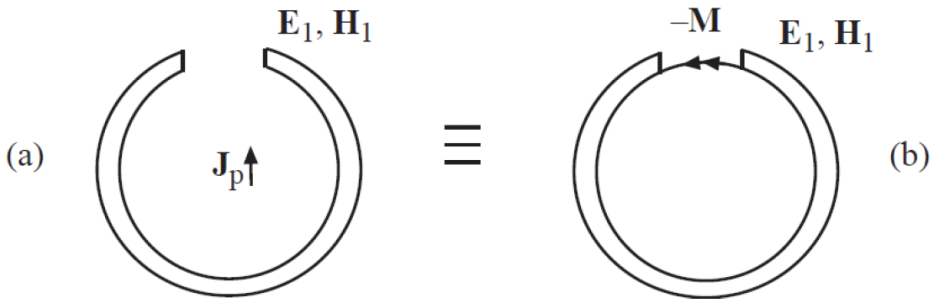


Figure 4.24: Equivalent problem of a waveguide with an aperture—radiating case–equivalence principle III.

By a similar argument, if we have a closed cavity with a magnetic wall over part of the cavity as in Figure 4.22a, it is completely equivalent to the case of Figure 4.22b, with zero field outside the waveguide. Consequently, a waveguide with a small aperture in Figure 4.24a generates equivalent field outside the waveguide as in the case of Figure 4.24b.

Therefore, for radiation due to an aperture in a waveguide, there are two equivalent problems denoted by Figure 4.23 and Figure 4.24. For the case of Figure 4.23, to find **J**, we have to first solve the closed waveguide problem denoted by Figure 4.21a. For the case of Figure 4.24, we have to first solve the problem denoted by Figure 4.22a to find **M**. It turns out that equivalence principle III is preferred over equivalence principle II because the radiation of a magnetic current in a small aperture is easier to calculate than the radiation of an electric current in a small aperture. For example, if the surface of the waveguide is flat enough, or that the aperture is small enough, we can replace the problem in Figure 4.25a

Figure 4.25: Equivalent problem of a magnetic current radiating in the vicinity of an aperture.

with that in Figure 4.25b.

When the PEC surface is flat, a horizontal magnetic dipole radiating in the aperture is the same as the dipole radiating in free space, as the dipole produces only horizontal magnetic field that satisfies the boundary condition on the flat PEC surface. So the source is oblivious of the presence of the flat PEC surface. By image theorem, a horizontal magnetic dipole radiating in free space is equivalent to one with half its original strength radiating on top of a PEC ground plane.

Figure 4.26: Equivalent problem of the **E** field around a waveguide wall partially covered by a small magnetic wall.

As mentioned before, in order to find **M**, we need to solve the closed problem with a magnetic wall patch as shown in Figure 4.22. If the magnetic wall patch is small enough, the field around the magnetic wall may be approximated with a static field solution. For example, if the field is predominantly electric, the field in the vicinity of the magnetic wall patch (assuming that the wall is reasonably flat) resembles that of Figure 4.26a. Due to the symmetry of the problem, it is equivalent to that of Figure 4.26b. What happens is that the electric field induces circulating magnetic current on the magnetic wall patch that expels the electric field. The circulating magnetic current generates a vertical electric dipole moment that expels the electric field from the magnetic disk. If the field is predominantly magnetic, the magnetic field around the magnetic wall patch looks like that in Figure 4.27a, which is equivalent to that in Figure 4.27b. The horizontal magnetic field induces a horizontal magnetic dipole moment on the magnetic wall patch.

To obtain the solution for Figure 4.26, we need to solve for the solution of a static electric field in the vicinity of an ellipsoid. Consider the problem shown in Figure 4.28. When the relative permittivity, ϵ_r of the ellipsoid is less than one, the electric field avoids the ellipsoid by skirting around it. The solution of the static electric field around the ellipsoid can be obtained in ellipsoidal coordinates. By letting one of the axes of the ellipsoid shrink to zero, the ellipsoid becomes a disk. Moreover, if we let $\epsilon_r = 0$, no field can penetrate the disk and the requisite solution for Figure 4.26 is obtained.

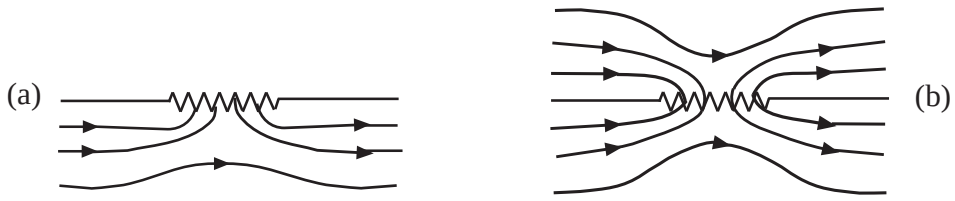


Figure 4.27: Equivalent problem of the **H** field around a waveguide wall partially covered by a small magnetic wall.

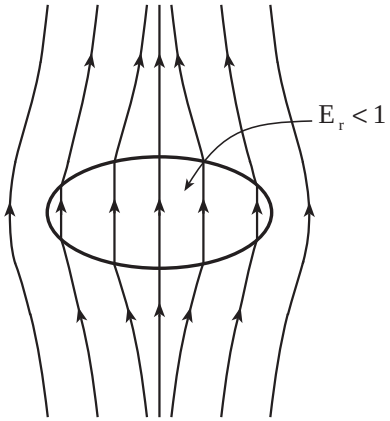


Figure 4.28: Equivalent problem of a static electric field around an ellipsoid. When the relative permittivity becomes zero, and the ellipsoid becomes a disk, the electric field around the disk is the same as the electric around a PMC disk.

By the same token, the solution needed for Figure 4.27 can be obtained by studying the static magnetic field around an ellipsoid with relative permeability μ_r larger than one. In this case, the magnetic field in the vicinity of the ellipsoid is attracted t the ellipsoid as shown in Figure 4.29. When we let the ellipsoid become a disk, and let $\mu_r \to \infty$, we obtain the requisite solution for Figure 4.29. These closed-form solutions can be used to obtain the coupling coefficients for small aperture coupling, as in Bethe coupling, in a waveguide. This work illustrates the genius of Hans Bethe.

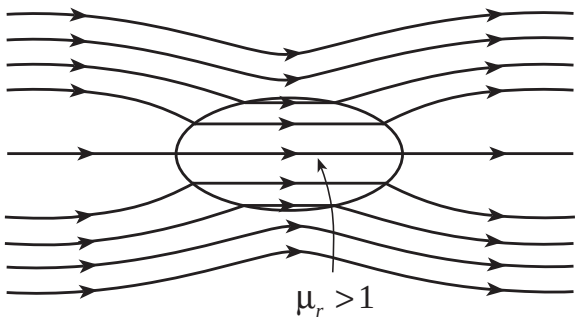


Figure 4.29: Equivalent problem of a static magnetic field around an ellipsoid. When the relative permeability becomes infinite, and the ellipsoid becomes a disk, the magnetic field around the disk is the same as the magnetic around a PMC disk.

Exercises for Chapter 4

Problem 4-1: By making use of reciprocity, prove the validity of Equation (4.1.7a) and (4.1.7b) of the text.

Problem 4-2: Given the Rayleigh quotient

$$\frac{\mathbf{a}^t \cdot \overline{\mathbf{A}} \cdot \mathbf{a}}{\mathbf{a}^t \cdot \overline{\mathbf{B}} \cdot \mathbf{a}},$$

find the solution that will minimize it. What equation does the variational solution solve?

Problem 4-3:

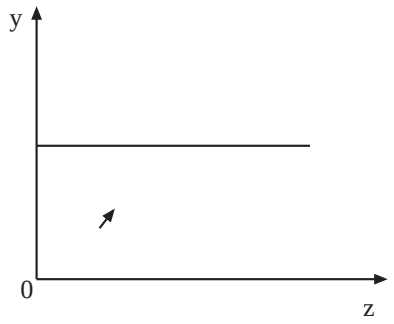


Figure 4.30: Problem 4-3

The dyadic Green's function of a infinitely long waveguide is given by

$$\overline{\mathbf{G}}(\mathbf{r},\mathbf{r}') = \sum_i \pi i \left[\frac{\mathbf{M}_{ei}(\pm k_{hiz},\mathbf{r})\mathbf{M}_{ei}(\mp k_{hiz},\mathbf{r}')}{k_{hiz}k_{his}^2 A_{hi}} + \frac{\mathbf{N}_{ei}(\pm k_{eiz},\mathbf{r})\mathbf{N}_{ei}(\mp k_{eiz},\mathbf{r}')}{k_{eiz}k_{eis}^2 A_{ei}} \right] - \frac{\hat{z}\hat{z}}{k_o^2}\delta(\mathbf{r}-\mathbf{r}'), \quad \begin{matrix} z > z' \\ z < z'. \end{matrix}$$

Using image theorem, find the dyadic Green's function of a waveguide with a shorting plane at $z = 0$.

Problem 4-4:

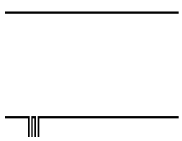


Figure 4.31: Problem 4-4

(a) For the geometry shown where a waveguide is fed by a coaxial probe flushed with the waveguide wall, write down the expression for the input admittance using the variational formula derived in the text assuming that the field at the aperture is that of the coaxial mode.

(b) Now, using mode-matching method, and assuming only a TEM mode in the coaxial waveguide, derive an expression for the input admittance using mode-matching method. Show that this result is the same as that in part (a). (Hint: Do not write out the dyadic Green's function explicitly. Leave it as a symbolic operator.)

Problem 4-5:

(a) For a probe in a waveguide excited by a dipole as shown, show that a variational expression for the input impedance at the base of the probe is [see e.g., Harrington]

$$Z_{in} = -\frac{i\omega\mu}{I^2}\langle \mathbf{J}_p, \overline{\mathbf{G}}, \mathbf{J}_p \rangle.$$

(b) If only the TE_{10} mode is propagating in a rectangular waveguide, we can approximate the dyadic Green's function only with the term associated with the TE_{10} mode. Assume that the probe current is of the form $\mathbf{J}_p = \hat{y}\sin[k(y-d)]$ so that the probe current is zero at the tip of the probe. Use the variational formula above to find an approximation to the probe input impedance with the simplifying approximation on the dyadic Green's

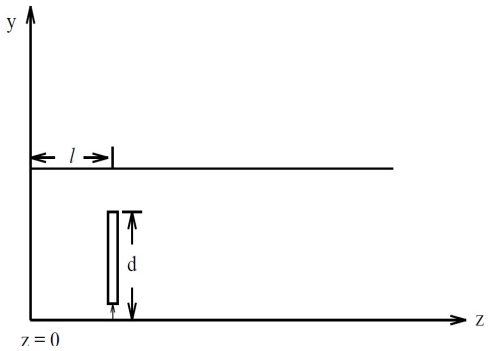


Figure 4.32: Problem 4-5

function. (Note that this approximation is only good for calculating the real part of the probe impedance. This is because the real part of the probe impedance is related to the real power radiated to infinity which is carried by the TE_{10} mode. The inductance of the probe will not be well approximated by this method because the probe inductance is associated with the singular field near the probe which can be well approximated only if we include the higher order evanescent modes in the Green's function.)

Problem 4-6:

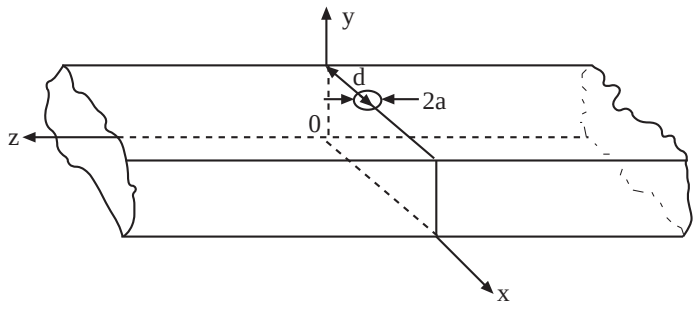


Figure 4.33: Problem 5-2

For a TE_{10} mode propagating in a rectangular waveguide with a small circular aperture of radius a on top of the waveguide, find the polarization, amplitude and phase of the induced electric and magnetic dipole moments at the aperture due to the field in the waveguide.
Problem 4-7: Derive the expressions for the equations for Bethe coupling in (4.4.22), and show that (4.4.23) does solve the equation above it.

Problem 4-8: Establish the equivalence principles I, II, and III using Green's theorem or

Huygens' principle for vector electromagnetic fields. In some of the cases, it will be useful to assume that the dyadic Green's function satisfies certain boundary conditions to reduce the size of the surface integral that needs to be performed.

Problem 4-9: Write down the electrostatic solution of the potential due to an ellipsoid in the presence of a constant electric field. Let one of its axes of the ellipsoid shrink to zero. What should the permittivity of the ellipsoid be in order to get the solution of Figures 4.26 and 4.27

Bibliography

[1] R.E. Collin, *Field Theory of Guided Waves,* IEEE Press, Piscataway, NJ, 1991.

[2] R.E. Collin, *Foundation for Microwave Engineering,* IEEE Press, Piscataway, NJ, 2001.

[3] W.C. Chew, *Waves and Fields in Inhomogeneous Media,* Van Nostrand Reinhold, New York, 1990, reprinted, IEEE Press, Piscataway, NJ, 1995.

[4] R.G. Harrington, *Time Harmonic Electromagnetic Field,* New York: McGraw-Hill, 1961.

[5] J.A. Kong, *Electromagnetic Wave Theory,* EMW Publishing, Cambridge, MA, 2000.

[6] E.C. Jordan and K.G. Balmain, *Electromagnetic Waves and Radiating Systems,* Second Edition. Prentice-Hall, 1968.

[7] W.C. Chew, Z.P. Nie, Q.H. Liu and Y.T. Lo, "Analysis of probe-fed microstrip disk antenna," *IEE Proceedings-H*, vol. 138, no. 2, pp. 185-191, April 1991.

[8] G.A. Deschamps, "Microstrip Microwave Antennas," presented at the 3rd USAF Symposium on Antennas, 1953.

[9] R.E. Munson, "Microstrip Phased Array Antennas, *Proc. of Twenty-Second Symp. on USAF Antenna Research and Development Program*, October 1972.

[10] Y.T. Lo, W.F. Richards, "Theory and experiment on microstrip antennas," *IEEE Trans. Ant. and Propagat.*, vol. AP-27, PP 137-145, March 1979.

[11] W. F. Richards, Y. T. Lo, and D. D. Harrison, "An improved theory for microstrip antennas and applications," *IEEE Trans. Antennas Propag.*, AP-29, pp. 38-46, Jan. 1981.

[12] W. C. Chew, J. A. Kong, and L. C. Shen, "Radiation characteristics of a circular microstrip antenna," *J. Appl. Phys.*, 51, 3907, 1980.

[13] D. M. Pozar, "Considerations for millimeter wave printed antennas, *IEEE Trans. Antennas Propagat.*, vol. AP-31, pp. 740-747, 1983.

[14] W.C. Chew and J.A. Kong, "Resonance of the axial-symmetric modes in microstrip disk resonators," *J. Math. Phys.*, vol. 21, no. 3, pp. 582-591, Mar. 1980.

[15] W.C. Chew and J.A. Kong, "Resonance of nonaxial symmetric modes in circular microstrip disk antenna," *J. Math. Phys.*, vol. 21, no. 10, pp. 2590-2598, Oct. 1980.

[16] W.C. Chew and J.A. Kong, "Asymptotic formula for the resonant frequencies of a circular microstrip antenna," *J. Appl. Phys.*, vol. 52, no. 8, pp. 5365-5369, Aug. 1981.

[17] J.W. Rayleigh, "In finding the correction for the open end of an organ-pipe," *Phil. Trans.*, 161, 77, 1870.

[18] W. Ritz, "ber eine neue Methode zur Lsung gewisser Variationsprobleme der mathematischen Physik," *J. Reine Angew. Math.*, 135, 1-61, 1908.

[19] P.S. Carter, "Circuit relations in radiating systems and applications to antenna problems," *Proc. IRE*, vol. 20, no. 6, pp. 1004-1041, 1932.

[20] K. Kurokawa, "The expansions of electromagnetic. fields in cavities," *IRE Trans. Microwave Theory Tech.*, vol. 6 , no. 2, pp. 178-187, 1958.

[21] H.A. Bethe, "Theory of diffraction by small holes," *Phys. Rev.*, vol. 66, pp. 163-182, 1944.

[22] S. B. Cohn, "Electric polarizability of apertures of arbitrary shape," *Proc. IRE,* vol. 40, pp. 1069-1071, 1952.

[23] N. Wiener, E. hopf, "Ueber eine Klasses singulärer Integralgleichungen," *S.-B. Deutsch. Akad. Wiss. Berlin KI. Math. Phys. Tech.*, pp. 696-706 1931.

[24] W. C. Chew and J. A. Kong, "Asymptotic formula for the capacitance of two oppositely charged discs, *Math. Proc. Camb. Philos. Soc.*, 89, pp. 373-384, 1981.

[25] T.N. Anderson, "Directional coupler design nomograms," *Microwave J.*, vol. 2, pp. 34-38, May 1959.

Chapter 5

Discontinuities in Waveguides

Discontinuities in waveguides cannot be avoided. When two waveguides of different sizes are connected together, the junction does not form a smooth transition introducing a discontinuity. Such discontinuities will reflect the waveguide mode. Moreover, infinitely many modes are needed at the discontinuity in order to match the boundary condition. But most of the modes "excited" by the presence of discontinuities are evanescent modes. They do not convect energy away from a junction. Hence, the higher-order modes serve to store energy. When most of the energy stored is in the magnetic field, the junction discontinuity effect is inductive, while if most of the stored energy is in the electric field, the effect is capacitive. Hence, simple equivalent models for the junction can be either a capacitor or an inductor. However, the calculation of these inductive and capacitive effect requires the use of some *tour de force* calculations, a subject that we will discuss in this Chapter.

However, waveguide junctions appear in more complex forms when applied to circulators and T junctions [1]. The analysis of these waveguide junctions is not amenable to analytic methods, and hence, much numerical methods have been invoked in analyzing them as is seen from the reference list in this chapter.

5.1 Transmission Line Equivalence of Waveguide

A wealth of engineering knowledge is built on circuit theory and transmission line theory. Due to our familiarity with circuit theory and transmission line theory, it is useful to relate the propagation of modes in a waveguide to transmission line theory [2–4].

For a hollow waveguide of arbitrary shape, it is easy to show that the ratio of the transverse components of the electric and magnetic fields are

$$Z^{TE} = \frac{\hat{z} \times \mathbf{E}_s}{\mathbf{H}_s} = \frac{\omega\mu_0}{k_z}, \quad Z^{TM} = \frac{\hat{z} \times \mathbf{E}_s}{\mathbf{H}_s} = \frac{k_z}{\omega\epsilon_0}. \tag{5.1.1}$$

These are called the wave impedances of a waveguide. Notice that they are mode dependent. If they are to be likened to the characteristic impedance of a transmission line, that it is natural to define a voltage which is proportional to $\mathbf{E}_s$ and a current which is proportional

to $\mathbf{H}_s$ for a transmission line equivalence. However, this proportionality constant should be chosen so that the time average power given by $\Re e[VI^*/2]$, is the same as the power flowing down the waveguide.

As an example, a mode propagating down a waveguide in the positive z direction may be expressed as [1]

$$\mathbf{E}_s = C_+ \mathbf{e}_s e^{-jk_z z} \tag{5.1.2a}$$

$$\mathbf{H}_s = C_+ \mathbf{h}_s e^{-jk_z z} \tag{5.1.2b}$$

where $\hat{z} \times \mathbf{e}_s = Z_w \mathbf{h}_s$ where Z_w is wave impedance of the particular mode under discussion. For a mode propagating in the negative z direction, the fields are given by

$$\mathbf{E}_s = C_- \mathbf{e}_s e^{jk_z z}, \tag{5.1.3a}$$

$$\mathbf{H}_s = -C_- \mathbf{h}_s e^{jk_z z}. \tag{5.1.3b}$$

The expression for the field can be replaced by equivalent voltage and current waves given by

$$V = V_+ e^{-jk_z z} + V_- e^{jk_z z}, \tag{5.1.4a}$$

$$I = I_+ e^{-jk_z z} - I_- e^{jk_z z}, \tag{5.1.4b}$$

where $V_+ = K_1 C_+$, $V_- = K_1 C_-$, $I_+ = K_2 C_+$, and $I_- = K_2 C_-$.

To ensure that the equivalent circuit carries the same power, we require that

$$\frac{1}{2} V_+ I_+^* = \frac{|C_+|^2}{2} \int_s (\mathbf{e}_s \times \mathbf{h}_s^*) \cdot \hat{z} dS \tag{5.1.5}$$

or that

$$K_1 K_2^* = \int_s (\mathbf{e}_s \times \mathbf{h}_s^*) \cdot \hat{z} dS. \tag{5.1.6}$$

The characteristic impedance of this equivalent transmission line is

$$Z_c = V_+ / I_+ = K_1 / K_2. \tag{5.1.7}$$

Notice that there exists no unique way of choosing K_1 and K_2. Therefore, $\mathbf{e}_s$ and $\mathbf{h}_s$ in (5.1.6) can be normalized such that the right hand side of (5.1.6) is 1. In this case, we require $K_1 K_2^* = 1$. Also, the characteristic impedance of the transmission line equivalence in (5.1.7) can be made equal to 1 since all Smith charts are given for normalized impedance values. Alternatively, one can choose $K_1/K_2 = Z_w$.

Armed with the transmission line model of a waveguide, much of the tools that are found in transmission line theory like Smith chart, and impedance matching techniques can be used. However, connecting two waveguides of different wave impedances and hence sizes gives rise to waveguide discontinuities. These discontinuities scatters a propagating mode into higher modes which are evanescent. Hence, they give rise to localized stored energy in their vicinity. Depending on if the stored energy is of electric or magnetic type, these discontinuities are often modelled by shunt capacitances and inductances which can be calculated. At other times, one may deliberately introduce discontinuities in a waveguide call diaphragms to give rise to a shunt capacitance or inductance for the purpose of matching to the load or designing filters.

[1] We shall use $e^{j\omega t}$ time convention for agreement with the time convention of circuit theory.

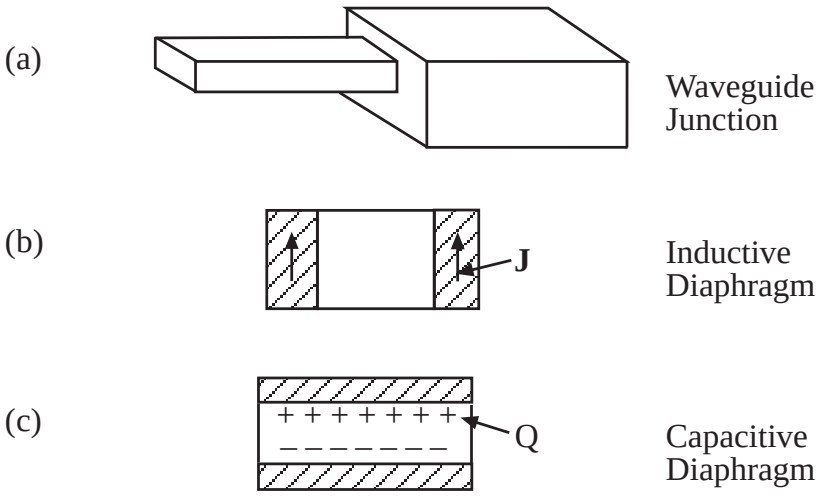


Figure 5.1: Different kinds of waveguide junctions. The inductive and capacitive diaphragms are for TE_{10} mode.

5.2 Waveguide Junction

Waveguide discontinuities have been studied by a number of workers [5,6]. Before the advent of digital computers, analytic and variational methods were used to study these discontinuities. With the advent of high speed digital computers, these discontinuities are routinely studied with numerical method, requiring the solution of a large system of linear algebraic equation (see reference list).

Here, we propose to characterize the scattering by a waveguide discontinuity by reflection and transmission operators using mode matching. Then, the generalization to an arbitrary number of discontinuities (see Figure 5.3) becomes routine. The waveguide discontinuity problem has also been studied previously, but a more generalized formulation, valid for arbitrarily-shaped waveguides, is presented here. The formulation is then related to the integral equation formulation, and a new criterion for convergence of the method is given and corroborated by numerical simulation.

Waveguide discontinuities arise in many occasions, for example, at a waveguide junction (Figure 5.1(a)), or when diaphragms are added to change the phase and amplitude of a wave propagating through a waveguide. A diaphragm can be of the inductive type as shown in Figure 1(b), or of the capacitive type as shown in Figure 1(c). An inductive diaphragm induces stored magnetic energy while a capacitive diaphragm induces stored electric energy. Moreover, a capacitive diaphragm changes the phase of a mode differently from an inductive diaphragm. Hence, they can be used as tuning elements for matching purposes in waveguides.

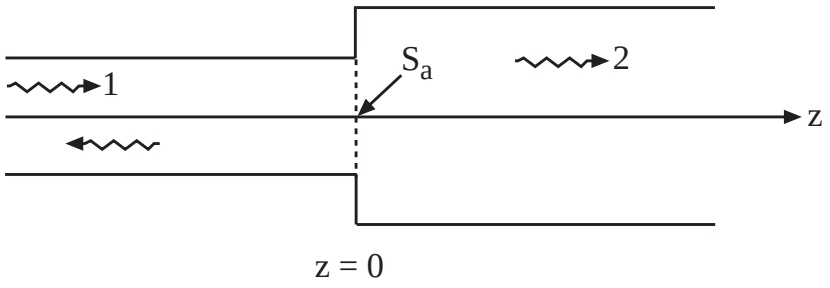


Figure 5.2: Mode matching at a general waveguide junction.

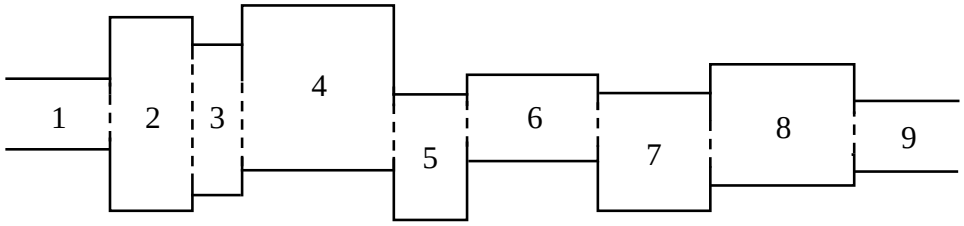


Figure 5.3: A waveguide with many junction discontinuities.

5.2.1 Mode Matching–Eigenmode Expansion Method

Waveguide discontinuity problems are generally solved by the method of mode matching.[2] In this method, the waves in waveguide 1 and waveguide 2 (see Figure 5.2) are expanded in terms of the modes of the waveguide. The amplitudes of the modes are found by matching the boundary conditions at the discontinuity. The field structure is quite complicated at the waveguide junction as the field has to bend to adjust to the boundary conditions. A multitude of modes is needed for matching the boundary condition. However, most of these modes are evanescent giving rise to localized field at the junction discontinuity. The local field can store electric field energy giving rise to a capacitive effect. If it stores magnetic field energy, it gives rise to an inductive effect.

Another important physics that happens at a waveguide junction is that if there is a pure waveguide mode impinging on the junction, due to the need to match boundary conditions, all higher order modes are excited at the junction giving rise to infinitely many reflected modes and transmitted modes. This is the physics of mode conversion.

In order to facilitate the ease for mode matching, we first develop a succinct and compact notation that can encompass all modes in a waveguide. If we have a superposition of modes in an arbitrarily-shaped, hollow waveguide, the z components of the fields can be written as

$$H_z = \sum_i H_i \psi_{hi}(\mathbf{r}_s) e^{ik_{hiz}z}, \qquad \text{(TE modes)} \tag{5.2.1}$$

[2]This section follows the development in [20].

$$E_z = \sum_i E_i \psi_{ei}(\mathbf{r}_s) e^{ik_{eiz}z}. \qquad \text{(TM modes)} \tag{5.2.2}$$

Here, $\psi_{hi}(\mathbf{r}_s)$ and $\psi_{ei}(\mathbf{r}_s)$ are solutions to the equation $(\nabla_s^2 + k_{is}^2)\psi_i(\mathbf{r}_s) = 0$ with Neumann and Dirichlet boundary conditions, respectively, on the waveguide wall. $\nabla_s^2 = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2}$, and k_{iz}^2 is the eigenvalue corresponding to the eigenmode $\psi_i(\mathbf{r}_s)$. The subscripts h and e are used to denote the Neumann problem (TE) and the Dirichlet problem (TM), respectively. For each mode, the transverse components of the fields can be found via the equations derivable from Maxwell's equations similar to those in Chapter 3:

$$\mathbf{e}_{is} = \frac{1}{k_{eis}^2}\frac{\partial}{\partial z}\nabla_s e_{iz} - \frac{i\omega\mu}{k_{his}^2}\hat{z} \times \nabla_s h_{iz}, \tag{5.2.3}$$

$$\mathbf{h}_{is} = \frac{1}{k_{his}^2}\frac{\partial}{\partial z}\nabla_s h_{iz} + \frac{i\omega\epsilon}{k_{eis}^2}\hat{z} \times \nabla_s e_{iz}, \tag{5.2.4}$$

where e_{iz} and h_{iz} are the z components of the electric field and magnetic field respectively for each individual mode. Consequently, $\mathbf{e}_{is}$ and $\mathbf{h}_{is}$ are the transverse components of the electric field and magnetic field respectively for each individual mode. Furthermore, $k_{eis}^2 = k^2 - k_{eiz}^2$ and $k_{his}^2 = k^2 - k_{hiz}^2$.

Using the above, we deduce that

$$\begin{aligned}\mathbf{E}_s &= \sum_i \mathbf{e}_{is} \\ &= \sum_i \left[\frac{E_i}{k_{eis}^2} ik_{eiz}\nabla_s\psi_{ei}(\mathbf{r}_s)e^{ik_{eiz}z} - \frac{i\omega\mu H_i}{k_{his}^2}\hat{z} \times \nabla_s\psi_{hi}(\mathbf{r}_s)e^{ik_{hiz}z}\right],\end{aligned} \tag{5.2.5}$$

$$\begin{aligned}\mathbf{H}_s &= \sum_i \mathbf{h}_{is} \\ &= \sum_i \left[\frac{i\omega\epsilon E_i}{k_{eis}^2}\hat{z} \times \nabla_s\psi_{ei}(\mathbf{r}_s)e^{ik_{eiz}z} + \frac{H_i}{k_{his}^2} ik_{hiz}\nabla_s\psi_{hi}(\mathbf{r}_s)e^{ik_{hiz}z}\right].\end{aligned} \tag{5.2.6}$$

We see that $\hat{z} \times \mathbf{H}_s$ is closely related to $\mathbf{E}_s$, i.e.,

$$\hat{z} \times \mathbf{H}_s = \sum_i \left[-\frac{i\omega\epsilon E_i}{k_{eis}^2}\nabla_s\psi_{ei}(\mathbf{r}_s)e^{ik_{eiz}z} + \frac{H_i}{k_{his}^2} ik_{hiz}\hat{z} \times \nabla_s\psi_{hi}(\mathbf{r}_s)e^{ik_{hiz}z}\right]. \tag{5.2.7}$$

Using vector notation, we can write $\mathbf{E}_s$ more compactly as

$$\begin{aligned}\mathbf{E}_s &= \sum_i \begin{bmatrix} \frac{\nabla_s\psi_{ei}(\mathbf{r}_s)}{k_{eis}^2} \\ \frac{\hat{z}\times\nabla_s\psi_{hi}(\mathbf{r}_s)}{k_{his}^2} \end{bmatrix}^t \begin{bmatrix} e^{ik_{eiz}z} & 0 \\ 0 & e^{ik_{hiz}z} \end{bmatrix} \begin{bmatrix} ik_{eiz}E_i \\ -i\omega\mu H_i \end{bmatrix} \\ &= \sum_i \underbrace{\overline{\psi}_i^t(\mathbf{r}_s)}_{2\times2} \cdot \underbrace{e^{i\overline{\mathbf{k}}_{iz}z}}_{2\times2} \cdot \underbrace{\mathbf{e}_i}_{2\times1},\end{aligned} \tag{5.2.8a}$$

where

$$\overline{\psi}_i(\mathbf{r}_s) = \begin{bmatrix} \nabla_s \psi_{ei}(\mathbf{r}_s)/k_{eis}^2 \\ \hat{z} \times \nabla_s \psi_{hi}(\mathbf{r}_s)/k_{his}^2 \end{bmatrix}, \qquad e^{i\overline{\mathbf{k}}_{iz}z} = \begin{bmatrix} e^{ik_{eiz}z} & 0 \\ 0 & e^{ik_{hiz}z} \end{bmatrix}, \tag{5.2.8b}$$

and

$$\mathbf{e}_i = \begin{bmatrix} ik_{eiz}E_i \\ -i\omega\mu H_i \end{bmatrix}. \tag{5.2.8c}$$

Similarly, we can write

$$\hat{z} \times \mathbf{H}_s = -\sum_i \underbrace{\overline{\psi}_i^t}_{2\times 2}(\mathbf{r}_s) \cdot \underbrace{\overline{\mathbf{g}}_i}_{2\times 2} \cdot \underbrace{e^{i\overline{\mathbf{k}}_{iz}z}}_{2\times 2} \cdot \underbrace{\mathbf{e}_i}_{2\times 1}, \tag{5.2.9}$$

where

$$\overline{\mathbf{g}}_i = \begin{bmatrix} \omega\epsilon/k_{eiz} & 0 \\ 0 & k_{hiz/\omega\mu} \end{bmatrix}.$$

Using the fact that one can write

$$\sum_n a_n \lambda_n b_n = \mathbf{a}^t \cdot \overline{\lambda} \cdot \mathbf{b}, \tag{5.2.10a}$$

where

$$\mathbf{a}^t = [a_1, a_2, a_3, \ldots], \qquad \mathbf{b}^t = [b_1, b_2, b_3, \ldots], \tag{5.2.10b}$$

$$\overline{\lambda} = \begin{bmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \lambda_3 & \\ & & & \ddots \end{bmatrix}, \tag{5.2.10c}$$

we can rewrite (5.2.8a) and (5.2.9) compactly as

$$\mathbf{E}_s = \underbrace{\overline{\mathbf{\Psi}}^t(\mathbf{r}_s)}_{2\times\infty} \cdot \underbrace{e^{i\overline{\mathbf{K}}z}}_{\infty\times\infty} \cdot \underbrace{\mathbf{e}}_{\infty\times 1}, \tag{5.2.11a}$$

$$\hat{z} \times \mathbf{H}_s = -\underbrace{\overline{\mathbf{\Psi}}^t(\mathbf{r}_s)}_{2\times\infty} \cdot \underbrace{\overline{\mathbf{G}}}_{\infty\times\infty} \cdot \underbrace{e^{i\overline{\mathbf{K}}z}}_{\infty\times\infty} \cdot \underbrace{\mathbf{e}}_{\infty\times 1}, \tag{5.2.11b}$$

where

$$\overline{\mathbf{\Psi}}^t(\mathbf{r}_s) = \left[\overline{\psi}_1^t(\mathbf{r}_s), \overline{\psi}_2^t(\mathbf{r}_s), \overline{\psi}_3^t(\mathbf{r}_s), \ldots\right], \tag{5.2.12a}$$

$$e^{i\overline{\mathbf{K}}z} = \begin{bmatrix} e^{i\overline{\mathbf{k}}_{1z}z} & & & \\ & e^{i\overline{\mathbf{k}}_{2z}z} & & \\ & & e^{i\overline{\mathbf{k}}_{3z}z} & \\ & & & \ddots \end{bmatrix}, \tag{5.2.12b}$$

$$\overline{\mathbf{G}} = \begin{bmatrix} \overline{\mathbf{g}}_1 & & & \\ & \overline{\mathbf{g}}_2 & & \\ & & \overline{\mathbf{g}}_3 & \\ & & & \ddots \end{bmatrix}, \tag{5.2.12c}$$

$$\mathbf{e}^t = [\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3, \ldots]. \tag{5.2.12d}$$

The above order of the elements of the matrices may be rearranged for bookkeeping purposes. For instance, one may prefer to group the TE modes and the TM modes separately rather than as a couplet in the above. In this case, for example,

$$\overline{\mathbf{G}} = \begin{bmatrix} g_1^{TM} & & & & & & \\ & g_2^{TM} & & & & & \\ & & \ddots & & & & \\ & & & g_1^{TE} & & & \\ & & & & g_2^{TE} & & \\ & & & & & \ddots & \end{bmatrix}. \tag{5.2.13a}$$

where,

$$g_i^{TM} = \omega\epsilon/k_{eiz}, \qquad g_i^{TE} = k_{hiz}/\omega\mu. \tag{5.2.13b}$$

Similar rearrangement need be done for (5.2.12a), (5.2.12b) and (5.2.12d) if this is in fact preferred.

With the compact way to write the transverse components of the fields as in (5.2.11a), we can write down with physical intuition the general solution for waveguides 1 and 2. In waveguide 1, there will be an incident as well as a reflected wave. Hence, we have

$$\mathbf{E}_{1s} = \overline{\mathbf{\Psi}}_1^t(\mathbf{r}_s) \cdot \left(e^{i\overline{\mathbf{K}}_1 z} \cdot \mathbf{e} + e^{-i\overline{\mathbf{K}}_1 z} \cdot \mathbf{e}_R\right), \tag{5.2.14a}$$

$$-\hat{z} \times \mathbf{H}_{1s} = \overline{\mathbf{\Psi}}_1^t(\mathbf{r}_s) \cdot \overline{\mathbf{G}}_1 \cdot \left(e^{i\overline{\mathbf{K}}_1 z} \cdot \mathbf{e} - e^{-i\overline{\mathbf{K}}_1 z} \cdot \mathbf{e}_R\right). \tag{5.2.14b}$$

In waveguide 2, we can write the fields as

$$\mathbf{E}_{2s} = \overline{\mathbf{\Psi}}_2^t(\mathbf{r}_s) \cdot e^{i\overline{\mathbf{K}}_2 z} \cdot \mathbf{e}_T, \tag{5.2.14c}$$

$$-\hat{z} \times \mathbf{H}_{2s} = \overline{\mathbf{\Psi}}_2^t(\mathbf{r}_s) \cdot \overline{\mathbf{G}}_2 \cdot e^{i\overline{\mathbf{K}}_2 z} \cdot \mathbf{e}_T. \tag{5.2.14d}$$

We can define $\overline{\mathbf{R}}$ and $\overline{\mathbf{T}}$ which are reflection and transmission operators such that

$$\mathbf{e}_R = \overline{\mathbf{R}} \cdot \mathbf{e}, \qquad \mathbf{e}_T = \overline{\mathbf{T}} \cdot \mathbf{e}. \tag{5.2.15}$$

These operators entail the physics of mode conversion at a waveguide junction. They are infinite dimensional matrices, or operators. Then

$$\mathbf{E}_{1s} = \overline{\mathbf{\Psi}}_1^t(\mathbf{r}_s) \cdot \left(e^{i\overline{\mathbf{K}}_1 z} + e^{-i\overline{\mathbf{K}}_1 z} \cdot \overline{\mathbf{R}}\right) \cdot \mathbf{e}, \tag{5.2.16a}$$

$$-\hat{z} \times \mathbf{H}_{1s} = \overline{\mathbf{\Psi}}_1^t(\mathbf{r}_s) \cdot \overline{\mathbf{G}}_1 \cdot \left(e^{i\overline{\mathbf{K}}_1 z} - e^{-i\overline{\mathbf{K}}_1 z} \cdot \overline{\mathbf{R}}\right) \cdot \mathbf{e}, \tag{5.2.16b}$$

and

$$\mathbf{E}_{2s} = \overline{\mathbf{\Psi}}_2^t(\mathbf{r}_s) \cdot e^{i\overline{\mathbf{K}}_2 z} \cdot \overline{\mathbf{T}} \cdot \mathbf{e}, \tag{5.2.17a}$$

$$-\hat{z} \times \mathbf{H}_{2s} = \overline{\mathbf{\Psi}}_2^t(\mathbf{r}_s) \cdot \overline{\mathbf{G}}_2 \cdot e^{i\overline{\mathbf{K}}_2 z} \cdot \overline{\mathbf{T}} \cdot \mathbf{e}. \tag{5.2.17b}$$

From the continuity of the tangential electric field, we have

$$\overline{\boldsymbol{\Psi}}_1^t(\mathbf{r}_s)\cdot\left(\overline{\mathbf{I}}+\overline{\mathbf{R}}\right)\cdot\mathbf{e}=\overline{\boldsymbol{\Psi}}_2^t(\mathbf{r}_s)\cdot\overline{\mathbf{T}}\cdot\mathbf{e},\quad \mathbf{r}_s\in S_a, \tag{5.2.18}$$

Furthermore, we require that tangential electric fields equal zero for $\mathbf{r}_s\notin S_a$ at the plane of the discontinuity.

The continuity of the magnetic field implies

$$\overline{\boldsymbol{\Psi}}_1^t(\mathbf{r}_s)\cdot\overline{\mathbf{G}}_1\cdot\left(\overline{\mathbf{I}}-\overline{\mathbf{R}}\right)\cdot\mathbf{e}=\overline{\boldsymbol{\Psi}}_2^t(\mathbf{r}_s)\cdot\overline{\mathbf{G}}_2\cdot\overline{\mathbf{T}}\cdot\mathbf{e},\quad \mathbf{r}_s\in S_a. \tag{5.2.19}$$

The unknowns to be sought here are $\overline{\mathbf{R}}$ and $\overline{\mathbf{T}}$.

The aperture S_a can be thought of as the cross-section of a waveguide with modes $\overline{\boldsymbol{\Psi}}_a^t(\mathbf{r}_s)$. We can then equate (5.2.18) to

$$\underbrace{\overline{\boldsymbol{\Psi}}_1^t(\mathbf{r}_s)}_{2\times\infty}\cdot\underbrace{\left(\overline{\mathbf{I}}+\overline{\mathbf{R}}\right)}_{\infty\times\infty}\cdot\underbrace{\mathbf{e}}_{\infty\times1}=\underbrace{\overline{\boldsymbol{\Psi}}_2^t(\mathbf{r}_s)}_{2\times\infty}\cdot\underbrace{\overline{\mathbf{T}}}_{\infty\times\infty}\cdot\underbrace{\mathbf{e}}_{\infty\times1}=\underbrace{\overline{\boldsymbol{\Psi}}_a^t(\mathbf{r}_s)}_{\infty\times N}\cdot\underbrace{\mathbf{a}}_{N\times1},\quad \mathbf{r}_s\in S_a, \tag{5.2.20}$$

where $\overline{\boldsymbol{\Psi}}_a^t(\mathbf{r}_s)=0$, $\mathbf{r}_s\notin S_a$. In this way, (5.2.18) plus the auxiliary condition after (5.2.18) are satisfied. For practical purposes, we choose $\mathbf{a}$ to be a vector of length N even though in theory, the summation should be infinite. This is obviated by the integral equation formulation in the next section. Hence, the dot product for the last term in (5.2.20) implies N-term summation while the rest of the dot products imply infinite summation. From this point onward, we shall denote inner products with infinite summations with double-dot products, and leave the single-dot product for inner product with a finite summation. Note that in general, the length of the vector $\mathbf{e}$ is determined by the number of incident modes. We shall assume that $\mathbf{e}$ is of infinite length, but the following formulation is also valid for $\mathbf{e}$ of finite length. It is to be noted that in (5.2.20), any complete set of basis functions instead of waveguide modes that can be used to expand the aperture field for $\mathbf{r}_s\in S_a$ will also suffice.

Multiplying equation (5.2.20) by $\overline{\boldsymbol{\Psi}}_1(\mathbf{r}_s)$ and integrate over $\mathbf{r}_s$, we get

$$\underbrace{\overline{\mathbf{D}}_1}_{\infty\times\infty}:\underbrace{\left(\overline{\mathbf{I}}+\overline{\mathbf{R}}\right)}_{\infty\times\infty}:\underbrace{\mathbf{e}}_{\infty\times1}=\underbrace{\overline{\mathbf{L}}_{1a}}_{\infty\times N}\cdot\underbrace{\mathbf{a}}_{N\times1}, \tag{5.2.21a}$$

where we define

$$\underbrace{\overline{\mathbf{D}}_i}_{\infty\times\infty}=\left\langle\underbrace{\overline{\boldsymbol{\Psi}}_i}_{\infty\times2},\underbrace{\overline{\boldsymbol{\Psi}}_i^t}_{2\times\infty}\right\rangle,\qquad\underbrace{\overline{\mathbf{L}}_{ia}}_{\infty\times N}=\left\langle\underbrace{\overline{\boldsymbol{\Psi}}_i}_{\infty\times2},\underbrace{\overline{\boldsymbol{\Psi}}_a^t}_{2\times N}\right\rangle. \tag{5.2.21b}$$

$\overline{\mathbf{D}}_i$ is diagonal due to mode orthogonality while $\overline{\mathbf{L}}_{ia}$ is in general non-square, infinite dimensional by N dimensional matrix. $\overline{\mathbf{D}}_i$ can be made into an identity matrix if the modes are made orthonormal. Similarly, multiplying (5.2.20) by $\overline{\boldsymbol{\Psi}}_2(\mathbf{r}_s)$ and integrating over $\mathbf{r}_s$, we have

$$\overline{\mathbf{D}}_2:\overline{\mathbf{T}}:\mathbf{e}=\overline{\mathbf{L}}_{2a}\cdot\mathbf{a}. \tag{5.2.22}$$

From (5.2.21a) and (5.2.22), we deduce that

$$\overline{\mathbf{R}}:\mathbf{e}=\overline{\mathbf{D}}_1^{-1}:\overline{\mathbf{L}}_{1a}\cdot\mathbf{a}-\mathbf{e}, \tag{5.2.23a}$$

$$\overline{\mathbf{T}} : \mathbf{e} = \overline{\mathbf{D}}_2^{-1} : \overline{\mathbf{L}}_{2a} \cdot \mathbf{a}. \tag{5.2.23b}$$

The above allows us to express the unknowns $\overline{\mathbf{R}}$ and $\overline{\mathbf{T}}$ in terms of the new unknown $\mathbf{a}$. As $\overline{\mathbf{D}}_i$ is diagonal, $\overline{\mathbf{D}}_i^{-1}$ is easily found even though it is infinite dimensional. In the above, (5.2.19) ensures the continuity of tangential magnetic field. To this end, upon substituting (5.2.23a) into (5.2.19), and rearranging terms, we have

$$\begin{aligned} 2\overline{\boldsymbol{\Psi}}_1^t(\mathbf{r}_s) : \overline{\mathbf{G}}_1 : \mathbf{e} =& \overline{\boldsymbol{\Psi}}_1^t(\mathbf{r}_s) : \overline{\mathbf{G}}_1 : \overline{\mathbf{D}}_1^{-1} : \overline{\mathbf{L}}_{1a} \cdot \mathbf{a} \\ &+ \overline{\boldsymbol{\Psi}}_2^t(\mathbf{r}_s) : \overline{\mathbf{G}}_2 : \overline{\mathbf{D}}_2^{-1} : \overline{\mathbf{L}}_{2a} \cdot \mathbf{a}, \qquad \mathbf{r}_s \in S_a. \end{aligned} \tag{5.2.24}$$

Remember that the above is derived from the continuity of the magnetic field which needs to be imposed only on S_a. Hence, we should weight the above equation with functions whose support is over S_a. Weighting the above equation by $\overline{\boldsymbol{\Psi}}_a(\mathbf{r}_s)$ where $\overline{\boldsymbol{\Psi}}_a(\mathbf{r}_s)$ is a vector of length N, we have, after using the definition for $\overline{\mathbf{L}}_{ia}$ given in (5.2.21b),

$$2\overline{\mathbf{L}}_{1a}^t : \overline{\mathbf{G}}_1 : \mathbf{e} = \left(\overline{\mathbf{L}}_{1a}^t : \overline{\mathbf{G}}_1 : \overline{\mathbf{D}}_1^{-1} : \overline{\mathbf{L}}_{1a} + \overline{\mathbf{L}}_{2a}^t : \overline{\mathbf{G}}_2 : \overline{\mathbf{D}}_2^{-1} : \overline{\mathbf{L}}_{2a} \right) \cdot \mathbf{a}. \tag{5.2.25}$$

Note that in general, $\overline{\mathbf{L}}_{ia}$ is nonsquare, but $\overline{\mathbf{L}}_{ia}^t : \overline{\mathbf{G}}_i : \overline{\mathbf{D}}_i^{-1} : \overline{\mathbf{L}}_{ia}$ is an $N \times N$ square matrix. Now, we can solve (5.2.25) (which is unlike (5.2.24)) for $\mathbf{a}$ giving

$$\mathbf{a} = \left(\overline{\mathbf{L}}_{1a}^t : \overline{\mathbf{G}}_1 : \overline{\mathbf{D}}_1^{-1} : \overline{\mathbf{L}}_{1a} + \overline{\mathbf{L}}_{2a}^t : \overline{\mathbf{G}}_2 : \overline{\mathbf{D}}_2^{-1} : \overline{\mathbf{L}}_{2a} \right)^{-1} \cdot 2 \left(\overline{\mathbf{L}}_{1a}^t : \overline{\mathbf{G}}_1 \right) : \mathbf{e}. \tag{5.2.26}$$

Substituting (5.2.26) in (5.2.23a), we can solve for $\overline{\mathbf{R}}$ and $\overline{\mathbf{T}}$ giving

$$\begin{aligned} \overline{\mathbf{R}} = \overline{\mathbf{D}}_1^{-1} : \overline{\mathbf{L}}_{1a} \cdot & \left(\overline{\mathbf{L}}_{1a}^t : \overline{\mathbf{G}}_1 : \overline{\mathbf{D}}_1^{-1} : \overline{\mathbf{L}}_{1a} \right. \\ & \left. + \overline{\mathbf{L}}_{2a}^t : \overline{\mathbf{G}}_2 : \overline{\mathbf{D}}_2^{-1} : \overline{\mathbf{L}}_{2a} \right)^{-1} \cdot 2\overline{\mathbf{L}}_{1a}^t : \overline{\mathbf{G}}_1 - \overline{\mathbf{I}}, \end{aligned} \tag{5.2.27a}$$

$$\begin{aligned} \overline{\mathbf{T}} = \overline{\mathbf{D}}_2^{-1} : \overline{\mathbf{L}}_{2a} \cdot & \left(\overline{\mathbf{L}}_{1a}^t : \overline{\mathbf{G}}_1 : \overline{\mathbf{D}}_1^{-1} : \overline{\mathbf{L}}_{1a} \right. \\ & \left. + \overline{\mathbf{L}}_{2a}^t : \overline{\mathbf{G}}_2 : \overline{\mathbf{D}}_2^{-1} : \overline{\mathbf{L}}_{2a} \right)^{-1} \cdot 2\overline{\mathbf{L}}_{1a}^t : \overline{\mathbf{G}}_1. \end{aligned} \tag{5.2.27b}$$

The above expressions simplify further with the following assumptions. If S_a is as large as the smaller waveguide, then $\overline{\boldsymbol{\psi}}_a(\mathbf{r}_s) = \overline{\boldsymbol{\psi}}_1(\mathbf{r}_s)$, $\overline{\mathbf{L}}_{ia} = \overline{\mathbf{L}}_{i1}$ and $\overline{\mathbf{L}}_{11} = \overline{\mathbf{D}}_1$. Consequently, the above simplifies to

$$\begin{aligned} \overline{\mathbf{R}} &= \left(\overline{\mathbf{D}}_1 \cdot \overline{\mathbf{G}}_1 + \overline{\mathbf{L}}_{21}^t : \overline{\mathbf{G}}_2 : \overline{\mathbf{D}}_2^{-1} : \overline{\mathbf{L}}_{21} \right)^{-1} \cdot 2\overline{\mathbf{D}}_1 \cdot \overline{\mathbf{G}}_1 - \overline{\mathbf{I}} \\ &= \left(\overline{\mathbf{D}}_1 \cdot \overline{\mathbf{G}}_1 + \overline{\mathbf{L}}_{21}^t : \overline{\mathbf{G}}_2 : \overline{\mathbf{D}}_2^{-1} : \overline{\mathbf{L}}_{21} \right)^{-1} \\ &\quad \cdot \left(\overline{\mathbf{D}}_1 \cdot \overline{\mathbf{G}}_1 - \overline{\mathbf{L}}_{21}^t : \overline{\mathbf{G}}_2 : \overline{\mathbf{D}}_2^{-1} : \overline{\mathbf{L}}_{21} \right), \end{aligned} \tag{5.2.28a}$$

$$\overline{\mathbf{T}} = \overline{\mathbf{D}}_2^{-1} : \overline{\mathbf{L}}_{21} \cdot \left(\overline{\mathbf{D}}_1 \cdot \overline{\mathbf{G}}_1 + \overline{\mathbf{L}}_{21}^t : \overline{\mathbf{G}}_2 : \overline{\mathbf{D}}_2^{-1} : \overline{\mathbf{L}}_{21} \right)^{-1} \cdot 2\overline{\mathbf{D}}_1 \cdot \overline{\mathbf{G}}_1. \tag{5.2.28b}$$

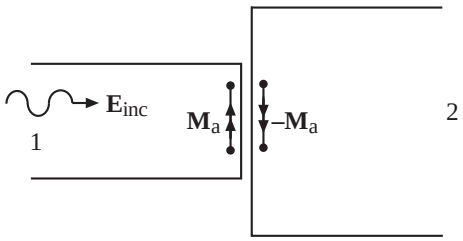


Figure 5.4: Equivalent problem for a waveguide discontinuity.

This is the case for the absence of the diaphragm. Note that $\overline{\mathbf{R}}$ is now $N \times N$ and $\overline{\mathbf{T}}$ is $\infty \times N$. This follows from that the basis functions used at the aperture is orthogonal to the higher order modes in waveguide 1, and hence, would not excite them.

If waveguides 1 and 2 are identical, but a diaphragm is present, (5.2.27a) simplifies to

$$\overline{\mathbf{R}} = \overline{\mathbf{D}}_1^{-1} : \overline{\mathbf{L}}_{1a} \cdot \left(\overline{\mathbf{L}}_{1a}^t : \overline{\mathbf{G}}_1 : \overline{\mathbf{D}}_1^{-1} : \overline{\mathbf{L}}_{1a} \right)^{-1} \cdot \overline{\mathbf{L}}_{1a}^t : \overline{\mathbf{G}}_1 - \overline{\mathbf{I}}, \tag{5.2.29a}$$

$$\overline{\mathbf{T}} = \overline{\mathbf{D}}_1^{-1} : \overline{\mathbf{L}}_{1a} \cdot \left(\overline{\mathbf{L}}_{1a}^t : \overline{\mathbf{G}}_1 : \overline{\mathbf{D}}_1^{-1} : \overline{\mathbf{L}}_{1a} \right)^{-1} \cdot \overline{\mathbf{L}}_{1a}^t : \overline{\mathbf{G}}_1. \tag{5.2.29b}$$

In this case, $\overline{\mathbf{I}} + \overline{\mathbf{R}} = \overline{\mathbf{T}}$. Here, $\overline{\mathbf{R}}$ and $\overline{\mathbf{T}}$ are again $\infty \times \infty$ matrices.

In general, $\overline{\mathbf{R}}$ and $\overline{\mathbf{T}}$ are nondiagonal matrices. Physically, this means that a mode coming into the discontinuity will excite many modes, giving rise to numerous reflected and transmitted modes. As mentioned before, this entails the physics of mode conversion. However, if in both waveguides, only the dominant modes are propagating, e.g., the TE_{10} mode in a rectangular waveguide, the excited higher order modes will be evanescent, and hence, localized around the discontinuity. These evanescent fields store electromagnetic energy, making the discontinuity either capacitive or inductive depending on if the energy is stored in the electric field or magnetic field.

Even though the double-dot products in the above should involve infinite summations, a practical implementation necessitates the truncation of the infinite summations. If the infinite summation is truncated at P terms, as shall be shown in the next section, $P \gg N$ to ensure the accuracy of the double-dot products (see also Problem 5-2).

If only a few elements of $\overline{\mathbf{R}}$ and $\overline{\mathbf{T}}$ are needed, to save on computer memory, the matrix $\overline{\mathbf{L}}_{ia}$, which is $P \times N$, need not be calculated and stored. Explicit expressions can usually be derived for the elements of $\overline{\mathbf{L}}_{ia}^t : \overline{\mathbf{G}}_i : \overline{\mathbf{D}}_i^{-1} : \overline{\mathbf{L}}_{ia}$, which is $N \times N$, so that only these matrices need be stored. Then, the corresponding elements of $\overline{\mathbf{L}}_{ia}$ are calculated to form the desired elements of $\overline{\mathbf{R}}$ and $\overline{\mathbf{T}}$.

Note that the above formulation is general enough that it encompasses the case of even if the waveguides are of different shapes.

5.2.2 Equivalence Principle and Integral Equation Formulation

The above formulation shows a derivation of the reflection and transmission operators of a waveguide discontinuity using a mode matching procedure. Their calculation involves $\infty \times N$ dimensional matrices. From the above equations, we see that the reflection and transmission operators could be calculated even if we do not truncate the infinite dimension of the matrices. The matrix that needs to be inverted is always $N \times N$. However, to save computation resources, it is necessary to truncate the infinite summations at some finite values. Many authors have suggested using P_1 modes only in waveguide 1 and P_2 modes only in waveguide 2 in the mode matching procedure [5]. Consequently, the resultant matrices in the above formulation, which are $\infty \times N$, become $P_1 \times N$ or $P_2 \times N$. A rule is often given for the ratio between P_1, P_2 and N for a convergent result. This issue is generally regarded as relative convergence. We shall give a new rule for the choice of P_1 and P_2 for this problem.

Alternatively, the problem on the choice of P_1, P_2, and N can be further enlightened by looking at an equivalent formulation of the above problem. This equivalent formulation can be achieved with the use of equivalence principle first, and later, deriving an integral equation for the problem.

In the equivalent problem, the waveguide discontinuity problem is divided into two waveguides with shorting planes at the location of the discontinuity (see Figure 5.4). Moreover, magnetic current $\mathbf{M}_a = \mathbf{E} \times \hat{n}$ are impressed at the location of the aperture. By the equivalence principle, if $\mathbf{E} \times \hat{n}$ is known at the aperture, the reflected and the transmitted fields are the same in the original problem and the equivalent problem. This equivalence problem can be easily derived using vector Green's theorem and vector Huygens principle.

If the magnetic dyadic Green's functions are known for the shorted waveguide 1 and waveguide 2, then the magnetic fields in waveguide 1 and waveguide 2 are

$$\mathbf{H}_1(\mathbf{r}) = \mathbf{H}_{inc}(\mathbf{r}) + i\omega\epsilon_1 \int_{S_a} dS' \overline{\mathbf{G}}_{1m}(\mathbf{r}, \mathbf{r}') \cdot \mathbf{M}_a(\mathbf{r}'), \tag{5.2.30a}$$

$$\mathbf{H}_2(\mathbf{r}) = -i\omega\epsilon_2 \int_{S_a} dS' \overline{\mathbf{G}}_{2m}(\mathbf{r}, \mathbf{r}') \cdot \mathbf{M}_a(\mathbf{r}'). \tag{5.2.30b}$$

Such a dyadic Green's function can be derived using the method outlined in the previous chapter, which expands the Green's function in terms of an infinite number of waveguide modes. Since tangential magnetic field is continuous across the aperture, we have

$$\hat{n} \times \mathbf{H}_{inc}(\mathbf{r}) = -i\omega\hat{n} \times \int_{S_a} dS' [\epsilon_2 \overline{\mathbf{G}}_{2m}(\mathbf{r}, \mathbf{r}') + \epsilon_1 \overline{\mathbf{G}}_{1m}(\mathbf{r}, \mathbf{r}')] \cdot \mathbf{M}_a(\mathbf{r}'), \quad \mathbf{r} \in S_a. \tag{5.2.31}$$

The above integral equation can be solved with the method of moments or Galerkin's method, and converted into an $N \times N$ matrix equation, by using N expansion functions and N testing functions. In this procedure, we let

$$\mathbf{M}_a(\mathbf{r}') = \sum_{n=1}^{N} a_n \mathbf{f}_n(\mathbf{r}') \tag{5.2.32}$$

By testing the above with $-\hat{n} \times \mathbf{f}_m(\mathbf{r})$, we obtain

$$\langle \mathbf{f}_m, \mathbf{H}_{inc} \rangle = -i\omega \sum_{n=1}^{\infty} \langle \mathbf{f}_m, \epsilon_2 \overline{\mathbf{G}}_{2m} + \epsilon_1 \overline{\mathbf{G}}_{1m}, \mathbf{f}_n \rangle a_n, \quad m = 1, \ldots, N \tag{5.2.33}$$

The above is a matrix equation of the form

$$\mathbf{b} = \overline{\mathbf{A}} \cdot \mathbf{a} \tag{5.2.34}$$

where,

$$[\mathbf{b}]_m = \langle \mathbf{f}_m, \mathbf{H}_{inc} \rangle, \quad [\overline{\mathbf{A}}]_{mn} = -i\omega \langle \mathbf{f}_m, \epsilon_2 \overline{\mathbf{G}}_{2m} + \epsilon_1 \overline{\mathbf{G}}_{1m}, \mathbf{f}_n \rangle, \quad [\mathbf{a}]_n = a_n \tag{5.2.35}$$

and

$$\langle \mathbf{f}_m, \overline{\mathbf{G}}, \mathbf{f}_n \rangle = \int_{S_a} dS\, \mathbf{f}_m(\mathbf{r}) \cdot \int_{S_a} dS'\, \overline{\mathbf{G}}(\mathbf{r}, \mathbf{r}') \cdot \mathbf{f}_n(\mathbf{r}') \tag{5.2.36}$$

$$\langle \mathbf{f}_m, \mathbf{H}_{inc} \rangle = \int_{S_a} dS\, \mathbf{f}_m(\mathbf{r}) \cdot \mathbf{H}_{inc}(\mathbf{r}) \tag{5.2.37}$$

As has been seen, the method of moments yields the optimal solution when N basis functions are used to solve (5.2.31). Equation (5.2.31) is an exact integral equation, and it is clear that the numbers of waveguide modes needed to approximate $\overline{\mathbf{G}}_{1m}$ and $\overline{\mathbf{G}}_{2m}$ should be infinite to render them as accurate as possible. This is equivalent to choosing P_1 and P_2 to be infinite in the previous formulation using mode matching.

5.2.3 Relative Convergence

If S_a is made as large as waveguide 1, as is the case discussed in Equation (5.2.28a) in the previous section, and the electric field of N modes from waveguide 1 is used to approximate $\mathbf{M}_a$ via $\mathbf{E}_1 \times \hat{n}$, then the integral

$$i\omega\epsilon_1 \int_{S_a} dS' \overline{\mathbf{G}}_{1m}(\mathbf{r}, \mathbf{r}') \cdot \mathbf{M}_a(\mathbf{r}') \tag{5.2.38}$$

would only yield N waveguide modes for the reflected field in waveguide 1 due to mode orthogonality. Hence, $P_1 = N$ in this case. What then is the number of modes required in the expansion of dyadic Green's function in the following integral?

$$i\omega\epsilon_2 \int_{S_a} dS' \overline{\mathbf{G}}_{2m}(\mathbf{r}, \mathbf{r}') \cdot \mathbf{M}_a(\mathbf{r}'). \tag{5.2.39}$$

This integral yields the transmitted field in waveguide 2. Since mode orthogonality does not apply here, in theory, an infinite number of modes is needed in $\overline{\mathbf{G}}_{2m}(\mathbf{r}, \mathbf{r}')$. But in practice, we need only a finite number of modes P_2. How large should P_2 be to accurately represent the field generated by (5.2.39)?

Further insight can be obtained by studying a simpler problem—the parallel plate waveguide problem. The problem is scalar in this case and a typical integral for (5.2.39) looks like

$$I = \int_0^{d_2} dx' \left[\sum_{m=0}^{\infty} \frac{e^{ik_{mz}|z|}}{k_{mz}} \cos\left(\frac{m\pi x}{d_2}\right) \cos\left(\frac{m\pi x'}{d_2}\right) \right] M_a(x'), \tag{5.2.40a}$$

where we have expanded the scalar Green's function as

$$G_{2m}(\mathbf{r}, \mathbf{r}') \propto \sum_{m=0}^{\infty} \frac{e^{ik_{mz}|z-z'|}}{k_{mz}} \cos\left(\frac{m\pi x}{d_2}\right) \cos\left(\frac{m\pi x'}{d_2}\right). \tag{5.2.40b}$$

Here, $k_{mz} = \sqrt{k_2^2 - \left(\frac{m\pi}{d_2}\right)^2}$ and d_2 is the separation of the parallel plate waveguide 2. $M_a(x')$, the aperture field, following the above method, is expanded in terms of the modes of waveguide 1, i.e.,

$$M_a(x') = \sum_{n=0}^{N} a_n \cos\left(\frac{n\pi x'}{d_1}\right), \tag{5.2.41}$$

where d_1 is the separation of parallel plate waveguide 1. Hence, (5.2.40a), after using (5.2.41), letting $z = z'$, and exchanging the order of summation and integration, becomes,

$$I = \sum_{m=0}^{\infty} \frac{1}{k_{mx}} \cos\left(\frac{m\pi x}{d_2}\right) \sum_{n=0}^{N} a_n \int_0^{d_2} dx' \cos\left(\frac{m\pi x'}{d_2}\right) \cos\left(\frac{n\pi x'}{d_1}\right). \tag{5.2.42}$$

The above integral could be integrated readily to yield

$$I = \sum_{n=0}^{N} a_n \sum_{m=0}^{\infty} \frac{1}{k_{mz}} \cos\left(\frac{m\pi x}{d_2}\right) f_{nm}, \tag{5.2.43a}$$

where

$$f_{nm} = \frac{1}{2} \left\{ \frac{\sin\left[\left(\frac{m\pi}{d_2} + \frac{n\pi}{d_1}\right) d_2\right]}{\frac{m\pi}{d_2} + \frac{n\pi}{d_1}} + \frac{\sin\left[\left(\frac{m\pi}{d_2} - \frac{n\pi}{d_1}\right) d_2\right]}{\frac{m\pi}{d_2} - \frac{n\pi}{d_1}} \right\}. \tag{5.2.43b}$$

In the above, the integral I is the result of the action of the Green's function on the expansion or basis functions at the aperture. The integral in (5.2.42) is the cosine transform of the basis functions in waveguide 1 using the cosine functions of waveguide 2. In other words, the basis function $\cos\left(\frac{n\pi x'}{d_1}\right)$ is expanded in terms of the functions $\cos\left(\frac{m\pi x'}{d_2}\right)$. The coefficient of the expansion peaks when $\frac{m\pi}{d_2} = \frac{n\pi}{d_1}$, or when the wavelengths of the two Fourier harmonics match or about equal.

When the integral (5.2.39) is used in (5.2.31) to match boundary condition, $\mathbf{r} \in S_a$, or $z = 0$ in (5.2.40a) and (5.2.43a) when used in (5.2.31). If we further test (or weight) (5.2.43a) by $\cos\left(\frac{n'\pi x}{d_1}\right)$ and integrate over x, as is required when the method of Galerkin is used to

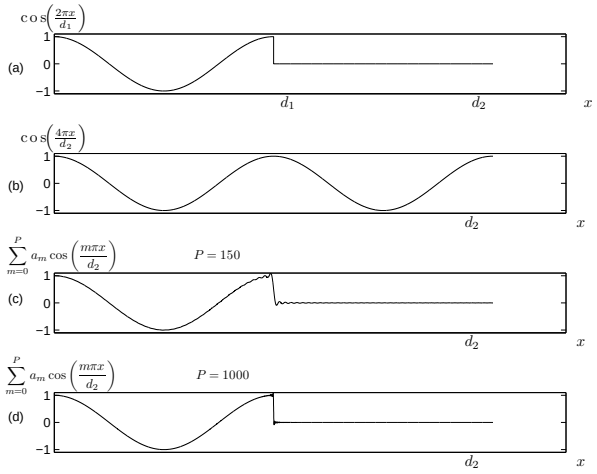


Figure 5.5: Convergence of the cosine Fourier series expansion of the mode of one waveguide with dimesion d_1 in terms of the modes of the second waveguide with dimension d_2.

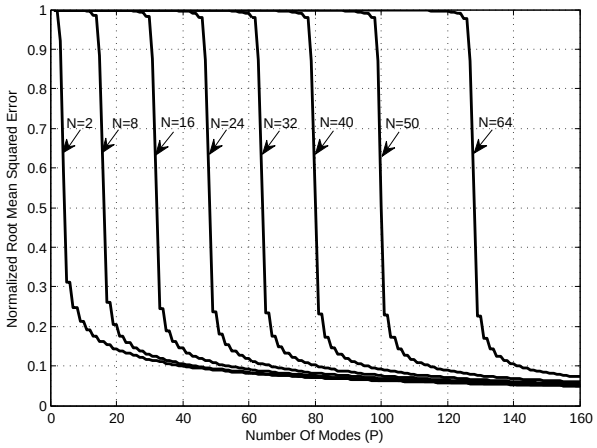


Figure 5.6: Error convergence of the cosine Fourier series expansion $\cos\left(\frac{N\pi x}{d_1}\right) = \sum_{m=0}^{P} a_m \cos\left(\frac{m\pi x}{d_2}\right)$ as a function of P with $d_2 = 2d_1$. Notice that as $P > Nd_2/d_1$, the error drops rapidly.

solve (5.2.31), then (5.2.43a) can be transformed to (with $z = 0$)

$$\begin{aligned} A &= \sum_{n=0}^{N} a_n \sum_{m=0}^{\infty} f_{nm} \frac{1}{k_{mz}} f_{n'm} \\ &\cong \sum_{n=0}^{N} a_n \sum_{m=0}^{P_2} f_{nm} \frac{1}{k_{mz}} f_{n'm}, \quad n' = 0, 1, \cdots, N. \end{aligned} \tag{5.2.44}$$

The above is the matrix representation of the Green's function using the eigenmodes from waveguide 1. Note that in the above, f_{nm} peaks at $m = \frac{d_2}{d_1} n$, and $f_{n'm}$ peaks at $m = \frac{d_2}{d_1} n'$. Hence, in order to evaluate A accurately in (5.2.44), the summation over m in (5.2.44) must at least be large enough so that the contribution from the peaks of $f_{nm} f_{n'm}$ are included. Since the largest n and n' are N, for the approximate summation in (5.2.44) to be accurate when the infinite summation in (5.2.44) is replaced by a summation over m from 0 to P_2, we require that

$$\frac{P_2}{d_2} \gg \frac{N}{d_1}. \tag{5.2.45}$$

From the above analysis, it is clear that

(a) if the dimension of the diaphragm region is d_a such that $d_a < d_1$, $d_a < d_2$, and

(b) if N basis functions are used to approximate the aperture field, and moreover,

(c) if truncated numbers of modes, P_1 and P_2 are used to represent the fields in waveguides 1 and 2, respectively,

then in order for this truncation to be accurate,

$$\frac{P_1}{d_1} \gg \frac{N}{d_a}, \qquad \frac{P_2}{d_2} \gg \frac{N}{d_a}. \tag{5.2.46}$$

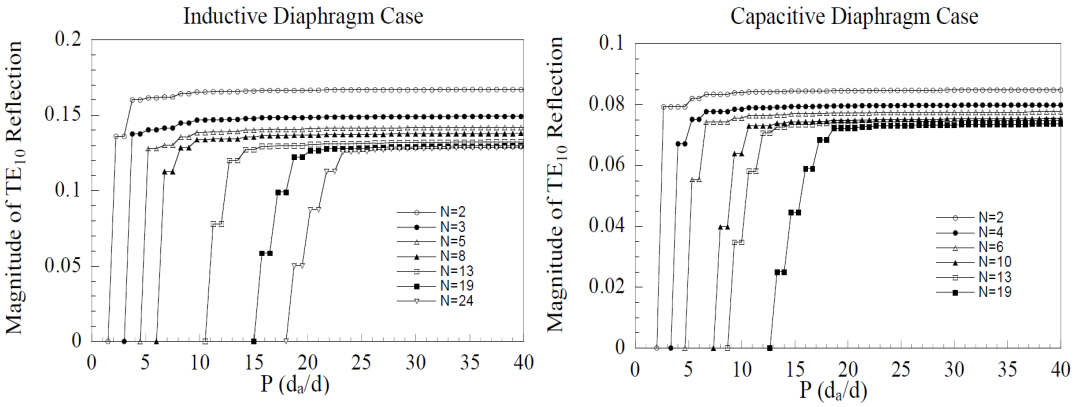
This point has also been noted by Orta et al[3] in the study of frequency selective surfaces. Note that P_1 and P_2 need not be related by a specific ratio, but if the same degree of accuracies is required of the fields in both waveguides, then

$$\frac{P_1}{d_1} \simeq \frac{P_2}{d_2}. \tag{5.2.47}$$

The inequalities in (5.2.46) can also be interpreted as that P_1 and P_2 should be chosen large enough so that the spectral components used in $\overline{\mathbf{G}}_{1m}$ and $\overline{\mathbf{G}}_{2m}$ are large enough to capture (or accurately represent) the dominant spectral components in $\mathbf{M}_a$.

The above analysis can be extended to the rectangular waveguide case. Since the x and y coordinates of a rectangular waveguides are separable, the x-spectral components can be considered separately from the y-spectral components. Similar inequalities as in (40) will hold separately for the x and y spectral components of the waveguide. For the case of a circular

[3]R. Orta, R. Tascone, and R. Zich, "Multiple dielectric loaded perforated screens as frequency selective surfaces," *IEE Proc.*, vol. 135, pt. H, no. 2, pp. 75-82, 1988.



(a) (b)

Figure 5.7: Plots showing the convergence of the reflection coefficient amplitude of TE_{10} mode of a rectangular waveguide when (5.2.46) is satisfied. The dimension of the waveguide is 0.8×0.6 wavelength. (a) The inductive diaphragm case. The dimension of the diaphragm is 0.6×0.6 wavelength. (b) The capacitive diaphragm case. The dimension of the diaphragm is 0.8×0.4 wavelength. In these cases, $P_1 = P_2 = P$ (re-plotted by F. Ling).

waveguide which is axially symmetric, a similar analysis will yield the inequality as in (5.2.46) but d_1, d_2, and d_a represent the diameter of the waveguides and aperture. This is because Bessel functions behave like sinusoidal functions when their arguments are large.

It is to be noted that other basis functions can be used to expand the aperture field in (5.2.41) other than the waveguide modes. In this case, f_{nm} will not be of the form given by (5.2.43a), but usually, a more complex form ensues. In this case P_1 and P_2 should be chosen large enough to capture the dominant spectral components in $\mathbf{M}_a$. The rule for choosing P_1 and P_2 will not be as simple as that given by (5.2.46). But for a fixed N, one should increase P_1 and P_2 so that they are large enough until the calculated amplitudes of the reflected and transmitted modes stabilize.

5.3 Numerical Examples

A program has been developed using the formulation of Section 5.2. This program yields the reflection and transmission operators due to a junction discontinuity with a diaphragm of two rectangular waveguides. This program is used to study the convergence of the TE_{10} mode reflection coefficient when the two waveguides are identical but separated by either an inductive diaphragm or a capacitive diaphragm. The dimension of the waveguide is 0.8×0.6 wavelength. The apertures are symmetrically located with dimension 0.6×0.6 wavelength for the inductive diaphragm, and 0.8×0.4 wavelength for the capacitive diaphragm.

A TE_{10} mode is assumed incident onto this junction discontinuity. For the inductive diaphragm, due to symmetry, only modes with x-variations will be excited. Figure 5.7(a) shows the amplitude of the reflection coefficient as a function of the number of waveguide modes when the number of basis function N in the aperture is kept fixed. It is seen that for a fixed N, the solution converges when (5.2.46) is satisfied. It is also seen that N has to be sufficiently large before the reflection coefficient amplitude is accurate. Note that in this case, $P\left(\frac{d_a}{d}\right)$ need not be very much larger than N for the calculation to stabilize. $P\left(\frac{d_a}{d}\right)$ has to be only a little larger than N to capture the dominant components of $\mathbf{M}_a$.

Figure 5.7(b) shows a similar convergence plot for the case of a capacitive diaphragm. Due to symmetry, only higher order modes with y-variation will be excited. Qualitatively, the result is similar to that of Figure 5.7(a).

Note from Figures 5.7(a) and 5.7(b) that if the conventional rule of choosing $\frac{P_1}{d_1} = \frac{N}{d_a} = \frac{P_2}{d_2}$ is used, fairly good result could be obtained when N is large enough. This is an entirely an artifact of choosing waveguide modes to expand the field at the aperture. The conventional wisdom of choosing P_1 and P_2 for this case will capture most of the spectral components (the aperture basis functions are expanded in terms of a Fourier series using the waveguide modes, and we call the amplitude of a term of this series a spectral component) this series of the basis functions at the aperture. If basis functions with edge conditions, for instance, are used, high spectral components will ensue and the conventional wisdom will not apply.

As a final note, it can be shown that the mode matching method always conserves energy irrespective of how many basis functions are used at the aperture. Thus the conservation of energy can be used to check the correctness of the implementation of the method, but not the accuracy. As seen in the previous simulation, an accurate solution can only be obtained by slowly increasing N while ensuring that (5.2.46) is satisfied. An accurate solution is obtained when the reflection and transmission coefficients cease to change with increasing N.

5.4 Solution to the Multiple Waveguide Junction Problem

Once the solution to the one waveguide junction problem is known, a two waveguide junction problem can be solved. Then, the N-waveguide junction problem can be sought recursively in the manner of [45]. The N-waveguide junction solution can be used to model complex waveguide junctions. It can also be used to model tapered waveguide junctions.

5.4.1 A Two-Waveguide-Junction Problem

Given a two-waveguide-junction geometry as shown, we can write down the solution in waveguide 1 for the electric field as consisting of

$$\mathbf{E}_{1s} = \overline{\mathbf{\Psi}}_1^t(\mathbf{r}_s) \cdot \left(e^{i\overline{\mathbf{K}}_1 z} + e^{-i\overline{\mathbf{K}}_1 z} \cdot \widetilde{\overline{\mathbf{R}}}_{12} \right) \cdot \mathbf{e} \tag{5.4.1}$$

where $\mathbf{e}$ is the amplitude of the incident modes, and $\widetilde{\overline{\mathbf{R}}}$ is the reflection operator including multiple reflections between junctions 1 and 2. The first term consists of modes travelling to the right while the second term consists of modes travelling to the left.

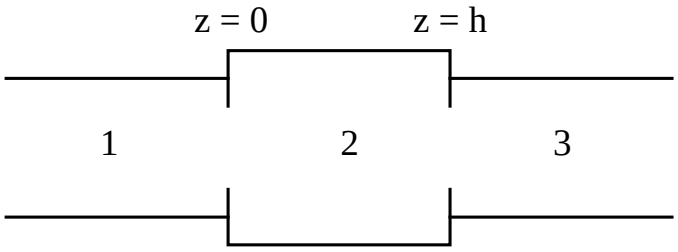


Figure 5.8: A two-waveguide junction problem.

In waveguide 2, we can write the solution for the electric field as

$$\mathbf{E}_{2s} = \overline{\mathbf{\Psi}}_2^t(\mathbf{r}_s) \cdot \left(e^{i\overline{\mathbf{K}}_2 z} \cdot \mathbf{A}_2 + e^{-i\overline{\mathbf{K}}_2 z} \cdot \mathbf{B}_2 \right). \tag{5.4.2}$$

In waveguide 3, the corresponding electric field is

$$\mathbf{E}_{3s} = \overline{\mathbf{\Psi}}_3^t(\mathbf{r}_s) \cdot e^{i\overline{\mathbf{K}}_3 z} \cdot \mathbf{A}_3. \tag{5.4.3}$$

A relationship can be easily established between the amplitudes $\mathbf{A}_2$ and $\mathbf{B}_2$ in waveguide 2, because the left-going wave in waveguide 2 is a consequence of the right-going wave in the same waveguide. Therefore, at $z = h$, we must have

$$e^{-i\overline{\mathbf{K}}_2 h} \cdot \mathbf{B}_2 = \overline{\mathbf{R}}_{23} \cdot e^{i\overline{\mathbf{K}}_2 h} \cdot \mathbf{A}_2, \tag{5.4.4}$$

where $\overline{\mathbf{R}}_{23}$ is the reflection operator for the single waveguide junction. Consequently, we can rewrite (5.4.2) as

$$\mathbf{E}_{2s} = \overline{\mathbf{\Psi}}_2^t(\mathbf{r}_s) \cdot \left(e^{i\overline{\mathbf{K}}_2 z} + e^{-i\overline{\mathbf{K}}_2 (z-h)} \cdot \overline{\mathbf{R}}_{23} \cdot e^{i\overline{\mathbf{K}}_2 h} \right) \cdot \mathbf{A}_2. \tag{5.4.5}$$

The right-going wave in waveguide 2 is a result of the transmission of right-going wave in waveguide 1 plus a reflection of the left-going wave in waveguide 2. Therefore, at $z = 0$, we must have

$$\mathbf{A}_2 = \overline{\mathbf{T}}_{12} \cdot \mathbf{e} + \overline{\mathbf{R}}_{21} \cdot \mathbf{B}_2. \tag{5.4.6}$$

Using (5.4.4) in (5.4.6), we can solve for $\mathbf{A}_2$ to yield

$$\mathbf{A}_2 = \left(\bar{\mathbf{I}} - \overline{\mathbf{R}}_{21} \cdot e^{i\overline{\mathbf{K}}_2 h} \cdot \overline{\mathbf{R}}_{23} \cdot e^{i\overline{\mathbf{K}}_2 h} \right)^{-1} \cdot \overline{\mathbf{T}}_{12} \cdot \mathbf{e} \tag{5.4.7}$$

Using (5.4.7) in (5.4.4), we can express $\mathbf{B}_2$ in terms of $\mathbf{e}$. Therefore, $\mathbf{A}_2$ and $\mathbf{B}_2$ are found in terms of $\mathbf{e}$.

The right-going wave in waveguide 3 is a consequence of a transmission of the right-going wave in waveguide 2. Therefore, at $z = h$, we must have

$$e^{i\overline{\mathbf{K}}_3 h} \cdot \mathbf{A}_3 = \overline{\mathbf{T}}_{23} \cdot e^{i\overline{\mathbf{K}}_2 h} \cdot \mathbf{A}_2. \tag{5.4.8}$$

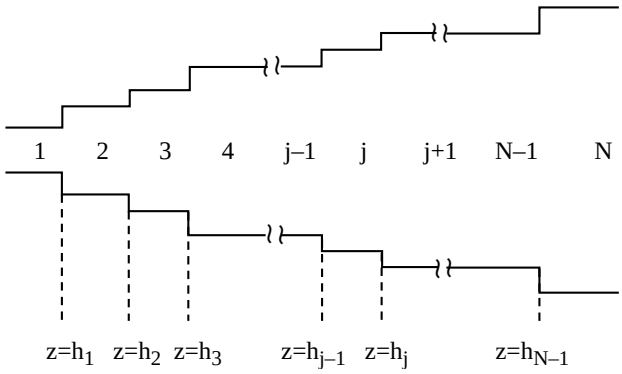


Figure 5.9: A multiple-junction waveguide.

Consequently, $\mathbf{A}_3$ can be found in terms of $\mathbf{e}$ via the use of (5.4.7) and (5.4.8).

The left-going wave in waveguide 1 is a consequence of the reflection of the right-going wave in waveguide 1 plus a transmission of the left-going wave in waveguide 2. Therefore, we can write

$$\widetilde{\overline{\mathbf{R}}}_{12} \cdot \mathbf{e} = \overline{\mathbf{R}}_{12} \cdot \mathbf{e} + \overline{\mathbf{T}}_{21} \cdot \mathbf{B}_2. \tag{5.4.9}$$

Using $\mathbf{B}_2$ from (5.4.4) and (5.4.7), we have

$$\begin{aligned} \widetilde{\overline{\mathbf{R}}}_{12} = \overline{\mathbf{R}}_{12} + \overline{\mathbf{T}}_{21} \cdot e^{i\overline{\mathbf{K}}_2 h} \cdot \overline{\mathbf{R}}_{23} \cdot e^{i\overline{\mathbf{K}}_2 h} \\ \cdot \left(\bar{\mathbf{I}} - \overline{\mathbf{R}}_{21} \cdot e^{i\overline{\mathbf{K}}_2 h} \cdot \overline{\mathbf{R}}_{23} \cdot e^{i\overline{\mathbf{K}}_2 h}\right)^{-1} \cdot \overline{\mathbf{T}}_{12}. \end{aligned} \tag{5.4.10}$$

Therefore, with $\mathbf{e}$, the amplitude of the incident modes, known, one can find the solution of the two waveguide junction problem in all the three waveguides. We call $\widetilde{\overline{\mathbf{R}}}_{12}$ the generalized reflection operator at the (1,2) junction that accounts for multiple reflections.

5.4.2 An N-Waveguide-Junction Problem

Given an N-waveguide-junction problem where the discontinuities of the waveguide are at $z = h_j$, $j = 1, \cdots, N$, one can write down the solution in region j as

$$\mathbf{E}_{js} = \overline{\mathbf{\Psi}}_j^t(\mathbf{r}_s) \cdot \left(e^{i\overline{\mathbf{K}}_j z} + e^{-i\overline{\mathbf{K}}_j (z-h_j)} \cdot \widetilde{\overline{\mathbf{R}}}_{j,j+1} \cdot e^{i\overline{\mathbf{K}}_j h_j}\right) \cdot \mathbf{A}_j. \tag{5.4.11}$$

The amplitude of the left-going wave in the above is written in such a way so that at $z = h_j$, the left-going wave amplitude is just related to the right-going wave amplitude by $\widetilde{\overline{\mathbf{R}}}_{j,j+1}$, the generalized reflection operator at the $(j, j+1)$ junction. Motivated by (5.4.10), we can write down the expression for the generalized reflection operator as

$$\begin{aligned} \widetilde{\overline{\mathbf{R}}}_{j,j+1} = \overline{\mathbf{R}}_{j,j+1} + \overline{\mathbf{T}}_{j+1,j} \cdot e^{i\overline{\mathbf{K}}_j (h_j - h_{j-1})} \cdot \widetilde{\overline{\mathbf{R}}}_{j+1,j+2} \cdot e^{i\overline{\mathbf{K}}_j (h_j - h_{j-1})} \\ \cdot \left(\bar{\mathbf{I}} - \overline{\mathbf{R}}_{j,j+1} \cdot e^{i\overline{\mathbf{K}}_j (h_j - h_{j-1})} \cdot \widetilde{\overline{\mathbf{R}}}_{j+1,j+2} \cdot e^{i\overline{\mathbf{K}}_j (h_j - h_{j-1})}\right)^{-1} \cdot \overline{\mathbf{T}}_{j,j+1}. \end{aligned} \tag{5.4.12}$$

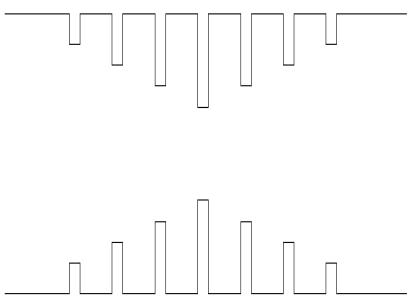


Figure 5.10: A multiple-junction waveguide used to design a filter.

Notice that we have replaced $\overline{\mathbf{R}}_{23}$ in (5.4.10) with a generalized reflection operator $\widetilde{\overline{\mathbf{R}}}_{j+1,j+2}$ at the $(j+1, j+2)$ junction because for a multiple-junction waveguide, multiple reflections to the right of the $(j, j+1)$ junction has to be accounted for.

Using (5.4.12), starting at the right-most junction, the generalized reflection operators in all the waveguides can be found. Next, Using an equation similar to (5.4.7), the amplitude $\mathbf{A}_j$ in waveguide j can be related to $\mathbf{A}_{j-1}$ in waveguide $j-1$ as

$$\begin{aligned} e^{i\overline{\mathbf{K}}_j h_{j-1}} \cdot \mathbf{A}_j = &\left(\bar{\mathbf{I}} - \overline{\mathbf{R}}_{j,j-1} \cdot e^{i\overline{\mathbf{K}}_j (h_j - h_{j-1})} \cdot \widetilde{\overline{\mathbf{R}}}_{j,j+1} \cdot e^{i\overline{\mathbf{K}}_j (h_j - h_{j-1})}\right)^{-1} \\ &\cdot \overline{\mathbf{T}}_{j-1,j} \cdot e^{i\overline{\mathbf{K}}_{j-1} h_{j-1}} \cdot \mathbf{A}_{j-1}. \end{aligned} \tag{5.4.13}$$

Starting with the left-most waveguide, the field solution for all the waveguides can be obtained.

5.4.3 Filter Design–A Resonance Tunneling Problem

Multiple junctions in a waveguide can be used to design filters. Junctions set up interference in a waveguide, giving rise to resonant modes. The resonant mode of a multiple junction waveguide is defined as one that a reflected mode can exist without the presence of the incident mode. This is the same as requiring that

$$\det\left(\widetilde{\overline{\mathbf{R}}}_{j,j+1}\right) = \infty \tag{5.4.14}$$

The above is equivalent to

$$\det\left(\bar{\mathbf{I}} - \overline{\mathbf{R}}_{j,j+1} \cdot e^{i\overline{\mathbf{K}}_j (h_j - h_{j-1})} \cdot \widetilde{\overline{\mathbf{R}}}_{j+1,j+2} \cdot e^{i\overline{\mathbf{K}}_j (h_j - h_{j-1})}\right) = 0 \tag{5.4.15}$$

The above is the generalized transverse resonance condition. One can vary the frequency ω until the above equation is satisfied. It yields the poles of the system as is the case of the Fabry-Perot etalon. The locations of the poles can be used to guide the design of filters as shown in Figures 5.10 and 5.11.

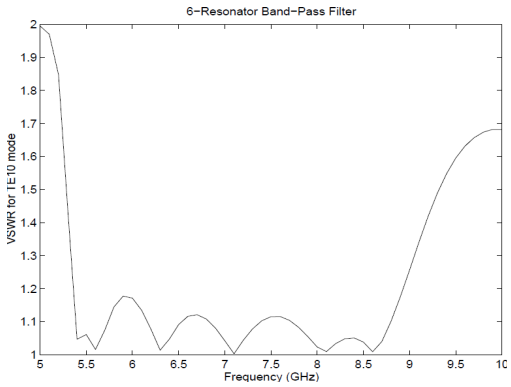


Figure 5.11: The pass-band response of the filter in Figure 5.10 with six resonators. The waveguide dimensions are 13.7 in by 6.22 in. The symmetric irises are 0.02 in thick. The widths of apertures are 1.026, 0.958, 0.898, and 0.870 in respectively from one end to the center. The cavity lengths are 0.626, 0.653, and 0.674 in respectively from one end to the center (courtesy of F. Ling).

5.5 Hybrid Junctions

A hybrid-T junction is a nifty device that when a TE_{10} mode is incident from port 1, the wave will couple to a mode in ports 2 and 3 but not 4. Similarly, a mode incident from port 4 will couple to only ports 2 and 3 but not 1. A mode incident from port 2 couples to ports 1 and 4 but not 3, while a mode incident from port 3 couples to ports 1 and 4 but not 2.

The understanding of the working of the hybrid-T, sometimes known as a magic-T, can be derived from symmetry arguments. For a TE_{10} mode incident at port 1, the $\mathbf{E}$ field is symmetrical about the plane that bisects the center of the waveguide of port 1. From the figure, the vertical component of the electric field is even symmetric about the plane of symmetry, while the horizontal component is odd symmetric. Hence, a TE_{10} mode cannot be excited in port 4.

For a TE_{10} mode incident from port 4, the horizontal component of the electric field is even symmetric while the vertical component is odd symmetric about the plane of symmetry. Hence, the modes excited in ports 2 and 3 are of opposite polarity. Furthermore, no TE_{10} mode can be coupled to port 1 because of the symmetry condition.

When ports 2 and 3 are terminated in matched loads so that only outgoing waves exist in them, ports 1 and 4 can be matched by adding matching elements so that S_{11} and S_{44} are zero. In this case, only incoming wave exists in port 1 or port 4 when ports 2 and 3 are matched. The remaining property of the magic-T can then be explained by time-reversal symmetry.

The magic-T is a linear and lossless device. By linearly superposing a mode incident at port 1 and another mode incident at port 4, one can obtain a mode exiting at port 3 but not a mode exiting at port 2. Hence, this linear superposition is a solution to Maxwell's equations. However, the time-reversed solution is also a solution to Maxwell's equations. The

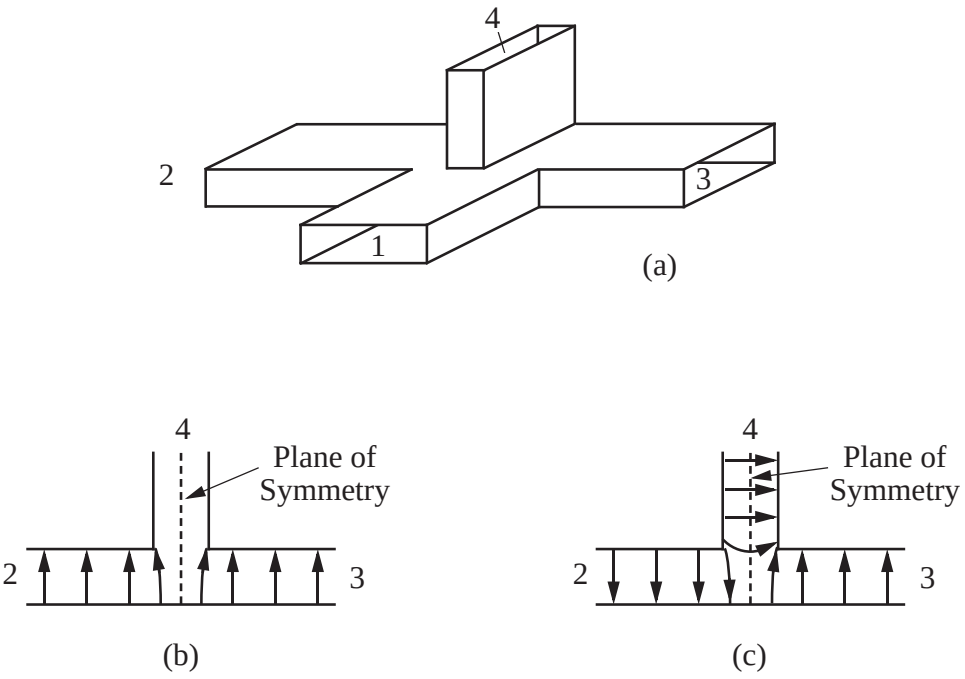


Figure 5.12: (a) A hybrid-T junction. (b) Electric field pattern in port 2 and port 3 with a TE_{10} mode incident from port 1. (c) Electric field pattern in ports 2 and 3 with a TE_{10} mode incident from port 4.

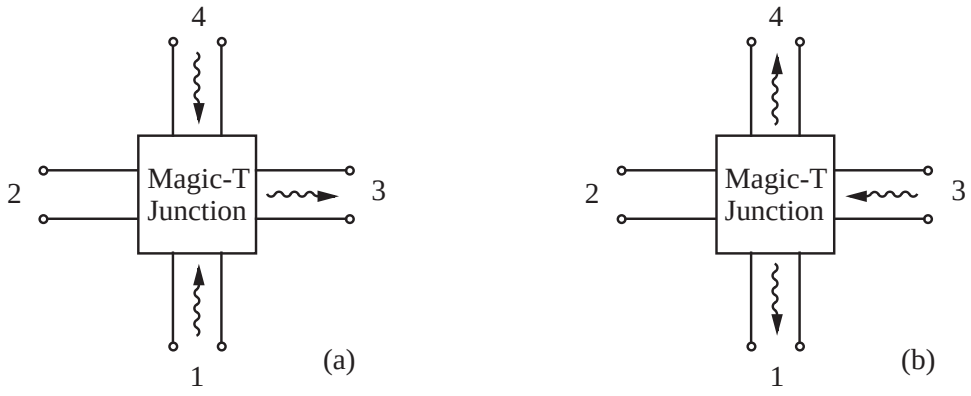


Figure 5.13: (a) Normal-time operation of a magic-T. (b) Reversed-time operation of a magic-T.

time-reversed solution corresponds to having ports 1 and 4 matched since there will only be outgoing waves at these ports. Furthermore, port 3 is matched since only incoming wave exists at this port, and magically, there is no coupling from port 3 to port 2! Similar argument leads to a matched port 2 with matched loads at ports 1 and 4, and no coupling from port 2 to port 3.

Another useful hybrid junction is the rat-race ring circuit. When a mode is incident in port 1, it will split evenly in two directions into two waves. Due to the choice in the size of the two ring, the waves will arrive out-of-phase at port 3. The clockwise travelling wave will suffer a perturbation at port 2 while the counterclockwise travelling wave will suffer a perturbation at port 4. Therefore, their amplitudes at port 3 are equal and out of phase, and there is no coupling to port 3. The phases of these two waves are equal at ports 2 and 4 even though their amplitudes may not be equal because the clockwise travelling wave and counterclockwise travelling wave are perturbed by different amount. Therefore, there will be coupling to ports 2 and 4.

By the same argument, port 2 will not couple to port 4. Port 3 will not couple to port 1 and port 4 will not couple to port 2 by reciprocity.

Hybrid junctions is used to make microwave impedance bridges, frequency discriminator circuits as in balanced mixers, circulators as well as other applications.

5.6 Periodic Structures

A periodic structure is useful as a filter or a slow-wave structure. The constructive and destructive interference phenomenon of a periodic structure can generate passbands and stopbands: A passband is a frequency band that allows the propagation of a wave with no attenuation while a stopband forbids the propagation of a wave. A periodic structure can be analyzed by the Bloch-Floquet theorem [2, 46, 47].

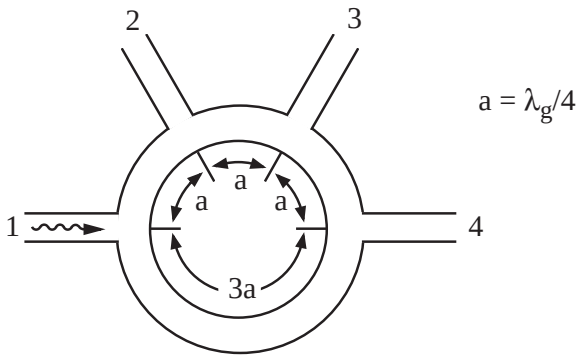


Figure 5.14: A rat-race ring-type hybrid junction.

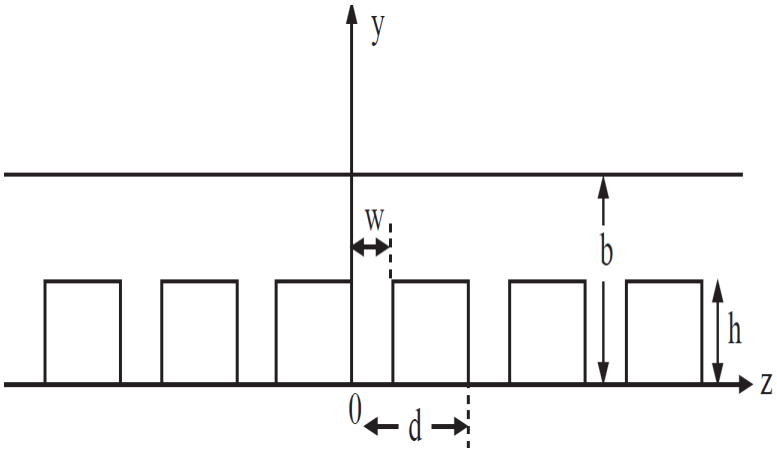


Figure 5.15: A periodic structure in a parallel plate waveguide.

If a wave is traveling inside a waveguide with a period d, it implies that it has to be of the form

$$\mathbf{E}(\mathbf{r}) = e^{ik_z z}\mathbf{E}_p(\mathbf{r}), \tag{5.6.1}$$

$$\mathbf{H}(\mathbf{r}) = e^{ik_z z}\mathbf{H}_p(\mathbf{r}), \tag{5.6.2}$$

where $\mathbf{E}_p(\mathbf{r})$ and $\mathbf{H}_p(\mathbf{r})$ are periodic functions in z having the property

$$\mathbf{E}_p(\mathbf{r} + \hat{z}nd) = \mathbf{E}_p(\mathbf{r}), \tag{5.6.3}$$

$$\mathbf{H}_p(\mathbf{r} + \hat{z}nd) = \mathbf{H}_p(\mathbf{r}), \tag{5.6.4}$$

where n is any positive or negative integer. The expression of a field in (5.6.1) and (5.6.2) is known as the Floquet theorem or the Bloch theorem. In other words, due to the periodicity of this structure, a solution $\mathbf{E}(\mathbf{r})$ when translated by a distance d in the z direction, must be itself save an added phase, viz.

$$\mathbf{E}(\mathbf{r} + \hat{z}nd) = \mathbf{E}(\mathbf{r})e^{ik_z nd}, \quad \mathbf{H}(\mathbf{r} + \hat{z}nd) = \mathbf{H}(\mathbf{r})e^{ik_z nd}. \tag{5.6.5}$$

Any periodic function can be expanded in terms of a Fourier series. Therefore, we can write

$$\mathbf{E}_p(\mathbf{r}) = \sum_{n=-\infty}^{\infty} \mathbf{E}_{pn}(\mathbf{r}_s)e^{i\frac{2n\pi}{d}z}, \tag{5.6.6}$$

$$\mathbf{H}_p(\mathbf{r}) = \sum_{n=-\infty}^{\infty} \mathbf{H}_{pn}(\mathbf{r}_s)e^{i\frac{2n\pi}{d}z}, \tag{5.6.7}$$

where $\mathbf{r}_s = \hat{x}x + \hat{y}y$. Consequently, we can rewrite (5.6.1) and (5.6.2) as

$$\mathbf{E}(\mathbf{r}) = \sum_{n=-\infty}^{\infty} \mathbf{E}_{pn}(\mathbf{r}_s)e^{i(k_z+\frac{2n\pi}{d})z} = \sum_{n=-\infty}^{\infty} \mathbf{E}_{pn}(\mathbf{r}_s)e^{ik_{nz}z}, \tag{5.6.8}$$

$$\mathbf{H}(\mathbf{r}) = \sum_{n=-\infty}^{\infty} \mathbf{H}_{pn}(\mathbf{r}_s)e^{i(k_z+\frac{2n\pi}{d})z} = \sum_{n=-\infty}^{\infty} \mathbf{H}_{pn}(\mathbf{r}_s)e^{ik_{nz}z}, \tag{5.6.9}$$

where $k_{nz} = k_z + \frac{2n\pi}{d}$. Each of the terms above is termed a Floquet mode, with a wavenumber k_{nz}. However, each Floquet mode cannot exist by itself–they have to exist together as a cluster. The phase velocity of the n-th Floquet mode is

$$v_{pn} = \frac{\omega}{k_{nz}} = \frac{\omega}{k_z + 2n\pi/d}, \tag{5.6.10}$$

while the group velocity is given to be

$$v_{gn} = \frac{d\omega}{dk_{nz}} = \left(\frac{dk_{nz}}{d\omega}\right)^{-1} = v_g, \tag{5.6.11}$$

The group velocity is independent of the harmonics. Since v_{pn} can be negative for some negative n, v_{pn} can be opposite in sign to v_g.

Let us assume that a TM mode is propagating in the parallel plate waveguide such that $\mathbf{H} = \hat{x}H_x$. There will be no depolarization since $\partial/\partial x = 0$, and this field component is parallel to the surfaces involved. Consequently, we have

$$H_x = \sum_{n=-\infty}^{\infty} h_n(y) e^{ik_{nz}z}, \tag{5.6.12}$$

Since $(\nabla^2 + {k_0}^2)H_x = 0$, we have

$$\frac{d^2 h_n(y)}{dy^2} + k_{ny}^2 h_n(y) = 0, \tag{5.6.13}$$

where $k_{ny}^2 = k_0^2 - k_{nz}^2$.

For $y > h$, the coefficient $h_n(y)$ can be expanded as

$$h_n(y) = a_n \cos\left[k_{ny}(y-b)\right], \tag{5.6.14}$$

Since the Neumann boundary condition $\hat{n} \cdot \nabla H_x = 0$ on a metallic surface, we have chosen a cosine in (5.6.14) so that this condition is satisfied.

Inside the corrugation, we can expand H_x in terms of the modes of a shorted parallel waveguide, or

$$H_x = \sum_{m=0}^{\infty} g_m(y) \cos\left(\frac{m\pi z}{w}\right), \tag{5.6.15}$$

By the same token, $g_m(y)$ satisfies

$$\frac{d^2 g_m(y)}{dy^2} + \gamma_{my}^2 g_m(y) = 0, \tag{5.6.16}$$

where $\gamma_{my}^2 = k_0^2 - \left(\frac{m\pi}{w}\right)^2$. Therefore, we derive that

$$g_m(y) = b_m \cos(\gamma_{my} y), \tag{5.6.17}$$

Consequently, we can write

$$H_x = \sum_{n=-\infty}^{\infty} a_n \cos[k_{ny}(y-b)] e^{ik_{nz}z}, \quad h < y < b, \tag{5.6.18}$$

$$H_x = \sum_{m=0}^{\infty} b_m \cos(\gamma_{my} y) \cos\left(\frac{m\pi z}{w}\right), \quad 0 < y < h. \tag{5.6.19}$$

The bottom equation is valid for $0 < z < w$, and it replicates itself with the correct phase shift for different slots. To find the guidance condition for nontrivial solutions to a_n and b_n, we match boundary condition across the $y = h$ interface. At this interface, the tangential component of the magnetic field is continuous, so is the tangential component of the electric

field. The latter is the same as $\frac{\partial H_x}{\partial y}$ and is continuous across this interface. Therefore, we arrive at

$$\sum_{n=-\infty}^{\infty} a_n \cos[k_{ny}(h-b)]e^{ik_{nz}z} = \sum_{m=0}^{\infty} b_m \cos\gamma_{my}h \cos\left(\frac{m\pi z}{w}\right), \quad 0<z<w \tag{5.6.20}$$

$$\sum_{n=-\infty}^{\infty} a_n k_{ny} \sin[k_{ny}(h-b)]e^{ik_{nz}z} = \begin{cases} \sum_{m=0}^{\infty} b_m \gamma_{my} \sin(\gamma_{my}h)\cos\left(\frac{m\pi z}{w}\right), 0<z<w, \\ 0 \qquad\qquad\qquad\qquad\qquad\qquad w<z<d \end{cases} \tag{5.6.21}$$

Equation (5.6.20) is a cosine series expansion for $0<z<w$. We can solve for b_m by multiplying (5.6.20) by $\cos(\frac{m\pi z}{w})$ and integrate from 0 to w to obtain

$$b_m \cos(\gamma_{my}h)\frac{w}{2}(1+\delta_{om}) = \sum_{n=-\infty}^{\infty} a_n \cos[k_{ny}(h-b)]\left\langle e^{ik_{nz}z}, \cos\left(\frac{m\pi z}{w}\right)\right\rangle, \tag{5.6.22}$$

where

$$\left\langle e^{ik_{nz}z}, \cos\left(\frac{m\pi z}{w}\right)\right\rangle = \int_0^w e^{ik_{nz}z}\cos\left(\frac{m\pi z}{w}\right)dz, \quad m=0,..., \tag{5.6.23}$$

and can be evaluated in closed form if needed.

In (5.6.21), by writing

$$e^{ik_{nz}z} = e^{ik_z z + i\frac{2n\pi}{d}z}, \tag{5.6.24}$$

it is apparently a Fourier series expansion. We can find the coefficients a_n as

$$a_n k_{ny}\sin[k_{ny}(h-b)]d = \sum_{m=0}^{\infty} b_m\gamma_{my}\sin(\gamma_{my}h)\left\langle \cos\left(\frac{m\pi z}{w}\right), e^{-ik_{nz}z}\right\rangle, \tag{5.6.25}$$

where

$$\left\langle \cos\left(\frac{m\pi z}{w}\right), e^{-ik_{nz}z}\right\rangle = \int_0^w \cos\left(\frac{m\pi z}{w}\right)e^{-ik_{nz}z}dz, \tag{5.6.26}$$

and n varies from $-\infty$ to $+\infty$. Equations (5.6.22) and (5.6.25) entail two infinite system of equations. Since γ_{my} and k_{ny} become large imaginary numbers when m and n are large, $\cos(\gamma_{my}h)$, $\sin(\gamma_{my}h)$, $\cos[k_{ny}(h-b)]$, and $\sin[k_{ny}(h-b)]$ become exponentially large. In this case, we can define

$$\hat{b}_m = b_m\cos(\gamma_{my}h), \tag{5.6.27}$$

$$\hat{a}_n = a_n\cos[k_{ny}(h-b)], \tag{5.6.28}$$

to rewrite (5.6.22) and (5.6.25) as

$$\hat{b}_m\frac{w}{2}(1+\delta_{om}) = \sum_{n=-\infty}^{\infty}\hat{a}_n\left\langle e^{ik_{nz}z}, \cos\left(\frac{m\pi z}{w}\right)\right\rangle, \tag{5.6.29}$$

$$\hat{a}_n k_{ny} \tan[k_{ny}(h-b)]d = \sum_{m=0}^{\infty} \hat{b}_m \gamma_{my} \tan(\gamma_{my} h) \left\langle \cos\left(\frac{m\pi z}{w}\right), e^{-ik_{nz}z} \right\rangle, \tag{5.6.30}$$

To solve the above, we need to truncate the infinite system. We can rewrite (5.6.29) and (5.6.30) as

$$\hat{b}_m \lambda_m = \sum_{n=-N}^{N} \hat{a}_n A_{mn}, \quad m = 0, ..., M, \tag{5.6.31}$$

$$\hat{a}_n \beta_n = \sum_{m=0}^{M} \hat{b}_m B_{nm}, \quad n = -N, ..., N. \tag{5.6.32}$$

The above can be written as matrix equations

$$\overline{\boldsymbol{\lambda}} \cdot \hat{\mathbf{b}} = \overline{\mathbf{A}} \cdot \hat{\mathbf{a}}, \quad \overline{\boldsymbol{\beta}} \cdot \hat{\mathbf{a}} = \overline{\mathbf{B}} \cdot \hat{\mathbf{b}}, \tag{5.6.33}$$

where $\overline{\boldsymbol{\lambda}}$ and $\overline{\boldsymbol{\beta}}$ are diagonal matrices, $\overline{\mathbf{A}}$ is a $(M+1)\times(2N+1)$ matrix, $\overline{\mathbf{B}}$ is a $(2N+1)\times(M+1)$ matrix, $\mathbf{a}$ is a length $(2N+1)$ vector while $\mathbf{b}$ is a length $(M+1)$ vector. Equation (5.6.33) can be rewritten as

$$\hat{\mathbf{b}} - \overline{\boldsymbol{\lambda}}^{-1} \cdot \overline{\mathbf{A}} \cdot \overline{\boldsymbol{\beta}}^{-1} \cdot \overline{\mathbf{B}} \cdot \hat{\mathbf{b}} = 0. \tag{5.6.34}$$

Nontrivial solution for $\mathbf{b}$ will exist if

$$\det(\overline{\mathbf{I}} - \overline{\boldsymbol{\lambda}}^{-1} \cdot \overline{\mathbf{A}} \cdot \overline{\boldsymbol{\beta}}^{-1} \cdot \overline{\mathbf{B}}) = 0. \tag{5.6.35}$$

The above is the guidance condition from which one can solve for k given k_z. It will be found that k does not always exist for all values for a given k_z. Also, it could be that for a certain window of k, no value exists for all k_z. Those are the stop band of the periodic structure.

5.6.1 Floquet Modes and Brillouin Zone

From the periodic structure theory, we notice that a wave propagating in a periodic waveguide is of the form

$$\phi(\mathbf{r}) = e^{ik_z z} \phi_p(\mathbf{r}), \tag{5.6.36}$$

where $\phi_p(\mathbf{r})$ is a periodic function in z with $\phi_p(\mathbf{r} + \hat{z}nd) = \phi_p(\mathbf{r})$, and

$$\phi_p(\mathbf{r}) = \sum_{n=-\infty}^{\infty} \phi_{pn}(\mathbf{r}_s) e^{i\frac{2n\pi}{d}z}. \tag{5.6.37}$$

The physical picture is that this wave is a cluster of Floquet modes propagating in unison through the waveguide in order to satisfy the boundary condition on the waveguide wall. In other words, all the Floquet modes of the form $\exp(i(k_z + 2n\pi/d))$ move in lock step with respect to each other. We can express (5.6.36) more explicitly in terms of the Floquet modes, namely,

$$\phi(\mathbf{r}) = \sum_{n=-\infty}^{\infty} \phi_{pn}(\mathbf{r}_s, k_z) e^{i\left(k_z + \frac{2n\pi}{d}\right)z}. \tag{5.6.38}$$

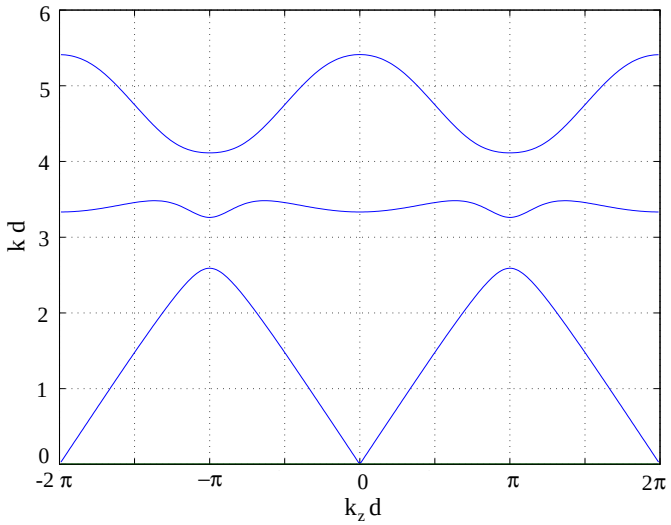


Figure 5.16: Dispersion ($k - k_z$) diagram of a corrugated parallel plate waveguide showing bandpass and bandstop sections ($w = d/2$, $b = d$, $h = 0.25b$) (courtesy of L.J. Jiang).

Now, if we let $k_z \to k_z + \frac{2m\pi}{d}$, where m is an integer, the above becomes

$$\begin{aligned}\phi(\mathbf{r}) &= \sum_{n=-\infty}^{\infty} \phi_{pn}\left(\mathbf{r}_s, k_z + \frac{2m\pi}{d}\right) e^{i\left[k_z + \frac{2(n+m)\pi}{d}\right]z} \\ &= \sum_{n'=-\infty}^{\infty} \phi_{p,n'-m}\left(\mathbf{r}_s, k_z + \frac{2m\pi}{d}\right) e^{i\left(k_z + \frac{2n'\pi}{d}\right)z}.\end{aligned} \tag{5.6.39}$$

It is clear that if (5.6.38) is an eigensolution that can satisfy the boundary condition on the waveguide wall, (5.6.39) can be easily made to satisfy the boundary condition on the waveguide wall if

$$\phi_{p,n'-m}\left(\mathbf{r}_s, k_z + \frac{2m\pi}{d}\right) = c\phi_{pn'}(\mathbf{r}_s, k_z)$$

where c is a multiplicative constant. Therefore, the cluster of Floquet modes is invariant with respect to the transform $k_z \to k_z + \frac{2m\pi}{d}$. Consequently, the dispersion diagram for a guided mode in a periodic structure is $\frac{2\pi}{d}$ periodic in k_z as shown in Figure 5.16. Each of this periodic zone is called the Brillouin zone.

Furthermore, we can show that if $\phi(\mathbf{r})$ is a solution, so is $\phi^*(\mathbf{r})$. Using this fact, one can show that the dispersion diagram has to be even symmetric about the origin.

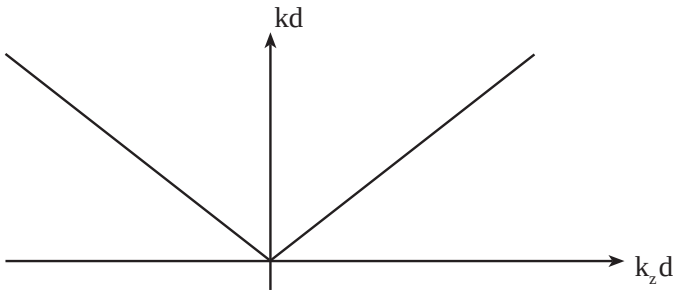


Figure 5.17: Dispersion ($k - k_z$) diagram of a parallel plate waveguide without periodic perturbation.

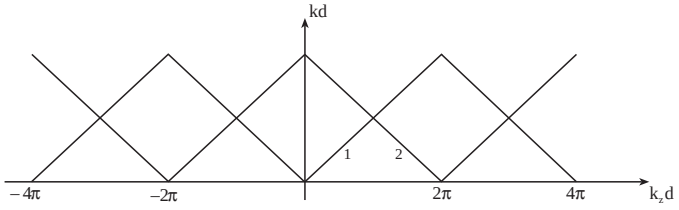


Figure 5.18: Dispersion ($k - k_z$) diagram of a parallel plate waveguide with weak or no periodic perturbation. The Floquet modes are generated, but they remain weakly coupled to each other.

5.7 Stop Band and Coupled-Mode Theory

The existence of pass bands and stop bands in a periodic waveguide is due to the coupling, constructive and destructive interference of the Floquet modes. To see how the interfering Floquet modes can yield a stop band when they are coupled, we use the coupled modes theory [48].

To do this analysis, we think of the corrugation in a waveguide as a perturbation. Before the corrugation perturbations are introduced, assume that we have only the forward and backward propagating TEM modes within the parallel-plate waveguide. The dispersion diagram is as shown in Figure 5.17. The moment a small periodic perturbation is introduced, higher order Floquet modes emerge in the waveguide and the dispersion diagram becomes periodic. When this happens, all the Brillouin zones are equivalent to each other, and none is preferred over the others, as can be seen from the analysis in the previous subsection.

If the perturbation is weak enough, little coupling occurs between the forward Floquet modes and the backward Floquet modes. The forward and backward Floquet modes carry energy in the forward and backward directions respectively due to the sign of their group velocities. The dispersion curve in the case of vanishingly small periodic perturbation is shown in Figure 5.18.

We can focus our attention on modes 1 and 2 and study the coupling behavior between them. The couple-mode equation for describing these two contra-propagating modes can be written as

$$\frac{d\phi_1}{dz} = ik_1\phi_1 + i\alpha_{12}\phi_2, \tag{5.7.1}$$

$$\frac{d\phi_2}{dz} = ik_2\phi_2 + i\alpha_{21}\phi_1, \tag{5.7.2}$$

α_{ij} is the coupling coefficient describing the coupling of energy between the two modes. If $\alpha_{ij} = 0$, coupling ceases to exist, and we have $k_z = k_1$ for mode 1, and $k_z = k_2$ for mode 2 as expected. More specifically, $k_1 = k$, and $k_2 = -k + \frac{2\pi}{d}$. We can rewrite (5.7.1) and (5.7.2) using matrix notation as

$$\frac{d}{dz}\boldsymbol{\phi} = i\ \overline{\mathbf{K}} \cdot \boldsymbol{\phi}, \tag{5.7.3}$$

where

$$\boldsymbol{\phi} = \begin{bmatrix}\phi_1\\ \phi_2\end{bmatrix}, \quad \overline{\mathbf{K}} = \begin{bmatrix}k_1 & \alpha_{12}\\ \alpha_{21} & k_2\end{bmatrix}. \tag{5.7.4}$$

But letting the solution

$$\boldsymbol{\phi} = e^{ik_z z}\mathbf{a_0}, \tag{5.7.5}$$

we have from (5.7.3) that

$$k_z\mathbf{a_0} = \overline{\mathbf{K}} \cdot \mathbf{a_0}, \tag{5.7.6}$$

or that $\mathbf{a_0}$ is the eigenvector of $\overline{\mathbf{K}}$, and then k_z is its eigenvalue given by

$$\det\left(k_z\overline{\mathbf{I}} - \overline{\mathbf{K}}\right) = 0. \tag{5.7.7}$$

Equation (5.7.7) yields

$$(k_1 - k_z)(k_2 - k_z) - \alpha_{12}\alpha_{21} = 0. \tag{5.7.8}$$

Solving the above for k_z, we have

$$k_z = \frac{(k_1 + k_2)}{2} \pm \frac{1}{2}\sqrt{(k_1 - k_2)^2 + \alpha_{12}\alpha_{21}}. \tag{5.7.9}$$

From energy conservation, we rewrite that

$$\frac{d}{dz}\left(|\phi_1|^2 - |\phi_2|^2\right) = 0, \tag{5.7.10}$$

since two modes propagate in the opposite directions, and hence, their energy flow cancel each other. Equation (5.7.10) is the same as

$$\frac{d}{dz}\boldsymbol{\phi}^\dagger \cdot \overline{\mathbf{S}} \cdot \boldsymbol{\phi} = 0, \tag{5.7.11}$$

where $\overline{\mathbf{S}} = \begin{bmatrix}1 & 0\\ 0 & -1\end{bmatrix}$. From (5.7.11), and using (5.7.3), we have

$$\begin{aligned}\frac{d}{dz}\boldsymbol{\phi}^\dagger \cdot \overline{\mathbf{S}} \cdot \boldsymbol{\phi} &= \left(\frac{d}{dz}\boldsymbol{\phi}^\dagger\right) \cdot \overline{\mathbf{S}} \cdot \boldsymbol{\phi} + \boldsymbol{\phi}^\dagger \cdot \overline{\mathbf{S}} \cdot \frac{d\boldsymbol{\phi}}{dz}\\ &= -i\boldsymbol{\phi}^\dagger \cdot \overline{\mathbf{K}}^\dagger \cdot \overline{\mathbf{S}} \cdot \boldsymbol{\phi} + i\boldsymbol{\phi}^\dagger \cdot \overline{\mathbf{S}} \cdot \overline{\mathbf{K}} \cdot \boldsymbol{\phi}\\ &= i\boldsymbol{\phi}^\dagger \cdot \left(\overline{\mathbf{S}} \cdot \overline{\mathbf{K}} - \overline{\mathbf{K}}^\dagger \cdot \overline{\mathbf{S}}\right) \cdot \boldsymbol{\phi} \quad = 0.\end{aligned} \tag{5.7.12}$$

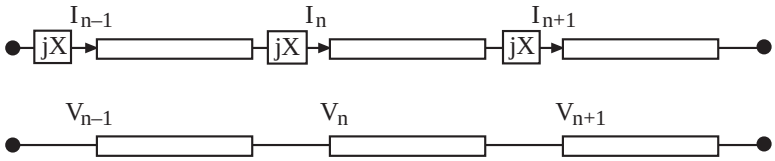


Figure 5.19: Circuit equivalence of a periodic structure.

In order for the above to be zero, it is required that $\alpha_{12} = -\alpha_{21}^*$. (If the two modes are co-propagating, carrying energy the same direction, the condition for energy conservation is $\alpha_{12} = \alpha_{21}^*$.)

For mode 1, $k_1 = k$, and for mode 2, $k_2 = -k + \frac{2\pi}{d}$. From (5.7.9), we have that

$$k_z = \frac{\pi}{d} \pm \sqrt{\left(k - \frac{\pi}{d}\right)^2 - |\alpha_{12}|^2}. \tag{5.7.13}$$

But in the dispersion diagram, we usually fix k_z and solve for k. Alternatively, we can invert Equation (5.7.13) to obtain

$$k = \frac{\pi}{d} \pm \sqrt{\left(k_z - \frac{\pi}{d}\right)^2 - |\alpha_{12}|^2} \tag{5.7.14}$$

In the vicinity of $k_z = \frac{\pi}{d}$, $k = \frac{\pi}{d} \pm |\alpha_{12}|$. In other words, k splits into two distinct values about $\frac{\pi}{d}$ due to forward and backward wave coupling.

5.7.1 Circuit Analysis of Periodic Structure

The above numerical analysis solves the problem exactly within numerical approximation. However, it offers little insight to the problem. As an approximation to the problem, we can think of the corrugated parallel plate waveguide problem as being a transmission line problem with shorted stubs connected in series.

The structure in Figure 5.19 is best analyzed by the chain matrix or the $ABCD$ transmission matrix. It can be shown that the chain matrix connecting the voltages of a section of transmission line of length d is

$$\begin{bmatrix} V_1 \\ I_1 \end{bmatrix} = \begin{bmatrix} \cos(kd) & jZ_c \sin(kd) \\ jY_c \sin(kd) & \cos(kd) \end{bmatrix} \begin{bmatrix} V_2 \\ I_2 \end{bmatrix}. \tag{5.7.15}$$

For a series reactance, the corresponding chain matrix is

$$\begin{bmatrix} V_1 \\ I_1 \end{bmatrix} = \begin{bmatrix} 1 & jX \\ 0 & 1 \end{bmatrix} \begin{bmatrix} V_2 \\ I_2 \end{bmatrix}. \tag{5.7.16}$$

Therefore, the chain matrix connecting V_n, I_n to V_{n+1}, I_{n+1} is

$$\begin{bmatrix} V_n \\ I_n \end{bmatrix} = \begin{bmatrix} \cos(kd) & jZ_c \sin(kd) \\ jY_c \sin(kd) & \cos(kd) \end{bmatrix} \begin{bmatrix} 1 & j \\ 0 & 1 \end{bmatrix} \begin{bmatrix} V_{n+1} \\ I_{n+1} \end{bmatrix}. \tag{5.7.17}$$

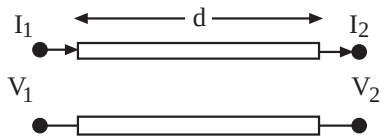


Figure 5.20: A section of a transmission line.

But if a wave propagates on a periodic structure, then

$$V_n = V_{n+1}e^{jk_z d}, \quad I_n = I_{n+1}e^{jk_z d}. \tag{5.7.18}$$

Consequently, (5.7.17) becomes

$$\left\{ \begin{bmatrix} \cos(kd) & jX\cos(kd) + jZ_c \sin(kd) \\ jY_c \sin(kd) & -XY_c \sin(kd) + \cos(kd) \end{bmatrix} - \begin{bmatrix} e^{jk_z d} & 0 \\ 0 & e^{jk_z d} \end{bmatrix} \right\} \begin{bmatrix} V_{n+1} \\ I_{n+1} \end{bmatrix} = 0. \tag{5.7.19}$$

The above will have a nontrivial solution only if

$$\det \begin{pmatrix} A - e^{jk_z d} & B \\ C & D - e^{jk_z d} \end{pmatrix} = 0, \tag{5.7.20}$$

The above is equivalent to

$$AD - BC - (A + D)e^{jk_z d} + e^{2jk_z d} = 0, \tag{5.7.21}$$

Since $AD - BC = 1$ for a reciprocal network, we have

$$\cos(k_z d) = \frac{A + D}{2}, \tag{5.7.22}$$

The above gives the guidance condition for the wave number of a wave propagating on a periodic structure. For the circuit of a series reactance loaded periodic structure, this becomes

$$\cos(k_z d) = \cos(kd) - \frac{XY_c}{2} \sin(kd), \tag{5.7.23}$$

If X is due to a shorted stub, then $X = Z_s \tan(kh)$, and we have

$$\cos(k_z d) = \cos(kd) - \frac{Z_s Y_c}{2} \tan(kh) \sin(kd), \tag{5.7.24}$$

Figure 5.21 plots the right-hand side of (5.7.24) which is $f(kd)$ as a function of kd, and the left-hand side of (5.7.24) which is just $\cos(k_z d)$. For every $k_z d$, we can read off several values of kd such that $\cos(k_z d) = f(kd)$. Hence we arrive at the following diagram for kd versus $k_z d$.

As can be seen from Figure 5.22, there are frequency bands at which real values of $k_z d$ could exist. These are the passbands. In Figure 5.22, we have only shown the first two bands.

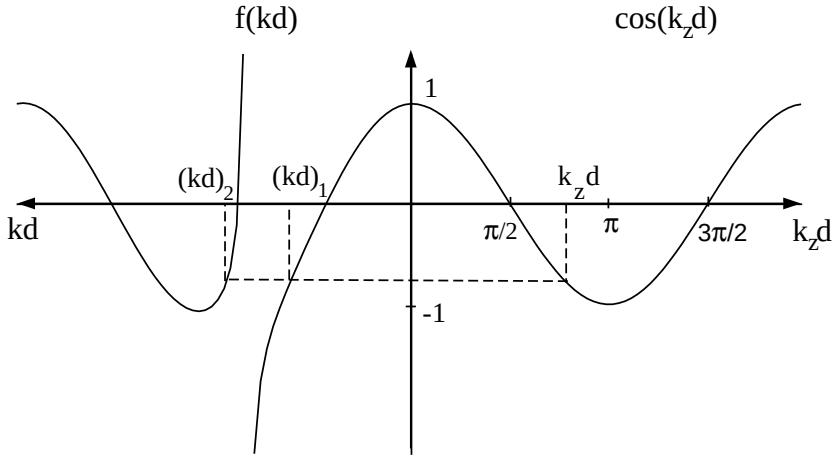


Figure 5.21: Plot of $\cos(k_z d)$ versus $k_z d$ and $f(kd)$ versus kd.

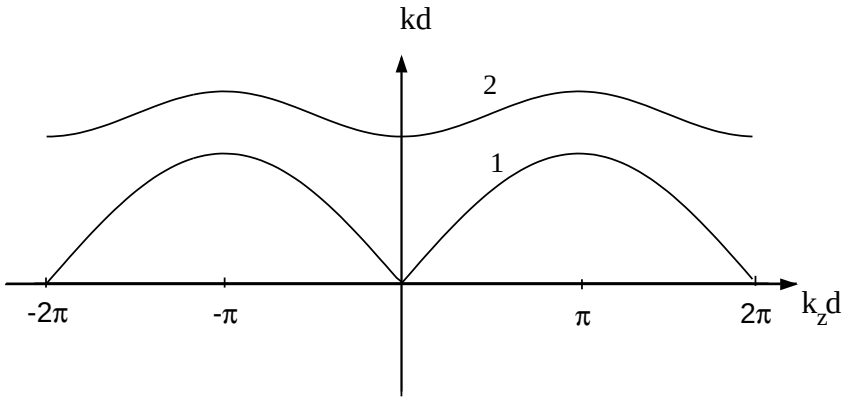


Figure 5.22: The $kd - k_z d$ dispersion diagram of a periodically loaded transmission line.

The band of frequencies at which no real values of $k_z d$ can exist is the stopband. Because of this property, a periodic structure can be used as a filter. Also, notice that the phase velocity is smaller than that of just a TEM mode propagating in a parallel plate waveguide, hence, the name slow-wave structure for a periodic waveguide. It is also possible to have the group velocity opposite in sign to the phase velocity as can be seen from the above diagram. This is used in a backward wave oscillator in microwave circuits.

5.8 Metamaterials

The area of metamaterials has been inspired by the suggestion of Vesalago [49] that if we have both ϵ negative and μ negative material (also called a double-negative or DNG material), then the wave will be a backward wave with the phase velocity traveling in the opposite direction to the group velocity. It was further suggested by Pendry [50] that a DNG material time reverses a field, and hence, can be used to make a perfect lens (or superlens). As a result there have been a flurry of activities in this field.

Unfortunately, many of the superlensing effect of DNG material disappears with the slightest amount of loss or imperfection. However, the excitement in search of the holy grail has inspired many new ideas and structures that could be of interest. Consequently, many other metamaterials have been proposed in recent years. Examples of these are epsilon negative materials (ENG), mu negative materials (MNG), zero index materials (ZIM), different effective index materials. In single negative materials (SNG), one can show that a surface plasmon polariton can be excited at the air-material interface. In ZIM, the wavelength of the field is infinite, and hence, many concepts prevailing in circuit theory can be applied in optical frequencies. Our ability to fabricate different index materials also gives rise to the field of transformation optics, which further yield the concept of cloaking. Since the fabrication of 3D bulk metamaterials has been difficult, there has been interest in meta-surfaces that can be fabricated easily by epitaxial techniques. Also, advances in nano-fabrication technology allow the fabrication of artificial atoms such as Cooper-pair boxes. Suggestions of quantum metamaterials have emerged.

Here, we will briefly review the physics of DNG materials. We can start with Maxwell's equations for regular materials in the frequency domain:

$$\begin{aligned} \nabla \times \mathbf{E} &= i\omega\mu\mathbf{H} \\ \nabla \times \mathbf{H} &= -i\omega\epsilon\mathbf{E} + \mathbf{J} \\ \nabla \cdot \epsilon\mathbf{E} &= \rho \\ \nabla \cdot \mu\mathbf{H} &= 0 \end{aligned} \tag{5.8.1}$$

If we were to change the sign of ϵ and μ in Equations (5.8.1), we arrive at:

$$\begin{aligned} \nabla \times \mathbf{E} &= -i\omega\mu\mathbf{H} \\ \nabla \times \mathbf{H} &= i\omega\epsilon\mathbf{E} + \mathbf{J} \\ \nabla \cdot \epsilon\mathbf{E} &= -\rho \\ \nabla \cdot \mu\mathbf{H} &= 0 \end{aligned} \tag{5.8.2}$$

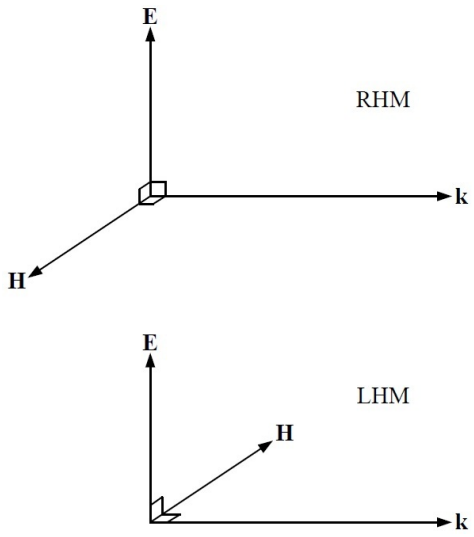


Figure 5.23: An RHM plane wave in a regular medium versus an LHM plane wave that propagates in a double negative medium.

Notice that once the solution to (5.8.1) is obtained, we can obtain the solution to (5.8.2) by changing the signs of $\mathbf{H}$, $\mathbf{J}$, and ρ. Hence power flow, which is defined to be $\mathbf{E} \times \mathbf{H}^*$, points in the opposite direction for solutions of Equations (5.8.2) compared to that of solutions of Equations (5.8.1). Also, the $\mathbf{E}$, $\mathbf{H}$ and $\mathbf{k}$ vectors form the left-hand rule (see Figure 5.23). Hence, a DNG medium is also called a left-handed medium (LHM) as opposed to the regular right-handed medium (RHM). As the plane wave has a k-vector that points in the opposite direction to that of power flow, and that $k = -\omega\sqrt{\epsilon\mu}$, a DNG material is also called a negative index material (NIM).

The fact that k-vector points in opposition to the power flow, it implies that the phase velocity is opposite in sign to the group velocity. This implies that the medium has to be dispersive as the phase velocity $v_{ph} = \omega/k$ while the group velocity is $v_{gr} = (dk/d\omega)^{-1}$. They can only be opposite of each other if $k = \omega\sqrt{\mu(\omega)\epsilon(\omega)}$.

We can also easily prove that LHM has to be frequency dispersive by *reductio ad absurdum* [53]. Assuming that the medium is frequency independent or non-dispersive. Then we can convert Maxwell's equations back to the time domain. Subsequently, a conservation law for power flow can be easily derived for a source-free region to be:

$$\nabla \cdot (\mathbf{E} \times \mathbf{H}) = -\frac{1}{2}\frac{\partial}{\partial t}\left\{\mu|\mathbf{H}|^2 + \epsilon|\mathbf{E}|^2\right\}$$

or in integral form via the use of Gauss' divergence theorem:

$$\oint_S d\mathbf{S} \cdot (\mathbf{E} \times \mathbf{H}) = -\frac{1}{2}\frac{\partial}{\partial t}\int_V dV\left\{\mu|\mathbf{H}|^2 + \epsilon|\mathbf{E}|^2\right\}$$

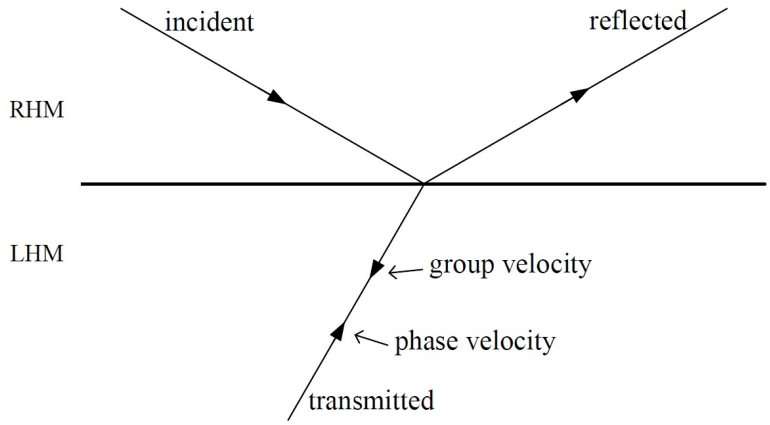


Figure 5.24: Negative refraction happens at an interface because of the need for phase matching and energy conservation.

In the above, when LHM and a non-dispersive RHM, such as vacuum, co-exist and mix with each other, ϵ and μ can have different signs for different regions. Hence, with a proper choice of surface and volume, the right-hand side of the above can be zero, while the left-hand side is non-zero violating energy conservation! However, if the medium is dispersive, as we have learned from Chapter 1, the above is not the correct expression for energy storage. Therefore, double negative materials can only be made over a narrow bandwidth when it co-exists with regular materials. Moreover, we still have to use $\mathbf{E} \times \mathbf{H}$ as the direction of power flow if the DNG has been made from regular materials. Consequently, the $\mathbf{k}$ vector of a plane wave points in opposite direction to that of power flow.

As a result, the group velocity of a DNG material has to be opposite to that of the phase velocity. At a simple interface, due to phase matching and energy conservation requirements, negative refraction occurs as shown in Figure 5.24. Phase matching requires the $\mathbf{k}$ vector to be aligned as shown, but energy conservation requires that the group velocity be opposite to that of the phase velocity. Because of negative refraction, it has proposed that negative refraction can be used for focusing [49].

5.8.1 Evanescent Amplification by a Matched DNG Slab

A matched DNG slab has the capability of amplifying evanescent wave, as shall be shown. This has been suggested to use a DNG slab as a super lens [50].

The transmission coefficient, as we have seen from the section on Fabry-Perot etalon is given by

$$\tilde{T} = \frac{T_{12}T_{21}e^{ik_{2z}d}}{1 - R_{21}^2 e^{2ik_{2z}d}} = \frac{(1 - R_{21}^2)e^{ik_{2z}d}}{1 - R_{21}^2 e^{2ik_{2z}d}} \tag{5.8.3}$$

with

$$R_{ij}^{TE} = \frac{\mu_j k_{iz} - \mu_i k_{jz}}{\mu_j k_{iz} + \mu_i k_{jz}}, \qquad R_{ij}^{TM} = \frac{\epsilon_j k_{iz} - \epsilon_i k_{jz}}{\epsilon_j k_{iz} + \epsilon_i k_{jz}}, \tag{5.8.4}$$

depending on whether we are calculating $\tilde{T}$ for a TE wave or a TM wave. Also, $T_{ij} = 1 + R_{ij}$ in the above, and we have assumed that the z axis is normal to the slab.

For a matched medium, we pick $\mu_2 = -\mu_1$ and $\epsilon_2 = -\epsilon_1$. For propagating wave, we pick the phase velocity in region 2 to be opposite to that of regions 1 and 3. In other words, $k_{2z} = -k_{1z}$. Then it is quite clear that $R_{12} = 0$, and

$$\tilde{T} = e^{-ik_{1z}h} \tag{5.8.5}$$

In fact, it can be shown that the above result is independent of whatever sign we pick for k_{2z}. Because of the backward wave nature of the wave in LHM, the above represents a phase advancement instead of phase retardation in RHM. The phase advancement compensates for the phase retardation from the source to the first slab interface, which can also be thought of as time reversal.

For the evanescent spectrum it can be shown that, irrespective of the branch of square root for choose for k_{2z}, the transmission coefficient is always

$$\tilde{T} = e^{\alpha_{1z}h} \tag{5.8.6}$$

where $\alpha_{1z} = \sqrt{k_1^2 - k_x^2}$ where k_x is the wavenumber parallel to the slab. Hence, the evanescent wave has been amplified after it has passed through the DNG slab. This can be used as a super lens for super resolution phenomena [50].

To understand this, we first look at the Weyl [15] identity

$$\frac{e^{ik_0 r}}{r} = \frac{i}{2\pi} \iint_{-\infty}^{\infty} dk_x dk_y \, \frac{e^{ik_x x + ik_y y + ik_z |z|}}{k_z},$$

where $k_x^2 + k_y^2 + k_z^2 = k_0^2$, or $k_z = (k_0^2 - k_x^2 - k_y^2)^{1/2}$. The above says that the field produced by a point source consists of both evanescent spectrum as well as the propagating spectrum. However, it is the evanescent spectrum that contains the high-resolution information of the point source. As we move away from the point source, the evanescent spectrum becomes smaller, and therefore, the high-resolution information is lost. However, if this evanescent spectrum can be reconstituted by using the DNG slab, then the high-resolution information can be regained. This is illustrated in Figure 5.25. However, this reconstitution of the evanescent spectrum is a highly unstable undertaking and is easily upset by loss or geometry imperfection [52, 53].

5.8.2 Composite Right-Left Handed Transmission Line

One interesting technology that has emerged in this area is the composite right-left handed (CRLH) transmission line [54, 55]. We will analyze this transmission line next using circuit theory. For the case shown in Figure 5.26, we can write down Kirchoff voltage law to get

$$V(z + \Delta z) - V(z) = -I(z) Z \Delta z \tag{5.8.7}$$

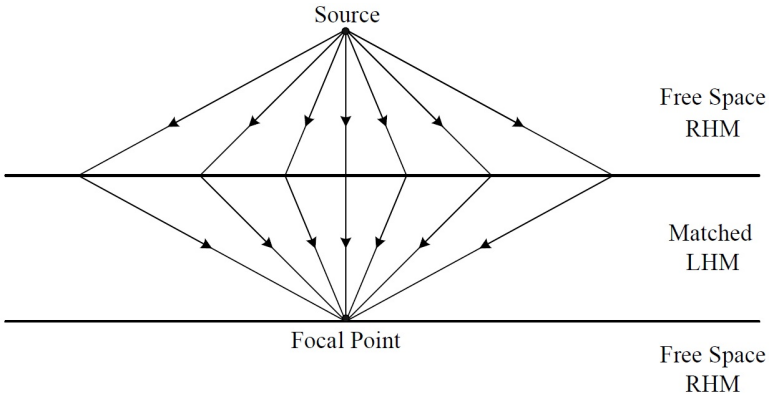


Figure 5.25: Super lens or super resolution effect of a matched DNG slab. In the LHM material, the arrows show the direction of power flow. But the phase velocities are opposite to the directions of the arrows.

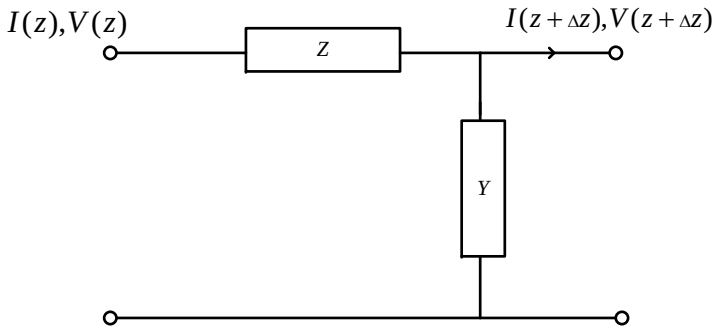


Figure 5.26: Lumped element model for deriving the CRLH transmission line equations.

$$I(z+\Delta z) - I(z) = -V(z+\Delta z)Y\Delta z \tag{5.8.8}$$

where Z and Y are per unit length impedance and admittance. The above becomes the Telegraphers equations when $\Delta z \to 0$

$$\frac{\partial V(z)}{\partial z} = -ZI(z) \tag{5.8.9}$$

$$\frac{\partial I(z)}{\partial z} = -YV(z) \tag{5.8.10}$$

or

$$\frac{\partial^2 V(z)}{\partial z^2} = ZYV(z) \tag{5.8.11}$$

$$\frac{\partial^2 I(z)}{\partial z^2} = ZYV(z) \tag{5.8.12}$$

or

$$\beta^2 = -YZ \tag{5.8.13}$$

$$\beta = \pm j\sqrt{YZ} \tag{5.8.14}$$

If

$$Z = \frac{1}{j\omega C_L} \tag{5.8.15}$$

$$Y = \frac{1}{j\omega L_L} \tag{5.8.16}$$

then above imitates a double negative material. We can pick one branch of the square root to get

$$\beta = -\frac{1}{\omega\sqrt{L_L C_L}} < 0 \tag{5.8.17}$$

$$Z_0 = \sqrt{\frac{Z}{Y}} = \sqrt{\frac{L_L}{C_L}} > 0 \tag{5.8.18}$$

$$v_{ph} = \frac{\omega}{\beta} = -\omega^2\sqrt{L_L C_L} < 0 \tag{5.8.19}$$

$$v_g = \frac{d\omega}{d\beta} = -\frac{\omega}{k} = \omega^2\sqrt{L_L C_L} > 0 \tag{5.8.20}$$

The above indicates that the group velocity is opposite to the phase velocity: it supports a backward wave.

For a CRLH transmission line, we have

$$Z = j\left(\omega L_R - \frac{1}{\omega C_L}\right) \tag{5.8.21}$$

$$Y = j\left(\omega C_R - \frac{1}{\omega L_L}\right) \tag{5.8.22}$$

Then if $\omega > \omega_{se} = \frac{1}{\sqrt{L_R C_L}}$, $\omega > \omega_{sh} = \frac{1}{\sqrt{L_L C_R}}$, where ω_{se} and ω_{sh} are the resonant frequencies of the series and shunt impedances respectively, then

$$\beta = \sqrt{\left(\omega L_R - \frac{1}{\omega C_L}\right)\left(\omega C_R - \frac{1}{\omega L_L}\right)} > 0 \tag{5.8.23}$$

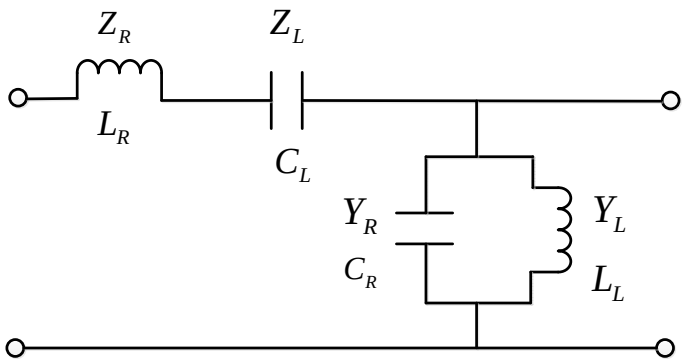


Figure 5.27: The lumped element model for a CRLH transmission line.

If $\omega < \omega_{se}$, $\omega < \omega_{sh}$, then

$$\beta = -\sqrt{\left(\frac{1}{\omega C_L} - \omega L_R\right)\left(\frac{1}{\omega L_L} - \omega C_R\right)} \tag{5.8.24}$$

and $\beta < 0$. A band gap exists for $\min(\omega_{se}, \omega_{sh}) < \omega < \max(\omega_{se}, \omega_{sh})$. If we make $\omega_{se} = \omega_{sh}$, the band gap disappears.

At $\omega = \omega_0 > 0$, there exists a wave with $\beta = 0$, or a constant phase wave. At this frequency, the series resonance becomes a short and the shunt resonance becomes an open. This idea can be used to design equi-phase loop antenna for RFID (radio frequency identification) and MRI (magnetic resonance imaging) applications. When $\beta < \omega/c$, the wave that propagates on the CRLH line leaks energy to the space around it: the structure can be used to make leaky-wave antennas.

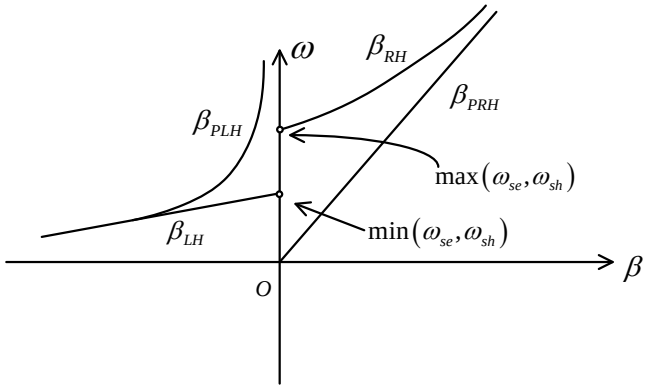


Figure 5.28: The ω-β diagrams of the pure RH (PRH) line, the pure LH (PLH) line, and that of the CRLH line. For CRLH, and band gap exists where the frequency ω has no solution for β or a propagating mode.

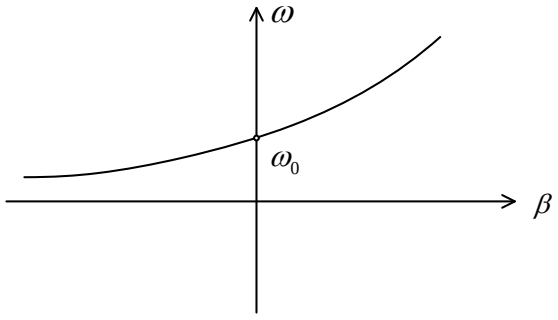


Figure 5.29: The choice of L_R, L_L, C_R, and C_L where $\omega_{se} = \omega_{sh}$, and the band gap disappears.

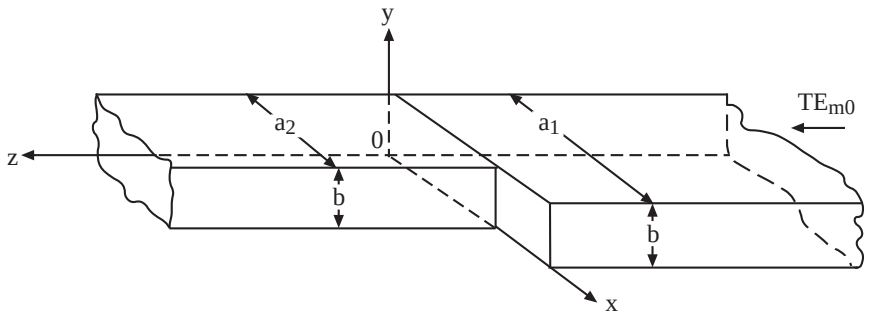


Figure 5.30: Problem 5-3

Exercises for Chapter 5

Problem 5-1: A waveguide junction is formed from two rectangular waveguides, one of which is 2 cm by 4 cm, and the other is 2.5 cm by 4 cm. Assume that a 5 GHz TE_{10} mode is incident on the waveguide junction from the smaller waveguide. Use the transmission line model, ascertain the approximate amplitudes of the reflection and transmission coefficients for the TE_{10} mode.

Problem 5-2: To understand the number of modes required for convergence in the mode-matching problem in a waveguide, consider the matching of the following two Fourier series:

$$f_1(x) = \sum_{n=-P_1}^{P_1} a_n e^{in\pi x/d_1}, \qquad 0 < x < d_1, \tag{5.8.25}$$

$$= 0, \qquad \text{otherwise}, \tag{5.8.26}$$

$$\hat{f}_1(x) = \sum_{n'=-P_2}^{P_2} b'_n e^{in'\pi x/d_2} = \begin{cases} f_1(x), & 0 < x < d_1, \\ 0, & d_1 < x < d_2, \end{cases} \tag{5.8.27}$$

where $d_2 > d_1$. Find the coefficents b'_n in terms of a_n. Show that in order for $\hat{f}_1(x)$ to approximate $f_1(x)$ well, we require that $P_2/d_2 >> P_1/d_1$.

Problem 5-3:

If a sum of TE_{m0} modes are incident at the discontinuity in the waveguide in Figure 5.30, due to the symmetry of the problem, only TE_{m0} modes are reflected and transmitted. Given that the incident modes are described by

$$H_z = \sum_m H_{m0} \cos\left(\frac{m\pi x}{a_1}\right) e^{ik_{mz}z}$$

find the transmission and reflection operators that describe mode conversions at the discontinuity.

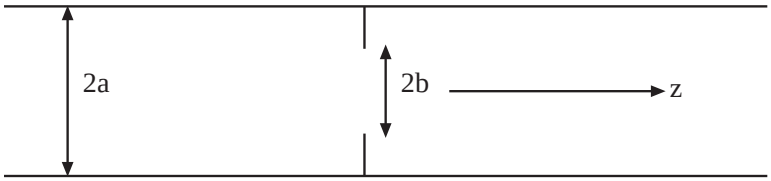


Figure 5.31: Problem 5-4

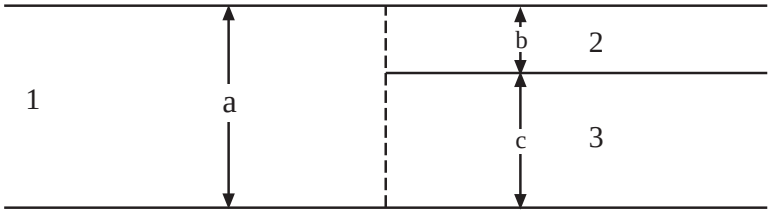


Figure 5.32: Problem 5-5

Problem 5-4:

A TE_{11} mode is propagating in a circular waveguide with radius a as shown in Figure 5.31. A circular diaphragm with radius b is placed at $z = 0$.

(a) Find the reflection and transmission operators due to the presence of this diaphragm.

(b) Give explicit expressions for the elements of the matrices involved in the description of the reflection and transmission operators.

Problem 5-5:

A parallel plate waveguide is bifurcated as shown in Figure 5.32. Assume a bunch of TM_n mode incident from the left of the waveguide.

(a) Using the formulation of this Chapter, find the reflection operator in waveguide 1, and transmission operators in waveguides 2 and 3.

The field at the waveguide discontinuity can either be approximated by the modes of waveguide 1, or the modes of waveguides 2 and 3.

(b) If the field at the discontinuity is approximated by N modes of waveguide 1, how many reflected modes are there in waveguide 1, and how many transmitted modes are there in waveguides 2 and 3.

(c) If the field at the discontinuity is approximated by M modes in waveguide 2 and P modes in waveguide 3, how many reflected modes are there in waveguide 1, and how many transmitted modes are there in waveguides 2 and 3. In this latter method of solving the problem, what should be the relative ratio of M and P in order to maintain the same accuracy for the transmitted field in waveguides 2 and 3. (Hint: If only

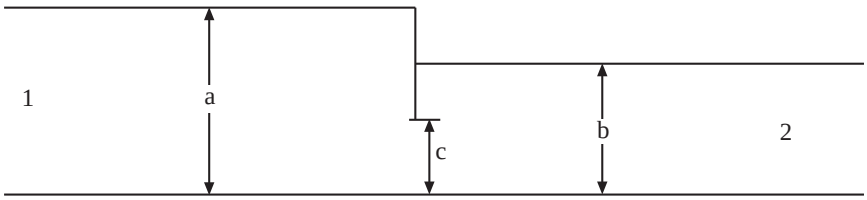


Figure 5.33: Problem 5-6

1 mode is assumed in waveguide 2 while 10 modes are assumed in waveguide 3 at the discontinuity, the accuracy of the transmitted field in waveguide 1 will not be as accurate as that in waveguide 2. So, a certain ratio needs be maintained to achieve the same order of accuracy. Think of this in terms of Fourier series expansions.)

Problem 5-6:
A parallel plate waveguide is shown in Figure 5.33. Assume a TM_1 mode incident from the left of the waveguide.

(a) Using the formulation of this Chapter, find the reflection operator in waveguide 1, and transmission operator in waveguide 2.

(b) If the field at the discontinuity is approximated by N modes of a waveguide with the same dimension as the aperture, how many reflected modes are there in waveguide 1, and how many transmitted modes are there in waveguide 2 in theory, in order to obtain a numerically accurate solution.

Problem 5-7: Show that a typical integral in Equation (5.2.39) looks like the expression in Equation (5.2.40a).

Problem 5-8: Expand (5.4.10) in a geometrical series, and give a physical explanation of each term of the geometrical series. Would this series always converge? What can you say about the norm of the reflection matrices?

Problem 5-9: Explain how you would construct a time reversed solution to Maxwell's equations once you have found a time dependent solution. Is the time reversed solution for a lossy waveguide a physical solution?

Problem 5-10: Show that when the Neumann boundary condition is satisfied for H_x in the corrugated periodic waveguide, then it is equivalent to tangential electric field being zero on the surface of the waveguide wall.

Problem 5-11: Evaluate Equation (5.6.23) and hence Equation (5.6.26) in closeform. Write a computer program to solve for different values of k_z for different k. Discuss how you would truncate the series involved in the equations.

Problem 5-12: Derive Equations (5.7.15) and (5.7.16).

Problem 5-13: Write a computer program to plot kd versus $k_z d$ according to Equation (5.7.24).

Problem 5-14: Derive the guidance condition for a transmission line that is periodically loaded with a shunt capacitance. Plot kd versus $k_z d$ for this circuit.

Bibliography

[1] J. Helszajn, *Waveguide Junction Circulators: Theory and Practice,* John Wiley, NY, 1998.

[2] R.E. Collin, *Field Theory of Guided Waves,* IEEE Press, Piscataway, NJ, 1991.

[3] R.E. Collin, *Foundation for Microwave Engineering,* IEEE Press, Piscataway, NJ, 2001.

[4] L. Matthaei, L. Young, and E. M. T. Jones, *Microwave Filters Impedance Matching Network, and Coupling Structures,* New York, McGraw-Hill, 1964.

[5] R. Mittra and S.W. Lee, *Analytical Techniques in the Theory of Guided Waves,* New York, Macmillan, 1971.

[6] E. Kühn, "A mode-matching method for solving field problem in waveguide and resonator circuits," *Arch. Elek. Übertragung*, vol. 27, 511518, 1973.

[7] L. Lewin, *Theory of Waveguides*, London: Newnes-Butterworths, 1975.

[8] M. Koshiba, M. Sato, and M. Suzuki, "Application of finite element method to H-plane waveguide discontinuities," *Electron. Lett.*, Vol. 18, 364365, 1982.

[9] H. Patzelt and F. Arndt, "Double-plane steps in rectangular waveguides and their application for transformers, irises and filters," *IEEE Trans. Microwave Theory and Techniques*, vol. MTT-30, pp. 771-776, May 1982.

[10] W.C. Chew, S. Barone, B. Anderson and C. Hennessy, "Diffraction of axisymmetric waves in a borehole by bed boundary discontinuities," *Geophysics*, vol. 49, no. 10, pp. 1586-1595, Oct. 1984.

[11] M. Koshiba and M. Suzuki, "Application of the boundary element method to waveguide discontinuities," *IEEE Trans. Microwave Theory Tech.*, Vol. 34, 301307, 1986.

[12] J. D. Wade and R. H. MacPhie, "Scattering at circular-to-rectangular waveguide junctions," *IEEE Trans. Microwave Theory Tech.*, vol. 34, pp. 10851091, Nov. 1986.

[13] F. Arndt, I. Ahrens, U. Papziner, U. Wiechmann, and R. Wilkeit, "Optimized E-plane T-junction series power dividers," *IEEE Trans. Microwave Theory Tech.*, vol. 35, 10521059, 1987.

[14] F. Alessandri, G. Bartolucci, R. Sorrentino, "Admittance matrix formulation of waveguide discontinuity problems: computer-aided design of branch guide directional couplers," *IEEE Trans. Microwave Theory and Techniques*, Vol. 36, No. 2, pp 394-403, Feb. 1988.

[15] R.R. Mansour, R.S.K. Tong, and R.H. McPhie, "Simplified description of the field distribution in finlines and ridge waveguides and its application to the analysis of E-plane discontinuities," *IEEE Trans. Microwave Theory and Techniques*, vol. MTT-36, pp. 1825-1832, Dec. 1988.

[16] T. Itoh, Ed., *Numerical Techniques for microwave and millimeter wave passive structures,* New York, Wiley, 1989.

[17] Q.H. Liu and W.C. Chew, "Numerical mode-matching method for the multiregion, vertically stratified media," *IEEE Trans. Antennas Propag.*, vol. AP-38, no. 4, pp. 498-506, Apr. 1990.

[18] M. Guglielmi and C. Newport, "Rigorous, multimode equivalent network represntation of inductive discontinuities," *IEEE Trans. Microwave Theory and Techniques*, Vol. 38, No. 11, pp 1651-1659, Nov. 1990.

[19] Q.H. Liu and W.C. Chew, "Analysis of discontinuities in planar dielectric waveguides: An eigenmode propagation method," *IEEE Trans. Micro. Theory Tech.*, vol. 39, no. 3, pp. 422-430, Mar. 1991.

[20] W. C. Chew, K. H. Lin, J. Friedrich and C. H. Chan, "Reflection and transmission operators for general discontinuities in waveguides," *J. Elect. Waves Appl.*, vol. 5, no. 8, pp. 819-834, 1991.

[21] K. Ise, K. Inoue, and M. Koshiba, "Three-dimensional finite-element method with edge elements for electromagnetic waveguide discontinuities," *IEEE Trans. Microwave Theory Tech.*, vol. 39, pp. 1289-1295, Aug. 1991.

[22] X. Liang, K. A. Zaki, and A. E. Atia, "Rigorous three plane mode-matching technique for characterizing waveguide T-junctions, and its application in multiplexer design," *IEEE Trans. Microwave Theory Tech.*, Vol. 39, 21382147, 1991.

[23] T. Sieverding and F. Arndt, "Field theoretic CAD of open or aperture matched T-junction coupled rectangular waveguide structures," *IEEE Trans. Microwave Theory Tech.*, vol. 40, 353362, 1992.

[24] R.R. Mansour and J. Dude, "Analysis of microstrip T-junction and its applications to the design of transfer switches," IEEE-MTT-S Dig., pp. 889-892, 1992.

[25] F. Alessandri, M. Mongiardo, and R. Sorrentino, "A technique for the fullwave automatic synthesis of waveguide components: application to fixed phase shifters," *IEEE Trans. Microwave Theory Tech.*, Vol. 40, 14841495, 1992.

[26] T. Sieverding and F. Arndt, "Modal analysis of the magic tee," *IEEE Microwave Guided Wave Lett.*, vol. 3, pp. 150-152, May 1993.

[27] R. Keller and F. Arndt, "Rigorous modal analysis of the asymmetric rectangular iris in circular waveguides," *IEEE Microwave Guided Wave Lett.*, vol. 3, pp. 185187, June 1993.

[28] J.H. Lee, H. J. Eom, J. W. Lee, and K. Yoshitomi, "Transverse electric mode scattering from rectangular grooves in parallel plate," *Radio Science*, vol. 29, 12151218, 1994.

[29] J.M. Rebollar, J. Esteban, and J. E. Page, "Fullwave analysis of three and four-port rectangular waveguide junctions," *IEEE Trans. Microwave Theory Tech.*, vol. 42, 256263, 1994.

[30] F. Alessandri, M. Mongiardo, and R. Sorrentino, "Rigorous mode matching analysis of mitered E-plane bends in rectangular waveguide," *IEEE Microwave and Guide Wave Lett.*, vol. 4, 408410, 1994.

[31] J.W. Lee and H. J. Eom, "TE-mode scattering from two junctions in H-plane waveguide," *IEEE Trans. Microwave Theory Tech.*, vol. 42, 601606, 1994.

[32] Z. Ma and E. Yamashita, "Port reflection coefficient method for solving multi-port microwave network problems," *IEEE Trans. Microwave Theory Tech.*, vol. 43, 331337, 1995.

[33] W. Pascher and R. Pregla, "Analysis of rectangular waveguide discontinuities by the method of lines," *IEEE Trans. Microwave Theory Tech.*, vol. 43, 416420, 1995.

[34] R.H. MacPhie and K.L. Wu, "Scattering at the junction of a rectangular waveguide and a larger circular waveguide," *IEEE Trans. Microwave Theory Tech.*, vol. 43, pp. 20412045, Sept. 1995.

[35] Z. Shen and R.H. MacPhie, "Scattering by a thick off-centered circular iris in circular waveguide," *IEEE Trans. Microwave Theory Tech.*, vol. 43, pp. 26392642, Nov. 1995.

[36] A. Weisshaar, M. Mongiardo, and V. K. Tripathi, "CAD-oriented equivalent circuit modelling of step discontinities in rectangular waveguides," *IEEE Microwave and Guided Wave Letters*, vol. 6, no. 4, pp. 171-173, April 1996. 38

[37] A.A. Melcon, G. Connor, M. Guglielmi, "New simple procedure for the computation of the multimode admittance or impedance matrix of planar waveguide junctions," *IEEE Trans. Microwave Theory and Techniques*, Vol. 44, No. 3, pp 413-416, March 1996

[38] C.T. Iatrou and M. Cavenago, "Field analysis of rectangular waveguide open junction," *IEEE Trans. Microwave Theory Tech.*, vol. 45, 165172, 1997.

[39] P. Matras, R. Bunger, and F. Arndt, "Mode scattering matrix of the general step discontinuity in elliptical waveguides," *IEEE Trans. Microwave Theory Tech.*, vol. 45, pp. 453457, Mar. 1997.

[40] S.P. Yeo and S.G. Teo, "Thick eccentric circular iris in circular waveguide," *IEEE Trans. Microwave Theory Tech.*, vol. 46, pp. 11771180, Aug. 1998.

[41] H. Jia, K. Yoshitomi, and K. Yasumoto, "Rigorous analysis of rectangular waveguide junctions by Fourier transform technique," *Progress in Electromagnetics Research*, PIER 20, 263282, 1998.

[42] S. Amari. J. Bornemann, A. Laisn, and R. Vahldieck, "Design and analysis of iris-coupled and dielectric loaded 1/8-cut TE01-mode microwave bandpass filters," *IEEE Trans. Microwave Theory Tech.* 49, 413-421, March 2001.

[43] K. Radhakrishnan and W.C. Chew, "Efficient analysis of waveguiding structures," in *Fast and Efficient Algorithms in Computational Electromagnetics,* W. Chew, J. Jin, E. Michielssen, and J. Song, editors, Artech House, Boston, 2001.

[44] Z.X Shen, C.K. Law, C. Qian, "Hybrid finite-element-modal-expansion method for matched magic T-junction," *IEEE Transactions on Magnetics*, vol. 38, no. 2, pp. 385-388, 2002.

[45] W.C. Chew, *Waves and Fields in Inhomogeneous Media,* Van Nostrand Reinhold, New York, 1990, reprinted, IEEE Press, Piscataway, NJ, 1995.

[46] C. Kittel, *Introduction to Solid State Physics,* 7th Ed., John Wiley & Sons, 1995.

[47] J.A. Kong, *Electromagnetic Wave Theory,* EMW Publishing, Cambridge, MA, 2000.

[48] H.A. Haus, *Electromagnetic Noise and Quantum Optical Measurements,* Berlin: Springer-Verlag, 2000.

[49] V. G. Veselago, "The electrodynamics of substances with simultaneously negative values of permittivity and permeability," *Soviet Physics USPEKI,* Vol. 10, No. 509, 1968.

[50] J. B. Pendry, "Negative refraction makes a perfect lens," *Phys. Rev. Lett.*, Vol 85, No. 3966, 2000.

[51] R. A. Shelby, D. R. Smith, S. C. Nemat-Nasser, and S. Schultz, "Microwave transmission through a two-dimensional, isotropic, left-handed metamaterial," *Applied Physics Letters,* Vol. 78, Issue 4, pp. 489-491, Jan. 22, 2001.

[52] J.R. Thomas and A. Ishimaru, "Transmission properties of material with relative permittivity and permeability close to 1, *Proc. SPIE*, vol. 4806, pp. 167-175, 2002.

[53] W. C. Chew, "Some reflections on double negative materials," *PIER* 51, pp. 1-26, 2005.

[54] A. Lai, T. Itoh, and C. Caloz, "Composite right/left-handed transmission line metamaterials," *IEEE Microwave Magazine*, 5 (3), 34-50, 2004.

[55] C. Caloz and T. Itoh, "Transmission line approach of left-handed (LH) structures and microstrip implementation of an artificial LH transmission line, *IEEE Trans. Antennas Propagat.*, vol. 52, no. 5, pp. 11591166, May 2004.

Chapter 6

Optical Waveguides

Optical waveguides are some of the most important waveguides. Their importance stems from the broad bandwidth and low loss deliverable by optical communication systems [1, 2]. As a result, optical fiber cables have replaced transmission line cables as submarine cables throughout the world for global communication [3]. The use of optical fiber for communication was first proposed by Kao and Hockham [4]. The success of Corning Glass Works on making fiber of loss below 20 dB/km in the 1970s spurred tremendous interest in the use of optical fiber. An interesting account of the history is given in Okoshi [5].

Optical fibers work by the physics of total internal reflection. Waves are confined within a waveguide by total internal reflection due to the presence of a dielectric interface. Hence, no metallic part is needed in the construction of such a waveguide. The presence of metallic component is deleterious at high frequencies due to the loss it incurs. Optical fibers now can have a loss as low as 0.2 dB/km [2].

In an optical waveguide, which usually is an open waveguide, a wave is guided along a structure, but the field is not enclosed completely in the structure—the field extends to infinity. However, outside the waveguide, the field is evanescent, an it decays exponentially away from the guiding structure; hence, the energy of the wave is still localized around the guiding structure. The guiding structure is often filled with inhomogeneous medium. Therefore, many properties of inhomogeneously filled waveguides are also true in open, optical waveguides.

Due to the importance of optical waveguides, there has been a tremendous amount of work on this subject. Much of the work can be found from the references for this chapter and the references therein. This chapter only serves to provide a sampling of some topics available in the vast literature. Optical waveguides are still under intensive research by many workers (see reference list). A recent development is the use of photonic crystals for optical waveguides. Artificial photonic crystals can generate stop bands (band gaps) in which the wave has to be evanescent in the crystal. The band gap structures are then used to trap waves inside the waveguide. An excellent overview of planar lightwave circuits is given by Okamoto [14].

6.1 Surface Waveguides–Dielectric Slab Waveguides

An example of an open waveguide is a dielectric slab waveguide. The waveguide is made with dielectric coating on a ground plane, or in the case of optical thin film waveguides, it is a coating of an optically more dense medium on top of an optically less dense substrate [5,7–15].

Due to the symmetry of the geometry, we can decompose the field inside such a waveguide into TM and TE types. The mode is guided by total internal reflection. This is only possible if $\epsilon_1 > \epsilon_0$ and $\epsilon_1 > \epsilon_2$. At total internal reflection, the fields in region 0 and 2 are evanescent, and hence they decay exponentially away from the structure. Therefore, most of the energy of the mode is still trapped and localized in the vicinity of the structure.

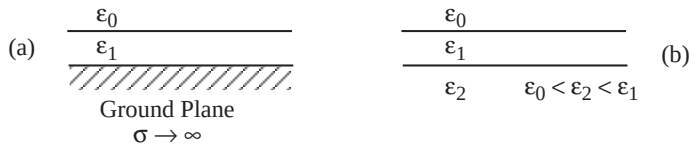


Figure 6.1: Geometry for dielectric slab waveguides.

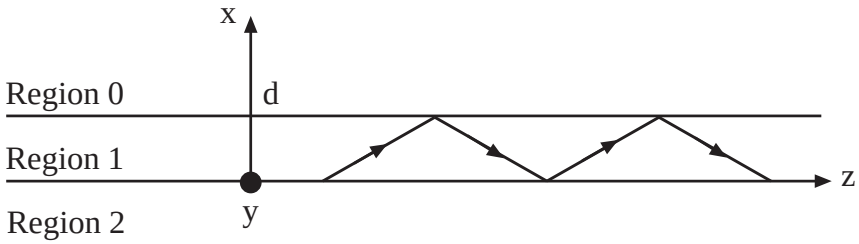


Figure 6.2: Bouncing waves in a dielectric slab waveguide.

If a TM wave is in a dielectric slab, we can write the field in region 1 as

$$\mathbf{H}_1 = \hat{y}\left[A_1 e^{ik_{1x}x} + B_1 e^{-ik_{1x}x}\right] e^{ik_z z}. \tag{6.1.1}$$

Equation (6.1.1) has the physical meaning that the wave in region 1 is representable as bouncing waves. At $x = 0$, the upgoing wave is the reflection of the downgoing wave; hence, we have

$$A_1 = R_{12}^{TM} B_1, \tag{6.1.2}$$

where R_{12}^{TM} is the TM reflection coefficient at the 1-2 interface. If there are subsurface layers, R_{12}^{TM} could be the generalized reflection coefficient that includes subsurface reflections. Otherwise, it is just the single interface, Fresnel reflection coefficient for a TM wave. At the upper interface at $x = d$, we require that the downgoing wave is a reflection of the upgoing wave, i.e.,

$$B_1 e^{-ik_{1x}d} = R_{10}^{TM} e^{ik_{1x}d} A_1. \tag{6.1.3}$$

For non-trivial A_1 and B_1, Equations (6.1.2) and (6.1.3) imply that

$$1 - R_{10}^{TM} R_{12}^{TM} e^{2ik_{1x}d} = 0. \tag{6.1.4}$$

The above is the guidance condition sometimes known as the transverse resonance condition for TM modes in a dielectric slab, with

$$R_{ij}^{TM} = \frac{\epsilon_j k_{ix} - \epsilon_i k_{jx}}{\epsilon_j k_{ix} + \epsilon_i k_{jx}}, \qquad k_{ix} = \sqrt{k_i^2 - k_z^2}. \tag{6.1.5}$$

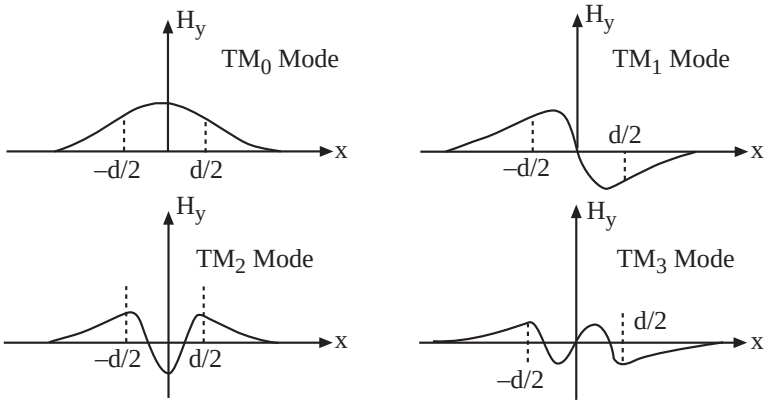


Figure 6.3: Field distributions for different modes in a dielectric slab waveguide.

Due to Equation (6.1.5), the guidance condition (6.1.4) can be expressed entirely as a function of k_z. We can either solve Equation (6.1.4) graphically or numerically on a computer. Once the value of k_z that satisfies (6.1.4) is found, it can be used in (6.1.2) or (6.1.3) to find a relationship between A_1 and B_1. The fields in region 0 and 2 can be found easily, i.e.,

$$\mathbf{H}_0 = \hat{y} T_{10}^{TM} A_1 e^{ik_{1x}d + ik_{0x}(x-d) + ik_z z}, \tag{6.1.6a}$$

$$\mathbf{H}_2 = \hat{y} T_{12}^{TM} B_1 e^{-ik_{2x}x + ik_z z}. \tag{6.1.6b}$$

In other words, the field in region 0 is a consequence of the transmission of the upgoing wave in region 1, while the field in region 2 is a consequence of the transmission of the downgoing wave in region 1. In the above, T_{ij} is a transmission coefficient with

$$T_{ij}^{TM} = 1 + R_{ij}^{TM}. \tag{6.1.7}$$

In order for a mode to be trapped, the field has to decay exponentially in the x direction. Therefore, k_{0x} and k_{2x} have to be pure imaginary. In other words, we can find the values of k_z for guidance only in the range $k_z > k_0$ and $k_z > k_2$.

For a symmetric waveguide where regions 0 and 2 are the same, the first few modes of the waveguides are as sketched in Figure 6.3. These modes are, in general, more well-trapped

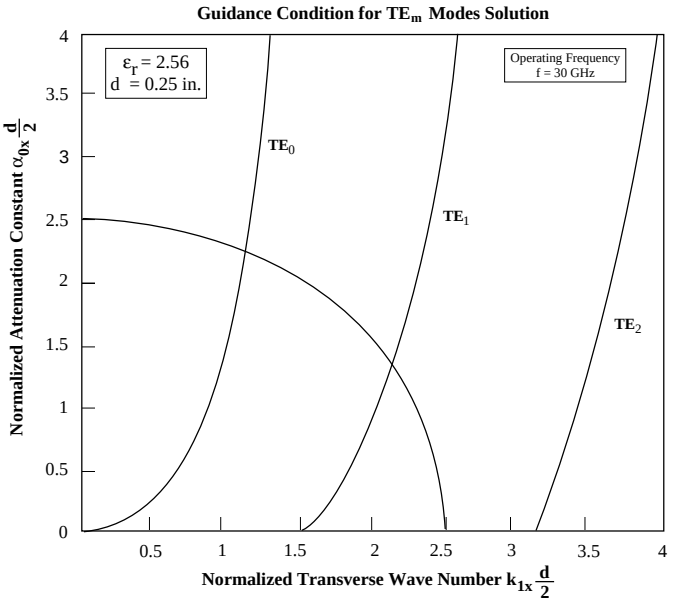


Figure 6.4: Graphical solution of the guided TM modes of a dielectric slab waveguide.

when the refractive index in region 1 is higher than those in regions 0 and 2, or when the frequency is high. We can study the guidance of a TE mode which is the dual of a TM mode in a similar fashion.

By a further manipulation of Equation (6.1.4), and using the definition of Fresnel reflection coefficients, Equation (6.1.4) can be written as

$$\alpha_{0x}\frac{d}{2} = \frac{\epsilon_0}{\epsilon_1} k_{1x}\frac{d}{2}\tan\left(\frac{k_{1x}d - m\pi}{2}\right), \tag{6.1.8}$$

where $\alpha_{0x} = \sqrt{k_z^2 - k_0^2}$, $k_{1x} = \sqrt{k_1^2 - k_z^2}$, and $m = 1, 2, 3, 4, \cdots$. The left-hand side of (6.1.8) can be expressed in terms of the $k_{1x}d$ variable, viz.,

$$\sqrt{k_z^2 - k_0^2}\frac{d}{2} = \sqrt{(k_1^2 - k_0^2)\left(\frac{d}{2}\right)^2 - \left(\frac{k_{1x}d}{2}\right)^2}, \tag{6.1.9}$$

which is the equation of a circle. Equation (6.1.8) can hence be solved graphically by plotting both sides of Equation (6.1.8) as a function of $k_{1z}d$. The TE mode case can be obtained by duality. The graphical-solution plots are shown in Figures 6.4 and 6.5.

From Figure 6.6, it is clear that the TM_0 mode has no cut-off, since continuity of the slope and field amplitude can be satisfied for all frequencies and yet the field is evanescent outside. This is not true of the higher order modes. Figure 6.7 shows the profile of the TM_1 mode for different frequencies.

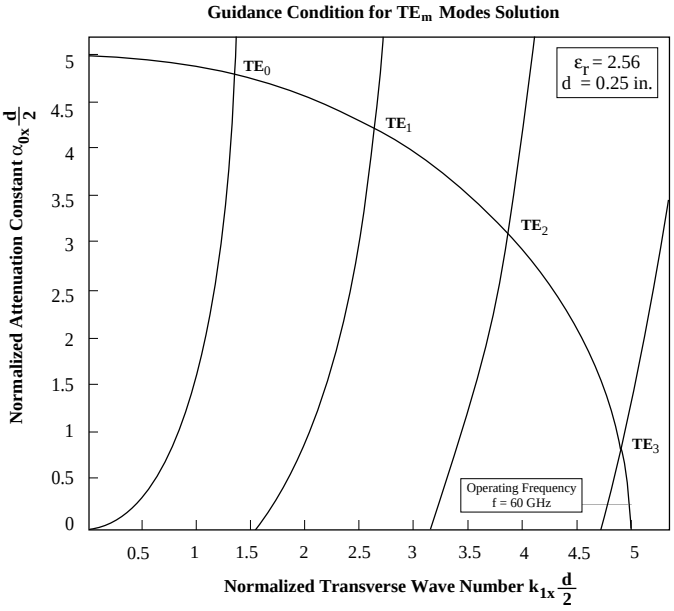


Figure 6.5: Graphical solution of the guided TE modes of a dielectric slab waveguide.

A mode in a dielectric slab waveguide has part of its energy outside the waveguide and inside the waveguide as shown in Figures 6.6 and 6.7. Hence, the group velocity of the mode is between that of the slab region and the outer region. For the TM_0 mode, when the frequency is very low, the mode is weakly evanescent outside the dielectric slab and the group velocity of a mode is closer to that of the outer region because most of the energy of the mode is outside the waveguide. When the frequency is high, the mode is strongly evanescent outside and most of the energy of the mode is trapped inside the slab. Hence, the group velocity of the mode is close to that of the dielectric slab. Therefore, the dispersion curve of the TM_0 mode is as shown in Figure 6.8.

To elaborate on the TM_1 mode, for high frequencies, the mode is well-trapped inside the dielectric waveguide. When the frequency is low such that $\lambda/4 = d/2$, the field ceases to be evanescent outside the dielectric slab. In this case, the field is a constant outside, and this is the mode at the cut-off frequency precisely. When $\lambda/4 > d/2$, the mode leaks energy to outside the slab, it becomes a leaky mode and is not guided at all. Figure 6.8 shows the dispersion curve of the TM_1 mode in a dielectric slab waveguide. Close to cut-off, the group velocity of the mode should be close to that of the outside region, and vice-versa for high frequencies. Figure 6.7 allows us to determine the cut-off frequency of the TM_1 mode quite readily.

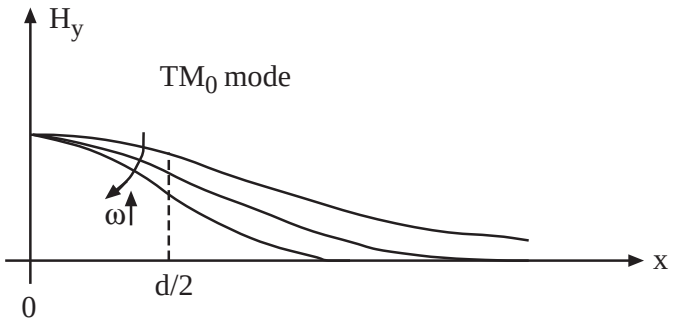


Figure 6.6: For increasing frequency, a mode is better trapped inside the dielectric slab. Only the field for $x > 0$ is sketched in the above.

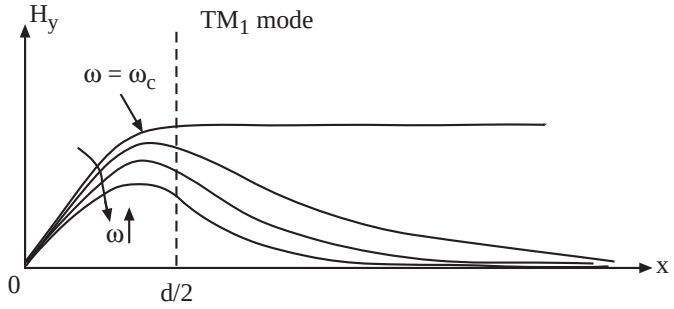


Figure 6.7: The profile of the TM_1 mode for different frequencies. When $\omega = \omega_c$, the cutoff frequency, the mode ceases to be evanescent outside the waveguide.

6.2 Circular Dielectric Waveguide

An optical fiber is a circular dielectric waveguide as shown in Figure 6.9. Usually, an optical fiber has a protective cladding as well. In the ensuing analysis, we will assume that the protective cladding is infinitely thick, i.e., letting $b \to \infty$ [11, 16, 17, 20].

In order for a mode to be guided, we require that $\epsilon_1\mu_1 > \epsilon_2\mu_2$. In other words, the light velocity in the core region has to be slower than the light velocity in the cladding. The field outside the core region is evanescent for a guided mode. Therefore, the value of b affects the guided mode little when b is large, especially if the mode is tightly bound to the core; hence, $b \to \infty$ is a good approximation.

In the optical fiber modes, except for the axisymmetric modes, the TE and TM fields are coupled to each other by the boundary conditions as in an inhomogeneously filled waveguide. The z components of the field characterize the TE and TM fields in the fiber. Moreover, they are decoupled in a homogeneous region, and are solutions to the wave equation in cylindrical

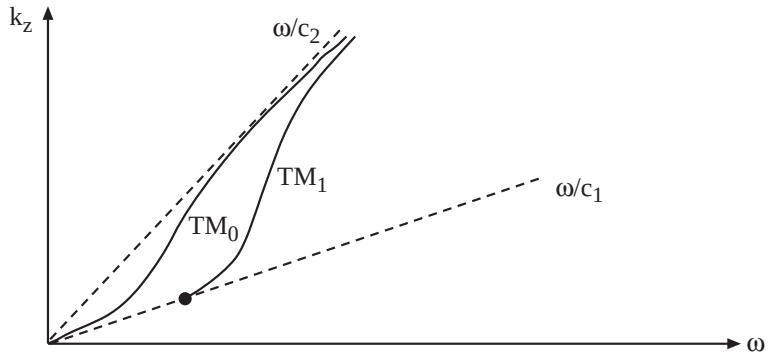


Figure 6.8: Dispersion of the modes in a dielectric slab waveguide.

coordinates for each of the homogeneous regions, i.e.,

$$\left[\frac{1}{\rho}\frac{\partial}{\partial\rho}\rho\frac{\partial}{\partial\rho}+\frac{1}{\rho^2}\frac{\partial^2}{\partial\phi^2}+\frac{\partial^2}{\partial z^2}+\omega^2\mu_j\epsilon_j\right]\begin{bmatrix}E_{jz}\\H_{jz}\end{bmatrix}=0. \tag{6.2.1}$$

The solution to (6.2.1) in region j is of the general form

$$\begin{Bmatrix}J_n(k_{j\rho}\rho)\\H_n^{(1)}(k_{j\rho}\rho)\end{Bmatrix}e^{\pm in\phi+ik_z z}, \tag{6.2.2}$$

where $k_{j\rho}^2+k_z^2=k_j^2$ and k_j is the wave number in region j. Here, $H_n^{(1)}(k_\rho\rho)$ is singular when $\rho\to 0$ while $J_n(k_\rho\rho)$ is regular. Therefore, for the $e^{in\phi}$ harmonic, the solution in the core region is

$$E_{1z}=E_1J_n(k_{1\rho}\rho)e^{ik_z z+in\phi}, \tag{6.2.3a}$$

$$H_{1z}=H_1J_n(k_{1\rho}\rho)e^{ik_z z+in\phi}. \tag{6.2.3b}$$

In the cladding region, when $b\to\infty$, we should only have outgoing waves. Therefore,

$$E_{2z}=E_2H_n^{(1)}(k_{2\rho}\rho)e^{ik_z z+in\phi}, \tag{6.2.4a}$$

$$H_{2z}=H_2H_n^{(1)}(k_{2\rho}\rho)e^{ik_z z+in\phi}. \tag{6.2.4b}$$

By the phase matching condition, k_z is the same in both regions. The choice of Hankel functions of the first kind in (6.2.4a) and (6.2.4b) stems from the fact that Hankel functions represent outgoing waves. The asymptotic expansions of the Hankel and Bessel functions are

$$H_n^{(1)}(k_\rho\rho)\sim\sqrt{\frac{2}{\pi k_\rho\rho}}e^{ik_\rho\rho-in\frac{\pi}{2}-i\frac{\pi}{4}},\qquad \rho\to\infty, \tag{6.2.5a}$$

$$J_n(k_\rho\rho)\sim\sqrt{\frac{2}{\pi k_\rho\rho}}\cos(k_\rho\rho-\frac{n\pi}{2}-\frac{\pi}{4}),\qquad \rho\to\infty. \tag{6.2.5b}$$

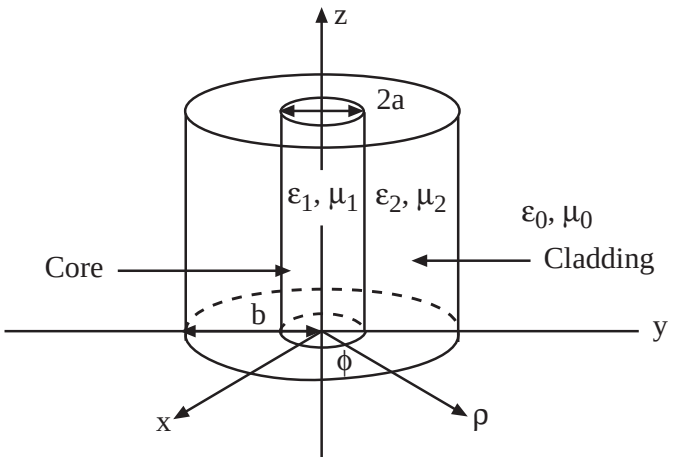


Figure 6.9: Geometry of an optical fiber—a circular dielectric waveguide.

Equation (6.2.5a) corresponds to an outgoing wave while (6.2.5b) corresponds to a standing wave. When a mode is confined in the core, $k_{2\rho} = \sqrt{k_2^2 - k_z^2} = i\alpha_2$ is always positive imaginary. Therefore, guidance is only possible if $k_z > k_2$.

At this point, we introduce the modified Bessel function $K_n(x)$ such that

$$H_n^{(1)}(i\alpha_2\rho) = \frac{2}{\pi i} e^{-i\frac{\pi n}{2}} K_n(\alpha_2\rho). \tag{6.2.6}$$

As such, Equations (6.2.4a) and (6.2.4b) can be written as

$$E_{2z} = E_2' K_n(\alpha_2\rho) e^{ik_z z + in\phi}, \tag{6.2.7a}$$

$$H_{2z} = H_2' K_n(\alpha_2\rho) e^{ik_z z + in\phi}, \tag{6.2.7b}$$

where $\alpha_2 = \sqrt{k_z^2 - k_2^2}$. When $\rho \to \infty$,

$$K_n(\alpha_2\rho) \sim \sqrt{\frac{\pi}{2\alpha_2\rho}} e^{-\alpha_2\rho}, \tag{6.2.8}$$

corresponding to an evanescent wave. A guided mode must satisfy the boundary conditions at the core-cladding interface. Therefore, we need to find E_ϕ and H_ϕ, the other tangential components of the field.

In the core region, they are

$$E_{1\phi} = \frac{1}{k_{1\rho}^2}\left[-\frac{nk_z}{\rho} E_1 J_n(k_{1\rho}\rho) - i\omega\mu_1 H_1 k_{1\rho} J_n'(k_{1\rho}\rho)\right] e^{in\phi + ik_z z}, \tag{6.2.9a}$$

$$H_{1\phi} = \frac{1}{k_{1\rho}^2}\left[-\frac{nk_z}{\rho} H_1 J_n(k_{1\rho}\rho) + i\omega\epsilon_1 E_1 k_{1\rho} J_n'(k_{1\rho}\rho)\right] e^{in\phi + ik_z z}. \tag{6.2.9b}$$

In the cladding region, they are

$$E_{2\phi} = -\frac{1}{\alpha_2^2}\left[-\frac{nk_z}{\rho}E_2' K_n(\alpha_2\rho) - i\omega\mu_2 H_2'\alpha_2 K_n'(\alpha_2\rho)\right] e^{in\phi+ik_z z}, \tag{6.2.10a}$$

$$H_{2\phi} = -\frac{1}{\alpha_2^2}\left[-\frac{nk_z}{\rho}H_2' K_n(\alpha_2\rho) + i\omega\epsilon_2 E_2'\alpha_2 K_n'(\alpha_2\rho)\right] e^{in\phi+ik_z z}. \tag{6.2.10b}$$

The continuity of the tangential components of the fields from (6.2.3a), (6.2.3b), (6.2.7a), (6.2.7b), (6.2.9a), (6.2.9b), and (6.2.10a), (6.2.10b) implies that

$$E_1 J_n(k_{1\rho}a) = E_2' K_n(\alpha_2 a), \tag{6.2.11a}$$

$$H_1 J_n(k_{1\rho}a) = H_2' K_n(\alpha_2 a), \tag{6.2.11b}$$

$$\begin{aligned}&\frac{1}{k_{1\rho}^2}\left[-\frac{nk_z}{a}E_1 J_n(k_{1\rho}a) - i\omega\mu_1 H_1 k_{1\rho} J_n'(k_{1\rho}a)\right]\\ &\quad = -\frac{1}{\alpha_2^2}\left[-\frac{nk_z}{a}E_2' K_n(\alpha_2 a) - i\omega\mu_2 H_2'\alpha_2 K_n'(\alpha_2 a)\right],\end{aligned} \tag{6.2.11c}$$

$$\begin{aligned}&\frac{1}{k_{1\rho}^2}\left[-\frac{nk_z}{a}H_1 J_n(k_{1\rho}a) + i\omega\epsilon_1 E_1 k_{1\rho} J_n'(k_{1\rho}a)\right]\\ &\quad = -\frac{1}{\alpha_2^2}\left[-\frac{nk_z}{a}H_2' K_n(\alpha_2 a) + i\omega\epsilon_2 E_2'\alpha_2 K_n'(\alpha_2 a)\right].\end{aligned} \tag{6.2.11d}$$

Substituting (6.2.11a) and (6.2.11b) into (6.2.11c) and (6.2.11d) to eliminate the unknowns E_1 and H_1, there are only two remaining unknowns E_2' and H_2'. We can form a 2×2 matrix equation with no driving term. For non-trivial solution, we set the determinant of the resultant matrix to zero to arrive at

$$\begin{aligned}&k_z^2 n^2\left(\frac{1}{k_{1\rho}^2 a^2} + \frac{1}{\alpha_2^2 a^2}\right)^2\\ &\quad = \omega^2\left[\frac{\mu_1 J_n'(k_{1\rho}a)}{k_{1\rho}aJ_n(k_{1\rho}a)} + \frac{\mu_2 K_n'(\alpha_2 a)}{\alpha_2 a K_n(\alpha_2 a)}\right]\\ &\qquad \left[\frac{\epsilon_1 J_n'(k_{1\rho}a)}{k_{1\rho}aJ_n(k_{1\rho}a)} + \frac{\epsilon_2 K_n'(\alpha_2 a)}{\alpha_2 a K_n(\alpha_2 a)}\right].\end{aligned} \tag{6.2.12}$$

Since $k_{1\rho} = \sqrt{k_1^2 - k_z^2}$, $\alpha_2 = \sqrt{k_z^2 - k_2^2}$, we can solve the above transcendental equation for k_z. Once the values of k_z that satisfies (6.2.12) are found, we can find the ratios of E_1/H_1 from (6.2.11a)–(6.2.11d). In particular,

$$\frac{E_1}{H_1} = \frac{nk_z}{i\omega}\left(\frac{1}{k_{1\rho}^2 a^2} + \frac{1}{\alpha_2^2 a^2}\right)\left(\frac{\epsilon_1 J_n'(k_{1\rho}a)}{k_{1\rho}aJ_n(k_{1\rho}a)} + \frac{\epsilon_2 K_n'(\alpha_2 a)}{\alpha_2 a K_n(\alpha_2 a)}\right)^{-1}. \tag{6.2.13}$$

This is the ratio of the TM wave amplitude to the TE wave amplitude inside the core.

In Equation (6.2.12), $\frac{K_n'(x)}{xK_n(x)}$ is not rapidly oscillating while $\frac{J_n'(x)}{xJ_n(x)}$ is rapidly oscillating. Moreover, it is a quadratic equation in terms of $\frac{J_n'(x)}{xJ_n(x)}$. We can solve Equation (6.2.12) for $\frac{J_n'(k_{1\rho}a)}{k_{1\rho}aJ_n(k_{1\rho}a)}$ giving

$$\begin{aligned}\frac{J_n'(k_{1\rho}a)}{k_{1\rho}aJ_n(k_{1\rho}a)} =& -\frac{1}{2}\left(\frac{\mu_2}{\mu_1}+\frac{\epsilon_2}{\epsilon_1}\right)\frac{K_n'(\alpha_2 a)}{\alpha_2 aK_n(\alpha_2 a)} \\ &\pm\left[\frac{1}{4}\left(\frac{\mu_2}{\mu_1}-\frac{\epsilon_2}{\epsilon_1}\right)^2\left(\frac{K_n'(\alpha_2 a)}{\alpha_2 aK_n(\alpha_2 a)}\right)^2\right. \\ &\left.+\frac{n^2k_z^2}{k_1^2}\left(\frac{1}{k_{1\rho}^2a^2}+\frac{1}{\alpha_2^2a^2}\right)^2\right]^{\frac{1}{2}}.\end{aligned} \tag{6.2.14}$$

The plus and minus signs give rise to two classes of solutions. We next make use of the recurrence relationship of Bessel functions

$$J_n'(x) = -J_{n+1}(x) + \frac{n}{x}J_n(x), \tag{6.2.15a}$$

$$J_n'(x) = J_{n-1}(x) - \frac{n}{x}J_n(x), \tag{6.2.15b}$$

to get

$$\begin{aligned}\frac{J_{n+1}(k_{1\rho}a)}{k_{1\rho}aJ_n(k_{1\rho}a)} =& \frac{1}{2}\left(\frac{\mu_2}{\mu_1}+\frac{\epsilon_2}{\epsilon_1}\right)\frac{K_n'(\alpha_2 a)}{\alpha_2 aK_n(\alpha_2 a)} \\ &+\left(\frac{n}{(k_{1\rho}a)^2}-R\right), \qquad \text{EH modes,}\end{aligned} \tag{6.2.16a}$$

$$\begin{aligned}\frac{J_{n-1}(k_{1\rho}a)}{k_{1\rho}aJ_n(k_{1\rho}a)} =& -\frac{1}{2}\left(\frac{\mu_2}{\mu_1}+\frac{\epsilon_2}{\epsilon_1}\right)\frac{K_n'(\alpha_2 a)}{\alpha_2 aK_n(\alpha_2 a)} \\ &+\left(\frac{n}{(k_{1\rho}a)^2}-R\right), \qquad \text{HE modes,}\end{aligned} \tag{6.2.16b}$$

where

$$R = \left[\frac{1}{4}\left(\frac{\mu_2}{\mu_1}-\frac{\epsilon_2}{\epsilon_1}\right)^2\left(\frac{K_n'(\alpha_2 a)}{\alpha_2 aK_n(\alpha_2 a)}\right)^2+\frac{n^2k_z^2}{k_1^2}\left(\frac{1}{k_{1\rho}^2a^2}+\frac{1}{\alpha_2^2a^2}\right)^2\right]^{\frac{1}{2}}. \tag{6.2.16c}$$

For the $n = 0$ case, or the axisymmetric case, the TE and TM fields are decoupled, and the TE and TM modes are the dual of each other. The guidance conditions are

$$\frac{J_1(k_{1\rho}a)}{k_{1\rho}aJ_0(k_{1\rho}a)} = -\frac{\epsilon_2}{\epsilon_1}\frac{K_1(\alpha_2 a)}{\alpha_2 aK_0(\alpha_2 a)}, \qquad \text{TM,} \tag{6.2.17a}$$

$$\frac{J_1(k_{1\rho}a)}{k_{1\rho}aJ_0(k_{1\rho}a)} = -\frac{\mu_2}{\mu_1}\frac{K_1(\alpha_2 a)}{\alpha_2 aK_0(\alpha_2 a)}, \qquad \text{TE.} \tag{6.2.17b}$$

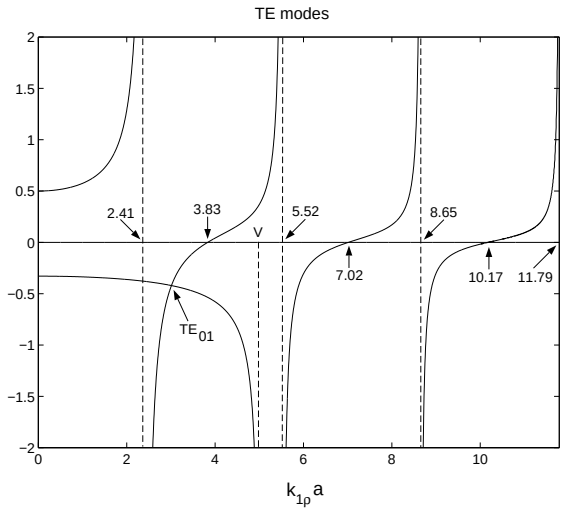


Figure 6.10: Graphical solution of the axially symmetric TE modes.

In the above, we have used $K_0'(x) = -K_1(x)$. Note that TE and TM waves are decoupled from (6.2.11a)–(6.2.11c), since the TE and TM waves can satisfy the boundary conditions separately. By writing

$$(\alpha_2 a)^2 = (k_1 a)^2 - (k_2 a)^2 - (k_{1\rho} a)^2 = V^2 - (k_{1\rho} a)^2, \tag{6.2.18}$$

(6.2.17a) can be solved in terms of $k_{1\rho}a$. To get a feeling for the behavior of the solution, it can be solved graphically. Since $\alpha_2 a > 0$ for a guided mode, from (6.2.18), note that we need only to consider the case where

$$0 < k_{1\rho}^2 a^2 < (k_1^2 - k_2^2)a^2 = V^2. \tag{6.2.19}$$

The right-hand side of (6.2.17a) is always negative. For the TE case, it has a value of $-\mu_2 K_1(V)/\mu_1 V K_0(V)$ at $k_{1\rho}a = 0$, and it has a value of $-\infty$, when $k_{1\rho}a \to V$. More precisely,

$$-\frac{\mu_2 K_1(\alpha_2 a)}{\mu_1 \alpha_2 a K_0(\alpha_2 a)} \sim \frac{2\mu_2}{\mu_1 (V^2 - k_{1\rho}^2 a^2) \ln(V^2 - k_{1\rho}^2 a^2)}, \qquad k_{1\rho}a \to V. \tag{6.2.20}$$

The left-hand side of (6.2.17a) starts from 1/2 at $k_{1\rho}a = 0$ and goes to infinity at the zeros of $J_0(k_{1\rho}a)$, and goes to zero at the zeros of $J_1(k_{1\rho})$. A sketch of the left-hand side and the right-hand side of (6.2.17a) is shown in Figure 6.10. The number of guided modes depends on $V = \sqrt{(k_1^2 - k_2^2)}a$, the normalized frequency. V can be increased by increasing the contrast between ϵ_1 and ϵ_2, by raising the frequency, or by increasing a. For $V < 2.405$, there could be no possible guided modes. Hence, all axisymmetric modes have a finite cut-off frequency.

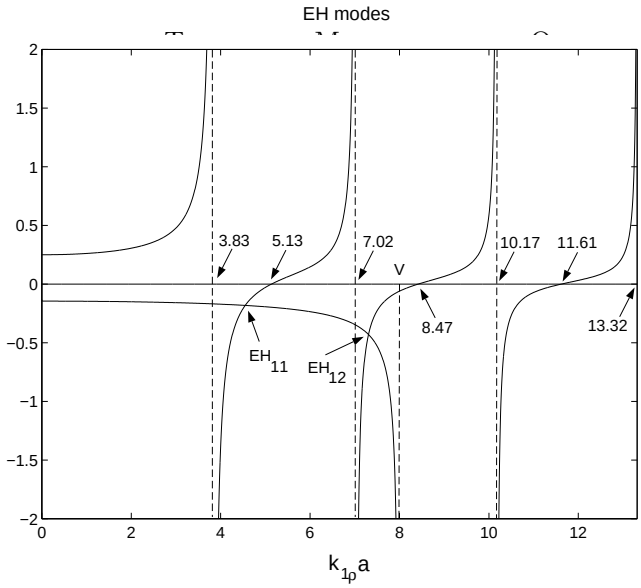


Figure 6.11: Graphical solution of the EH modes.

When $n = 1$, Equations (6.2.16a), (6.2.16b), become

$$\frac{J_2(k_{1\rho}a)}{k_{1\rho}aJ_1(k_{1\rho}a)} = \frac{1}{2}\left(\frac{\mu_2}{\mu_1} + \frac{\epsilon_2}{\epsilon_1}\right)\frac{K_1'(\alpha_2 a)}{\alpha_2 a K_1(\alpha_2 a)} + \left(\frac{1}{(k_{1\rho}a)^2} - R\right), \quad \text{EH} \tag{6.2.21a}$$

$$\frac{J_0(k_{1\rho}a)}{k_{1\rho}aJ_1(k_{1\rho}a)} = -\frac{1}{2}\left(\frac{\mu_2}{\mu_1} + \frac{\epsilon_2}{\epsilon_1}\right)\frac{K_1'(\alpha_2 a)}{\alpha_2 a K_1(\alpha_2 a)} + \left(\frac{1}{(k_{1\rho}a)^2} - R\right), \quad \text{HE}. \tag{6.2.21b}$$

The left and the right-hand side of (6.2.21a) is shown in Figures 6.11 and 6.12. The right-hand side of (6.2.21a) goes to infinity at $k_{1\rho}a \to V$, while the left-hand side goes to zero and infinity as before. Note that all the EH modes have finite cut-off frequencies while the HE_{11} mode does not have a cut-off frequency. Therefore, the HE_{11} mode is the dominant mode in an optical fiber. If (i) the optical fiber is small enough, or (ii) the frequency is low enough, or (iii) when the contrast is very low, it is the only mode propagating for the single mode operation of the optical fiber.

Usually, in a waveguide, an EH notation is used to denote a mode where E_z dominates over H_z, or the TM component dominates over the TE component. However, due to a quirk in the history of optical fibers, the EH notation is used to denote a mode whose TE component dominates over its TM component, and vice versa for the HE notation.

A parameter of interest is the axial wave number k_z as a function of the normalized frequency. In Figure 6.13, we plot k_z/k_0 versus V. At very low frequencies, or near the cut-off of a mode, k_z/k_0 approaches the refractive index n_2 of the cladding. This is because the mode is not well confined, and most of the energy of the mode is in medium 2. Hence, the mode phase velocity is close to that of medium 2. At higher normalized frequency, the mode is well confined, and it propagates in medium 1. Hence, k_z/k_0 approaches n_1 which is the refractive index of medium 1.

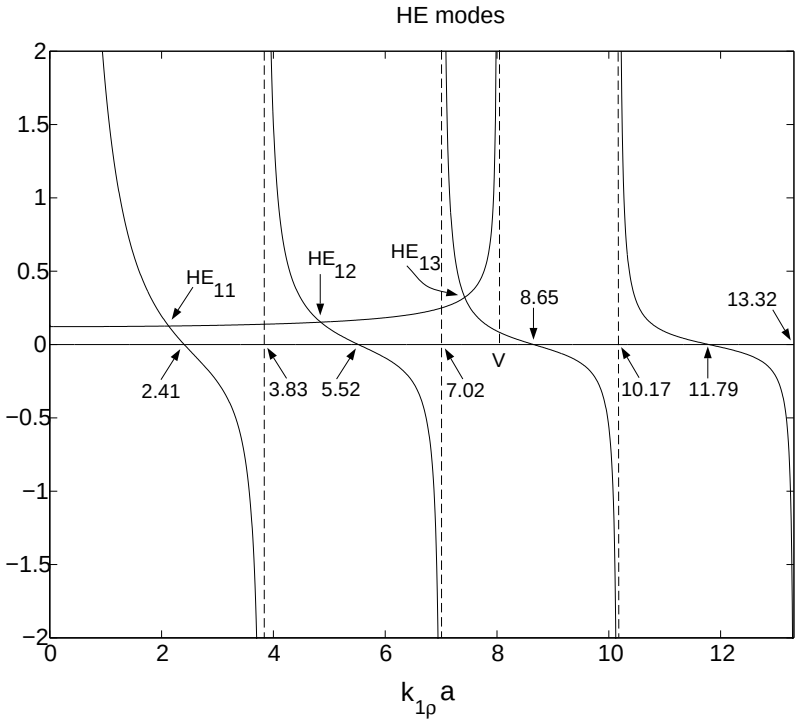


Figure 6.12: Graphical solution of the HE modes.

From Equation (6.2.12), it is clear that when we find a solution, a dual solution also exists. If $\mu_1/\mu_2 = \epsilon_1/\epsilon_2$, then a solution and its dual are degenerate. However, usually, $\mu_1/\mu_2 = 1$; therefore, a solution and its dual are not degenerate. However, when $\epsilon_1/\epsilon_2 \to 1$ as in the case of a weak-contrast optical fiber, a solution and its dual are near degenerate. The dispersion curve for the weak contrast case is shown in Figure 6.14.

6.3 Weak-Contrast Optical Fiber

When the contrast between the cladding and the core of the fiber is small, the analysis of the guided mode inside a fiber can be greatly simplified. In this case, polarization effect at the interface of the fiber is unimportant, and scalar wave theory can be applied [11, 14, 18].

The vector wave equations governing electromagnetic fields in an inhomogeneous waveguide are

$$\nabla \times \mu_r^{-1} \nabla \times \mathbf{E} - \omega^2 \mu_0 \epsilon_0 \epsilon_r \mathbf{E} = 0, \tag{6.3.1a}$$

$$\nabla \times \epsilon_r^{-1} \nabla \times \mathbf{H} - \omega^2 \mu_0 \epsilon_0 \mu_r \mathbf{H} = 0, \tag{6.3.1b}$$

where $\mu_r = \mu(\mathbf{r})/\mu_0 = f_1(\mathbf{r})$, and $\epsilon_r = \epsilon(\mathbf{r})/\epsilon_0 = f_2(\mathbf{r})$. If we find a solution to (6.3.1a), the solution to the dual problem is obtained by letting $\mathbf{E} \to -\mathbf{H}$, $\mathbf{H} \to \mathbf{E}$, $\mu_r \to \epsilon_r$, $\epsilon_r \to \mu_r$.

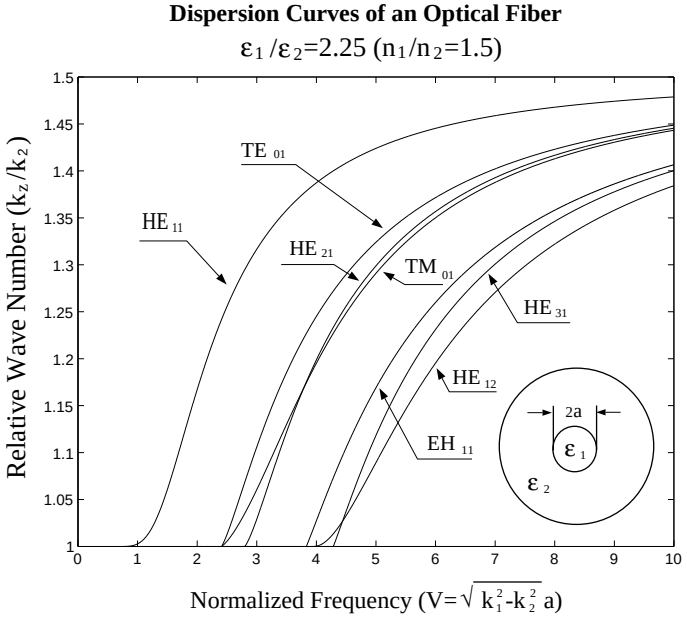


Figure 6.13: Dispersion curves for various modes of an optical fiber for the high contrast case.

However, if $f_1(\mathbf{r}) \neq f_2(\mathbf{r})$, the dual problem corresponds to a different waveguide with $\mu_r = f_2(\mathbf{r})$ and $\epsilon_r = f_1(\mathbf{r})$. In order for a dual problem to be itself, we require that $f_1(\mathbf{r}) = f_2(\mathbf{r})$. If this is the case, then a mode and its dual are degenerate. Therefore, we can associate every mode in a waveguide with a dual mode. However, $f_1(\mathbf{r}) = 1$ usually, while $f_2(\mathbf{r}) \neq 1$; therefore, a mode is not degenerate with its dual.

However, in the case of a weak-contrast optical fiber, $\epsilon_r \to 1$; hence, a mode is near degenerate with its dual. Furthermore, we can show that the vector nature of the wave is unimportant. If $\mu_r = 1$, we can rewrite (6.3.1a) and (6.3.1b) as

$$\nabla \times \nabla \times \mathbf{E} - \omega^2 \mu_0 \epsilon_0 \epsilon_r \mathbf{E} = 0, \tag{6.3.2a}$$

$$\nabla \times \nabla \times \mathbf{H} - (\nabla \ln \epsilon_r) \times \nabla \times \mathbf{H} - \omega^2 \mu_0 \epsilon_0 \epsilon_r \mathbf{H} = 0. \tag{6.3.2b}$$

In the above $\nabla \ln \epsilon_r = \frac{\nabla \epsilon_r}{\epsilon_r}$ is the effect of the polarization charges at the dielectric interface. If the transverse and longitudinal components of (6.3.2a) and (6.3.2b) are extracted, the following equations ensue:

$$\nabla^2 \mathbf{E}_s + \nabla_s [\nabla_s \ln \epsilon_r \cdot \mathbf{E}_s] + k^2 \mathbf{E}_s = 0, \tag{6.3.3a}$$

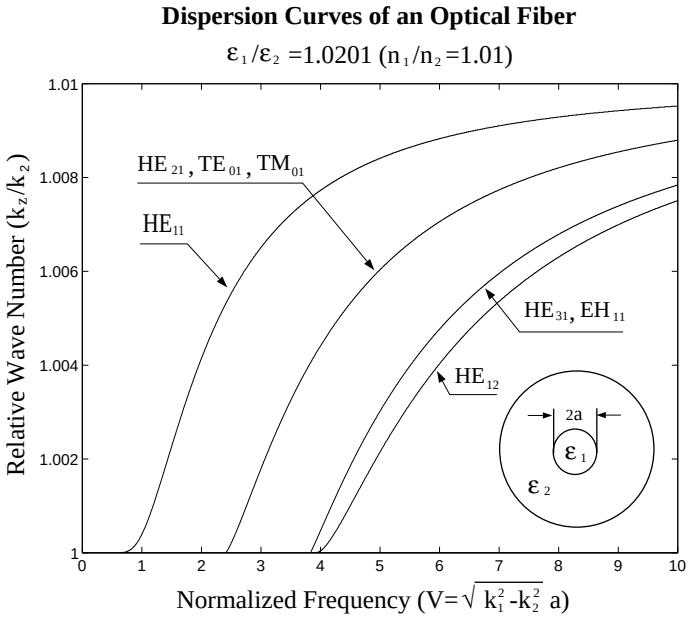


Figure 6.14: Dispersion curves for various modes of an optical fiber for the low contrast case.

$$\nabla^2 E_z + ik_z(\nabla_s \ln \epsilon_r) \cdot \mathbf{E}_s + k^2 E_z = 0, \tag{6.3.3b}$$

and

$$\nabla^2 \mathbf{H}_s + (\nabla_s \ln \epsilon_r) \times (\nabla_s \times \mathbf{H}_s) + k^2 \mathbf{H}_s = 0, \tag{6.3.4a}$$

$$\nabla^2 \mathbf{H}_z + (\nabla_s \ln \epsilon_r) \times (\nabla_s \times \mathbf{H}_z) + (\nabla_s \ln \epsilon_r) \times (\hat{z} i k_z \times \mathbf{H}_s) + k^2 \mathbf{H}_z = 0. \tag{6.3.4b}$$

In the limit when $\epsilon_r \to 1$, the polarization charge terms in (6.3.3a), (6.3.3b), (6.3.4a), and (6.3.4b) can be ignored with respect to the other terms, yielding

$$\nabla^2 \mathbf{E}_s + k^2 \mathbf{E}_s = 0, \tag{6.3.5a}$$

$$\nabla^2 E_z + k^2 E_z = 0, \tag{6.3.5b}$$

$$\nabla^2 \mathbf{H}_s + k^2 \mathbf{H}_s = 0. \tag{6.3.5c}$$

$$\nabla^2 H_z + k^2 H_z = 0. \tag{6.3.5d}$$

In other words, the wave guidance problem by a fiber of weak contrast reduces to a scalar problem. Also, from the above equations, it is apparent that $\nabla^2 \sim -k^2$, and hence, when $\omega \to \infty$, the polarization terms in (6.3.3a) and (6.3.4a) are much smaller than the first and the last terms. From (6.3.3b) and (6.3.4b), it is seen that the z components of the fields are induced by their transverse components; hence they are much smaller than the transverse components. Moreover, when the contrast $\epsilon_r = 1$, the guided mode in the fiber becomes a

TEM mode with $k = k_z$, and $E_z = H_z = 0$. This represents the leading order solution when $\epsilon_r = 1$.

As mentioned above, when $\epsilon_r > 1$, we see from Equation (6.3.3b) that a nonzero E_z is induced by the presence of $\mathbf{E}_s$. Moreover, by comparing terms, and assuming high frequency, $E_z \sim |\nabla \ln \epsilon_r \cdot \mathbf{E}_s|/k \approx \ln \epsilon_r |\mathbf{E}_s|/(ka) \ll |\mathbf{E}_s|$, because $\nabla \ln \epsilon_r \approx \ln \epsilon_r / a$. Therefore, when $\omega \to \infty$, $|E_z| \ll |\mathbf{E}_s|$ (see Problem 6-8.). By the same token, $|H_z| \ll |\mathbf{H}_s|$ when $\omega \to \infty$. Therefore, when the contrast is very weak, and the frequency is very high, the mode is quasi-TEM, namely, $\nabla_s \cdot \epsilon_r \mathbf{E}_s = -ik_z \epsilon_r E_z \approx 0$ and $\nabla_s \cdot \mathbf{H}_s = -ik_z H_z \approx 0$, since $k_z E_z \ll |\mathbf{E}_s|/a$ and $k_z H_z \ll |\mathbf{H}_s|/a$. From the aforementioned analysis, it is clear that (6.3.5a) and (6.3.5c) are the equations to solve when $\omega \to \infty$ and the contrast small.

Equations (6.3.5a) and (6.3.5c) are equivalent to

$$(\nabla^2 + k^2)\phi = 0 \tag{6.3.6}$$

where ϕ is either E_x, E_y, H_x, or H_y. For example, we can let

$$\phi = \begin{cases} AJ_n(k_{1\rho}\rho)e^{in\phi + ik_z z}, & \rho < a, \\ BK_n(\alpha_2 \rho)e^{in\phi + ik_z z}, & \rho > a, \end{cases} \tag{6.3.7}$$

The boundary conditions for ϕ at the interface where $k^2 = \omega^2 \mu_0 \epsilon_0 \epsilon_r$ displays a step discontinuity is

$$\phi_1 = \phi_2, \tag{6.3.8a}$$

$$\hat{n} \cdot \nabla \phi_1 = \hat{n} \cdot \nabla \phi_2. \tag{6.3.8b}$$

These boundary conditions are derivable from Equation (6.3.6) alone.

Imposing the above boundary conditions at $\rho = a$ for the weak contrast optical fiber, whose field is given by (6.3.7), we have

$$AJ_n(k_{1\rho}a) = BK_n(\alpha_2 a), \tag{6.3.9a}$$

$$Ak_{1\rho}J_n'(k_{1\rho}a) = B\alpha_2 K_n'(\alpha_2 a). \tag{6.3.9b}$$

The above yields

$$\frac{k_{1\rho}J_n'(k_{1\rho}a)}{J_n(k_{1\rho}a)} = \frac{\alpha_2 K_n'(\alpha_2 a)}{K_n(\alpha_2 a)}. \tag{6.3.10}$$

Using the recurrence relationship that $J_n'(x) = -J_{n+1}(x) + \frac{n}{x}J_n(x)$, and that $K_n'(x) = -K_{n+1}(x) + \frac{n}{x}K_n(x)$, we can transform the above to

$$\frac{k_{1\rho}J_{n+1}(k_{1\rho}a)}{J_n(k_{1\rho}a)} = \frac{\alpha_2 K_{n+1}(\alpha_2 a)}{K_n(\alpha_2 a)}. \tag{6.3.11}$$

Similarly, using the recurrence relationship that $J_n'(x) = J_{n-1}(x) - \frac{n}{x}J_n(x)$, and $K_n'(x) = -K_{n-1}(x) - \frac{n}{x}K_n(x)$, we have

$$\frac{k_{1\rho}J_{n-1}(k_{1\rho}a)}{J_n(k_{1\rho}a)} = -\frac{\alpha_2 K_{n-1}(\alpha_2 a)}{K_n(\alpha_2 a)}. \tag{6.3.12}$$

Comparing with Equations (6.2.16a) and (6.2.16b), we note that now there are half as many solutions as before. This is because when $\epsilon_r \to 1$, $\mathrm{HE}_{n+1,m}$ and $\mathrm{EH}_{n-1,m}$ modes are degenerate. We can solve (6.3.11) graphically as before. The modes thus found are designated the LP_{nm} mode. The lowest order mode is the LP_{01} mode which is the degenerate case of the HE_{11} mode. LP here stands for "linearly polarized." The LP modes are fragile, as they are actually the linear superpositions of the degenerate HE and EH modes. The moment the contrast increases, this degeneracy splits into two modes again.[1] Figure 6.15 shows the dispersion curves of the LP_{nm} modes. In the figure, $b = \left(\frac{k_z}{k_2} - n_2\right)/(n_1 - n_2)$, and V is the normalized frequency.

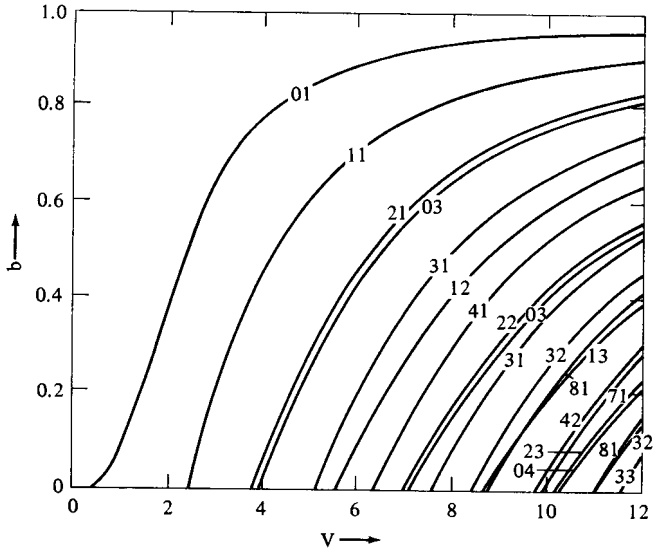


Figure 6.15: Dispersion curves for a weak contrast optical fiber [18].

These modes are termed "weakly guided" modes, but it is a misnomer. When V is large, a mode can still be tightly confined to the waveguide as demonstrated by the phase velocity approaching that of a core. Also, E_x and E_y are not independent of each other since $\nabla_s \cdot \epsilon_r \mathbf{E}_s \approx 0$. The same statement applies to H_x and H_y. Hence, these modes are not linearly polarized.

6.4 Perturbation Formula for Dielectric Waveguides

If we know the solution to a waveguide geometry, and wish to change the phase velocity of the waveguiding mode, dielectric material can be added to achieve the purpose. A perturbative

[1]The author thanks Erhan Kudeki for this discussion.

approach can be used to analyze such problem [15, 21, 22].

For the unperturbed problem, the electromagnetic field satisfies

$$\nabla \times \mathbf{E}_0 = i\omega\mu_0 \mathbf{H}_0, \tag{6.4.1}$$

$$\nabla \times \mathbf{H}_0 = -i\omega\epsilon_0 \mathbf{E}_0. \tag{6.4.2}$$

Now, let us change the permittivity and the permeability of the waveguide. The new electromagnetic field satisfies

$$\nabla \times \mathbf{E} = i\omega\mu \mathbf{H}, \tag{6.4.3}$$

$$\nabla \times \mathbf{H} = -i\omega\epsilon \mathbf{E}, \tag{6.4.4}$$

where μ and ϵ are new. Taking the divergence of $\mathbf{E}_0^* \times \mathbf{H}$ and $\mathbf{E} \times \mathbf{H}_0^*$, we have

$$\nabla \cdot (\mathbf{E}_0^* \times \mathbf{H}) = -i\omega\mu_0 \mathbf{H} \cdot \mathbf{H}_0^* + i\omega\epsilon \mathbf{E}_0^* \cdot \mathbf{E}, \tag{6.4.5a}$$

$$\nabla \cdot (\mathbf{E} \times \mathbf{H}_0^*) = i\omega\mu \mathbf{H} \cdot \mathbf{H}_0^* - i\omega\epsilon_0 \mathbf{E}_0^* \cdot \mathbf{E}. \tag{6.4.5b}$$

Adding the above equations, we have

$$\nabla \cdot (\mathbf{E}_0^* \times \mathbf{H} + \mathbf{E} \times \mathbf{H}_0^*) = i\omega\delta\mu \mathbf{H} \cdot \mathbf{H}_0^* + i\omega\delta\epsilon \mathbf{E}_0^* \cdot \mathbf{E}, \tag{6.4.6}$$

where $\delta\mu = \mu - \mu_0$, and $\delta\epsilon = \epsilon - \epsilon_0$. Since the unperturbed field has $e^{ik_{0z}z}$ dependence, while the perturbed field has $e^{ik_z z}$ dependence, we have

$$\begin{aligned} \nabla \cdot (\mathbf{E}_0^* \times \mathbf{H} + \mathbf{E} \times \mathbf{H}_0^*) = {} & \nabla_s \cdot (\mathbf{E}_0^* \times \mathbf{H} + \mathbf{E} \times \mathbf{H}_0^*) \\ & + \hat{z} i(k_z - k_{0z}) \cdot (\mathbf{E}_0^* \times \mathbf{H} + \mathbf{E} \times \mathbf{H}_0^*). \end{aligned} \tag{6.4.7a}$$

Integrating Equation (6.4.6) over the cross-section of the waveguide, making use of (6.4.7a), we have

$$k_z - k_{0z} = \omega \frac{\iint\limits_S dS[\delta\mu \mathbf{H} \cdot \mathbf{H}_0^* + \delta\epsilon \mathbf{E} \cdot \mathbf{E}_0^*]}{\iint\limits_S dS[\mathbf{E}_0^* \times \mathbf{H} + \mathbf{E} \times \mathbf{H}_0^*]}. \tag{6.4.7b}$$

Equation (6.4.7a) is exact at this point, but is not very useful because $\mathbf{E}$ and $\mathbf{H}$ are unknowns on the right-hand side. However, when the perturbation is small, we can approximate $\mathbf{E} \simeq \mathbf{E}_0$ and $\mathbf{H} \simeq \mathbf{H}_0$, and (6.4.7a) becomes

$$k_z - k_{0z} \simeq \omega \frac{\iint\limits_S dS[\delta\mu |\mathbf{H}_0|^2 + \delta\epsilon |\mathbf{E}_0|^2]}{2\Re e \iint\limits_S dS[\mathbf{E}_0 \times \mathbf{H}_0^*]}. \tag{6.4.8}$$

The time average power flow in the unperturbed waveguide is given by

$$\langle P_f \rangle = \frac{1}{2} \Re e \iint dS[\mathbf{E}_0 \times \mathbf{H}_0^*]. \tag{6.4.9}$$

The time average energy stored per unit length is

$$\langle W \rangle = \langle W_e \rangle + \langle W_m \rangle = 2\langle W_e \rangle = \frac{1}{2} \iint\limits_S dS\epsilon |\mathbf{E}|^2. \tag{6.4.10}$$

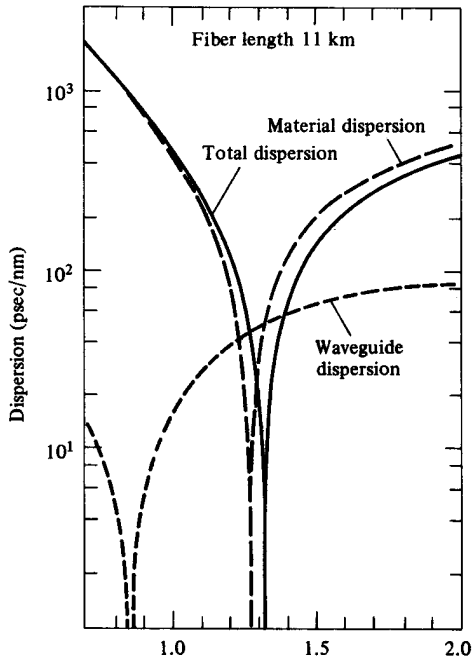


Figure 6.16: Dispersion in a single-mode graded index optical fiber. (From W.A. Gambling et al, "Zero total dispersion in graded-index single-mode fibers," *Electron. Lett.*, v. 15, p. 474, 1979.).

In the case of a step-index optical fiber, where a perturbation is in the dielectric constant, which is uniformly $\delta\epsilon_1$ in the core region and is uniformly $\delta\epsilon_2$ in the cladding region, we can rewrite (6.4.8) as

$$\delta k_z \simeq \omega \frac{\iint\limits_{S_1} dS\delta\epsilon_1|\mathbf{E}_1|^2 + \iint\limits_{S_2} dS\delta\epsilon_2|\mathbf{E}_2|^2}{4\langle P_f\rangle}, \tag{6.4.11}$$

where $\mathbf{E}_1$ is the original field in the core region, and $\mathbf{E}_2$ is the original field in the cladding region. Since $\delta\epsilon_i$ is constant in region i, we can rewrite (6.4.11) as

$$\delta k_z \simeq \omega \frac{\delta\epsilon_1\epsilon_1^{-1}\langle W_{e1}\rangle + \delta\epsilon_2\epsilon_2^{-1}\langle W_{e2}\rangle}{\langle P_f\rangle}, \tag{6.4.12}$$

where $\langle W_{e1}\rangle$ is the time average energy stored in the electric field in region 1 while $\langle W_{e2}\rangle$ is that for region 2. Since the total power flow in each region is $2v_g\langle W_{ei}\rangle$, we can rewrite (6.4.12) as

$$\delta k_z \simeq \frac{\omega}{2v_g}[\delta\epsilon_1\epsilon_1^{-1}\Gamma_1 + \delta\epsilon_2\epsilon_2^{-1}\Gamma_2], \tag{6.4.13}$$

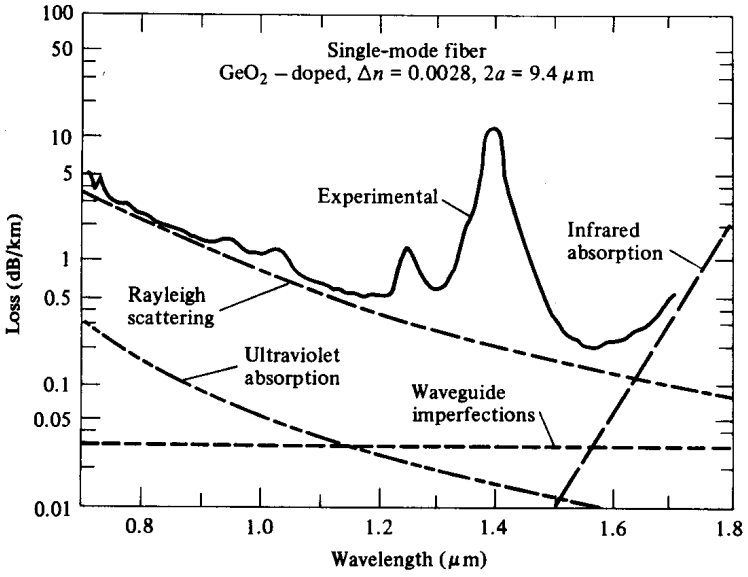


Figure 6.17: Different loss mechanisms in a germanosilicate single-mode optical fiber. Updated figure can be found in Saleh and Teich [62].

where Γ_i is the fraction of power flow in region i.

For a weak contrast optical fiber, we can further approximate the above with $\epsilon_1 \simeq \epsilon_2$, $v_g \simeq c_1 \simeq c_2$,

$$\delta k_z \simeq \frac{k_1}{2\epsilon_1}[\delta\epsilon_1\Gamma_1 + \delta\epsilon_2\Gamma_2]. \tag{6.4.14}$$

Defining an effective refractive index $n = k_z/k_0$, then $\delta n = \delta k_z/k_0$, and the above can be rewritten as

$$\delta n \simeq \frac{1}{2n_1}[\delta n_1^2\Gamma_1 + \delta n_2^2\Gamma_2], \tag{6.4.15}$$

or that

$$\delta n \simeq [\delta n_1\Gamma_1 + \delta n_2\Gamma_2], \tag{6.4.16}$$

for a weak contrast optical fiber. Hence, the change in the effective refractive index is proportional to the change in the refractive index in each region, weighted by the fraction of the power flow in each region.

6.5 Mode Dispersion in an Optical Fiber

Once the axial wave member k_z of our optical fiber is found, we can define an effective refractive index given by

$$n = \frac{k_z}{k_0}. \tag{6.5.1}$$

In other words,

$$k_z = n\frac{\omega}{c_0}, \tag{6.5.2}$$

where n is a function of n_1, n_2 and ω or $n = n(n_1, n_2, \omega)$. The signal in an optical fiber travels at the group velocity. Hence, we can study how the group velocity depends on n_1, n_2 and ω.

$$\frac{1}{v_g} = \frac{dk_z}{d\omega} = \frac{\omega}{c_0}\left[\frac{\partial n}{\partial n_1}\frac{dn_1}{d\omega} + \frac{\partial n}{\partial n_2}\frac{dn_2}{d\omega} + \frac{\partial n}{\partial \omega}\right] + \frac{n}{c_0}. \tag{6.5.3}$$

This first two terms in the square bracket come from the materials making up n_1 and n_2, and the last term in the square bracket is a consequence of the waveguide geometry. From (6.4.16), we deduce that

$$\frac{\partial n}{\partial n_1} = \Gamma_1, \qquad \frac{\partial n}{\partial n_2} = \Gamma_2, \tag{6.5.4}$$

This first two terms for a weak contrast optical fiber, and Γ_1 and Γ_2 are the fractions of power flow in regions 1 and 2. With the approximation that

$$\frac{dn_1}{d\omega} \simeq \frac{dn_2}{d\omega} = \left(\frac{\partial n}{\partial \omega}\right)_m, \tag{6.5.5}$$

where the subscript m stands for dispersion from material property, we can rewrite (6.5.3) as

$$\frac{1}{v_g} = \frac{dk_z}{d\omega} = \frac{\omega}{c_0}\left[\left(\frac{\partial n}{\partial \omega}\right)_m + \left(\frac{\partial n}{\partial \omega}\right)_w\right] + \frac{n}{c_0}, \tag{6.5.6}$$

where the subscript w stands for dispersion from waveguide geometry. With $\omega = 2\pi c_0/\lambda$, we have $\omega/d\omega = -\lambda/d\lambda$, and the above can be rewritten as

$$\frac{1}{v_g} = -\frac{\lambda}{c_0}\left[\left(\frac{\partial n}{\partial \lambda}\right)_m + \left(\frac{\partial n}{\partial \lambda}\right)_w\right] + \frac{n}{c_0}. \tag{6.5.7}$$

A measure of dispersion along an optical fiber of length L is the group velocity dispersion D defined as

$$D = L^{-1}\frac{\partial T}{\partial \lambda}, \tag{6.5.8}$$

where T is the travel time taken by a pulse to traverse the length L of the optical fiber. Since $T = L/v_g$, we have

$$D = \frac{\partial}{\partial \lambda}\frac{1}{v_g} = -\frac{\lambda}{c_0}\left[\left(\frac{\partial^2 n}{\partial \lambda^2}\right)_m + \left(\frac{\partial^2 n}{\partial \lambda^2}\right)_w\right]. \tag{6.5.9}$$

In order to have least pulse distortion, we should operate at a frequency where both the material dispersion, $\left(\frac{\partial^2 n}{\partial \lambda^2}\right)_m$, and the waveguide dispersion, $\left(\frac{\partial^2 n}{\partial \lambda^2}\right)_w$ are small. For GeO_2-doped silica, the material dispersion passes through a minimum at $\lambda = 1.3\mu m$. The waveguide dispersion can be altered by altering a as well as n_1 and n_2. By choosing a core diameter between 4 and 5 μm, and relative refractive index difference of $(n_1 - n_2)/n_1 > 0.004$, the wavelength of the minimum group velocity dispersion can be shifted to 1.5 to 1.6 μm region where the loss is lowest (see Figure 6.16 and Figure 6.17). The unit of D is usually in picosecond per nanometer for a given fiber lenght L. In the case of Figure 6.16, the fiber length is 11 km.

6.6 A Rectangular Dielectric Waveguide

When a dielectric waveguide is rectangular in shape, there is no closed form solution to the problem. The eigenmodes of the waveguide have to be found numerically. We shall discuss two methods of solving for the eigenmodes and eigenvalues of the rectangular dielectric waveguide.

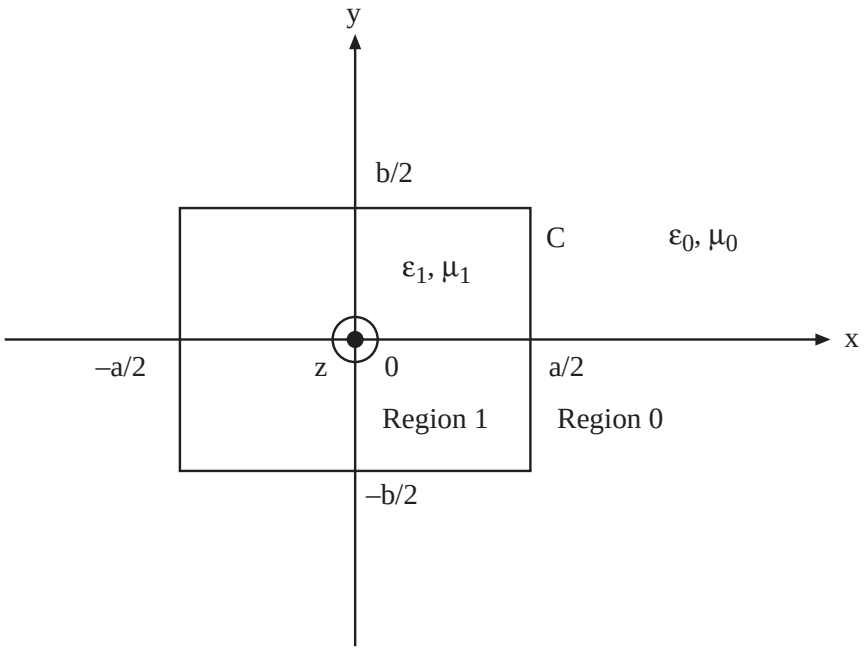


Figure 6.18: The geometry of a rectangular dielectric waveguide.

6.6.1 Harmonic Expansion Method

In this method [24] [2], we assume that the waveguide is piecewise homogeneous in each region. Inside the waveguide, the field can be decomposed into TE and TM to z waves, each of which satisfies the following equations,

$$\text{TE}: \quad (\nabla_s^2 + k_{1s}^2) H_{1z}(\mathbf{r}) = 0, \quad \mathbf{r} \in \text{region 1}, \tag{6.6.1}$$

$$\text{TM}: \quad (\nabla_s^2 + k_{1s}^2) E_{1z}(\mathbf{r}) = 0, \quad \mathbf{r} \in \text{region 1} \tag{6.6.2}$$

where $k_{1s}^2 = \omega^2 \mu_1 \epsilon_1 - k_z^2$ and $\nabla_s^2 = \partial^2/\partial x^2 + \partial^2/\partial y^2$. We assume all the fields to have $e^{ik_z z}$ dependence due to the phase matching condition. In region 0, the fields satisfy similar

[2] This method is similar to the method of Rayleigh hypothesis described in [25]. It eventually will have ill-conditioning problem when the number of unknowns is large. Mercatili [26] also analyzed this problem approximately, although with a different approach.

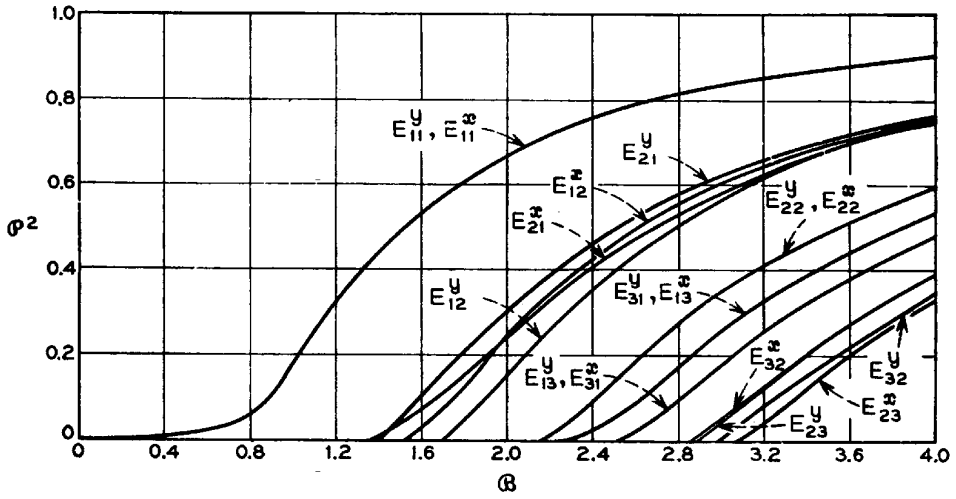


Figure 6.19: Dispersion curves of different modes of a rectangular dielectric waveguide where $(\epsilon_1/\epsilon_0)^{1/2} = 1.5$, and where the aspect ratio $a/b = 1$ [24].

equations

$$\text{TE}: \quad (\nabla_s^2 + k_{0s}^2) H_{0z}(\mathbf{r}) = 0, \quad \mathbf{r} \in \text{region 0}, \tag{6.6.3}$$

$$\text{TM}: \quad (\nabla_s^2 + k_{0s}^2) E_{0z}(\mathbf{r}) = 0, \quad \mathbf{r} \in \text{region 0}. \tag{6.6.4}$$

The TE and TM waves will be coupled by the boundary conditions at the dielectric boundary. In region 1, the general solution is

$$\begin{aligned} \begin{bmatrix} E_{1z} \\ H_{1z} \end{bmatrix} &= \sum_{n=-\infty}^{\infty} \begin{bmatrix} a_n \\ b_n \end{bmatrix} J_n(k_{1s}\rho) e^{in\phi + ik_z z} \\ &= \sum_{n=-\infty}^{\infty} \begin{bmatrix} a_n \\ b_n \end{bmatrix} \Re g\psi_n(k_{1s}, \mathbf{r}_s), \end{aligned} \tag{6.6.5}$$

while in region 0, it is

$$\begin{aligned} \begin{bmatrix} E_{0z} \\ H_{0z} \end{bmatrix} &= \sum_{n=-\infty}^{\infty} \begin{bmatrix} c_n \\ d_n \end{bmatrix} H_n^{(1)}(k_{0s}\rho) e^{in\phi + ik_z z} \\ &= \sum_{n=-\infty}^{\infty} \begin{bmatrix} c_n \\ d_n \end{bmatrix} \psi_n(k_{0s}, \mathbf{r}_s), \end{aligned} \tag{6.6.6}$$

where $\Re g\psi_n(k_s, \mathbf{r}_s) = J_n(k_s\rho)e^{in\phi}$, and $\psi_n(k_s, \mathbf{r}_s) = H_n^{(1)}(k_s\rho)e^{in\phi}$. We need to derive the transverse components of the fields in order to match boundary conditions. Using the follow-

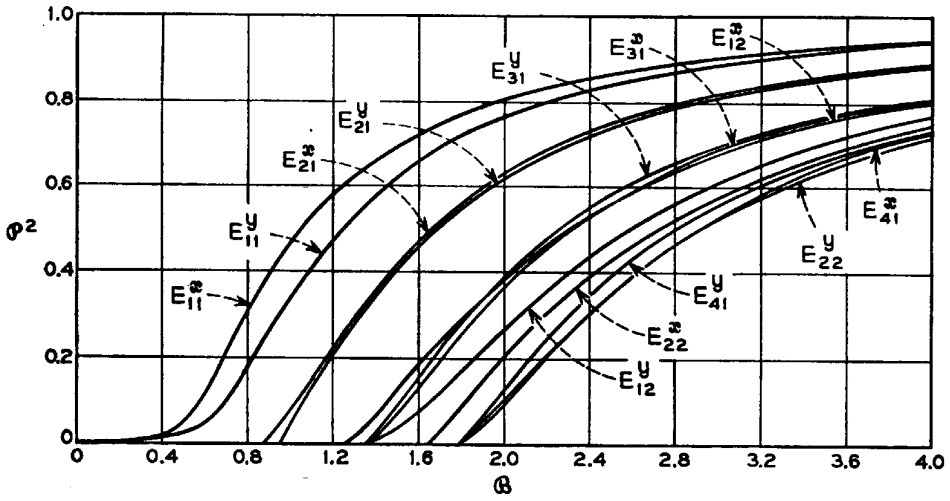


Figure 6.20: Dispersion curves of different modes of a rectangular dielectric waveguide where $(\epsilon_1/\epsilon_0)^{1/2} = 1.5$, and where the aspect ratio $a/b = 2$ [24].

ing equations,

$$\mathbf{E}_s = \frac{i}{k_s^2}[k_z \nabla_s E_z - \omega\mu \hat{z} \times \nabla_s H_z], \tag{6.6.7a}$$

$$\mathbf{H}_s = \frac{i}{k_s^2}[k_z \nabla_s H_z + \omega\epsilon \hat{z} \times \nabla_s E_z], \tag{6.6.7b}$$

we deduce that in region 1,

$$\mathbf{E}_{1s} = \sum_{n=-\infty}^{\infty} \frac{i}{k_{1s}^2}[k_z a_n \nabla_s \Re g\psi_n(k_{1s}, \mathbf{r}_s) - \omega\mu_1 b_n \hat{z} \times \nabla_s \Re g\psi_n(k_{1s}, \mathbf{r}_s)], \tag{6.6.8a}$$

$$\mathbf{H}_{1s} = \sum_{n=-\infty}^{\infty} \frac{i}{k_{1s}^2}[k_z b_n \nabla_s \Re g\psi_n(k_{1s}, \mathbf{r}_s) + \omega\epsilon_1 a_n \hat{z} \times \nabla_s \Re g\psi_n(k_{1s}, \mathbf{r}_s)], \tag{6.6.8b}$$

and in region 0,

$$\mathbf{E}_{0s} = \sum_{n=-\infty}^{\infty} \frac{i}{k_{0s}^2}[k_z c_n \nabla_s \psi_n(k_{0s}, \mathbf{r}_s) - \omega\mu_0 d_n \hat{z} \times \nabla_s \psi_n(k_{0s}, \mathbf{r}_s)], \tag{6.6.9a}$$

$$\mathbf{H}_{0s} = \sum_{n=-\infty}^{\infty} \frac{i}{k_{0s}^2}[k_z d_n \nabla_s \psi_n(k_{0s}, \mathbf{r}_s) + \omega\epsilon_0 c_n \hat{z} \times \nabla_s \psi_n(k_{0s}, \mathbf{r}_s)]. \tag{6.6.9b}$$

The boundary condition requires that tangential **E** and **H** be continuous across the dielectric interface. Matching the z components of the fields, we have

$$\sum_{n=-\infty}^{\infty} a_n \Re g \psi_n(k_{1s}, \mathbf{r}_s) = \sum_{n=-\infty}^{\infty} c_n \psi_n(k_{0s}, \mathbf{r}_s), \quad \mathbf{r}_s \in C, \tag{6.6.10a}$$

$$\sum_{n=-\infty}^{\infty} b_n \Re g \psi_n(k_{1s}, \mathbf{r}_s) = \sum_{n=-\infty}^{\infty} d_n \psi_n(k_{0s}, \mathbf{r}_s), \quad \mathbf{r}_s \in C. \tag{6.6.10b}$$

By defining a unit vector $\hat{c}$ to be pointing along the circumference of the waveguide C, we can equate the tangential components in (6.6.8a) and (6.6.9a) on C. By doing so, we have

$$\begin{aligned} &\sum_{n=-\infty}^{\infty} \frac{1}{k_{1s}^2} \left[k_z a_n \hat{c} \cdot \nabla_s \Re g \psi_n(k_{1s}, \mathbf{r}_s) - \omega \mu_1 b_n \hat{c} \cdot \hat{z} \times \nabla_s \Re g \psi_n(k_{1s}, \mathbf{r}_s) \right] \\ &= \sum_{n=-\infty}^{\infty} \frac{1}{k_{0s}^2} \left[k_z c_n \hat{c} \cdot \nabla_s \psi_n(k_{0s}, \mathbf{r}_s) - \omega \mu_0 d_n \hat{c} \cdot \hat{z} \times \nabla_s \psi_n(k_{0s}, \mathbf{r}_s) \right], \end{aligned} \tag{6.6.11a}$$

$$\begin{aligned} &\sum_{n=-\infty}^{\infty} \frac{1}{k_{1s}^2} \left[k_z b_n \hat{c} \cdot \nabla_s \Re g \psi_n(k_{1s}, \mathbf{r}_s) + \omega \epsilon_1 a_n \hat{c} \cdot \hat{z} \times \nabla_s \Re g \psi_n(k_{1s}, \mathbf{r}_s) \right] \\ &= \sum_{n=-\infty}^{\infty} \frac{1}{k_{0s}^2} \left[k_z d_n \hat{c} \cdot \nabla_s \psi_n(k_{0s}, \mathbf{r}_s) + \omega \epsilon_0 c_n \hat{c} \cdot \hat{z} \times \nabla_s \psi_n(k_{0s}, \mathbf{r}_s) \right]. \end{aligned} \tag{6.6.11b}$$

Equations (6.6.10a) and (6.6.11a) can be written as

$$\sum_{n=-\infty}^{\infty} \Re g \psi_n(k_{1s}, \mathbf{r}_s) \mathbf{a}_n = \sum_{n=-\infty}^{\infty} \psi_n(k_{0s}, \mathbf{r}_s) \mathbf{c}_n, \quad \mathbf{r}_s \in C, \tag{6.6.12a}$$

$$\begin{aligned} &\sum_{n=-\infty}^{\infty} \frac{1}{k_{1s}^2} \begin{bmatrix} k_z \hat{c} \cdot \nabla_s \Re g \psi_n & -\omega \mu_1 \hat{c} \cdot \hat{z} \times \nabla_s \Re g \psi_n \\ \omega \epsilon_1 \hat{c} \cdot \hat{z} \times \nabla_s \Re g \psi_n i & k_z \hat{c} \cdot \nabla_s \Re g \psi_n \end{bmatrix} \cdot \mathbf{a}_n \\ &\quad = \sum_{n=-\infty}^{\infty} \frac{1}{k_{0s}^2} \begin{bmatrix} k_z \hat{c} \cdot \nabla_s \psi_n & -\omega \mu_0 \hat{c} \cdot \hat{z} \times \nabla_s \psi_n \\ \omega \epsilon_0 \hat{c} \cdot \hat{z} \times \nabla_s \psi_n & k_z \hat{c} \cdot \nabla_s \psi_n \end{bmatrix} \cdot \mathbf{b}_n, \quad \mathbf{r}_s \in C \end{aligned} \tag{6.6.12b}$$

where $\mathbf{a}_n^t = [a_n, b_n]$, $\mathbf{c}_n^t = [c_n, d_n]$. We can truncate the infinite summation to range from $-N$ to $+N$. In this case, the a_n, b_n, c_n, and d_n will constitute $4(2N+1)$ unknowns. The method of point matching can be used to convert (6.6.12a) and (6.6.12b) into matrix equations,

$$\sum_{n=-N}^{N} \Re g \psi_n(k_{1s}, \mathbf{r}_{ms}) \mathbf{a}_n = \sum_{n=-N}^{N} \psi_n(k_{0s}, \mathbf{r}_{ms}) \mathbf{c}_n, \quad m = -N, \ldots, +N, \tag{6.6.13a}$$

$$\begin{aligned} &\frac{k_{0s}^2}{k_{1s}^2} \sum_{n=-N}^{N} \Re g \overline{\psi}_n(k_{1s}, \mathbf{r}_{ms}) \cdot \mathbf{a}_n \\ &\qquad = \sum_{n=-N}^{N} \overline{\psi}_n(k_{0s}, \mathbf{r}_{ms}) \cdot \mathbf{c}_n, \quad m = -N, \ldots, N \end{aligned} \tag{6.6.13b}$$

where

$$\overline{\psi}_n = \begin{bmatrix} k_z\hat{c}\cdot\nabla_s\psi_n & -\omega\mu\hat{c}\cdot\hat{z}\times\nabla_s\psi_n \\ \omega\epsilon\hat{c}\cdot\hat{z}\times\nabla_s\psi_n & k_z\hat{c}\cdot\nabla_s\psi_n \end{bmatrix}. \tag{6.6.14}$$

Equations (6.6.13a) are matrix equations of the form

$$\overline{\mathbf{A}}\cdot\mathbf{a} = \overline{\mathbf{A}}^{(1)}\cdot\mathbf{b}, \tag{6.6.15a}$$

$$\overline{\mathbf{B}}\cdot\mathbf{a} = \overline{\mathbf{B}}^{(1)}\cdot\mathbf{b}. \tag{6.6.15b}$$

Eliminating $\mathbf{a}$ from the above yields

$$[\overline{\mathbf{B}}^{-1}\cdot\overline{\mathbf{B}}^{(1)} - \overline{\mathbf{A}}^{-1}\cdot\overline{\mathbf{A}}^{(1)}]\cdot\mathbf{b} = \overline{\mathbf{M}}\cdot\mathbf{b} = 0. \tag{6.6.16}$$

Since $\overline{\mathbf{A}}$ and $\overline{\mathbf{B}}$ are functions of $k_{1s} = \sqrt{k_1^2 - k_z^2}$, and $\overline{\mathbf{A}}^{(1)}$ and $\overline{\mathbf{B}}^{(1)}$ are functions of $k_{0s} = \sqrt{k_0^2 - k_z^2}$, the matrix in (6.6.16) is a function of k_z. Nontrivial solutions exist for $\mathbf{b}$, and hence $\mathbf{a}$, (i.e., the field) only if

$$\det\left(\overline{\mathbf{M}}(k_z)\right) = 0. \tag{6.6.17}$$

Equation (6.6.17) allows us to solve for the wavenumber k_z of a guided mode. The above method is in general, applicable to waveguides of arbitrary shapes. For rectangular waveguides, symmetry may be exploited to reduce the extend of the summation in (6.6.12a), and hence the number of unknowns.

In this method, since we are only assuming standing wave in region 1 and outgoing wave in region 0, it is not valid if the waveguide is of very distorted shapes. As mentioned earlier, this method is similar to the Rayleigh's hypothesis method of solving scattering problem [25].

Because of the hybrid nature of the modes, and the variation in the aspect ratio of a rectangular dielectric waveguide, the classification of modes in a rectangular dielectric waveguide is a complex subject. The EH_{mn} and HE_{mn} notations have been adopted by some workers to denote the TM-like and TE-like nature, respectively, of the modes. However, the EH_{mn} and HE_{mn} notations do not indicate if a mode is x-polarized or y-polarized. Hence, another notation is E^x_{mn} or E^y_{mn} to denote if the mode's electric field is predominantly x or y polarized. Yet, another notation, H^x_{mn} or H^y_{mn}, is used to denote if the magnetic field is predominantly x or y polarized, A combination of E^x_{mn}, E^y_{mn}, and H^x_{mn}, H^y_{mn} has also been suggested to denote the TM-like or TE-like nature of a mode. The subscripts mn in the above denote that the mode has m maxima in the x direction and n maxima in the y direction.

Figure 6.19 shows the dispersion curves for a rectangular dielectric waveguide with unity aspect ratio, $a/b = 1$, and Figure 6.20 shows the case when the aspect ratio, $a/b = 2$. In the figures,

$$P^2 = [(k_z/k_0)^2 - 1]/[(k_1/k_0)^2 - 1],$$

and

$$B = 2b/\lambda_0[(k_1/k_0)^2 - 1]^{1/2}.$$

Hence, P^2 is an indication of if the mode is trapped in the cladding or the core. If the mode is trapped in the core, then its energy is mainly in the core, and k_z is close to k_1, and conversely, if its energy is in the cladding. And B is the normalized dimension of the waveguide with respect to the wavelength, and hence, is a normalized frequency. The unity aspect ratio causes some of the modes to be degenerate. In Figure 6.21, the intensity field plot is displayed for several modes of a rectangular dielectric waveguide.

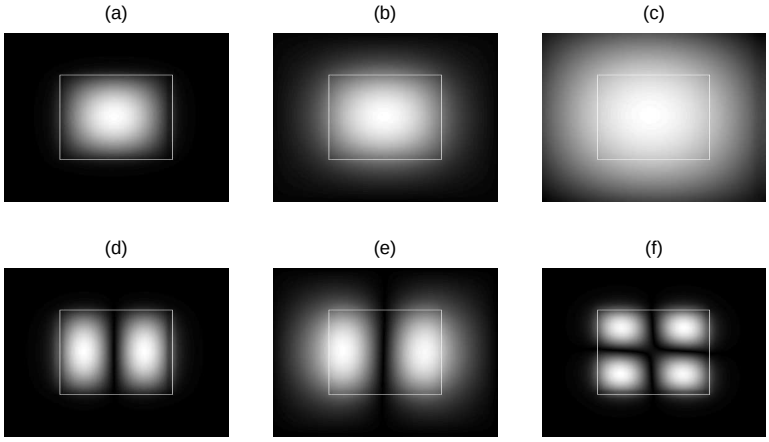


Figure 6.21: (Top row) Field intensity plots of the E_{11}^y for different degrees of mode confinement: (a) $P^2 = 0.91$, (b) $P^2 = 0.66$, and (c) $P^2 = 0.08$. (Bottom row) E_{12}^y for (d) $P^2 = 0.85$, (e) $P^2 = 0.56$, and E_{22}^y mode for (f) $P^2 = 0.66$ (courtesy of K. Radhakrishnan [30]).

6.6.2 Variational Method

The harmonic expansion method does not work if the rectangular waveguide is a part of a substrate. In such a case, a more versatile method like the variational method,[3] or the finite element method [28] should be adopted for an inhomogeneous waveguide.[4] A variational expression for the propagation constant of a waveguide mode can be derived from the vector wave equation governing the fields of the waveguide. We have shown in Chapter 3 that the equations governing the electromagnetic field in an inhomogeneously filled waveguide are

$$\mu\nabla_s \times \mu^{-1}\nabla_s \times \mathbf{E}_s - \nabla_s\epsilon^{-1}\nabla_s \cdot \epsilon\mathbf{E}_s - k^2\mathbf{E}_s + k_z^2\mathbf{E}_s = 0, \tag{6.6.18}$$

$$\epsilon\nabla_s \times \epsilon^{-1}\nabla_s \times \mathbf{H}_s - \nabla_s\mu^{-1}\nabla_s \cdot \mu\mathbf{H}_s - k^2\mathbf{H}_s + k_z^2\mathbf{H}_s = 0. \tag{6.6.19}$$

We can dot multiply (6.6.18) by $\hat{z} \times \mathbf{H}_s$ and integrate the resultant equation over the cross-section of the waveguide to yield

$$\begin{aligned}\int_S dS\hat{z} \times \mathbf{H}_s \cdot \mu\nabla_s \times \mu^{-1}\nabla_s \times \mathbf{E}_s - \int_S dS\hat{z} \times \mathbf{H}_s \cdot \nabla_s\epsilon^{-1}\nabla_s \cdot \epsilon\mathbf{E}_s \\ - \int_S dSk^2\hat{z} \cdot (\mathbf{H}_s \times \mathbf{E}_s) + k_z^2 \int_S dS\hat{z} \cdot (\mathbf{H}_s \times \mathbf{E}_s) = 0.\end{aligned} \tag{6.6.20}$$

[3] This section follows the analysis in [27].
[4] Alternatively, it can be solved by the numerical mode matching method [29].

Using the identity that $\nabla_s \cdot (\mathbf{A} \times \mathbf{B}) = \mathbf{B} \cdot \nabla_s \times \mathbf{A} - \mathbf{A} \cdot \nabla_s \times \mathbf{B}$, we have

$$\begin{aligned} \hat{z} \times \mathbf{H}_s \cdot \mu \nabla_s \times \mu^{-1} \nabla_s \times \mathbf{E}_s &= \nabla_s \cdot [(\mu^{-1} \nabla_s \times \mathbf{E}_s) \times (\hat{z} \times \mu \mathbf{H}_s)] \\ &\quad + \nabla_s \times (\hat{z} \times \mu \mathbf{H}_s) \cdot \mu^{-1} \nabla_s \times \mathbf{E}_s \\ &= \nabla_s \cdot [(\mu^{-1} \nabla_s \times \mathbf{E}_s) \times (\hat{z} \times \mu \mathbf{H}_s)] \\ &\quad + (\nabla_s \cdot \mu \mathbf{H}_s) \hat{z} \cdot \mu^{-1} \nabla_s \times \mathbf{E}_s. \end{aligned} \tag{6.6.21}$$

Using $\nabla_s \cdot \phi \mathbf{A} = (\nabla_s \phi) \cdot \mathbf{A} + \phi \nabla_s \cdot \mathbf{A}$, we have

$$\hat{z} \times \mathbf{H}_s \cdot \nabla_s \epsilon^{-1} \nabla_s \cdot \epsilon \mathbf{E}_s = \nabla_s \cdot [(\hat{z} \times \mathbf{H}_s) \epsilon^{-1} \nabla_s \cdot \epsilon \mathbf{E}_s] - \nabla_s \cdot (\hat{z} \times \mathbf{H}_s) \epsilon^{-1} \nabla_s \cdot \epsilon \mathbf{E}_s. \tag{6.6.22}$$

Equations (6.6.21) and (6.6.22) can be used to simplify the first two integrals in (6.6.20). After substituting them into (6.6.20), the divergence can be converted into a line integral on C, the outermost domain of the waveguide. This integral becomes zero by virtue of the boundary condition on C, or when $C \to \infty$. Consequently, Equation (6.6.20) becomes

$$\begin{aligned} &\int_S dS (\nabla_s \cdot \mu \mathbf{H}_s) \hat{z} \cdot \mu^{-1} \nabla_s \times \mathbf{E}_s - \int_S dS \hat{z} \cdot \nabla_s \times \mathbf{H}_s \epsilon^{-1} \nabla_s \cdot \epsilon \mathbf{E}_s \\ &\qquad - \int_S dS k^2 \hat{z} \cdot (\mathbf{H}_s \times \mathbf{E}_s) + k_z^2 \int dS \hat{z} \cdot (\mathbf{H}_s \times \mathbf{E}_s) = 0. \end{aligned} \tag{6.6.23}$$

Applying the same operation to (6.6.19), or by duality, we have

$$\begin{aligned} &\int_S dS (\nabla_s \cdot \epsilon \mathbf{E}_s) \hat{z} \cdot \epsilon^{-1} \nabla_s \times \mathbf{H}_s - \int_S dS \hat{z} \cdot \nabla_s \times \mathbf{E}_s \mu^{-1} \nabla_s \cdot \mu \mathbf{H}_s \\ &\qquad - \int_S dS k^2 \hat{z} \cdot (\mathbf{E}_s \times \mathbf{H}_s) + k_z^2 \int dS \hat{z} \cdot (\mathbf{E}_s \times \mathbf{H}_s) = 0. \end{aligned} \tag{6.6.24}$$

We note that Equations (6.6.23) and (6.6.24) are identical. If we write Equations (6.6.18) and (6.6.19) as

$$\mathcal{L}_e \cdot \mathbf{E}_s + k_z^2 \mathbf{E}_s = 0, \tag{6.6.25}$$

$$\mathcal{L}_h \cdot \mathbf{H}_s + k_z^2 \mathbf{H}_s = 0 \tag{6.6.26}$$

where $\mathcal{L}_e$ and $\mathcal{L}_h$ are the differential operators in (6.6.18) and (6.6.19), then, Equations (6.6.23) and (6.6.24) are the consequences of

$$\langle \hat{z} \times \mathbf{H}_s, \mathcal{L}_e \cdot \mathbf{E}_s \rangle + k_z^2 \langle \hat{z} \times \mathbf{H}_s, \mathbf{E}_s \rangle = 0, \tag{6.6.27}$$

$$\langle \hat{z} \times \mathbf{E}_s, \mathcal{L}_h \cdot \mathbf{H}_s \rangle + k_z^2 \langle \hat{z} \times \mathbf{E}_s, \mathbf{H}_s \rangle = 0. \tag{6.6.28}$$

Hence (6.6.25) and (6.6.26) are transpose equation of each other. A variational expression for k_z^2 is

$$k_z^2 = -\frac{\langle \hat{z} \times \mathbf{H}_s, \mathcal{L}_e \cdot \mathbf{E}_s \rangle}{\langle \hat{z} \times \mathbf{H}_s, \mathbf{E}_s \rangle} = -\frac{\langle \hat{z} \times \mathbf{E}_s, \mathcal{L}_h \cdot \mathbf{H}_s \rangle}{\langle \hat{z} \times \mathbf{E}_s, \mathbf{H}_s \rangle}. \tag{6.6.29}$$

In the above, $\langle \mathbf{A}, \mathbf{B} \rangle = \int_S dS \mathbf{A} \cdot \mathbf{B}$. The above also imply that the (H_x, H_y) formulation is the same as the (E_x, E_y) formulation, if solved variationally.

We can take the first variation of (6.6.29) by letting $\mathbf{E}_s = \mathbf{E}_{se} + \delta\mathbf{E}_s$, $\mathbf{H}_s = \mathbf{H}_{se} + \delta\mathbf{H}_s$, where $\mathbf{E}_{se}$ and $\mathbf{H}_{se}$ are the exact solutions. Then, after cross-multiplying the first equation in (6.6.29) and taking its first variation, we have

$$\begin{aligned} &\langle \hat{z} \times \mathbf{H}_{se}, \mathcal{L}_e \cdot \mathbf{E}_{se} \rangle + \langle \hat{z} \times \delta\mathbf{H}_s, \mathcal{L}_e \cdot \mathbf{E}_{se} \rangle + \langle \hat{z} \times \mathbf{H}_{se}, \mathcal{L}_e \cdot \delta\mathbf{E}_s \rangle \\ &\quad + k_{ze}^2 [\langle \hat{z} \times \mathbf{H}_{se}, \mathbf{E}_{se} \rangle + \langle \hat{z} \times \delta\mathbf{H}_s, \mathbf{E}_{se} \rangle + \langle \hat{z} \times \mathbf{H}_{se}, \delta\mathbf{E}_s \rangle] \\ &\quad + \delta k_z^2 \langle \hat{z} \times \mathbf{H}_{se}, \mathbf{E}_{se} \rangle + \ldots = 0. \end{aligned} \tag{6.6.30}$$

The leading order terms cancel as a consequence of (6.6.27). Similarly, as a consequence of (6.6.25),

$$\langle \hat{z} \times \delta\mathbf{H}_s, \mathcal{L}_e \cdot \mathbf{E}_{se} \rangle + k_{ze}^2 \langle \hat{z} \times \delta\mathbf{H}_s, \mathbf{E}_{se} \rangle = 0. \tag{6.6.31}$$

Because (6.6.23) and (6.6.24) are identical, we can show that

$$\langle \hat{z} \times \mathbf{A}, \mathcal{L}_h \cdot \mathbf{H}_s \rangle = -\langle \hat{z} \times \mathbf{H}_s, \mathcal{L}_e \cdot \mathbf{A} \rangle \tag{6.6.32}$$

where $\mathbf{A}$ is an arbitrary vector satisfying the boundary conditions on C, or if it vanishes when $C \to \infty$. As a result,

$$\langle \hat{z} \times \mathbf{H}_{se}, \mathcal{L}_e \cdot \delta\mathbf{E}_s \rangle + k_{ze}^2 \langle \hat{z} \times \mathbf{H}_{se}, \delta\mathbf{E}_s \rangle = 0, \tag{6.6.33}$$

and $\delta k_z^2 = 0$. Therefore, the first order variation in k_z^2 vanishes, implying the stationarity of (6.6.29). The same thing can be shown for the other equation in (6.6.29).

Since (6.6.29) is variational, a Rayleigh-Ritz procedure can be adopted to obtain an optimal solution for it. We let

$$\mathbf{E}_s = \sum_{n=1}^{N} a_n \mathbf{E}_{ns}, \quad \mathbf{H}_s = \sum_{m=1}^{N} b_m \mathbf{H}_{ms}. \tag{6.6.34}$$

Substituting into (6.6.29), we have

$$k_z^2 = -\frac{\sum_{n=1}^{N} \sum_{m=1}^{N} a_n b_m \langle \hat{z} \times \mathbf{H}_{ms}, \mathcal{L}_e \cdot \mathbf{E}_{ns} \rangle}{\sum_{n=1}^{N} \sum_{m=1}^{N} a_n b_m \langle \hat{z} \times \mathbf{H}_{ms}, \mathbf{E}_{ns} \rangle}, \tag{6.6.35}$$

or

$$k_z^2 = -\frac{\mathbf{b}^t \cdot \overline{\mathbf{A}} \cdot \mathbf{a}}{\mathbf{b}^t \cdot \overline{\mathbf{M}} \cdot \mathbf{a}} \tag{6.6.36}$$

where the mn element of the matrices $\overline{\mathbf{A}}$ and $\overline{\mathbf{M}}$ are

$$A_{mn} = \langle \hat{z} \times \mathbf{H}_{ms}, \mathcal{L}_e \cdot \mathbf{E}_{ns} \rangle, \tag{6.6.37a}$$

$$M_{mn} = \langle \hat{z} \times \mathbf{H}_{ms}, \mathbf{E}_{ns} \rangle. \tag{6.6.37b}$$

The optimal values of $\mathbf{a}$ and $\mathbf{b}$ in (6.6.36) are obtained by requiring the first variation of k_z^2 to vanish, or that

$$\begin{aligned}\mathbf{b}_0^t \cdot \overline{\mathbf{A}} \cdot \mathbf{a}_0 + k_{z0}^2 \mathbf{b}_0^t \cdot \overline{\mathbf{M}} \cdot \mathbf{a}_0 + \delta\mathbf{b}^t \cdot \overline{\mathbf{A}} \cdot \mathbf{a}_0 + k_{z0}^2 \delta\mathbf{b}^t \cdot \overline{\mathbf{M}} \cdot \mathbf{a}_0 \\ + \mathbf{b}_0^t \cdot \overline{\mathbf{A}} \cdot \delta\mathbf{a} + k_{z0}^2 \mathbf{b}_0^t \cdot \overline{\mathbf{M}} \cdot \delta\mathbf{a} = 0.\end{aligned} \tag{6.6.38}$$

The leading order terms cancel each other by virtue of (6.6.36). The first order term vanishes if

$$\overline{\mathbf{A}} \cdot \mathbf{a}_0 + k_{z0}^2 \overline{\mathbf{M}} \cdot \mathbf{a}_0 = 0, \tag{6.6.39a}$$

$$\overline{\mathbf{A}}^t \cdot \mathbf{b}_0 + k_{z0}^2 \overline{\mathbf{M}}^t \cdot \mathbf{b}_0 = 0. \tag{6.6.39b}$$

Equations (6.6.39a) and (6.6.39b) are matrix eigenvalue problems. They have the same set of eigenvalues. Hence, we need only to solve one of them. If $\overline{\mathbf{A}}$ and $\overline{\mathbf{M}}$ are $N \times N$ matrices, in general, there will be N eigenvalues and N eigenvectors $\mathbf{a}$ and $\mathbf{b}$. It can be shown easily that for two eigenvectors $\mathbf{a}_i$ and $\mathbf{b}_j$ corresponding to two distinct eigenvalues,

$$\mathbf{b}_j^t \cdot \overline{\mathbf{M}} \cdot \mathbf{a}_i = D_i \delta_{ij}. \tag{6.6.40}$$

In other words, they are $\overline{\mathbf{M}}$ orthogonal.

Equations (6.6.39a) and (6.6.39b) are exactly the equations one would obtain if one applies the Petrov-Galerkin method, or the method of weighted residuals to the differential Equations (6.6.25) and (6.6.26) using $\mathbf{E}_{ns}$ and $\mathbf{H}_{ns}$ as expansion functions, respectively, and using $\hat{n} \times \mathbf{H}_{ms}$ and $\hat{n} \times \mathbf{E}_{ms}$ as weighting functions, respectively. Equations (6.6.27) and (6.6.28) are also the variational integrals to be used in the finite element method. In finite element, a finite domain basis function is used. For example, the fields E_x and E_y can be written as a linear superposition of pyramidal functions, with polygonal base. The pyramids overlap with each other, and they form a piecewise linear approximation of the field between the nodal values. Alternatively, edge elements can be used to model the electric field where tangential components of the field are guaranteed to be continuous across edges [28]. Examples of a finite element mesh and a pyramidal function are shown in Figure 6.22.

6.7 Discontinuities in Dielectric Waveguides

Discontinuities in dielectric waveguides have been studied by a number of workers [25, 29, 31–36]. We have previously studied discontinuities in closed, hollow waveguides. There, we used the mode matching method to derive the solution of wave scattering by discontinuities. Because we have closed waveguides, only discrete modes exist [25]. However, when a waveguide is open, the number of modes that a waveguide has is uncountably infinite. Furthermore, there exists a set of modes which forms a continuum of modes. These modes carry energy to infinity and hence, are called the radiation modes. We shall address the mode matching method for such an open waveguide. This method is important, for instance, in ascertaining reflection loss at the facet of a heterojunction laser [32, 36].

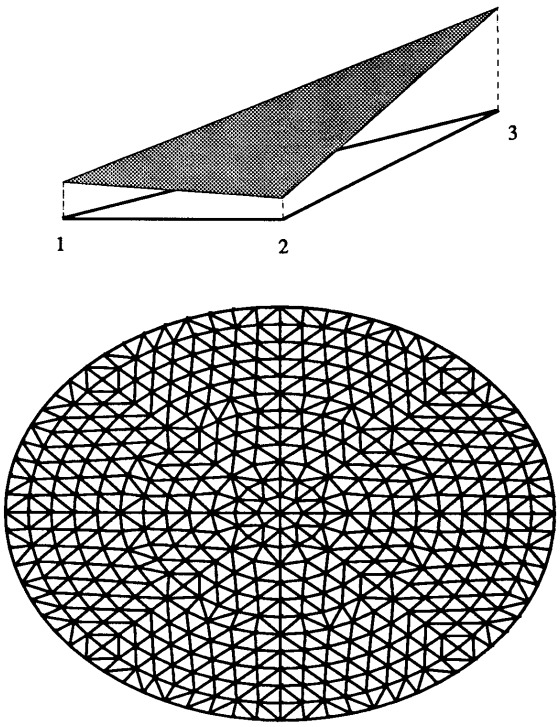


Figure 6.22: A basis function and a finite element mesh.

6.7.1 Reflection at a Laser Facet

Consider a laser facet as shown in Figure 6.23. The solution to this problem is important for the design of lasers as the reflectivity at the facet of a laser cavity determines the Q of the laser cavity. The reflectivity of such a facet can be found by mode-matching.

Consider a TE polarized mode with electric field polarized in the y direction. A part of the energy of the mode will be transmitted yielding radiation modes for $z > 0$. For $z < 0$, the mode will be reflected. Moreover, part of the energy of the reflected mode will be converted into other reflected modes giving rise to "mode conversion," just as discontinuities in uniform waveguides (see Chapter 5).

Assume that the incident mode to be of the form

$$E_{iy} = a_m E_y(m, x) e^{ik_{mz} z} \tag{6.7.1}$$

where $E_y(m, x)$ describes the transverse field distribution of the m-th mode, and k_{mz} is its corresponding wave number in the z direction. Then, the reflected modes can be expressed

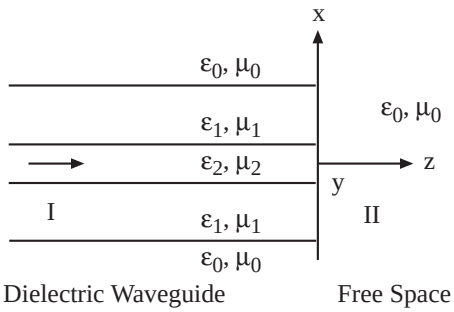


Figure 6.23: A laser facet where a mode is reflected.

as

$$E_{ry} = \sum_{m'=1}^{\infty} R_{m'm} a_m E_y(m', x) e^{-ik_{m'z}z}, \tag{6.7.2}$$

where $R_{m'm}$ is a reflection operator which is the unknown to be sought. Its off-diagonal components account for the physics of mode conversion.

For $z > 0$, the field can be expressed as

$$E_{ty} = \int_{-\infty}^{\infty} dk_x e^{ik_x x + ik_z z} e_t(k_x), \tag{6.7.3}$$

where $k_z = \sqrt{k^2 - k_x^2}$. For a fixed z, the above is essentially a Fourier transform integral.

At this point, $R_{m'm}$ and $e_t(k_x)$ are unknowns yet to be sought. They can be found by matching boundary conditions at $z = 0$. Requiring that tangential $\mathbf{E}$ field be continuous at $z = 0$, i.e., $E_{iy} + E_{ry} = E_{ty}$, we have

$$E_y(m, x) + \sum_{m'=1}^{\infty} R_{m'm} a_m E_y(m', x) = \int_{-\infty}^{\infty} dk_x e^{ik_x x} e_t(k_x) \tag{6.7.4}$$

The magnetic field is derived from Faraday's law $\nabla \times \mathbf{E} = i\omega\mu\mathbf{H}$. That is, the tangential magnetic field is H_x and is derivable from

$$H_x = \frac{-1}{i\omega\mu(x)} \frac{\partial}{\partial z} E_y \tag{6.7.5}$$

Consequently, in region I, we have

$$\begin{aligned} H_{ix} + H_{ry} = & -\frac{k_{mz}}{\omega\mu(x)} a_m E_y(m, x) e^{ik_{mz}z} \\ & + \sum_{m'=1}^{\infty} R_{m'm} a_m \frac{k_{m'z}}{\omega\mu(x)} E_y(m', x) e^{-ik_{m'z}z}, \end{aligned} \tag{6.7.6}$$

and in region II, we have

$$H_{ty} = -\int_{-\infty}^{\infty} dk_x \frac{k_z}{\omega\mu_0} e^{ik_x x + ik_z z} e_t(k_x). \tag{6.7.7}$$

Matching boundary condition for magnetic field at $z = 0$, we have

$$\begin{aligned} \frac{k_{mz}}{\omega\mu(x)} a_m E_y(m,x) - \sum_{m'=1}^{\infty} R_{m'm} a_m \frac{k_{m'z}}{\omega\mu(x)} E_y(m',x) \\ = \int_{-\infty}^{\infty} dk_x \frac{k_z}{\omega\mu_0} e^{ik_x x} e_t(k_x). \end{aligned} \tag{6.7.8}$$

By Fourier inverse transforming (6.7.4),[5] we obtain that

$$\tilde{E}_y(m,k_x) + \sum_{m'=1}^{\infty} R_{m'm} \tilde{E}_y(m',k_x) = e_t(k_x) \tag{6.7.9}$$

where

$$\tilde{E}_y(m,k_x) = \int_{-\infty}^{\infty} dx e^{-ik_x x} E_y(m,x). \tag{6.7.10}$$

The mode orthogonality relationship for an inhomogeneous waveguide is that

$$-\int_{-\infty}^{\infty} dx E_y(n,x) H_x^*(m,x) = C_n \delta_{nm} \tag{6.7.11}$$

Since $H_x(m,x) = -\frac{k_{mz}}{\omega\mu} E_y(m,x)$, this is equivalent to

$$\int_{-\infty}^{\infty} dx \frac{k_{mz}^*}{\omega\mu(x)} E_y(n,x) E_y^*(m,x) = C_n \delta_{nm}. \tag{6.7.12}$$

For a lossless waveguide, μ is real, and we can normalize the modes such that

$$\int_{-\infty}^{\infty} dx \frac{E_y(n,x) E_y^*(m,x)}{\mu(x)} = \delta_{nm}. \tag{6.7.13}$$

Then, C_n in (6.7.12) is k_{mz}^*/ω.

[5]This is the same as testing the above equation with $e^{-ik_x x}$ and integrate.

Multiplying (6.7.8) by $E_y^*(n,x)$ and integrating over x, we have

$$\begin{aligned} a_m k_{mz}\delta_{nm} - \sum_{m'=1}^{\infty} R_{m'm} a_m k_{m'z}\delta_{nm'} \\ = \int_{-\infty}^{\infty} dk_x k_z e_t(k_x) \left[\int_{-\infty}^{\infty} dx e^{-ik_x x} \frac{E_y(n,x)}{\mu_0}\right]^* \\ = \int_{-\infty}^{\infty} dk_x k_z e_t(k_x) \tilde{e}_y^*(n,k_x) \end{aligned} \tag{6.7.14}$$

where

$$\tilde{e}_y(n,k_x) = \int_{-\infty}^{\infty} dx e^{-ik_x x} \frac{E_y(n,x)}{\mu_0}. \tag{6.7.15}$$

Using (6.7.9) for $e_t(k_x)$ in (6.7.14), we have

$$\begin{aligned} a_m k_{mz}\delta_{nm} - R_{nm} a_m k_{nz} = \int_{-\infty}^{\infty} dk_x k_z \tilde{E}_y(m,k_x) \tilde{e}_y^*(n,k_x) \\ + \sum_{m'=1}^{\infty} R_{m'm} a_m \int_{-\infty}^{\infty} dk_x k_z \tilde{E}_y(m',k_x) \tilde{e}_y^*(n,k_x). \end{aligned} \tag{6.7.16}$$

The series summation in (6.7.16) can be truncated and (6.7.16) can then be solved as a matrix equation for the unknown $R_{m'm}$. By so doing, we obtain

$$\overline{\mathbf{K}}_z \cdot \mathbf{a} - \overline{\mathbf{K}}_z \cdot \overline{\mathbf{R}} \cdot \mathbf{a} = \overline{\mathbf{A}} \cdot \mathbf{a} + \overline{\mathbf{A}} \cdot \overline{\mathbf{R}} \cdot \mathbf{a}, \tag{6.7.17}$$

where

$$A_{nm} = \int_{-\infty}^{\infty} dk_x k_z \tilde{E}_y(m,k_x) \tilde{e}_y^*(n,k_x).$$

Equation (6.7.17) can be solved easily for $\overline{\mathbf{R}} \cdot \mathbf{a}$ or $\overline{\mathbf{R}}$.

The analysis above assumes that all the modes in the dielectric waveguide region are discrete. In actual fact, continuum modes exist and the discrete summations in (6.7.2) will have to be augmented by a continuous summation which is an integral. An analysis involving such an integral is difficult and the continuum modes can be discretized by putting metallic boundaries far away from the dielectric waveguide.

6.7.2 Determination of the Modes

The modes in the dielectric waveguide region are the natural solution of the wave equation [25]

$$\left[\mu_r \frac{\partial}{\partial x} \mu_r^{-1} \frac{\partial}{\partial x} + k^2(x) - k_z^2\right] E_y = 0. \tag{6.7.18}$$

If the dielectric waveguide has a piecewise constant or step profile, these modes can be found in closed form.

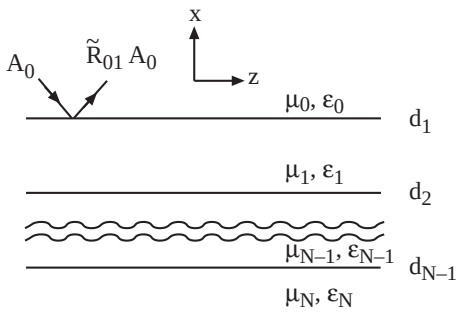


Figure 6.24: Waves in a layered medium.

For a particular mode, due to phase matching, the waves in all regions have $e^{ik_z z}$ dependence. Hence, in the i-th layer,

$$E_{iy} = e_{iy}(x)e^{ik_z z}. \tag{6.7.19}$$

Moreover, since each layer is homogeneous, $e_{iy}(x)$ is a linear superposition of upgoing and downgoing waves. More specifically,

$$e_{iy}(x) = A_i \left[e^{-ik_{ix}x} + \tilde{R}_{i,i+1} e^{2ik_{ix}d_i + ik_{ix}x} \right]. \tag{6.7.20}$$

The generalized Fresnel reflection coefficient can be found recursively via

$$\tilde{R}_{i,i+1} = \frac{R_{i,i+1} + \tilde{R}_{i+1,i+2} e^{2ik_{i+1,x}(d_{i+1}-d_i)}}{1 + R_{i,i+1}\tilde{R}_{i+1,i+2} e^{2ik_{i+1,x}(d_{i+1}-d_i)}} \tag{6.7.21}$$

where R_{ij} is the local Fresnel reflection coefficient. The amplitudes A_i can be found via the recursion relation

$$A_i e^{ik_{iz}d_{i-1}} = \frac{T_{i-1,i} A_{i-1} e^{ik_{i-1,z}d_{i-1}}}{1 - R_{i,i-1}\tilde{R}_{i,i+1} e^{2ik_{iz}(d_i - d_{i-1})}} \tag{6.7.22}$$

A guided mode is defined as a solution to (6.7.18) without an external excitation. Hence, it corresponds to a nonzero $\tilde{R}_{01}$ even when A_0, the amplitude of the external excitation, is zero. In other words, the guided modes by the layered region corresponds to the poles of $\tilde{R}_{01}$. Hence, they can be found by searching for the roots of $\left[\tilde{R}_{01}(k_z)\right]^{-1}$.

The continuum modes are called radiation modes, so called because they carry energy to infinity. There are two classes of radiation mode, one with real k_{0x}, and the other with real k_{Nx}. For the radiation modes with real k_{0x}, their expression in region 0 is

$$e_{0y}(x) = A_0 \left[e^{-ik_{0x}x} + \tilde{R}_{01} e^{2ik_{0x}d_0 + ik_{0x}x} \right], \tag{6.7.23}$$

where $k_{0x}^2 + k_z^2 = k_0^2$. Here, k_{0x} has to be real in order for $e_{0y}(x)$ to be bounded when $x \to \infty$. Notice that in order for k_{0x} to be real, k_z lies along the locus as shown in the complex k_z plane. The field of the radiation mode in every layer can be found by using the recursion relation as before.

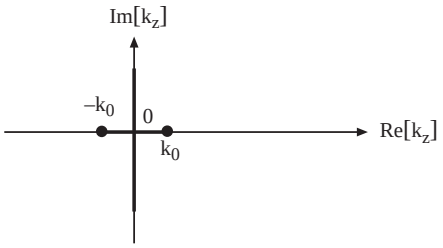


Figure 6.25: The distribution of k_z corresponding to the radiation modes with real k_{ox}.

The other class of radiation modes with real k_{Nx}, has both upgoing and downgoing wave in region N. Their expression in region N is

$$e_{Ny}(x) = A_N \left[e^{ik_{Nx}x} + \tilde{R}_{N,N-1} e^{-2ik_{Nx}d_{N-1} - ik_{Nx}x} \right] \tag{6.7.24}$$

Their field in every layer can be found by a similar recursive relation. Hence, in general, the radiation modes of an open dielectric waveguide is expressible as

$$E_y(x,y) = \int_0^\infty dk_{0x} e_y(x, k_{0x}) e^{ik_z z} + \int_0^\infty dk_{Nx} e_y(x, k_{Nx}) e^{ik_z z}. \tag{6.7.25}$$

6.8 Analyzing Weak Contrast Optical Fiber with WKB method

When the profile of an optical fiber is slowly varying and the radius of the core is large compared to wavelength, the WKB (Wentzel, Kramer and Brillouin) method can be applied to the analysis of the optical fiber [1,5]. The WKB method is often discussed in many books on quantum mechanics, and also discussed in [25].

If ϕ in Equation (6.3.6) is written as

$$\phi(\mathbf{r}) = R(\rho) e^{ik_z z + in\phi}, \tag{6.8.1}$$

then the equation for $R(\rho)$ is

$$\frac{d^2}{d\rho^2} R(\rho) + \frac{1}{\rho} \frac{d}{d\rho} R(\rho) - \frac{n^2}{\rho^2} R(\rho) + [k^2(\rho) - k_z^2] R(\rho) = 0. \tag{6.8.2}$$

By letting

$$\hat{R}(\rho) = \sqrt{\rho} R(\rho), \tag{6.8.3}$$

if follows that

$$\begin{aligned} \frac{d^2\hat{R}(\rho)}{d\rho^2} &= \frac{d}{d\rho}\left[\frac{1}{2}\rho^{-\frac{1}{2}}R(\rho) + \rho^{\frac{1}{2}}R'(\rho)\right] \\ &= -\frac{1}{4}\rho^{-\frac{3}{2}}R(\rho) + \rho^{-\frac{1}{2}}R'(\rho) + \rho^{\frac{1}{2}}R''(\rho). \end{aligned} \tag{6.8.4}$$

Multiplying (6.8.2) by $\rho^{\frac{1}{2}}$, and using (6.8.4) in the resultant (6.8.2), we have

$$\frac{d^2\hat{R}(\rho)}{d\rho^2} + [E - V(\rho)]\,\hat{R}(\rho) = 0 \tag{6.8.5}$$

where

$$\begin{aligned} E &= k^2(\infty) - k_z^2, \\ V(\rho) &= k^2(\infty) - k^2(\rho) + \frac{\left(n^2 - \frac{1}{4}\right)}{\rho^2}. \end{aligned} \tag{6.8.6}$$

Note that $V(\rho) \to 0$, when $\rho \to \infty$. Equation (6.8.5) is the same as the Schrödinger's equation for describing the motion of a particle in quantum mechanics. Here, $V(\rho)$ is the potential well, and E is the energy of the particle. A particle is bound in the potential well when $E < 0$, or when k_z is real and that $k_z > k(\infty)$. This corresponds to a guided mode in the optical fiber, because $k_\rho(\infty) = \sqrt{k^2(\infty) - k_z^2} = i\alpha(\infty)$. When $n = 0$, the function $V(\rho)$ is as shown in Figure 1. When n is large, $V(\rho)$ may not be negative at all, and no bound state or guided mode can exist. When $E > 0$, $k_z < k(\infty)$, the mode is not bound and it radiates energy to infinity and becomes a radiation mode. This is because $\alpha(\infty)$ is not real anymore and the field is not evanescent outside the fiber.

6.8.1 The WKB Method

To analyze Equation (6.8.5) with the WKB method [25], we rewrite it as

$$\hat{R}''(\rho) + k_\rho^2(\rho)\hat{R}(\rho) = 0 \tag{6.8.7}$$

where $k_\rho^2(\rho) = E - V(\rho)$. Note that $k_\rho^2 \sim \omega^2$ when $\omega \to \infty$. Hence k_ρ^2 becomes a large parameter in the high frequency limit. By foreseeing that the solution of (6.8.7) may look like a plane wave, we let

$$\hat{R}(\rho) = Ae^{i\omega\tau(\rho)}. \tag{6.8.8}$$

Then it follows that

$$\hat{R}''(\rho) = \left\{i\omega\tau''(\rho) - [\omega\tau'(\rho)]^2\right\} Ae^{i\omega\tau(\rho)}, \tag{6.8.9}$$

and (6.8.7) becomes

$$i\omega\tau''(\rho) - [\omega\tau'(\rho)]^2 + k_\rho^2(\rho) = 0. \tag{6.8.10}$$

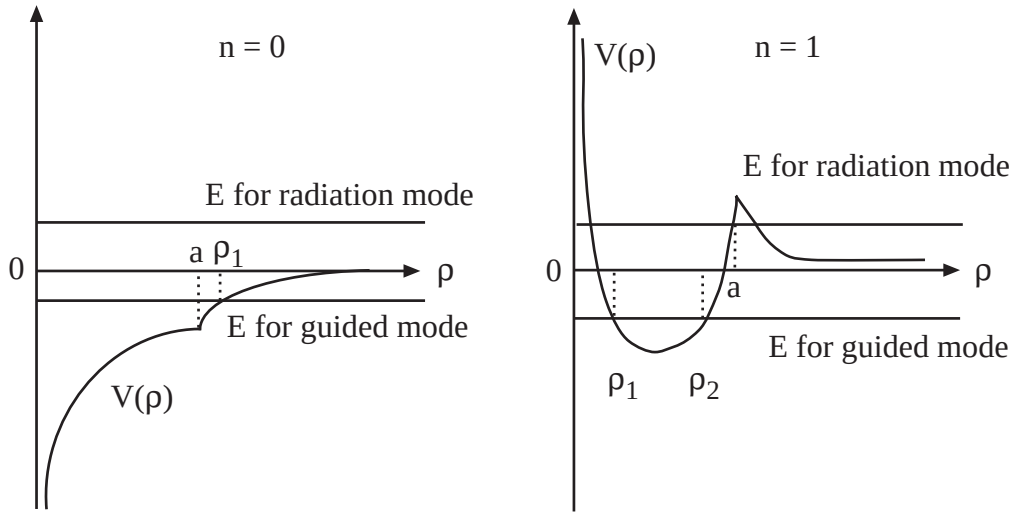


Figure 6.26: $V(\rho)$ for $n = 0$ and $n = 1$.

Using the perturbation method, we expand $\tau(\rho)$ in a perturbation series, namely,

$$\tau(\rho) = \tau_0(\rho) + \frac{1}{\omega}\tau_1(\rho) + \ldots, \qquad \omega \to \infty \tag{6.8.11}$$

where we have used $1/\omega$ as the small parameter. Note that

$$i\omega\tau_0''(\rho) \ll k_\rho^2(\rho), \qquad \omega \to \infty \tag{6.8.12}$$

because $k_\rho^2 \sim \omega^2$. Consequently, using (6.8.11) in (6.8.10), and collecting leading order terms when $\omega \to \infty$, we have

$$[\omega\tau_0'(\rho)]^2 = k_\rho^2(\rho) = \omega^2 s_\rho^2(\rho) \tag{6.8.13}$$

where we define $s_\rho = k_\rho/\omega$ to be the slowness of a wave. The above is known as the eikonal equation. Consequently, on solving (6.8.13), we arrive at

$$\tau_0(\rho) = \pm \int_{\rho_0}^{\rho} d\rho' s_\rho(\rho') + C_0. \tag{6.8.14}$$

Next, by collecting first order terms in (6.8.10) after substitution with (6.8.11), we have

$$i\omega\tau_0''(\rho) - 2\omega\tau_0'(\rho)\tau_1'(\rho) = 0. \tag{6.8.15}$$

The above is known as the transport equation. Solving this yields

$$\tau_1(\rho) = \frac{i}{2}\ln\tau_0'(\rho) + C_1 = \frac{i}{2}\ln s_\rho(\rho) + C_{1\pm}. \tag{6.8.16}$$

Consequently, using (6.8.14) and (6.8.16) in (6.8.11), we have

$$\tau(\rho) = \pm \int_{\rho_0}^{\rho} d\rho' s_\rho(\rho') + \frac{i}{2\omega} \ln s_\rho(\rho) + C_\pm, \tag{6.8.17}$$

or that the approximate solution to (6.8.5) is of the form

$$\hat{R}(\rho) \sim \frac{A_+}{\sqrt{s_\rho}} \exp\left(i\omega \int_{\rho_0}^{\rho} s_\rho(\rho') d\rho' \right) + \frac{A_-}{\sqrt{s_\rho}} \exp\left(-i\omega \int_{\rho_0}^{\rho} s_\rho(\rho') d\rho' \right). \tag{6.8.18}$$

The first term in (6.8.18) corresponds to a right-going wave because its phase is increasing with distance. By the same token, the second term in (6.8.18) is a left-going wave. Moreover, the integral in the exponent elucidates the physical picture that the phase gained by a wave gong from ρ_0 to ρ is proportional to

$$\omega \int_{\rho_0}^{\rho} s_\rho(\rho') d\rho' \tag{6.8.19}$$

which is the integral summation of all the phases gained locally at ρ' over the range from ρ_0 to ρ. This physical picture is true only if the multiple reflections of the wave can be neglected as it is propagating. Furthermore, Equation (6.8.12) shows that this physical picture, which corresponds to the leading order solution, is correct only if

$$\omega s'_\rho(\rho) \ll \omega^2 s_\rho^2(\rho). \tag{6.8.20}$$

Hence, this picture breaks down if the frequency is not high, or if $s_\rho(\rho) \simeq 0$. The factor of $1/\sqrt{s_\rho}$ in (6.8.18) is necessary for energy conservation. It is related to the wave impedance of the wave, and hence, it alters the amplitude of the wave to conserve energy.

Note that when $s_\rho^2 < 0$, corresponding to when $E < V(\rho)$, or when the wave becomes evanescent, the above analysis is still valid. Hence, the above analysis is valid for the field of a guided mode in the region when $E > V(\rho)$, and in the region where $E < V(\rho)$, but not in vicinity of the region where $E = V(\rho)$. In the last case, $s_\rho(\rho) \simeq 0$. This happens, for instance, in the guided mode case when $\rho = \rho_1$ and $\rho = \rho_2$ in Figure 6.26(b). In region II, we have bouncing waves, and regions I and III, the waves are evanescent. Hence, at $\rho = \rho_1$ and $\rho = \rho_2$, the waves are critically refracted, and they are also known as the turning points.

6.8.2 Solution in the Vicinity of a Turning Point

In the vicinity of $\rho = \rho_1$ in Figure 1(a), $k_\rho^2(\rho)$ can be approximated by a linear function, i.e.,

$$k_\rho^2 \simeq \omega^2 \Omega(\rho_1 - \rho), \qquad \rho \to \rho_1, \tag{6.8.21}$$

i.e., k_ρ^2 is proportional to ω^2. Hence, around ρ_1, Equation (6.8.7) becomes

$$\hat{R}''(\rho) + \omega^2 \Omega(\rho_1 - \rho)\hat{R}(\rho) = 0. \tag{6.8.22}$$

Next by letting $\eta = \omega^{\frac{2}{3}}\Omega^{\frac{1}{3}}(\rho_1 - \rho)$, (6.8.22) becomes

$$\left[\frac{d^2}{d\eta^2} + \eta\right]\hat{R}(\eta) = 0, \tag{6.8.23}$$

which is the Airy equation. The general solution to the above equation is of the form

$$\hat{R}(\eta) = C_1 A_i(-\eta) + C_2 B_i(-\eta) \tag{6.8.24}$$

where $A_i(-\eta)$ and $B_i(-\eta)$ are special functions called the Airy functions. Since $\eta = \omega^{\frac{2}{3}}\Omega^{\frac{1}{3}}(\rho_1 - \rho)$, $\eta \to \infty$ when $\rho \ll \rho_1$ and $\eta \to -\infty$ when $\rho \gg \rho_1$. The asymptotic expansions of Airy functions can be used to approximate them when their arguments are large. Therefore, when $\rho \gg \rho_1$, $\eta \to -\infty$, and we have

$$A_i(-\eta) \sim \frac{1}{2}\pi^{-\frac{1}{2}}(-\eta)^{-\frac{1}{4}}e^{-\frac{2}{3}(-\eta)^{\frac{3}{2}}}, \qquad \eta \to -\infty, \tag{6.8.25a}$$

$$B_i(-\eta) \sim \pi^{-\frac{1}{2}}(-\eta)^{-\frac{1}{4}}e^{\frac{2}{3}(-\eta)^{\frac{3}{2}}}, \qquad \eta \to -\infty. \tag{6.8.25b}$$

$A_i(-\eta)$ corresponds to an exponentially decaying wave while $B_i(-\eta)$ corresponds to an exponentially growing wave. Since we cannot have an exponentially growing wave to the right of ρ_1, we must have $C_2 = 0$. Hence, in the vicinity of the turning point $\rho = \rho_1$,

$$\hat{R}(\eta) = C_1 A_i(-\eta). \tag{6.8.26}$$

when $\rho \ll \rho_1$, $\eta \to -\infty$, and we have

$$A_i(-\eta) \sim \pi^{-\frac{1}{2}}\eta^{-\frac{1}{4}}\sin\left(\frac{2}{3}\eta^{\frac{3}{2}} + \frac{\pi}{4}\right), \qquad \eta \to +\infty, \tag{6.8.27}$$

which corresponds to a standing wave resulting from a superposition of incident and reflected waves on the left of the turning point.

6.8.3 Asymptotic Matching

The guidance condition of the modes in Figure 6.26(b) can be found by asymptotic matching. We shall illustrate asymptotic matching with the simpler case in Figure 6.26(a). In this method, we seek the solutions in the region where $0 < \rho < \rho_1$, $\rho > \rho_1$, and solutions in the vicinity of $\rho = 0$ and $\rho = \rho_1$. For $0 < \rho < \rho_1$ and $\rho > \rho_1$, we can use the WKB solutions, while for ρ in the vicinity of $\rho = 0$ and $\rho = \rho_1$, we need to use some special function solutions. Even though the WKB solutions are not valid at $\rho = 0$ and $\rho = \rho_1$, but when $\omega \to \infty$, the WKB solutions are valid in the vicinity of these points. By using the large argument expansions of the special function solutions, overlapping regions of validity of the solutions exist, and they can be matched to each other to find the unknowns and the guidance condition of the waveguide.

The WKB solutions are also the geometrical optics solutions which are valid when the frequency is high. Hence, they are similar to the ray-optics solutions. Ray optics solutions break down at caustics where rays bunch together. It turns out that $\rho = 0$ and $\rho = \rho_1$ are

caustics where rays bunch together. Hence, special function solutions are needed at these caustic points. It is worthwhile to notice that a ray undergoes a 90° phase shift at a caustic, and this phenomenon will be observed in the later derivation.

The solution to the left of $\rho = \rho_1$ is given by

$$\hat{R}(\rho) \sim \frac{A_+}{\sqrt{s_\rho}} \exp(i\omega \int_0^\rho s_\rho(\rho')d\rho') + \frac{A_-}{\sqrt{s_\rho}} \exp(-i\omega \int_0^\rho s_\rho(\rho')d\rho') \tag{6.8.28}$$

When $\rho \to 0$, assuming that $k^2(\rho)$ tends to a constant $k^2(0)$, then the solution

$$\hat{R}(\rho) = A\sqrt{\rho} J_0(k_s\rho), \qquad \rho \simeq 0, \tag{6.8.29}$$

where $k_s = \sqrt{k^2(0) - k_z^2}$. Hence k_s is proportional to ω.

When the frequency is high, $k_s\rho \gg 1$ for $\rho \neq 0$, and (6.8.29) can be approximated by

$$\hat{R}(\rho) \sim A\sqrt{\frac{2}{\pi k_s}} \cos\left(k_s\rho - \frac{\pi}{4}\right), \qquad k \to \infty. \tag{6.8.30}$$

In the limit when $\omega \to \infty$, $k_\rho^2(\rho)$ defined for (6.8.7) becomes

$$k_\rho^2 \sim k^2(\rho) - k_z^2, \qquad \omega \to \infty. \tag{6.8.31}$$

This is even valid when $\rho \simeq 0$ as long as $k^2\rho^2 \to \infty$. Hence,

$$k_\rho^2 \simeq k^2(0) - k_z^2 = k_s^2, \qquad \omega \to \infty, \qquad \rho \simeq 0. \tag{6.8.32}$$

In this limit, $s_\rho = k_s/\omega$. Then, Equation (6.8.28) in the vicinity of $\rho \simeq 0$ becomes

$$\hat{R}(\rho) \sim A_+\sqrt{\frac{\omega}{k_s}} e^{ik_s\rho} + A_-\sqrt{\frac{\omega}{k_s}} e^{-ik_s\rho}. \tag{6.8.33}$$

Comparing (6.8.30) and (6.8.33), we require that

$$A_+ = \frac{A}{\sqrt{2\pi\omega}} e^{-i\frac{\pi}{4}}, \tag{6.8.34a}$$

$$A_- = \frac{A}{\sqrt{2\pi\omega}} e^{+i\frac{\pi}{4}}. \tag{6.8.34b}$$

The $-90°$ phase shift between A_- and A_+ is reminiscent of an optical ray going through a caustic. Consequently, (6.8.28) becomes

$$\hat{R}(\rho) = \frac{A}{\sqrt{2\pi\omega s_\rho}} \cos\left[\omega \int_0^\rho s_\rho(\rho')d\rho' - \frac{\pi}{4}\right]. \tag{6.8.35}$$

In the vicinity of $\rho = \rho_1$, the phase integral can be approximated by

$$\int_0^\rho s_\rho(\rho')d\rho' = \int_0^{\rho_1} s_\rho(\rho')d\rho' + \int_{\rho_1}^\rho s_\rho(\rho')d\rho'. \tag{6.8.36}$$

From (6.8.21), $s_\rho(\rho') \simeq \Omega^{\frac{1}{2}}(\rho_1-\rho)^{\frac{1}{2}}$ when $\rho \simeq \rho_1$. Therefore,

$$\int_{\rho_1}^{\rho} s_\rho(\rho')d\rho' = \Omega^{\frac{1}{2}} \int_0^{(\rho_1-\rho)} \sqrt{x}dx = \frac{2}{3}\Omega^{\frac{1}{2}}(\rho_1-\rho)^{\frac{3}{2}}, \tag{6.8.37}$$

and

$$\hat{R}(\rho) \simeq \frac{A}{(2\pi\omega)^{\frac{1}{2}}\Omega^{\frac{1}{4}}(\rho-\rho_1)^{\frac{1}{4}}} \cos\left[\frac{2}{3}\omega\Omega^{\frac{1}{2}}(\rho_1-\rho)^{\frac{3}{2}} + \phi - \frac{\pi}{4}\right], \tag{6.8.38}$$

where

$$\phi = \omega \int_0^{\rho_1} s_\rho(\rho')d\rho'. \tag{6.8.39}$$

Using the definition of η in (6.8.26), and approximating $A_i(-\eta)$ with (6.8.27), we have

$$\hat{R}(\rho) \simeq C_1 \frac{1}{\sqrt{\pi}\omega^{\frac{1}{6}}\Omega^{\frac{1}{12}}(\rho_1-\rho)^{\frac{1}{4}}} \sin\left[\frac{2}{3}\omega\Omega^{\frac{1}{2}}(\rho_1-\rho)^{\frac{3}{2}} + \frac{\pi}{4}\right], \tag{6.8.40}$$

when $\omega\Omega^{\frac{1}{2}}(\rho_1-\rho)^{\frac{3}{2}} \gg 1$.

Comparing (6.8.38) with (6.8.40), in order for the solution in region for which $\rho < \rho_1$ and the solution for which $\rho \simeq \rho_1$ to agree with each other, we must have

$$C_1 = \frac{(-1)^m A}{\sqrt{2}\omega^{\frac{1}{3}}\Omega^{\frac{1}{6}}}, \tag{6.8.41a}$$

$$\phi = \omega \int_0^{\rho_1} s_\rho(\rho')d\rho' = m\pi, \tag{6.8.41b}$$

where $m = 1, 2, 3, \ldots$. Equation (6.8.41b) is the guidance condition for the $n = 0$ mode of an optical fiber. The phase shift at the $\rho = 0$ caustic is -90° while the phase shift at the $\rho = \rho_1$ caustic is $+90^\circ$. Hence, (6.8.41b) resembles the guidance condition for a parallel-plate waveguide. Similar procedures can be used to find the guidance condition for the $n = 1, 2, 3, \ldots$ modes. When $n \sim O(ka)$ then there will be two turning points both at $\rho = \rho_1$ and $\rho = \rho_2$. The analysis will be slightly different from the above.

Notice that in the above, the WKB solution for $\rho > \rho_1$, was not used other than that it is exponentially decaying. The reason is that if there is no turning point beyond $\rho = \rho_1$, the reflection of the bouncing waves in the optical fiber is determined by the turning point at $\rho = \rho_1$ only. A WKB solution presents a physical picture of a wave propagating without reflection. Hence, a turning point is the only place where a wave is reflected. To obtain the magnitude of the evanescent wave for $\rho > \rho_1$, asymptotic matching can be used.

6.9 Effective Index Method

In integrated optics waveguides, the refractive index profile is often obtained by doping. Hence, the variation of the refractive index is weak. In such a case, an approximate method

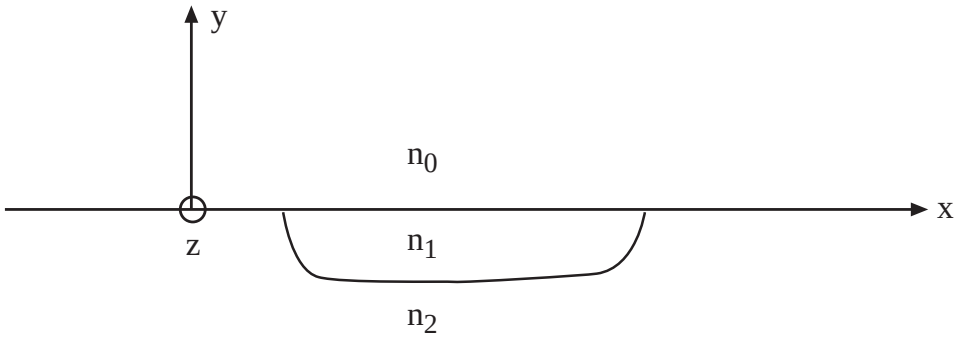


Figure 6.27: A typical integrated optics waveguide cross-section.

called the effective index method may be used to analyze the waveguiding structure. The method was proposed by Knox and Toulios [37], and has been used by many workers [38, 39]. Assuming that $\mu_r = 1$, the equations governing the electromagnetic field in such an inhomogeneous structure are exactly given by

$$\nabla^2\mathbf{E} + \nabla(\nabla \ln \epsilon_r \cdot \mathbf{E}) + k_0^2\epsilon_r\mathbf{E} = 0, \tag{6.9.1}$$

$$\nabla^2\mathbf{H} + \nabla \ln \epsilon_r \times \nabla \times \mathbf{H} + k_0^2\epsilon_r\mathbf{H} = 0. \tag{6.9.2}$$

6.9.1 Effective Index Concept

When a mode is propagating in a waveguide, say the optical fiber, with $e^{ik_z z}$ dependance, we can define an effective index n_e such that

$$k_z = k_0 n_e \tag{6.9.3}$$

Hence, a TEM wave propagating in a homogeneous medium with this effective index will have $k = k_0 n_e$ equal to the k_z of the guided mode. This concept can be extended to other structures, including a wave propagating in a slab or layered waveguide

To explain the effective index method, one considers a one-dimensional problem first as shown in Figure 6.28 where ϵ is a function of y only. Then if we consider TE wave, which can be characterized by H_y, the governing equation is

$$\nabla^2 H_y + k_0^2\epsilon_r(y)H_y = 0 \tag{6.9.4}$$

The above can be solved by the separation of variables, by letting

$$H_y(x, y, z) = Y(y)h(x, z) \tag{6.9.5}$$

Substituting (6.9.5) into (6.9.4) leads to

$$h(x, z)\frac{\partial^2}{\partial y^2}Y(y) + Y(y)\left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial z^2}\right)h(x, z) + k_0^2\epsilon_r Y(y)h(x, z) = 0 \tag{6.9.6}$$

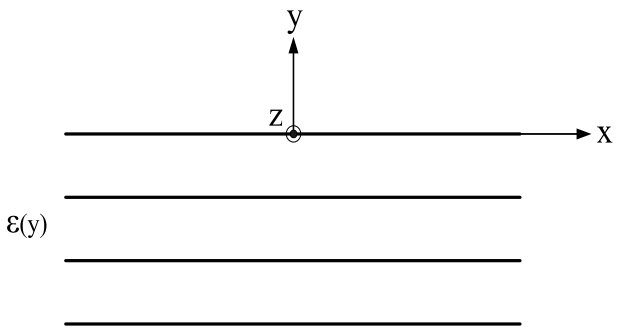


Figure 6.28: A layered medium waveguide.

Dividing by $Y(y)h(x,z)$ yields

$$Y^{-1}\frac{\partial^2}{\partial y^2}Y + h^{-1}\left(\frac{\partial^2}{\partial x^2}+\frac{\partial^2}{\partial z^2}\right)h + k_0^2\epsilon_r = 0 \tag{6.9.7}$$

The above can be satisfied by letting

$$\left(\frac{\partial^2}{\partial x^2}+\frac{\partial^2}{\partial z^2}\right)h(x,z) = -k_0^2 n^2 h(x,z) \tag{6.9.8}$$

$$\frac{\partial^2}{\partial y^2}Y(y) + {k_0}^2\epsilon_r(y)Y(y) = k_0^2 n^2 Y(y) \tag{6.9.9}$$

where $k_0^2 n^2$ is the separation constant.

Equation (6.9.9) is an eigenvalue problem with eigenvalue $k_0^2 n^2$ and eigenfunction $Y(y)$. Here, $Y(y)$ corresponds to guided mode or eigenmode in the layered medium with eigenvalue $k_0^2 n^2$. Equation (6.9.8) represents a wave traveling in the (x,z) direction with effective index n, which is a constant independent of (x,z). When there are more than one eigenmode, then one can consider the dominant eigenmode. In principle, each eigenmode can be considered separately even though the effective index for each of them could be different.

6.9.2 Quasi-TE polarization

If we assume that $\mathbf{E}$ is polarized predominantly in the xy direction, and that $\epsilon_r(x,y,z)$ is a slowly varying function of x and z, then, $\nabla \ln \epsilon_r$ is predominantly $\hat{y}$ directed. Extracting the y component of (6.9.2) to characterize a TE to y wave, we have

$$\nabla^2 H_y + k_0^2 \epsilon_r H_y \approx 0 \tag{6.9.10}$$

Next, we assume that

$$H_y = Y(x,y,z)h(x,z), \tag{6.9.11}$$

where $\partial Y/\partial x \simeq 0$, $\partial Y/\partial z \simeq 0$. In other words, $Y(x, y, z)$ is a slowly varying function of x and z. Equation (6.9.11) is motivated by the separation of variables as described in the previous section. As noted before, the problem is completely separable if ϵ_r is a function of y only. But here, we assume that the problem is approximately separable.

Then,

$$\nabla^2 H_y \simeq h\frac{\partial^2 Y}{\partial y^2} + Y\left(\frac{\partial^2 h}{\partial x^2} + \frac{\partial^2 h}{\partial z^2}\right). \tag{6.9.12}$$

Consequently, (6.9.11) becomes

$$h(x,z)\frac{\partial^2}{\partial y^2}Y(y) + Y(y)\left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial z^2}\right)h(x,z) + k_0^2\epsilon_r Y(y)h(x,z) \approx 0 \tag{6.9.13}$$

Again, motivated by the separation of variables, we let

$$\left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial z^2}\right)h(x,z) + k_0^2 n^2(x,z)h(x,z) = 0 \tag{6.9.14}$$

where $n(x, z)$ is an effective index that is a slowly varying function of x and z only. Furthermore, we require that

$$\frac{\partial^2}{\partial y^2}Y(x,y,z) + k_0^2\epsilon_r(x,y,z)Y(x,y,z) = k_0^2 n^2(x,z)Y(x,y,z). \tag{6.9.15}$$

For every fixed x and z, the above is a one-dimensional eigenequation with eigenvalue

$$k_0^2 n^2(x,z).$$

In principle, there are infinitely many eigenvalues and eigenfunctions associated with (6.9.15). The solutions of equation (6.9.15) are the modes of the structure. We can assume one mode only, or that only the fundamental mode is important. If $\epsilon_r(x, y, z)$ is piecewise constant in y, then transcendental equations can be derived to yield $k_0^2 n^2(x, z)$. Hence, if $\epsilon_r(x, y, z)$ is a slowly varying function of x and z, $Y(x, y, z)$ is in fact a slowly varying function of x and z validating the assumption (6.9.11) and the self-consistency of the method.

6.9.3 Quasi-TM Polarization

In this case, we assume that the magnetic field is predominantly polarized in the xz plane. Such a wave field can be characterized by the E_y component of the electric field. Extracting the y component of (6.9.1), we have

$$\nabla^2 E_y + \frac{\partial}{\partial y}\left[\left(\frac{\partial}{\partial y}\ln\epsilon_r\right)E_y\right] + k_0^2\epsilon_r E_y \approx 0. \tag{6.9.16}$$

In arriving at the approximate equation above, we assume that

$$\frac{\partial}{\partial y}\ln\epsilon_r \gg \nabla_s \ln\epsilon_r \tag{6.9.17}$$

where $\nabla_s = \hat{x}\frac{\partial}{\partial x} + \hat{z}\frac{\partial}{\partial z}$. In the above, (6.9.16) can be rewritten as

$$\left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial z^2} + \epsilon_r \frac{\partial}{\partial y}\epsilon_r^{-1}\frac{\partial}{\partial y} + k_o^2\epsilon_r\right)\epsilon_r E_y \approx 0$$

to resemble that in [40, (2.1.7)].

Again, if we assume that

$$E_y = Y(x, y, z)e(x, z), \tag{6.9.18}$$

where $\partial Y/\partial x \simeq 0$ and $\partial Y/\partial z \simeq 0$, then, motivated by the separation of variables, we let

$$\frac{\partial^2 Y}{\partial y^2} + \frac{\partial}{\partial y}\left[\left(\frac{\partial}{\partial y}\ln\epsilon_r\right)Y\right] + k_0^2\epsilon_r Y \approx k_0^2 n^2 Y. \tag{6.9.19}$$

The presence of the extra term in the equation for TM polarization compared to that for the TE polarization is because in the TM polarization, the electric field has a component normal to the interface that induces polarization charges at the interface.

The above could be rewritten as

$$\frac{\partial}{\partial y}\frac{1}{\epsilon_r}\frac{\partial}{\partial y}\epsilon_r Y + k_0^2\epsilon_r Y \approx k_0^2 n^2 Y. \tag{6.9.20}$$

For a fixed x and z, it is a one dimensional eigen-equation for the propagation of TM eigenmodes in a layered medium. Here, $k_0^2 n^2$ is the eigenvalue of the problem. It is seen that this effective index n now is function of x and z. If ϵ_r is a slowly varying function of x and z, so would the index n.

Consequently, the equation governing $e(x, z)$ is

$$\left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial z^2}\right)e(x, z) + k_0^2 n^2(x, z)e(x, z) = 0, \tag{6.9.21}$$

where $n(x, z)$ is the effective index obtained by solving (6.9.20). Again, (6.9.21) is now reduced to a two-dimensional equation.

The effective index method is equivalent to replacing the y variation of the field with a single-mode approximation, and the propagation of this single mode in the xz plane is governed by the two-dimensional equations (6.9.14) and (6.9.21), for different polarizations.

6.10 The Beam-Propagation Method

When the inhomogeneity of a waveguide is weakly varying, an efficient method of deriving a solution is to use the beam-propagation method. It was proposed by Fleck, Morris, and Feit [41] first for atmospheric wave propagation where the refractive index is often tenous. Later, it was adapted for optical waveguide analysis [42,43]. Consider a scalar wave equation governed by

$$\left[\nabla_s^2 + \frac{\partial^2}{\partial z^2} + k_0^2 n^2(x, y, z)\right]\phi(x, y, z) = 0, \tag{6.10.1}$$

where $\nabla_s^2 = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2}$. Assuming that $n(x, y, z)$ is a weak or slowly varying function of space, or (x, y, z), then the above equation can be factorized as

$$\left(\frac{\partial}{\partial z} + i\sqrt{k_0^2 n^2 + \nabla_s^2}\right)\left(\frac{\partial}{\partial z} - i\sqrt{k_0^2 n^2 + \nabla_s^2}\right)\phi(\mathbf{r}) \approx 0. \tag{6.10.2}$$

It is to be noted that two new concepts are embedded in the above expression. A function of an operator such as $f(\nabla_s^2)$ is also regarded as an operator, and it has meaning only when it operates on a function which is the eigenfunction of ∇_s^2. An eigenfunction of the ∇_s^2 operator is $e^{ik_x x + ik_y y}$ since $\nabla_s^2 e^{ik_x x + ik_y y} = -k_s^2 e^{ik_x x + ik_y y}$ where $-k_s^2 = -k_x^2 - k_y^2$ is the eigenvalue. A function can always be approximately by a Taylor series such as $f(x) = f(0) + xf'(0) + \frac{1}{2}x^2 f''(0) + \ldots$ assuming that $f(x)$ is analytic at $x = 0$. Then, using its Taylor series expansion,

$$\begin{aligned} f(\nabla_s^2)e^{ik_x x + ik_y y} &= \left(f(0) + \nabla_s^2 f'(0) + \frac{1}{2}\nabla_s^4 f''(0) + \ldots\right)e^{ik_x x + ik_y y} \\ &= \left(f(0) + \nabla_s^2 f'(0) + \frac{1}{2}\nabla_s^4 f''(0) + \ldots\right)e^{ik_x x + ik_y y} \\ &= \left(f(0) - k_s^2 f'(0) + \frac{1}{2}k_s^4 f''(0) + \ldots\right)e^{ik_x x + ik_y y} \\ &= f(-k_s^2)e^{ik_x x + ik_y y} \end{aligned} \tag{6.10.3}$$

So in general

$$f(\mathcal{A})\mathbf{v}_i = f(\lambda_i)\mathbf{v}_i \tag{6.10.4}$$

where $\mathbf{v}_i$ is an eigenvector of the operator $\mathcal{A}$ with eigenvalue λ_i.

Another concept is the commutativity of operators. In general,

$$(\mathcal{A} + \mathcal{B})(\mathcal{A} - \mathcal{B}) = \left(\mathcal{A}^2 - \mathcal{B}^2\right) \tag{6.10.5}$$

only if $\mathcal{AB} = \mathcal{BA}$, or only if $\mathcal{A}$ and $\mathcal{B}$ commute. The commutator of $\mathcal{A}$ and $\mathcal{B}$ is defined as $[\mathcal{A}, \mathcal{B}] = \mathcal{AB} - \mathcal{BA}$. Hence, if $\mathcal{A}$ and $\mathcal{B}$ commute, their commutator is zero.

But in the above (6.10.2), the operator $\frac{\partial}{\partial z}$ and $\sqrt{k_0^2 n^2 + \nabla_s^2}$ do not commute because n is a function of z. Hence, the above is not an exact factorization, but is a good approximation if n is a slowly varying function of z.

Therefore, a solution to

$$\frac{\partial}{\partial z}\phi(\mathbf{r}) = i\sqrt{k_0^2 n^2 + \nabla_s^2}\phi(\mathbf{r}) = i\wp\phi(\mathbf{r}) \tag{6.10.6}$$

is also an approximate solution to (6.10.1). In the above, $\wp = \sqrt{k_0^2 n^2 + \nabla_s^2}$ is to be interpreted as an operator.

Here, (6.10.6) is also a one-way wave equation as it describes the propagation of the wave in one direction only. Hence, multiply reflected wave is not accounted for in the above approximation.

Equation (6.10.6) is not any easier to solve compared to (6.10.1). To simplify it, we need to make a paraxial approximation. This assumes that

$$|\nabla_s^2 \phi| \ll |k_0^2 n^2 \phi|, \tag{6.10.7}$$

or that the transverse variation of ϕ is much smaller than its longitudinal variation. In other words, the wave is propagating almost parallel to the axis of the waveguide. If we let $n = n_0 + \delta n$, then

$$\wp \cong \sqrt{k_0^2 n_0^2 + \nabla_s^2} \left(1 + \frac{k_0^2 n_0 \delta n}{k_0^2 n_0^2 + \nabla_s^2} + \dots\right). \tag{6.10.8}$$

With the assumption (6.10.7), we can approximate (6.10.8) as

$$\wp \cong \sqrt{k_0^2 n_0^2 + \nabla_s^2} + k_0 \delta n. \tag{6.10.9}$$

The approximation (6.10.9) is judiciously tailored so that the first term is independent of space, and the second term is independent of the operator ∇_s.

Since $\wp$ is a function of z, Equation (6.10.6) still cannot be solved easily. However, if $\wp$ is assumed to be independent of z within a small Δz, then, we can write the solution to (6.10.6) as

$$\begin{aligned} \phi(x, y, z + \Delta z) &= e^{i\wp \Delta z} \phi(x, y, z) \\ &\cong e^{ik_0 \delta n \Delta z + i\sqrt{k_0^2 n_0^2 + \nabla_s^2} \Delta z} \phi(x, y, z) \\ &\cong e^{ik_0 \delta n \Delta z} e^{i\sqrt{k_0^2 n_0^2 + \nabla_s^2} \Delta z} \phi(x, y, z). \end{aligned} \tag{6.10.10}$$

The last expression is an approximation because in general, $\exp(\mathcal{A} + \mathcal{B}) \neq \exp(\mathcal{A}) + \exp(\mathcal{B})$ unless $\mathcal{A}$ and $\mathcal{B}$ commute. It can be verified by representing these operators with their Taylor series expansions.

Now, for a fixed z, using Fourier expansion, we can write

$$\phi(x, y, z) = \frac{1}{(2\pi)^2} \iint_{-\infty}^{\infty} d\mathbf{k}_s e^{i\mathbf{k}_s \cdot \mathbf{r}_s} \tilde{\phi}(\mathbf{k}_s, z) \tag{6.10.11}$$

where $\mathbf{k}_s = \hat{x} k_x + \hat{y} k_y$, $\mathbf{r}_s = \hat{x} x + \hat{y} y$ and $e^{i\mathbf{k}_s \cdot \mathbf{r}_s}$ is an eigenfunction of the ∇_s^2 operator. By so doing, we have expanded $\phi(x, y, z)$ as a linear superposition or integral summation of the eigenfunctions of ∇_s^2 operator. Substituting (6.10.11) into (6.10.10), we have

$$\phi(x, y, z + \Delta z) = \frac{1}{(2\pi)^2} e^{ik_0 \delta n \Delta z} \iint_{-\infty}^{\infty} d\mathbf{k}_s e^{i\mathbf{k}_s \cdot \mathbf{r}_s} e^{i\sqrt{k_0^2 n_0^2 - k_s^2} \Delta z} \tilde{\phi}(\mathbf{k}_s, z). \tag{6.10.12}$$

Equation (6.10.12) is the fundamental equation of the beam-propagation method. To implement it, one first takes the field $\phi(x, y, z)$ at a $z =$ constant plane and Fourier transform it to get $\tilde{\phi}(\mathbf{k}_s, z)$. Then one multiplies the result by a plane-wave propagator $e^{i\sqrt{k_0^2 n_0^2 - k_s^2} \Delta z}$ in the Fourier space. Next, a Fourier inverse transform is performed on the propagated result. Subsequently, the field at each (x, y) location is added a phase of $k_0 \delta n \Delta z$ to yield the field at $\phi(x, y, z + \Delta z)$.

Alternatively, one can write

$$\sqrt{k_0^2 n_0^2 - k_s^2} = k_0 n_0 - \frac{k_s^2}{\sqrt{k_0^2 n_0^2 - k_s^2} + k_0 n_0}, \tag{6.10.13}$$

where the second term is much smaller than the first term if the wave is paraxial. Therefore, we can let

$$\phi(x, y, z) = w(x, y, z) e^{ik_0 n_0 z}, \tag{6.10.14}$$

and the beam-propagation equation for $w(x, y, z)$ is then

$$w(x, y, z + \Delta z) = \frac{1}{(2\pi)^2} e^{ik_0 \delta n \Delta z} \iint_{-\infty}^{\infty} d\mathbf{k}_s e^{i\mathbf{k}_s \cdot \mathbf{r}_s} e^{ip(k_s)\Delta z} \tilde{\phi}(\mathbf{k}_s, z), \tag{6.10.15}$$

where

$$p(k_s) = \frac{k_s^2}{\sqrt{k_0^2 n_0^2 - k_s^2} + k_0 n_0}. \tag{6.10.16}$$

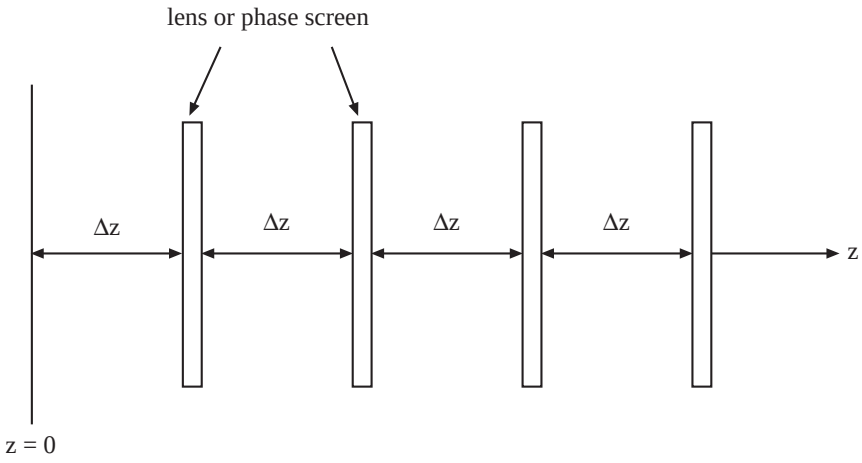


Figure 6.29: The physical interpretation of the beam-propagation method.

The Fourier transform and inverse transform in the beam-propagation method can be implemented efficiently using fast Fourier transform (FFT) which requires $O(N \log N)$ floating-point operations. Hence, albeit approximate, the beam-propagation method can be implemented efficiently. As it solves a first order equation (6.10.6), it only accounts for the forward component of the wave, but no reflections.

The physical interpretation of (6.10.12) or (6.10.15) is that the wave is first propagated through a homogeneous space using a homogeneous-space propagator. Then, the wave is passed through a lens or a phase screen which alters the phase of the wave at each (x, y) position. Therefore, Equation (6.10.12) or (6.10.15) represents physically the concatenation of a series of lenses or phase screens in a homogeneous space.

6.11 Ray Tracing Method

Multimode fibers correspond to the case where the core size is much larger than the wavelength. Hence, the fiber can be engineered by assuming that the light is a ray that bounces around in the fiber. Therefore, ray-tracing method can be used to solve the optical fiber problem. This method is described by Okoshi [5] and Senior [1] and many references therein.

The vector wave equation for an inhomogeneous medium can be written as

$$\nabla^2\mathbf{E} - \nabla(\nabla \ln \epsilon_r \cdot \mathbf{E}) + k_0^2\epsilon_r\mathbf{E} = 0. \tag{6.11.1}$$

The above equation can be approximated when the frequency is high, so that the wavelength of the wave is much smaller than the lengthscale of the variation of the inhomogeneity. Then we can assume that $\mathbf{E}(\mathbf{r})$ resembles a plane wave locally. In other words, $\mathbf{E}(\mathbf{r})$ could be more aptly described with a ray picture. In this case,

$$\mathbf{E}(\mathbf{r}) = \mathbf{e}(\mathbf{r})e^{ik_0\phi(\mathbf{r})}, \tag{6.11.2}$$

where $\phi(\mathbf{r})$ and $\mathbf{e}(\mathbf{r})$ are slowly varying but $e^{ik_0\phi(\mathbf{r})}$ is rapidly varying when $k_0 \to \infty$. Therefore,

$$|\nabla\mathbf{E}(\mathbf{r})| \sim |k_0\mathbf{E}(\mathbf{r})|, \qquad |\nabla^2\mathbf{E}(\mathbf{r}) \sim |k_0^2\mathbf{E}(\mathbf{r})|, \tag{6.11.3}$$

when $k_0 \to \infty$. Hence, Equation (6.11.1) can be approximated by

$$\nabla^2\mathbf{E}(\mathbf{r}) + k_0^2\epsilon_r\mathbf{E}(\mathbf{r}) = 0. \tag{6.11.4}$$

The above implies that the polarization term in (6.11.1) is unimportant in the high-frequency limit if the postulated form for $\mathbf{E}(\mathbf{r})$ given by (6.11.2) is true. However, if there is a jump discontinuity in ϵ_r, the polarization term could still be important, as its derivatives give rise to singular terms. However, if ϵ_r is a slowly varying function, it can be safely ignored.

Taking the Laplacian of $\mathbf{E}$, we have

$$\nabla^2\mathbf{E}(\mathbf{r}) \simeq \mathbf{e}(\mathbf{r})\nabla^2 e^{ik_0\phi(\mathbf{r})} = -k_0^2(\nabla\phi)^2\mathbf{E}(\mathbf{r}), \tag{6.11.5}$$

after ignoring higher order terms. Using (6.11.5) in (6.11.4) yields

$$(\nabla\phi)^2 = \epsilon_r = n^2 \tag{6.11.6}$$

or

$$|\nabla\phi| = n. \tag{6.11.7}$$

$\nabla\phi$ is the direction in which the phase in the wave in (6.11.2) is varying most rapidly. Hence, it is the direction of the wave in (6.11.2), and also the direction at which the ray is pointing. A unit vector along the ray direction is given by

$$\hat{s} = \nabla\phi/|\nabla\phi| = \nabla\phi/n. \tag{6.11.8}$$

Defining a point in a ray by the position vector $\mathbf{r}$, and that the ray is parametrized by the variable s which is the path length along a ray, then $\hat{s} = d\mathbf{r}/ds$, and (6.11.8) becomes

$$n\hat{s} = n\frac{d\mathbf{r}}{ds} = \nabla\phi. \tag{6.11.9}$$

Also, taking the gradient of (6.11.6) gives

$$2(\nabla\nabla\phi)\cdot\nabla\phi = 2n\nabla n. \tag{6.11.10}$$

After making use of (6.11.9), the above becomes

$$(\nabla\nabla\phi)\cdot\hat{s} = \nabla n. \tag{6.11.11}$$

Furthermore,

$$\begin{aligned}\frac{d}{ds}(n\hat{s}) = \frac{d}{ds}(\nabla\phi) &= \left(\frac{dx}{ds}\frac{\partial}{\partial s} + \frac{dy}{ds}\frac{\partial}{\partial y} + \frac{dz}{ds}\frac{\partial}{\partial z}\right)\nabla\phi \\ &= \frac{d\mathbf{r}}{ds}\cdot\nabla\nabla\phi = \hat{s}\cdot\nabla\nabla\phi.\end{aligned} \tag{6.11.12}$$

But $\hat{s}\cdot\nabla\nabla\phi = (\nabla\nabla\phi)\cdot\hat{s}$. Hence, from (6.11.11) and (6.11.12), one gets

$$\frac{d}{ds}(n\hat{s}) = \nabla n, \qquad \frac{d}{ds}\left(n\frac{d\mathbf{r}}{ds}\right) = \nabla n, \tag{6.11.13}$$

which are the fundamental equations of ray tracing.

In addition to the above equations, if one assumes that

$$\mathbf{H}(\mathbf{r}) = \mathbf{h}(\mathbf{r})e^{ik_0\phi(\mathbf{r})}, \tag{6.11.14}$$

it can be shown easily from Maxwell's equations that when $k_0 \to \infty$,

$$\nabla\phi\times\mathbf{h}(\mathbf{r}) \simeq c\epsilon\mathbf{e}(\mathbf{r}), \tag{6.11.15a}$$

$$\nabla\phi\times\mathbf{e}(\mathbf{r}) \simeq -c\mu\mathbf{h}. \tag{6.11.15b}$$

Hence,

$$\nabla\phi\cdot\mathbf{e} \simeq \nabla\phi\cdot\mathbf{h} \simeq 0. \tag{6.11.16}$$

The above equations indicate that the wave is locally a plane wave in the high frequency limit.

6.11.1 Ray Tracing Equations in an Optical Fiber

In an optical fiber with axial symmetry and uniformity in the z direction, then $dn/d\phi = 0$, and $dn/dz = 0$. A point in a ray can be described by

$$\mathbf{r} = \hat{\rho}\rho + \hat{z}z, \tag{6.11.17}$$

in cylindrical coordinates. In the above, $\hat{\rho}$ is a function of ϕ. Hence, $\mathbf{r}$ is a function of (ρ,ϕ,z). Extracting the $\hat{\rho}$ component of (6.11.13) gives

$$\hat{\rho}\cdot\frac{d}{ds}\left(n\frac{d\mathbf{r}}{ds}\right) = \frac{d}{d\rho}n. \tag{6.11.18}$$

But

$$\frac{d\mathbf{r}}{ds} = \frac{d}{ds}(\hat{\rho}\rho + \hat{z}z) = \hat{\rho}\frac{d\rho}{ds} + \rho\frac{d\hat{\rho}}{ds} + \hat{z}\frac{dz}{ds}, \tag{6.11.19}$$

and

$$\frac{d\hat{\rho}}{ds} = \frac{d}{ds}(\hat{x}\cos\phi + \hat{y}\sin\phi) = -\hat{\phi}\frac{d\phi}{ds}. \tag{6.11.20}$$

Therefore,

$$\hat{\rho}\cdot\frac{d}{ds}\left(n\frac{d\mathbf{r}}{ds}\right) = \hat{\rho}\cdot\frac{d}{ds}\left(n\hat{\rho}\frac{d\rho}{ds} - n\hat{\phi}\rho\frac{d\phi}{ds}\right). \tag{6.11.21}$$

However,

$$\frac{d\hat{\phi}}{ds} = \frac{d}{ds}(\hat{x}\sin\phi - \hat{y}\cos\phi) = \hat{\rho}\frac{d\phi}{ds}. \tag{6.11.22}$$

Consequently,

$$\hat{\rho}\cdot\frac{d}{ds}\left(n\frac{d\mathbf{r}}{ds}\right) = \frac{d}{ds}\left(n\frac{d\rho}{ds}\right) - n\rho\left(\frac{d\phi}{ds}\right)^2 = \frac{d}{d\rho}n. \tag{6.11.23}$$

Similarly, the $\hat{\phi}$ component of (6.11.13) can be extracted to obtain

$$\begin{aligned}\hat{\phi}\cdot\frac{d}{ds}\left(n\frac{d\mathbf{r}}{ds}\right) &= \hat{\phi}\cdot\frac{d}{ds}\left(n\hat{\rho}\frac{d\rho}{ds} - n\hat{\phi}\rho\frac{d\phi}{ds}\right)\\ &= -n\left(\frac{d\rho}{ds}\right)\left(\frac{d\phi}{ds}\right) - \frac{d}{ds}\left(n\rho\frac{d\phi}{ds}\right) = 0.\end{aligned} \tag{6.11.24}$$

Extracting the $\hat{z}$ component of (6.11.13) yields

$$\frac{d}{ds}\left(n\frac{dz}{ds}\right) = 0. \tag{6.11.25}$$

Equation (6.11.25) can be readily integrated to yield

$$ds = \frac{n}{n_0\cos\theta_i}dz. \tag{6.11.26}$$

where n_0 is the value of n at the initial point, and θ_i is the angle of the ray with the z-axis initially. Replacing n/ds in (6.11.24) with (6.11.26) gives

$$\frac{d\rho}{ds}\frac{d\phi}{dz} + \frac{d}{ds}\left(\rho\frac{d\phi}{dz}\right) = 0, \tag{6.11.27}$$

which can be rewritten as

$$\frac{d}{ds}\left(\rho^2\frac{d\phi}{dz}\right) = 0. \tag{6.11.28}$$

Equation (6.11.28) can be integrated to yield

$$\rho^2\frac{d\phi}{dz} = C_0. \tag{6.11.29}$$

To integrate (6.11.23), one multiplies it by n and use (6.11.26) to replace n/ds with $n_0 \cos\theta_i/dz$ to obtain

$$\frac{d^2}{dz^2}\rho - \rho\left(\frac{d\phi}{dz}\right)^2 = \frac{1}{2n_0^2\cos^2\theta_i}\frac{d}{d\rho}n^2. \tag{6.11.30}$$

Using (6.11.29) for $\frac{d\phi}{dz}$ yields

$$\frac{d^2}{dz^2}\rho = \frac{1}{\rho^3}C_0^2 + \frac{1}{2n_0^2\cos^2\theta_i}\frac{d}{d\rho}n^2. \tag{6.11.31}$$

The above equation can be integrated with respect to ρ. The left-hand side is

$$\begin{aligned}\int_{\rho_0}^{\rho}\frac{d^2}{dz'^2}\rho' d\rho' &= \int_0^z \frac{d^2\rho'}{dz'^2}\frac{d\rho'}{dz'}dz' = \frac{1}{2}\int_0^z \frac{d}{dz'}\left(\frac{d\rho'}{dz'}\right)^2 dz' \\ &= \frac{1}{2}\left(\frac{d\rho}{dz}\right)^2 - \frac{1}{2}D_0.\end{aligned} \tag{6.11.32}$$

where $D_0 = \left(\frac{d\rho}{dz}\right)^2_{z=0}$. The right hand side of (6.11.31) yields

$$\begin{aligned}&\left(-\frac{1}{2\rho^2}C_0^2 + \frac{1}{2n_0^2\cos^2\theta_i}n^2\right)\Bigg|_{\rho_0}^{\rho} \\ &= \left[1-\left(\frac{\rho_0}{\rho}\right)^2\right]\frac{C_0^2}{2\rho_0^2} + \frac{1}{2\cos^2\theta_i}\left(\frac{n^2}{n_0^2}-1\right).\end{aligned} \tag{6.11.33}$$

As a result, one gets

$$\left(\frac{d\rho}{dz}\right)^2 = \left[1-\left(\frac{\rho_0}{\rho}\right)^2\right]\frac{C_0^2}{\rho_0^2} + \frac{1}{\cos^2\theta_i}\left(\frac{n^2}{n_0^2}-1\right) + D_0, \tag{6.11.34}$$

or

$$z = \int_{\rho_0}^{\rho} d\rho \left\{\left[1-\left(\frac{\rho_0}{\rho}\right)^2\right]\left(\frac{C_0}{\rho_0}\right)^2 + \frac{1}{\cos^2\theta_i}\left(\frac{n^2}{n_0^2}-1\right) + D_0\right\}^{-\frac{1}{2}}. \tag{6.11.35}$$

Equation (6.11.35) is the basic equation for computing the ray path in an optical fiber, given the initial condition C_0, D_0, $\cos\theta_i$, n_0 and ρ_0 at $z = 0$.

6.11.2 Determination of Initial Conditions

One can assume that an optical ray enters an optical fiber at $z = 0$ at (x_0, y_0) and the direction of the ray is pointed at $\hat{s}_0$, where

$$\hat{s}_0 = \hat{x}\sin\theta_i\cos\phi_i + \hat{y}\sin\theta_i\sin\phi_i + \hat{z}\cos\theta_i. \tag{6.11.36}$$

If $d\phi/dz$ at $(x_0, y_0, z = 0)$ of the ray can be found, then C_0 in (6.11.29) can be found. The $\hat{\phi}_0$ component of $\hat{s}_0$ can be found by $\hat{s}_0 \cdot \hat{\phi}_0$ which is

$$\begin{aligned} \hat{s}_0 \cdot \hat{\phi}_0 &= \hat{s}_0 \cdot [-\hat{x} \sin\phi_0 + \hat{y} \cos\phi_0] \\ &= -\sin\phi_0 \sin\theta_i \cos\phi_i + \cos\phi_0 \sin\theta_i \sin\phi_i. \end{aligned} \tag{6.11.37}$$

Hence,

$$\rho_0 d\phi = \hat{s}_0 \cdot \hat{\phi}_0 ds = \hat{s}_0 \cdot \hat{\phi}_0 / \cos\theta_i dz. \tag{6.11.38}$$

Therefore,

$$C_0 = \rho_0 \frac{\hat{s}_0 \cdot \hat{\phi}_0}{\cos\theta_i}. \tag{6.11.39}$$

If $d\,\rho\,dz$ of the ray at $(x_0, y_0, z = 0)$ is known, then D_0 in (6.11.32) is known. But

$$d\rho = \hat{s}_0 \cdot \hat{\rho}_0 ds = \frac{\hat{s}_0 \cdot \hat{\rho}_0}{\cos\theta_i} dz. \tag{6.11.40}$$

Therefore,

$$\begin{aligned} D_0 = \left(\frac{d\rho}{dz}\right)^2_{z=0} = \frac{(\hat{s}_0 \cdot \hat{\rho}_0)^2}{\cos^2\theta_i} &= \frac{\sin^2\theta_i (\cos\phi_i \cos\phi_0 + \sin\phi_i \sin\phi_0)^2}{\cos^2\theta_i} \\ &= \tan^2\theta_i \cos^2(\phi_i - \phi_0) \end{aligned} \tag{6.11.41}$$

The above provides sufficient initial conditions to launch a ray pointing at $\hat{s}_0$ from the point $(x_0, y_0, z = 0)$. A ray which is only propagating radially, i.e., $\hat{s}_0 \cdot \hat{\phi}_0 = C_0 = 0$, is known as a meridional ray. A ray that propagates in both the $\hat{\rho}$ and $\hat{\phi}$ directions is known as a skew ray. A skew ray that propagates at a constant distance from the fiber axis is called a helical ray.

The ray equation (6.11.35) can be used to derive index profile so that the axial velocity of a ray is independent of the launch condition. In this manner, the dispersion of the fiber will be minimized.

Exercises for Chapter 6

Problem 6-1:

Show that the guidance condition for a dielectric slab of thickness d, permittivity ϵ_1 and permeability μ_1 suspended in air can be simplified to

$$\alpha_{0x}\frac{d}{2} = \frac{\mu_0}{\mu_1}k_{1x}\frac{d}{2}\tan\left(\frac{k_{1x}d - m\pi}{2}\right),$$

for the TE case. In the above, $\alpha_{0x} = \sqrt{k_z^2 - k_0^2}$, $k_{1x} = \sqrt{k_1^2 - k_z^2}$ and m is an integer. Plot the right and the left hand side of the equation as a function of $k_{1x}d$ for m even and m odd to obtain graphical solutions to the above equation.

Problem 6-2: Find the phase velocity of the TM_1 mode of a symmetric dielectric slab waveguide at cutoff. By analyzing the phase velocity in the vicinity of cutoff, find the group velocity analytically. Explain the answers.

Problem 6-3: For a circular dielectric waveguide of radius a, find the cutoff frequencies of the TE_{01} mode, the EH_{11} mode and the HE_{12} mode. Which is the next higher order mode to the HE_{11} mode? If $a = 10\mu$, $n_1 = 1.6$ and $n_2 = 1.5$, what is the bandwidth for single mode propagation in the optical fiber? Explain why the usable bandwidth of an optical fiber is not this bandwidth.

Problem 6-4: Find the phase velocity of the TE_{01} mode of an optical fiber near cutoff. Also, find the group velocity near cutoff analytically. Explain what you have observed about the answer.

Problem 6-5:

(a) Show that the ratio of E_z to H_z in the core region of an optical fiber is given by

$$H_1/E_1 = \frac{-nk_z}{i\omega}\left(\frac{1}{(k_{1\rho}a)^2} + \frac{1}{(\alpha_2 a)^2}\right) \left(\frac{\mu_1 J_n'(k_{1\rho}a)}{k_{1\rho}aJ_n(k_{1\rho}a)} + \frac{\mu_2 K_n'(\alpha_2 a)}{\alpha_2 aK_n(\alpha_2 a)}\right)^{-1}$$

(b) Using the equation for the guidance conditions of the EH and HE modes, show that the ratio E_1/H_1 is in fact larger for the HE modes compared to the EH modes.

Problem 6-6:

(a) Starting with the equation for the guidance condition of the modes in a step-index fiber (nonweakly guiding), by assuming that $\mu_1 = \mu_2$ and that $\epsilon_1 \approx \epsilon_2$, show that the guidance condition for the weak contrast optical fiber mode can be derived as

$$\frac{k_{1\rho}J_{n\pm1}(k_{1\rho}a)}{J_n(k_{1\rho}a)} = \pm\frac{\alpha_2 K_{n\pm1}(\alpha_2 a)}{K_n(\alpha_2 a)}$$

Show that the above two equations are equivalent, and hence the modes they define are degenerate.

(b) From the derivation, which of the EH mode is degenerate with the HE mode when the contrast of the fiber tends to zero?

(c) Even though in the weak contrast fiber approximation, the equation indicates that the x component and the y component of the electric field are decoupled, they are actually weakly coupled in this limit. Hence, it is not possible to conceive some modes to have E_x only or E_y only. The LP_{11} mode is such a mode. To test your physical insight, sketch the electric field on the xy plane on the cross section of a weak contrast optical fiber.

Problem 6-7: Write a computer program to solve for the k_z of the HE_{11}, HE_{21}, TE_{01}, TM_{01}, EH_{11}, HE_{31} and HE_{12} modes. That is, produce the dispersion curve for the first three families of modes shown in Figure 6.2.5 of the text. (a) First, generate the dispersion curves when $n_1/n_2 = 1.5$. (b) Second, generate the dispersion curves when $n_1/n_2 = 1.01$.

Problem 6-8: Using $1/\omega$ as a small parameter, expand $\mathbf{E}_s$ and E_z as perturbation series in a weak contrast optical fiber, and show that $E_z \sim |E_s|/\omega$.

Problem 6-9: Explain why the Rayleigh scattering loss and ultraviolet absorption loss diminish with wavelength in Figure 6.17, while the infrared absorption loss increases with wavelength. Is it reasonable to assume that waveguide imperfection loss does not alter with wavelength?

Problem 6-10: For the harmonic expansion method, why is the assumption of standing wave inside the waveguide and outgoing wave outside the waveguide not a valid assumption in Subsection 6.6.1?

Problem 6-11:
A circular waveguide is loaded with a circular dielectric rod of diameter d as shown in the above figure. Find the change in the propagation constant k_z of the TE_{11} mode due to the presence of this rod using a perturbation calculation.

Problem 6-12: The vector wave equation governing the propagation of waves in an anisotropic medium with reflection symmetry can be shown to be

$$\hat{z} \times \overline{\mu}_s \cdot \hat{z} \times \nabla_s \times \mu_{zz}^{-1} \nabla_s \times \mathbf{E}_s + \nabla_s \epsilon_{zz}^{-1} \nabla_s \cdot \overline{\epsilon}_s \cdot \mathbf{E}_s - \omega^2 \hat{z} \times \overline{\mu}_s \cdot \hat{z} \times \overline{\epsilon}_s \cdot \mathbf{E}_s - k_z^2 \mathbf{E}_s = 0.$$

(a) Derive a variational expression for the guided wave number k_z^2.

(b) Using Rayleigh-Ritz procedure, derive a matrix equation for the guided wave numbers of the waveguide.

Problem 6-13:
The harmonic expansion method, in theory, can be used for dielectric waveguides of any shapes. However, when the waveguide has symmetry about the x and the y axes as shown in the above figure, the z components of the fields are either even or odd about the the x and the y axes.

(a) Proof that if E_z (H_z) is even about the x or the y axes, H_z (E_z) has to be odd about the x or the y axes, and vice versa.

(b) Because of this symmetry, we need not have to solve for the solution of the waveguide in the full space, but only in a quadrant of the full space. Use the harmonic expansion method, together with point matching, derive the guidance conditions for the modes for which E_z is even about the x axis and odd about the y axis. Give a reason why this method of solution is preferable.

Problem 6-14: Assume a parallel waveguide terminated abruptly so that the modes of the waveguide will radiate into free space. Repeat the derivation of Subsection 6.7.1 for the terminated parallel plate waveguide with a flange as shown.

Problem 6-15:

(a) In the WKB method, explain why the prefactor of $1/\sqrt{s_\rho}$ term is necessary for energy conservation.

(b) Repeat the analysis of Subsection 6.8.1 for a planar geometry. Is the result of Equation (6.8.18) much different compared to the planar geometry case? Explain why.

Problem 6-16:

A planar waveguide has a index profile as shown. It is terminated by a perfect electric conductor at the $x = 0$ surface. Use the WKB method, write down the guidance condition for such a waveguide.

Problem 6-17:

(a) If $\epsilon_r(z)$ in Equation (6.9.15) of Subsection 6.9.1 for fixed x, y is described by a symmetric dielectric slab. Find the transcendental equation from which the eigenvalues of Equation (6.9.15) can be found.

(b) Repeat the same for Equation (6.9.20) of Subsection 6.9.2.

Problem 6-18: In the beam-propagation method, the equation of propagation is sometimes written as

$$\phi(x, y, z + \Delta z) \cong e^{ik_0 \delta n \Delta z/2} e^{i\sqrt{k_0^2 n_0^2 + \nabla_s^2}\Delta z} e^{ik_0 \delta n \Delta z/2} \phi(x, y, z).$$

as opposed to Equation (6.10.10) of Section 10. Explain if there is any advantage of writing the propagation equation as shown above.

Problem 6-19: Describe how you would solve the ray-tracing equations (6.11.13) given in Section 6.11 numerically.

Problem 6-20: Write a computer program to compute Equation (6.11.35) of Subsection 6.11.1 for ray tracing in an optical fiber.

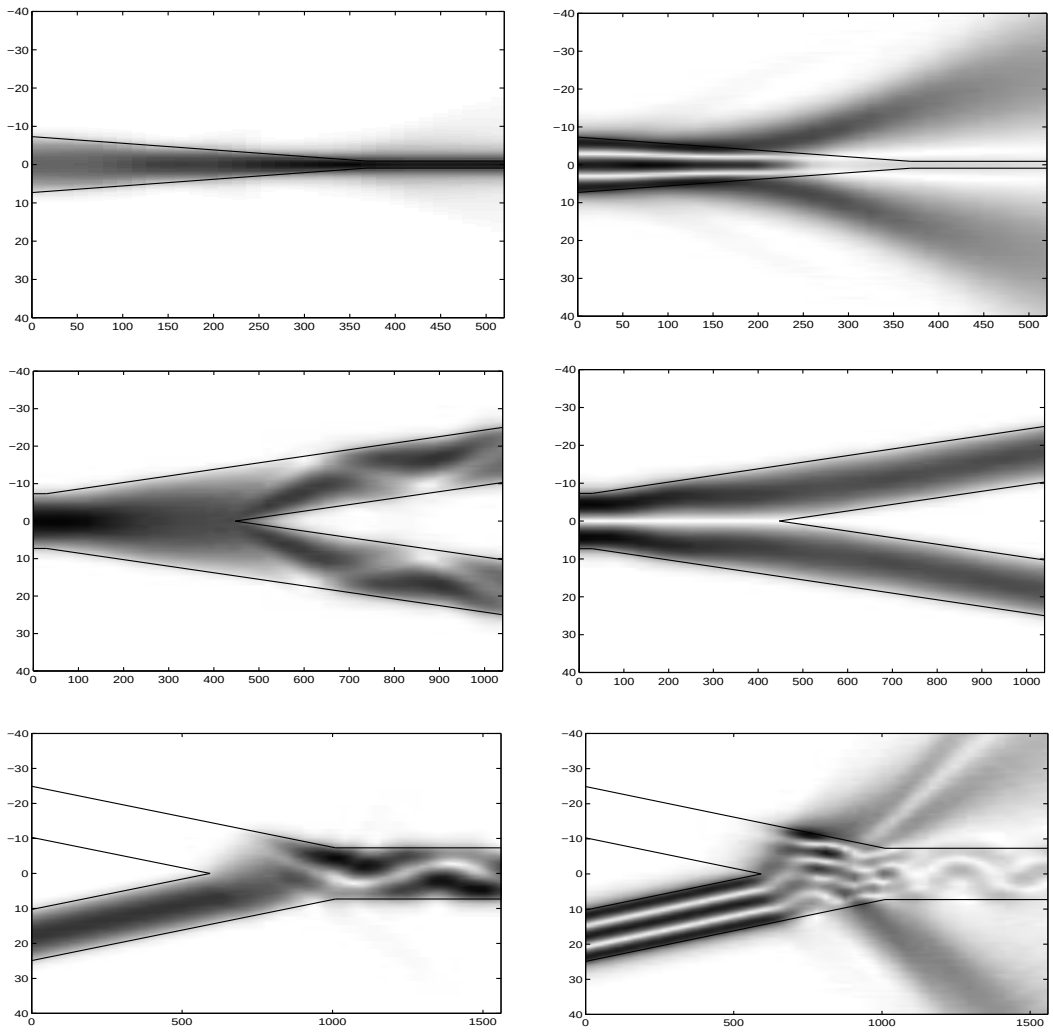


Figure 6.30: Examples of numerical simulations using the beam-propagation method for different kinds of optical waveguides. Some waveguide transitions give rise to much radiation loss, while some do not. Also, depending on the mode profile, the radiation at the waveguide transition is different. For instance, in the top figure, the fundamental mode (top left) couples more smoothly than the higher-order mode (top right) with two minima (Courtesy of F. Teixeira).

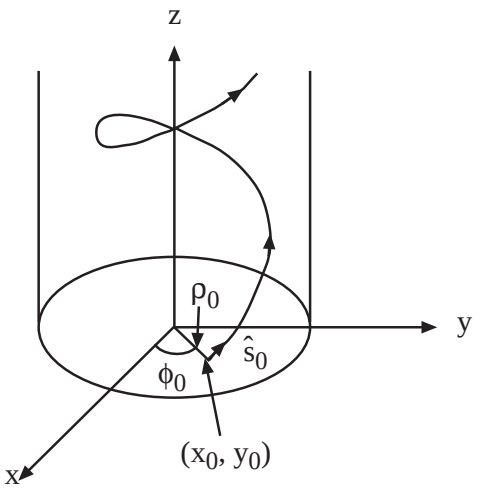


Figure 6.31: Launching of a skew ray pointing at $\hat{s}_0$ at $(x_0, y_0, z = 0)$.

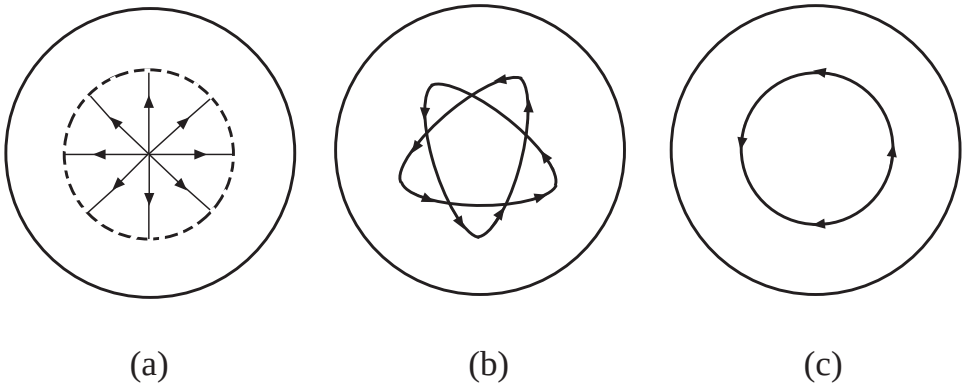


Figure 6.32: Cross-sectional pictures of (a) a meridional ray, (b) a complex skew ray, and (c) a skew ray that is helical.

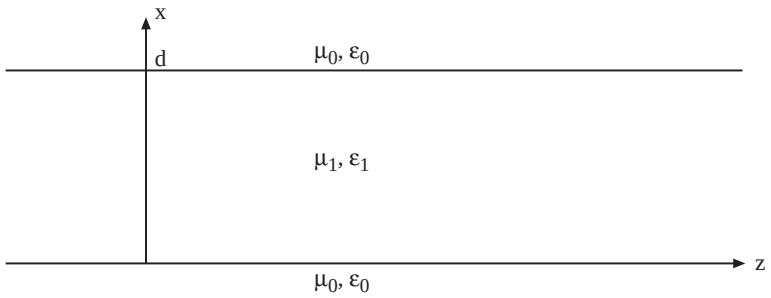


Figure 6.33: Problem 6-1

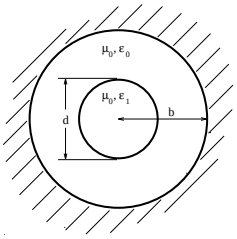


Figure 6.34: Problem 6-10

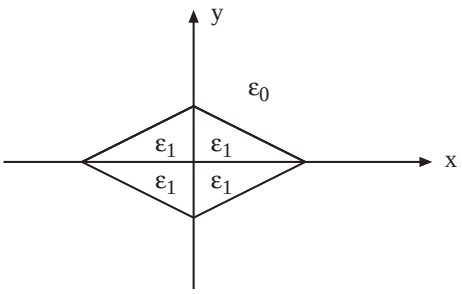


Figure 6.35: Problem 6-12

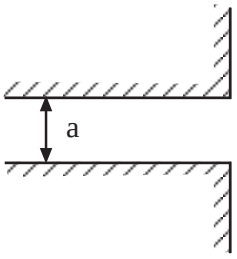


Figure 6.36: Problem 6-13

Figure 6.37: Problem 6-15

Bibliography

[1] J.M. Senior, *Optical Fiber Communications, Principles and Practice,* 2nd Edition, Publisher: Prentice Hall International, Hertfordshire, UK, 1992.

[2] G.P. Agrawal, *Fiber-Optic Communication Systems*, Wiley-Interscience, 3rd edition, 2002.

[3] P.K. Runge, "Undersea lightwave systems," *AT&T Technical Journal*, vol. 71, no. 1, pp. 5-13, 1992.

[4] K.C. Kao and G.A. Hockham, "Dielectric-fiber surface waveguides for optical frequencies," *Proc. IEE (London)*, 113, pp. 1151-1158, 1966.

[5] T. Okoshi, *Opitcal Fibers,* Academic Press, New York, 1982.

[6] J.E. Midwinter, *Optical Fibers for Transmission,* John Wiley & Sons, NY, 1979.

[7] D. Marcuse, Ed., *Integrated Optics*, IEEE Press, NY, 1973.

[8] D. Marcuse, *Theory of Optical Waveguides,* Academic Press, NY, 1974.

[9] H.A. Haus, *Waves and Fields in Optoelectronics,* Prentice Hall, 1983.

[10] A.W. Snyder and J.D. Love, *Optical Waveguide Theory,* London U.K: Chapman and Hall, 1983.

[11] A. Yariv, *Optical Electronics,* Holt, Rinehart, and Winston, New York, 1985.

[12] P.-A. Blanger, *Optical Fiber Theory a Supplement to Applied Electromagnetism,* World Scientific, 1993.

[13] S.L. Chuang, *Physics of Optoelectronic Devices,* Wiley-Interscience, 1995.

[14] K. Okamoto, *Fundamentals of Optical Waveguides,* Academic Press, 2000.

[15] J.A. Kong, *Electromagnetic Wave Theory,* EMW Publishing, Cambridge, MA, 2000.

[16] E. Snitzer, "Cylindrical dielectric waveguide modes," *J. Opt. Soc. Am.*, 51, pp. 491-498, 1961.

[17] A.W. Snyder, "Asymptotic expression for eigenfunctions and eigenvalues of dielectric optical waveguides," *IEEE Trans. Microwave Theory Tech.*, MTT-17, pp. 1130-1138, 1969.

[18] D. Gloge, "Weakly guiding fibers," *App. Opt.*, 10, 2252-2258, 1971.

[19] D. Gloge and E.A.J. Mercatili, "Multimode theory of graded-core fibers," *Bell Syst. Tech. J.*, 52, pp. 1563-1578, 1973.

[20] J.G. Dil and H. Blok, "Propagation of electromagnetic surface waves in a radially inhomogeneous optical waveguide," *Opto-Electronics (London)*, 5, pp. 415-428, 1973.

[21] M. Hashimoto, "A perturbation method for the analysis of wave propagation in inhomogeneous dielectric waveguides with perturbed media," *IEEE Trans. Microwave Theory and Tech.*, vol. 24, no. 9, pp. 559-566, 1976.

[22] H.A. Haus, *Electromagnetic Noise and Quantum Optical Measurements,* Berlin: Springer-Verlag, 2000.

[23] T. Miya, Y. Terunuma, T. Hosaka and T. Miyashita, "Ultimate low-loss single-mode fiber at 1.55 μm," *Electron Lett.*, v. 15, p. 106, 1979.

[24] J.E Goell, "A circular-harmonic computer analysis of rectangular dielectric waveguides," *Bell Syst. Tech. J.*, v. 48, pp. 2133-2160, 1969.

[25] W.C. Chew, *Waves and Fields in Inhomogeneous Media,* Van Nostrand Reinhold, New York, 1990, reprinted, IEEE Press, Piscataway, NJ, 1995.

[26] E.A.J. Mercatili, "Dielectric rectangular waveguide and directional coupler for integrated optics," *Bell Syst. Tech. J.*, 48, pp. 2071-2102, 1969.

[27] W. C. Chew and M. Nasir, "A variational analysis of anisotropic, inhomogeneous dielectric waveguides," *IEEE Trans. Microwave Theory Techniques*, vol. 37, no. 4, pp. 661-668, Apr. 1989.

[28] J. Jin, *The finite element method in electromagnetics*, John Wiley & Sons, Inc., New York, 1993.

[29] W.C. Chew, "Analysis of optical and millimeter wave dielectric waveguide," *J. Elect. Waves Appl.*, vol. 3, no. 4, pp. 359-377, 1989.

[30] K. Radhakrishnan and W.C. Chew, "An efficient Krylov subspace based algorithm to solve the dielectric waveguide problem," *IEEE Trans. Micro. Theory Tech.*, vol. 49, no. 7, pp. 1345-1347, July 2001.

[31] C.M. Angulo, "Diffraction of surface waves by a semi-infinite dielectric slab," *IRE Trans. Antennas Propag.*, vol. AP-7, pp. 261-274, 1957.

[32] T. Ikegami, "Reflectivity of mode of facet and oscillation mode in double heterostructure injection lasers," *IEEE J. Quantum Electron.*, QE-8,470-476, 1972.

[33] T. Rozzi, "Rigorous analysis of the step discontinuities in planar dielectric waveguides," *IEEE Trans. Microwave Theory Tech.*, vol. MTT-26, pp. 738-746, 1978.

[34] M. Pudensi and L. Ferreira, "Method to calculate the reflection and transmission of guided waves," *J. Opt. Soc. Am.*, 72, 126-130, 1982.

[35] Q.H. Liu and W.C. Chew, "Numerical mode-matching method for the multiregion, vertically stratified media," *IEEE Trans. Antennas Propag.*, vol. AP-38, no. 4, pp. 498-506, Apr. 1990.

[36] C. M. Herzinger, C.C. Lu, T.A. De Temple and W.C. Chew, "Semiconductor waveguide facet reflectivity problem," *IEEE J. Quantum Elec.*, 29, 2273-2281, 1993.

[37] R.M. Know and P.P. Toulios, "Integrated circuits for the millimeter through optical frequency range," *Symposium on Submillimeter Waves,* Polytechnic Institute of Brooklyn, pp. 497-516, 1970.

[38] T. Tamir, *Integrated Optics*, Springer-Verlag, Berlin, 1975.

[39] J. Buus, "The effective index method and its application to semiconductor lasers," *IEEE J. Quantum Electronics,* vol. 18, no. 7, pp. 1083-1089, Jul 1982.

[40] W.C. Chew, *Waves and Fields in Inhomogeneous Media,* Van Nostrand Reinhold, New York, 1990, reprinted, IEEE Press, Piscataway, NJ, 1995.

[41] J.A. Fleck, J.R. Morris, and M.D. Feit, "Time-dependent propagation of high energy laser beams through the atmosphere," *Appl. Phys.*, 10, pp. 129-160, 1976.

[42] M.D. Feit and J.A. Fleck, "Light propagation in graded-index optical fibers," *Appl. Opt.*, vol. 17, pp. 3990-3997, 1978.

[43] W.P. Huang and C.L. Xu, "Simulation of three-dimensional optical waveguides by a full-vector beam propagation method," *IEEE J. Quantum Electron.*, vol. 29, pp. 26392649, Oct. 1993.

[44] A. Hochman and Y. Leviatan, "Calculation of confinement losses in photonic crystal fibers by use of a source-model technique," *J. Opt. Soc. Am. B* , vol. 22, no. 2, p. 474, February 2005.

[45] T. Matsui, M. Ozaki, and K. Yoshino, "Tunable laser action in a dye-doped nematic liquid-crystal waveguide under holographic excitation based on electric-field-induced TM guided-mode modulation," *J. Opt. Soc. Am. B* , vol. 21, no. 9, p. 1651, September 2004.

[46] J.K.S. Poon, J. Scheuer, Y. Xu, and A. Yariv, "Designing coupled-resonator optical waveguide delay lines," *J. Opt. Soc. Am. B* , vol. 21, no. 9, p. 1665, September 2004.

[47] Y. Lu, S. Liu, G. Zhang, R. Guo, N. Zhu, and L. Yang, "Waveguides and directional coupler induced by white-light photovoltaic dark spatial solitons," *J. Opt. Soc. Am. B* , vol. 21, no. 9, p. 1674, September 2004.

[48] N. Malkova and V. Gopalan, "Resonant light propagation through 90-bend waveguide based on a strained two-dimensional photonic crystal," *J. Opt. Soc. Am. B* , vol. 21, no. 9, p. 1679, September 2004.

[49] C. Kappel, A. Selle, M.A. Bader, and G. Marowsky, "Resonant double-grating waveguide structures as inverted Fabry-Perot interferometers," *J. Opt. Soc. Am. B* , vol. 21, no. 6, p. 1127, June 2004.

[50] M. Mohebbi, "Dispersion of femtosecond laser pulses in hollow fibers," *J. Opt. Soc. Am. B* , vol. 21, no. 5, p. 893, May 2004.

[51] E. Miyai and S. Noda, "Structural dependence of coupling between a two-dimensional photonic crystal waveguide and a wire waveguide," *J. Opt. Soc. Am. B* , vol. 21, no. 1, p. 67, January 2004.

[52] P.E. Barclay, K. Srinivasan, and O. Painter, "Design of photonic crystal waveguides for evanescent coupling to optical fiber tapers and integration with high-Q cavities," *J. Opt. Soc. Am. B* , vol. 20, no. 11, p. 2274, November 2003.

[53] P. Bienstman, S. Assefa, S.G. Johnson, J.D. Joannopoulos, G.S. Petrich, and L.A. Kolodziejski, "Taper structures for coupling into photonic crystal slab waveguides," *J. Opt. Soc. Am. B* , vol. 20, no. 9, p. 1817, September 2003.

[54] O. Skorka, J. Salzman, and S. Zamir, "Coupled waveguides in GaN-based lasers," *J. Opt. Soc. Am. B* , vol. 20, no. 9, p. 1822, September 2003.

[55] T. Yang, Y. Sugimoto, S. Lan, N. Ikeda, Y. Tanaka, and K. Asakawa, "Transmission properties of coupled-cavity waveguides based on two-dimensional photonic crystals with a triangular lattice of air holes," *J. Opt. Soc. Am. B* , vol. 20, no. 9, p. 1922, September 2003.

[56] A. Zakery, Y. Ruan, A. V. Rode, M. Samoc, and B. Luther-Davies, "Low-loss waveguides in ultrafast laser-deposited As_2S_3 chalcogenide films," *J. Opt. Soc. Am. B* , vol. 20, no. 9, p. 1844, September 2003.

[57] J. Limeres, M. L. Calvo, J. M. Enoch, and V. Lakshminarayanan, "Light scattering by an array of birefringent optical waveguides: theoretical foundations," *J. Opt. Soc. Am. B* , vol. 20, no. 7, p. 1542, July 2003.

[58] A.M. Ljungstrm and T.M. Monro, "Exploration of self-writing and photosensitivity in ion-exchanged waveguides," *J. Opt. Soc. Am. B* , vol. 20, no. 6, p. 1317, June 2003.

[59] M. Huang and X. Yan, "Thermal-stress effects on the temperature sensitivity of optical waveguides," *J. Opt. Soc. Am. B* , vol. 20, no. 6, p. 1326, June 2003.

[60] Z.-Y. Li and K.-M. Ho, "Waveguides in three-dimensional layer-by-layer photonic crystals," *J. Opt. Soc. Am. B* , vol. 20, no. 5, p. 801, May 2003.

[61] S. Hadjiloucas, R.K.H. Galvão, J.W. Bowen, R. Martini, M. Brucherseifer, H.P.M. Pellemans, P.H. Bolvar, H. Kurz, J. Digby, G.M. Parkhurst, and J.M. Chamberlain, "Measurement of propagation constant in waveguides with wideband coherent terahertz spectroscopy," *J. Opt. Soc. Am. B* , vol. 20, no. 2, p. 391, February 2003.

[62] B.E.A. Saleh and M.C. Teich, *Fundamentals of photonics*, second edition, New York: Wiley, 2007.

Chapter 7

Microwave Integrated Circuits

Due to the advent of integrated circuits at microwave frequencies, microwave integrated circuits waveguides are omnipresent in microwave technologies. Their ease of fabrication, low cost, conformal nature have made them extremely popular. Moreover, they are easily integrated with other circuits. Some examples are shown in Figure 7.1. More recently, microwave integrated circuits have been used to study cavity QED (quantum electrodynamics) that plays an important role in quantum computing [1]. Hence, they are being used at the frontier of scientific investigations as well.

The microstrip line has a long history. Since its appearance before World War II, it has been continuously worked on. The early work uses quasi-TEM approximation which does not account for dispersive effects in the line [2–8]. The use of this waveguide at higher frequencies calls for the analysis accounting for dispersive effect [9–15,42]. Other works related to analysis of microstrip integrated circuits are [18–26]. Recent typical works in this area are given in [28–30].

7.1 Quasi-TEM Approximation

The integrated circuits waveguiding structures cannot support a TEM mode, for if they do, the phase matching condition will be violated at the interface between the inhomogeneities. This is because a TEM wave has the phase velocity of the medium in which the wave is traveling.

However, if the wavelength under consideration is much larger than the transverse structure of the waveguide, we can show that the fundamental mode of such a structure is almost TEM or quasi-TEM. A fundamental mode is the mode that is propagating when $\omega \to 0$ or $\lambda \to \infty$.

We can write Maxwell's equations by separating out the transverse and longitudinal components as

$$\nabla_s \times \mathbf{E}_s = ik\eta \mathbf{H}_z, \tag{7.1.1}$$

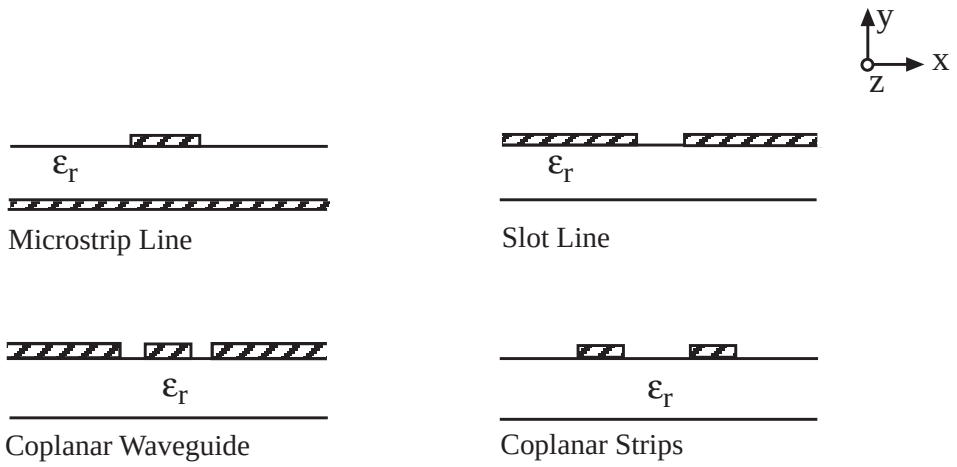


Figure 7.1: Different kinds of microwave integrated circuits waveguides.

$$\nabla_s \times \mathbf{H}_s = -ik\eta^{-1}\mathbf{E}_z, \tag{7.1.2}$$

$$\frac{\partial}{\partial z}\hat{z} \times \mathbf{E}_s - \hat{z} \times \nabla_s E_z = ik\eta\mathbf{H}_s, \tag{7.1.3}$$

$$\frac{\partial}{\partial z}\hat{z} \times \mathbf{H}_s - \hat{z} \times \nabla_s H_z = -ik\eta^{-1}\mathbf{E}_s, \tag{7.1.4}$$

where $\eta = \sqrt{\frac{\mu}{\epsilon}}$. For a structure whose dominant transverse dimension is much smaller than the wavelength, the transverse variation of the field would be more rapid than the longitudinal variation. The transverse variation of the field has to be fast enough for the field to match the boundary condition, namely, the x and y variations will vary on the length scale of δ. To emphasize this fact, we can perform a coordinate stretching transformation [31] by letting

$$x = \delta x', \qquad y = \delta y'. \tag{7.1.5}$$

Under such a coordinate stretching transformation,

$$\nabla_s \to \frac{1}{\delta}\nabla'_s, \tag{7.1.6}$$

Equations (7.1.3) and (7.1.4) become

$$ik_z\delta\hat{z} \times \mathbf{E}_s - \hat{z} \times \nabla'_s E_z = ik\eta\delta\mathbf{H}_s, \tag{7.1.7}$$

$$ik_z\delta\hat{z} \times \mathbf{H}_s - \hat{z} \times \nabla'_s H_z = -ik\eta^{-1}\delta\mathbf{E}_s, \tag{7.1.8}$$

where we have assumed $e^{ik_z z}$ dependence of the field. When $\delta/\lambda \to 0$, then, $k\delta \to 0$. Since $k_z = \sqrt{k^2 - k_s^2} < k$, we also expect $k_z\delta \to 0$, when $\delta/\lambda \to 0$. Therefore, in the long wavelength limit, from the above equations, it is seen that

$$(E_z, \eta H_z) \sim O(k\delta)(\mathbf{E}_s, \eta\mathbf{H}_s). \tag{7.1.9}$$

In other words, $|E_z| \ll |\mathbf{E}_s|$, $|H_z| \ll |\mathbf{H}_s|$, implying that the field is quasi-TEM. Consequently, we can write (7.1.1) and (7.1.2) as

$$\nabla_s \times \mathbf{E}_s = 0, \tag{7.1.10}$$

$$\nabla_s \times \mathbf{H}_s = 0. \tag{7.1.11}$$

From the divergence equation for source free region, it implies that

$$\nabla \cdot \epsilon \mathbf{E} = \nabla_s \cdot \epsilon \mathbf{E}_s + \partial_z \epsilon E_z = 0 \tag{7.1.12}$$

Since the E_z component is much smaller that the transverse component, one gets

$$\nabla_s \cdot \epsilon \mathbf{E}_s = 0, \tag{7.1.13}$$

Similarly, one gets

$$\nabla_s \cdot \mu \mathbf{H}_s = 0. \tag{7.1.14}$$

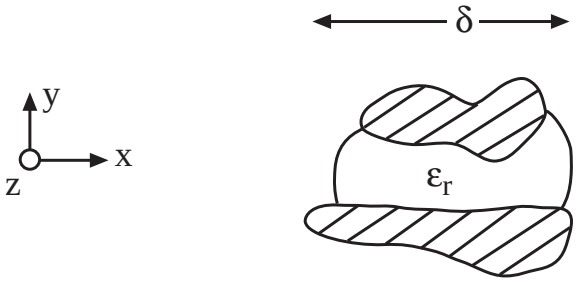


Figure 7.2: A generic geometry for the analysis of microwave integrated circuits waveguide.

Consequently, the transverse field of a quasi-TEM mode is essentially static. Because of this, the waveguide can be analyzed as if a transversely static TEM mode is propagating on it. We can solve the transverse electrostatic problem to find the line capacitance C of the line. The line inductance L can be obtained by solving the magnetostatic problem. Then, k_z, the axial wavenumber, can be found by

$$k_z = \omega\sqrt{LC}. \tag{7.1.15}$$

The above analysis indicates that when the wavelength is long, the axial variation of the field is slow compared to the transverse variation. The transverse variation of the field has to be such that the field can match the boundary condition on the metallic conductors, which is much smaller in dimension than the wavelength. Hence, the transverse variation of the field must balance itself resulting in equations (7.1.10) to (7.1.14), which are the static equations.

Please be noted that the reason for the quasi-TEM field here is quite different from that in the weak-contrast optical fiber. In the weak-contrast optical fiber, the reason for quasi-TEM is the paraxial nature of the wave as the frequency increases and the unimportance of the polarization term comparatively. In the weak-contrast optical fiber, the transverse dimension can be on the order or wavelength or larger, and yet the field is quasi-TEM.

7.1.1 Microstrip Line Capacitance—The Spectral Domain Approach

We shall discuss how the line capacitance of a microstrip line can be found. There is no closed-form solution for such a class of problem. When $\epsilon_0 = \epsilon_1$, one may solve such problems by conformal mapping. When $h/w \ll 1$, the problem can be solved by asymptotic matching [32]. However, to get an accurate value of C for all w/h, a numerical analysis is preferable.

To find the line capacitance of the microstrip line, one can first solve for the charge distribution on the line. Then, the capacitance can be easily found from the equation $Q = CV$, where Q is the total charge per unit length on the line, and V is the voltage applied between the strip and the ground plane. To find the charge distribution, we can first formulate its governing equation for the charge distribution. A potential ϕ can be defined such that $\mathbf{E}_s = -\nabla_s \phi$ and $\nabla_s^2 \phi = 0$, because $\mathbf{E}_s$ is an electrostatic field in the long wavelength limit.

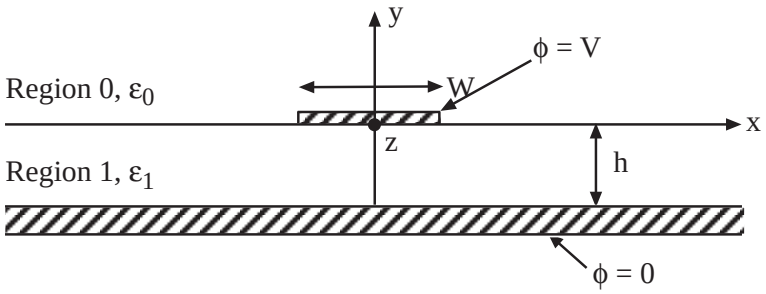


Figure 7.3: Geometry for deriving the line capacitance of a microstrip line.

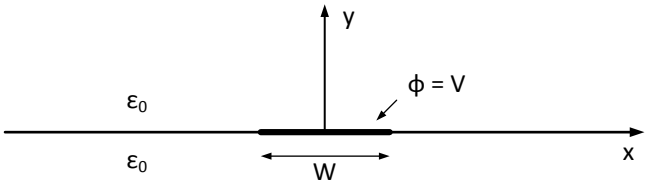


Figure 7.4: A charged strip hanging in free space assumed to be infinitesimally thin.

Before formulating the integral equation for the geometry shown in Figure 7.3, let us first consider the case of a charged strip suspended in free space shown in Figure 7.4. Using Fourier transforms, which are the gist of the spectral domain approach, the potential can be written as

$$\phi_{0\pm}(x,y) = \frac{1}{2\pi}\int_{-\infty}^{\infty} d\lambda e^{i\lambda x}\tilde{\phi}_{0\pm}(\lambda, y). \tag{7.1.16}$$

where the subscript $+$ and $-$ denote $y > 0$ and $y < 0$, respectively. Since $\nabla_s^2 \phi_{0\pm}(x,y) = 0$,

we deduce that

$$\frac{d^2}{dy^2}\tilde{\phi}_{0\pm}(\lambda, y) - \lambda^2 \tilde{\phi}_{0\pm}(\lambda, y) = 0, \tag{7.1.17}$$

or that

$$\tilde{\phi}_{0\pm}(\lambda, y) = A_{0\pm} e^{-|\lambda| y} + B_{0\pm} e^{|\lambda| y}. \tag{7.1.18}$$

Furthermore, $\tilde{\phi}_{0+}(\lambda, y) = 0$, $y \to \infty$, and $\tilde{\phi}_{0-}(\lambda, y) = 0$, $y \to -\infty$, imply that $B_{0+} = A_{0-} = 0$. But since $\phi_{0+}(x, y = 0) = \phi_{0-}(x, y = 0)$ for all x, we have $\tilde{\phi}_{0+}(\lambda, y = 0) = \tilde{\phi}_{0-}(\lambda, y = 0)$ implying that $A_{0+} = B_{0-}$. Consequently, we have

$$\phi_{0\pm}(x, y) = \frac{1}{2\pi} \int_{-\infty}^{\infty} d\lambda e^{i\lambda x} A_0(\lambda) e^{\mp|\lambda| y}, \qquad \begin{matrix} y > 0 \\ y < 0 \end{matrix}. \tag{7.1.19}$$

The charge on the strip is given by

$$\sigma(x) = -\epsilon_0 \left[\frac{\partial \phi_{0+}}{\partial y}(x, y = 0) - \frac{\partial \phi_{0-}}{\partial y}(x, y = 0) \right], \tag{7.1.20}$$

or that

$$\sigma(x) = \frac{\epsilon_0}{\pi} \int_{-\infty}^{\infty} d\lambda e^{i\lambda x} |\lambda| A_0(\lambda). \tag{7.1.21}$$

Defining the Fourier transform of $\sigma(x)$ as $\tilde{\sigma}(\lambda)$, it is seen that

$$A_0(\lambda) = \frac{\tilde{\sigma}(\lambda)}{2\epsilon_0 |\lambda|}, \tag{7.1.22}$$

or, in general, the potential becomes

$$\phi_{0\pm}(x, y) = \frac{1}{4\pi\epsilon_0} \int_{-\infty}^{\infty} d\lambda e^{i\lambda x \mp |\lambda| y} \frac{\tilde{\sigma}(\lambda)}{|\lambda|}, \qquad \begin{matrix} y > 0 \\ y < 0 \end{matrix}. \tag{7.1.23a}$$

or

$$\phi_0(x, y) = \frac{1}{4\pi\epsilon_0} \int_{-\infty}^{\infty} d\lambda e^{i\lambda x - |\lambda y|} \frac{\tilde{\sigma}(\lambda)}{|\lambda|}. \tag{7.1.23b}$$

If we now place the charged strip over a dielectric half space, the potential in region 0 becomes

$$\phi_0(x, y) = \frac{1}{4\pi\epsilon_0} \int_{-\infty}^{\infty} d\lambda \frac{\tilde{\sigma}(\lambda)}{|\lambda|} e^{i\lambda x} [e^{-|\lambda y|} + r_{01} e^{-|\lambda|(y+d) - |\lambda d|}], \tag{7.1.24}$$

where r_{01} is a reflection coefficient relating the reflected potential $e^{-|\lambda| y}$ to the incident potential $e^{-|\lambda y|}$ at $y = -d$. In region 1, the potential is

$$\phi_1(x, y) = \frac{1}{4\pi\epsilon_0} \int_{-\infty}^{\infty} d\lambda \frac{\tilde{\sigma}(\lambda)}{|\lambda|} e^{i\lambda x} t_{01} e^{|\lambda|(y+d) - |\lambda d|}, \tag{7.1.25}$$

where t_{01} is a transmission coefficient relating the transmitted potential$e^{|\lambda| y}$ to the incident potential $e^{-|\lambda y|}$. The continuity of potential at $y = -d$ requires that

$$1 + r_{01} = t_{01}. \tag{7.1.26}$$

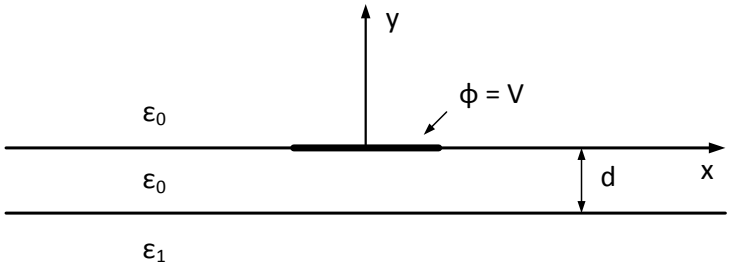


Figure 7.5: A charged strip, assumed to be infinitesimally thin, over a dielectric half-space.

The continuity of normal electric flux at $y = -d$ implies that

$$\epsilon_0(1 - r_{01}) = \epsilon_1 t_{01}. \tag{7.1.27}$$

Solving (7.1.26) and (7.1.27) yields

$$r_{01} = \frac{\epsilon_0 - \epsilon_1}{\epsilon_0 + \epsilon_1}, \qquad t_{01} = \frac{2\epsilon_0}{\epsilon_0 + \epsilon_1}. \tag{7.1.28}$$

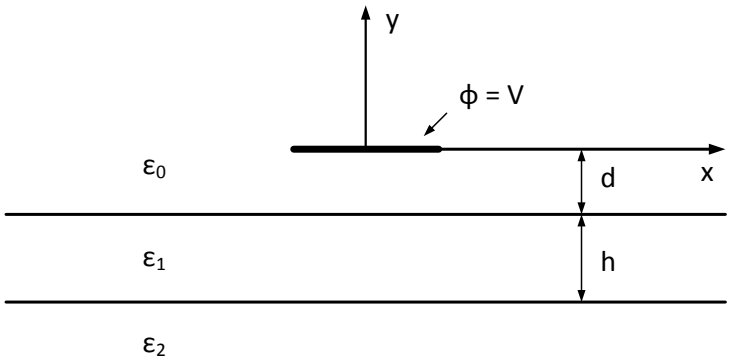


Figure 7.6: A charged strip over a dielectric slab.

For the geometry of Figure 7.6, we can find the total reflection coefficient using a geometric series expansion [33, 34] to obtain

$$\tilde{r}_{01} = r_{01} + \frac{t_{01} r_{12} t_{10} e^{-2|\lambda|h}}{1 - r_{10} r_{12} e^{-2|\lambda|h}}. \tag{7.1.29}$$

Equation (7.1.29) can be written as a recursive relation if we have more subsurface layers. In this case, the potential in region 1 can be written as

$$\phi_0(x,y) = \frac{1}{4\pi\epsilon_0}\int_{-\infty}^{\infty} d\lambda \frac{\tilde{\sigma}(\lambda)e^{i\lambda x}}{|\lambda|}[e^{-|\lambda y|} + \tilde{r}_{01}e^{-|\lambda|(y+d)-|\lambda d|}]. \tag{7.1.30}$$

For the geometry of Figure 7.3, $d = 0$, $r_{12} = -1$; therefore,

$$\phi_0(x,y) = \frac{1}{4\pi\epsilon_0}\int_{-\infty}^{\infty} d\lambda \frac{\tilde{\sigma}(\lambda)e^{i\lambda x}}{|\lambda|}e^{-|\lambda|y}(1+\tilde{r}_{01}), \qquad y > 0, \tag{7.1.31}$$

where

$$1 + \tilde{r}_{01} = t_{01} - \frac{t_{01}t_{10}e^{-2|\lambda|h}}{1 - r_{01}e^{-2|\lambda|h}} = t_{01}\frac{1 - e^{-2|\lambda|h}}{1 - r_{01}e^{-2|\lambda|h}}. \tag{7.1.32}$$

Requiring that the potential be V on the strip, then the integral equation for $\tilde{\sigma}(\lambda)$ can be written as

$$\frac{1}{2\pi}\int_{-\infty}^{\infty} d\lambda \tilde{G}(\lambda)\tilde{\sigma}(\lambda)e^{i\lambda x} = V, \qquad |x| < \frac{w}{2}, \tag{7.1.33a}$$

$$\sigma(x) = \frac{1}{2\pi}\int_{-\infty}^{\infty} d\lambda \tilde{\sigma}(\lambda)e^{i\lambda x} = 0, \qquad |x| > \frac{w}{2}, \tag{7.1.33b}$$

where $\tilde{\sigma}(\lambda)$ is the Fourier transform of $\sigma(x)$, and

$$\tilde{G}(\lambda) = \frac{1}{|\lambda|(\epsilon_0 + \epsilon_1)}\frac{1 - e^{-2|\lambda|h}}{1 - \frac{\epsilon_0 - \epsilon_1}{\epsilon_0 + \epsilon_1}e^{-2|\lambda|h}}. \tag{7.1.34}$$

Equations (7.1.33a) and (7.1.33b) are also known as dual integral equations. Dual integral equations in general do not have closed-form solutions, except for semi-infinite structures. In that case, they can be solved using the Wiener-Hopf technique [39, 40].

In order to solve (7.1.33b), we use Galerkin's method and let

$$\sigma(x) = \sum_{n=0}^{N} a_n f_n(x), \tag{7.1.35}$$

where $f_n(x) = 0$, $|x| > \frac{w}{2}$, and that the Fourier transform of $f_n(x)$ exists. Furthermore, $f_n(x)$, $n = 1, \ldots, \infty$ is complete for $|x| < \frac{w}{2}$. Then, we deduce that

$$\tilde{\sigma}(\lambda) = \sum_{n=0}^{N} a_n \tilde{f}_n(\lambda), \tag{7.1.36}$$

where $\tilde{f}_n(\lambda)$ is the Fourier transform of $f_n(x)$. Substituting (7.1.36) into (7.1.33a), we have

$$\frac{1}{2\pi}\sum_{n=0}^{N} a_n \int_{-\infty}^{\infty} d\lambda \tilde{G}(\lambda)\tilde{f}_n(\lambda)e^{i\lambda x} = V. \tag{7.1.37}$$

To remove the x dependence in (7.1.37), we multiply it by $f_m(x)$ and integrate over x, to obtain

$$\sum_{n=0}^{N} a_n \int_{-\infty}^{\infty} d\lambda \tilde{f}_m(-\lambda)\tilde{G}(\lambda)\tilde{f}_n(\lambda) = 2\pi V \int_{-w/2}^{w/2} f_m(x)dx. \tag{7.1.38}$$

The above is a matrix equation of the form

$$\sum_{n=0}^{N} A_{mn}a_n = b_m, \tag{7.1.39}$$

from which we can solve for a_n's. Once a_n's are known, we can find $\sigma(x)$ from (7.1.35). Then

$$Q = \int_{-w/2}^{w/2} \sigma(x)dx,$$

and the line capacitance is given by $C = Q/V$.

For the geometry considered in Figure 7.3, the quasi-TEM mode is even symmetric about $x = 0$. Therefore, we need only to pick even functions for our basis functions in (7.1.35). Then, by the theory of Fourier transform, $\tilde{f}_n(\lambda) = \tilde{f}_n(-\lambda)$ if $f_n(x)$ is an even function. Furthermore, since $\tilde{G}(-\lambda) = \tilde{G}(\lambda)$, (7.1.38) becomes

$$\sum_{n=0}^{N} a_n \int_{0}^{\infty} d\lambda \tilde{f}_m(\lambda)\tilde{G}(\lambda)\tilde{f}_n(\lambda) = 2\pi V \int_{0}^{w/2} f_m(x)dx. \tag{7.1.40}$$

Clearly, A_{mn} is a symmetric matrix.

Equation (7.1.33a) has a different meaning if it is transformed back to x-space, i.e.,

$$\int_{-w/2}^{w/2} dx' G(x - x')\sigma(x') = V, \qquad |x| < \frac{w}{2}, \tag{7.1.41}$$

where $G(x)$ is the inverse Fourier transform of $\tilde{G}(\lambda)$. And $G(x)$ could be thought of as the Green's function generating the potential due to a line of point surface charge at $y = 0$. Hence, the convolution of $G(x)$ with $\sigma(x)$ gives the potential. However, closed-form expression does not exist for $G(x)$ in general. A more convenient method to calculate the integral (7.1.41) is in the spectral domain as in (7.1.33a) where $\tilde{G}(\lambda)$ exists in closed-form.

Once C is known, we can estimate k_z via (7.1.15). To find L, we make use of the fact that if $\epsilon_1 = \epsilon_0$, a pure TEM mode propagates on the microstrip line. In this case,

$$k_z = k_0 = \omega\sqrt{L_0C_0}. \tag{7.1.42}$$

Therefore, L_0 can be found once C_0, the line capacitance with $\epsilon_1 = \epsilon_0$, is known. C_0 can be found by solving the integral equation above. Since L_0, which is obtained by solving the

magnetostatic problem, remains unchanged when $\epsilon_1 \neq \epsilon_0$, we have

$$k_z = \sqrt{\frac{C}{C_0}} k_0. \tag{7.1.43}$$

An effective relative dielectric constant can be defined such that

$$\epsilon_{re} = \frac{C}{C_0}. \tag{7.1.44}$$

It is the dielectric constant with which one can fill the space homogeneously around a microstrip line to yield the same line capacitance as the inhomogeneously filled microstrip line.

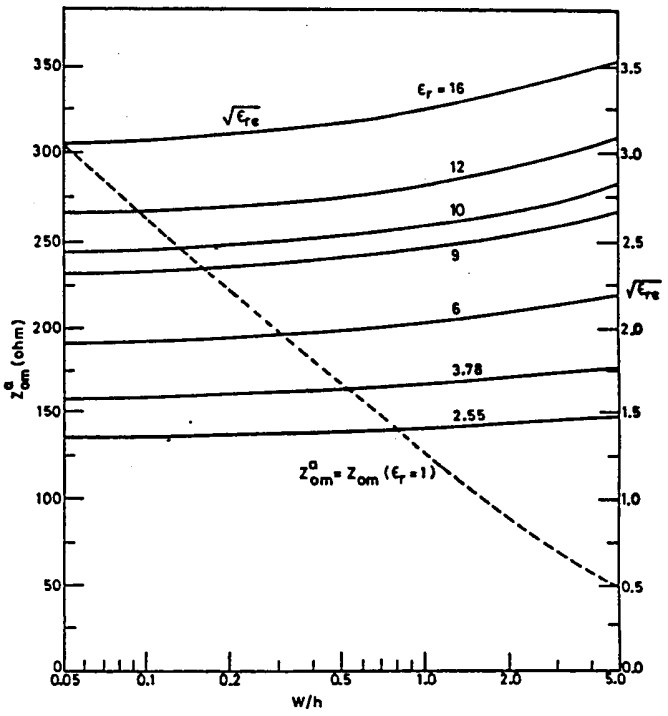


Figure 7.7: Characteristic impedance of a microstrip line as a function of line dimension w/h—quasi-static approximation (from H.A. Wheeler [2].)

A characteristic impedance can also be defined as

$$Z_0 = \sqrt{\frac{L_0}{C}} = \sqrt{\frac{\mu_0 \epsilon_0}{C_0 C}} = \frac{Z_0^a}{\sqrt{\epsilon_r}}, \tag{7.1.45}$$

where Z_0^a is the characteristic impedance of the air-filled microstrip line.

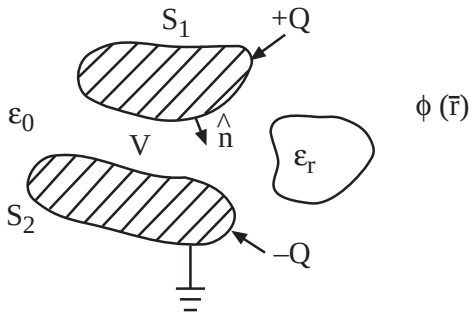


Figure 7.8: Geometry for the analysis of bounds for capacitance.

7.1.2 Variational Expressions and Bounds for Capacitance

There exists bounds for the capacitance between two conductors due to the minimum energy principle [35–38]. Due to the prevalence use of capacitance extraction software in the computer chip industry, this could be of importance.

The exact solution of Laplace's equation minimizes the energy in the potential as shall be shown. The capacitance can be related to the energy stored in the potential or chage in two ways, i.e.

$$W = \frac{1}{2}C\Phi^2 = \frac{1}{2}\frac{Q^2}{C}, \tag{7.1.46}$$

where W is the energy stored in the potential Φ, the voltage difference between the two conductors. Then if W is not accurately known, the exact capacitance C_e is bounded by

$$\frac{2W}{\Phi^2} \geq C_e \geq \frac{Q^2}{2W} \tag{7.1.47}$$

if either Φ is kept constant of if Q is kept constant on the metallic conductor. The above assertion shall be proved in the following.

In the first inequality in (7.1.47), the potential satisfies the boundary condition exactly, but does not satisfy Laplace's equation exactly in the space in between the conductors. It can be rewritten as

$$\frac{\int\limits_V \epsilon|\mathbf{E}|^2 dV}{\Phi^2} = \frac{\int\limits_V \epsilon|\nabla\phi|^2 dV}{\Phi^2} \geq C_e. \tag{7.1.48}$$

If we let $\phi = \phi_e + \delta\phi$ such that $\delta\phi = 0$ on S_1 and S_2, because ϕ satisfies the boundary condition exactly, then

$$\int\limits_V \epsilon|\nabla\phi|^2 = \int\limits_V \epsilon|\nabla\phi_e|^2 dV + 2\int\limits_V \epsilon\nabla\phi_e \cdot \nabla\delta\phi W + \int\limits_V \epsilon|\nabla\delta\phi|^2 dV. \tag{7.1.49}$$

But

$$\begin{aligned}\int_V \epsilon\nabla\phi_e\cdot\nabla\delta\phi dV &= \int_V \nabla\cdot(\delta\phi\epsilon\nabla\phi_e)dV - \int_V \delta\phi(\nabla\cdot\epsilon\nabla\phi_e)dV \\ &= -\int_{S_1+S_2}\hat{n}\cdot(\delta\phi\epsilon\nabla\phi_e)dS - \int_V \delta\phi(\nabla\cdot\epsilon\nabla\phi_e)dV.\end{aligned} \tag{7.1.50}$$

The first term in (7.1.50) vanishes because $\delta\phi = 0$ on S_1 and S_2, and the second term vanishes because $\nabla\cdot\epsilon\nabla\phi_e = 0$, in V. Therefore, if $\phi = \phi_e + \delta\phi$ where $\delta\phi$ is the error field in V, then

$$\int_V \epsilon|\nabla\phi|^2 dV = \int_V \epsilon|\nabla\phi_e|^2 dV + \int_V \epsilon|\nabla\delta\phi|^2 dV > \int_V \epsilon|\nabla\phi_e|^2 dV, \tag{7.1.51}$$

implying the bound in (7.1.48) for a constant Φ.

The second inequality in (7.1.47) can be written as

$$\frac{\int_V \epsilon|\mathbf{E}|^2 dV}{Q^2} = \frac{\int_V \epsilon|\nabla\phi|^2 dV}{Q^2} \geq C_e^{-1} \tag{7.1.52}$$

where Q is constant but ϕ satisfies Laplace's equation exactly in V but does not satisfy the boundary condition. Similar to (7.1.49), we let $\phi = \phi_e + \delta\phi$, but now, $\delta\phi \neq 0$ on S_1 and S_2. Similar to (7.1.50), we obtain that the second term in (7.1.49) is

$$\begin{aligned}\int_V \epsilon\nabla\phi_e\cdot\nabla\delta\phi dV &= \int_V \nabla\cdot(\epsilon\phi_e\nabla\delta\phi)dV - \int_V \phi_e\nabla\cdot\epsilon\nabla\delta\phi dV \\ &= -\int_{S_1+S_2}\hat{n}\cdot(\epsilon\phi_e\nabla\delta\phi)dS - \int_V \phi_e\nabla\cdot\epsilon\nabla\delta\phi dV.\end{aligned} \tag{7.1.53}$$

The second term in (7.1.53) vanishes because $\nabla\cdot\epsilon\nabla\delta\phi = 0$. Since $\phi_e = \Phi$ on S_1 and $\phi_e = 0$ on S_2, the first term becomes

$$-\Phi\int_{S_1}\hat{n}\cdot\epsilon\nabla\delta\phi dS. \tag{7.1.54}$$

Since $-\hat{n}\cdot\epsilon\nabla\delta\phi = \delta\sigma$, the error surface charge density on S_1, it integrates to zero because we assume that Q is a constant so that $\delta Q = 0$. Therefore, the second term in (7.1.49) vanishes too for this case. Consequently, we obtain the bound in (7.1.52).

Equation (7.1.52) is also Thompson's theorem [36] which says that a set of charge Q always adjusts itself on a metallic conductor such that the energy stored in the electric field is minimized. The inequality (7.1.52) assumes that the potential ϕ satisfies $\nabla\cdot\epsilon\nabla\phi = 0$ exactly in V but does not satisfy the boundary condition. Hence, it is more appropriate to write (7.1.52) in terms of surface integrals. To this end, we express using integration by parts that

$$\int_V \epsilon|\nabla\phi|^2 dV = -\int_{S_1+S_2} dS\phi\ \epsilon\ \hat{n}\cdot\nabla\phi = \int_{S_1+S_2} dS\phi\sigma. \tag{7.1.55}$$

In this manner, (7.1.52) becomes

$$C_e^{-1} \leq \frac{\int\limits_{S_1+S_2} dS\phi\sigma}{Q^2}. \tag{7.1.56}$$

A potential which satisfies $\nabla \cdot \epsilon \nabla \phi = 0$ exactly in V can be obtained by the Green's function method if the Green's function is known exactly. Hence,

$$\phi(\mathbf{r}) = \int\limits_{S_1+S_2} dS' G(\mathbf{r}, \mathbf{r}')\sigma(\mathbf{r}'), \tag{7.1.57}$$

or that (7.1.56) can be expressed as

$$C_e^{-1} \leq \int\limits_{S_1+S_2} dS\sigma(\mathbf{r}) \int\limits_{S_1+S_2} dS' G(\mathbf{r}, \mathbf{r}')\sigma(\mathbf{r}')/Q^2, \tag{7.1.58}$$

or

$$C_e^{-1} \leq \frac{\langle \sigma, G, \sigma \rangle}{Q^2}. \tag{7.1.59}$$

where

$$\langle \sigma, G, \sigma \rangle = \int\limits_{S_1+S_2} dS\sigma(\mathbf{r}) \int\limits_{S_1+S_2} dS' G(\mathbf{r}, \mathbf{r}')\sigma(\mathbf{r}')$$

The expressions that

$$C = \int\limits_V dV\epsilon|\nabla\phi|^2/\Phi^2 \tag{7.1.60}$$

and

$$C^{-1} = \langle \sigma, G, \sigma \rangle / Q^2 \tag{7.1.61}$$

are both variational expressions. Rayleigh-Ritz procedure can be applied to (7.1.60) to solve for ϕ which is the same as applying the finite element method to solve $\nabla \cdot \epsilon \nabla \phi = 0$. Applying Rayleigh-Ritz procedure to solve (7.1.61) for σ is the same as applying Galerkin's method to solve the integral equation

$$\int\limits_{S_1+S_2} dS' G(\mathbf{r}, \mathbf{r}')\sigma(\mathbf{r}') = \begin{cases} V, & \mathbf{r} \in S_1 \\ 0, & \mathbf{r} \in S_2 \end{cases}. \tag{7.1.62}$$

7.2 Microstrip Line—A Frequency Dependent Theory

The quasi-TEM model of the microstrip line triumphs in predicting the value of k_z or the phase velocity in the long wavelength limit. However, k_z is in general frequency dispersive. In order to have a k_z that is valid at high frequencies as well, we need to solve the full wave solution to the microstrip line problem. To do this, an integral equation can be formulated from Maxwell's equation with no approximation. From the integral equation, the guidance condition of the microstrip line can be solved for.

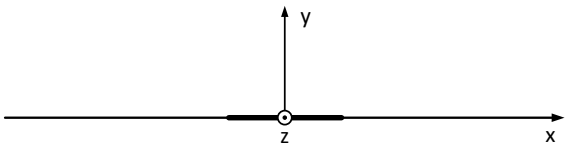


Figure 7.9: An infinitesimally thin strip hanging in free space.

7.2.1 Derivation of the Integral Equation

Before deriving an integral equation for the microstrip line, let us consider first a metallic strip carrying a current suspended in free space. The field around the strip can be decomposed into TE to y and TM to y fields by H_{0y} and E_{0y}, respectively. Using Fourier transforms, the fields can be written as

$$E_{0y}(x,y,z) = e^{ik_z z}\frac{1}{2\pi}\int_{-\infty}^{\infty} dk_x e^{ik_x x}\tilde{E}_{0y}(k_x,y), \tag{7.2.1}$$

$$H_{0y}(x,y,z) = e^{ik_z z}\frac{1}{2\pi}\int_{-\infty}^{\infty} dk_x e^{ik_x x}\tilde{H}_{0y}(k_x,y). \tag{7.2.2}$$

In the above, we assume that the fields have $e^{ik_z z}$ variation. Since $(\nabla^2 + k_0^2)E_{0y} = 0$, and $(\nabla^2 + k_0^2)H_{0y} = 0$, it follows that

$$\left(\frac{d^2}{dy^2} + k_y^2\right)\tilde{E}_{0y}(k_x,y) = 0, \tag{7.2.3a}$$

$$\left(\frac{d^2}{dy^2} + k_y^2\right)\tilde{H}_{0y}(k_x,y) = 0, \tag{7.2.3b}$$

where $k_y^2 = k_0^2 - k_x^2 - k_z^2$. This implies that

$$\tilde{E}_{0y}(k_x,y) = e_{0+}e^{ik_y y} + e_{0-}e^{-ik_y y}, \tag{7.2.4a}$$

$$\tilde{H}_{0y}(k_x,y) = h_{0+}e^{ik_y y} + h_{0-}e^{-ik_y y}. \tag{7.2.4b}$$

However, for $y > 0$, only upward going waves exist, while for $y < 0$, only have downward going waves exist. Therefore, $e_{0-} = h_{0-} = 0$, $y > 0$, $e_{0+} = h_{0+} = 0$, $y < 0$. Consequently, (7.2.1) and (7.2.2) become

$$E_{0y}(x,y,z) = \frac{e^{ik_z z}}{2\pi}\int_{-\infty}^{\infty} dk_x e^{ik_x x}e_{0\pm}e^{\pm ik_y y}, \qquad \begin{matrix} y>0 \\ y<0 \end{matrix}, \tag{7.2.5a}$$

$$H_{0y}(x,y,z) = \frac{e^{ik_z z}}{2\pi}\int_{-\infty}^{\infty} dk_x e^{ik_x x}h_{0\pm}e^{\pm ik_y y}, \qquad \begin{matrix} y>0 \\ y<0 \end{matrix}. \tag{7.2.5b}$$

On the other hand, if the strip is carrying electric current, then $E_{0y}(x, y+, z) = -E_{0y}(x, y-, z)$ and $H_{0y}(x, y+, z) = H_{0y}(x, y-, z)$ for all x. Therefore, $e_{0+} = -e_{0-} = e_0$, $h_{0+} = h_{0-} = h_0$, and we have

$$E_{0y}(x, y, z) = \pm \frac{e^{ik_z z}}{2\pi} \int_{-\infty}^{\infty} dk_x e^{ik_x x + ik_y |y|} e_0(k_x), , \tag{7.2.6a}$$

$$H_{0y}(x, y, z) = \frac{e^{ik_z z}}{2\pi} \int_{-\infty}^{\infty} dk_x e^{ik_x x + ik_y |y|} h_0(k_x).. \tag{7.2.6b}$$

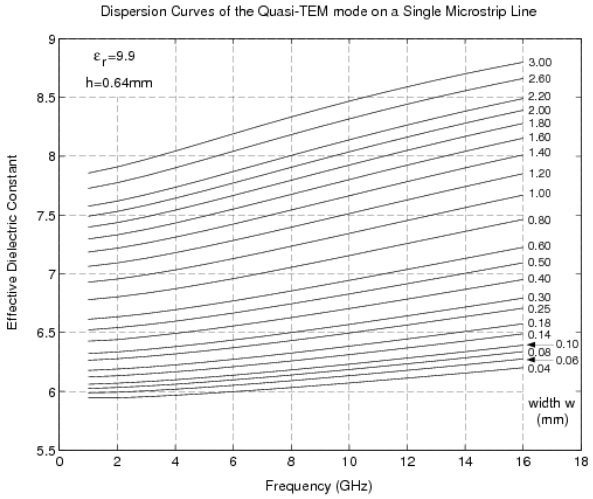


Figure 7.10: Frequency dependent effective dielectric constant of the quasi-TEM mode on single microstrip lines (from Jansen [14], reproduced by G. Papadopoulous.)

If now, the strip is on top of a stratified half space as shown in Figure 7.10, the field in region 0 can be written as

$$E_{0y}(x, y, z) = \frac{e^{ik_z z}}{2\pi} \int_{-\infty}^{\infty} dk_x e^{ik_x x} e_0(k_x) \left[\pm e^{ik_y |y|} - \tilde{R}^{TM} e^{ik_y (y+d) + ik_y |d|} \right], \tag{7.2.7a}$$

$$H_{0y}(x, y, z) = \frac{e^{ik_z z}}{2\pi} \int_{-\infty}^{\infty} dk_x e^{ik_x x} h_0(k_x) \left[e^{ik_y |y|} + \tilde{R}^{TE} e^{ik_y (y+d) + ik_y |d|} \right]. \tag{7.2.7b}$$

In order to relate e_0 and h_0 to the current on the strip, we need to derive the transverse components of the field. Since each spectral component in (7.2.7a) and (7.2.7b) consists of waves with $e^{\pm ik_y y}$ dependence, the fields transverse to y for each spectral component are

$$\tilde{\mathbf{E}}_s = \frac{1}{k_s^2} \left[\frac{\partial}{\partial y} \nabla_s \tilde{E}_y - i\omega\mu_0 \hat{y} \times \nabla_s \tilde{H}_y \right] \tag{7.2.8a}$$

$$\tilde{\mathbf{H}}_s = \frac{1}{k_s^2} \left[\frac{\partial}{\partial y} \nabla_s \tilde{H}_y + i\omega\epsilon_0 \hat{y} \times \nabla_s \tilde{E}_y \right]. \tag{7.2.8b}$$

where the tilde denotes the spectral component, $k_s^2 = k_0^2 - k_y^2$, and $\nabla_s = \hat{x}\frac{\partial}{\partial x} + \hat{z}\frac{\partial}{\partial z}$. Applying (7.2.8b) to (7.2.7a) and (7.2.7b), we have

$$\begin{aligned}\mathbf{H}_s = \frac{e^{ik_z z}}{2\pi}\int_{-\infty}^{\infty} dk_x \frac{e^{ik_x x}}{k_s^2}\{&-h_0(k_x)k_y\mathbf{k}_s[\pm e^{ik_y|y|} + \tilde{R}^{TE}e^{ik_y(y+d)+ik_y|d|}] \\ &- \omega\epsilon_0 e_0(k_x)\hat{y}\times\mathbf{k}_s[\pm e^{ik_y|y|} - \tilde{R}^{TM}e^{ik_y(y+d)+ik_y|d|}]\}.\end{aligned} \tag{7.2.9}$$

The discontinuity in $\mathbf{H}_s$ at $y = 0$ equals the current on the strip. More precisely, $\mathbf{J}_s = \hat{y}\times[\mathbf{H}_s(y=0+) - \mathbf{H}_s(y=0-)]$. Hence,

$$\begin{aligned}\mathbf{J}_s &= \frac{e^{ik_z z}}{2\pi}\int_{-\infty}^{\infty} dk_x \frac{e^{ik_x x}}{k_s^2}[-2h_0 k_y \hat{y}\times\mathbf{k}_s + 2\omega\epsilon_0 e_0 \mathbf{k}_s] \\ &= \frac{e^{ik_z z}}{2\pi}\int_{-\infty}^{\infty} dk_x \frac{e^{ik_x x}}{k_s}[\mathbf{k}_s, -\hat{y}\times\mathbf{k}_s]\begin{bmatrix}2\omega\epsilon_0 e_0/k_s \\ 2k_y h_0/k_s\end{bmatrix} \\ &= \frac{e^{ik_z z}}{2\pi}\int_{-\infty}^{\infty} dk_x \overline{\mathbf{F}}(k_x, x)\cdot\mathbf{K}_s(k_x),\end{aligned} \tag{7.2.10}$$

where $\overline{\mathbf{F}}(k_x, x) = \frac{e^{ik_x x}}{k_s}[\mathbf{k}_s, -\hat{y}\times\mathbf{k}_s]$, $\mathbf{K}_s^t(k_x) = 2[\omega\epsilon_0 e_0, k_y h_0]/k_s$. Here, $\overline{\mathbf{F}}(k_x, x)$ is a 2×2 matrix and $\mathbf{K}_s^t(k_x)$ is a 2×1 vector.

Similarly, $\mathbf{E}_s$ can be derived to be

$$\begin{aligned}\mathbf{E}_s = \frac{e^{ik_z z}}{2\pi}\int_{-\infty}^{\infty} dk_x \frac{e^{ik_x x}}{k_s^2}\{&-k_y\mathbf{k}_s e_0[e^{ik_y|y|} - \tilde{R}^{TM}e^{ik_y(y+d)+ik_y|d|}] \\ &+ \omega\mu_0 h_0 \hat{y}\times\mathbf{k}_s[e^{ik_y|y|} + \tilde{R}^{TE}e^{ik_y(y+d)+ik_y|d|}]\}.\end{aligned} \tag{7.2.11}$$

We can rewrite (7.2.11) as

$$\begin{aligned}\mathbf{E}_s &= \frac{e^{ik_z z}}{2\pi}\int_{-\infty}^{\infty} dk_x \frac{e^{ik_x x}}{k_s}[\mathbf{k}_s, -\hat{y}\times\mathbf{k}_s] \\ &\qquad \begin{bmatrix}-k_y e_0[e^{ik_y|y|} - \tilde{R}^{TM}e^{ik_y(y+d)+ik_y|d|}]/k_s \\ -\omega\mu_0 h_0[e^{ik_y|y|} + \tilde{R}^{TE}e^{ik_y(y+d)+ik_y|d|}]/k_s\end{bmatrix} \\ &= \frac{e^{ik_z z}}{2\pi}\int_{-\infty}^{\infty} dk_x \overline{\mathbf{F}}(k_x, x)\cdot\overline{\mathbf{G}}(k_x, y)\cdot\mathbf{K}_s(k_x),\end{aligned} \tag{7.2.12}$$

where

$$\overline{\mathbf{G}}(k_x, y) = \begin{bmatrix}-\frac{k_y}{2\omega\epsilon_0}[e^{ik_y|y|} - \tilde{R}^{TM}e^{ik_y(y+d)+ik_y|d|}] & 0 \\ 0 & -\frac{\omega\mu_0}{2k_y}[e^{ik_y|y|} + \tilde{R}^{TE}e^{ik_y(y+d)+ik_y|d|}]\end{bmatrix}. \tag{7.2.13}$$

For a microstrip line, $d = 0$, and from the boundary condition $\mathbf{E}_s = 0$ for $|x| < w/2$ and $y = 0$. Therefore, the integral equation is

$$\mathbf{E}_s(x, y=0, z) = \frac{e^{ik_z z}}{2\pi}\int_{-\infty}^{\infty}\overline{\mathbf{F}}(k_x, x)\cdot\overline{\mathbf{G}}_0(k_x)\cdot\mathbf{K}_s(k_x)dk_x = 0, \qquad |x| < \frac{w}{2}, \tag{7.2.14}$$

where $\mathbf{K}_s(k_x)$ is the unknown to be sought, and

$$\overline{\mathbf{G}}_0(k_x) = -\begin{bmatrix} \frac{k_y}{2\omega\epsilon_0}[1-\tilde{R}^{TM}] & 0 \\ 0 & \frac{\omega\mu_0}{2k_y}[1+\tilde{R}^{TE}] \end{bmatrix}. \tag{7.2.15}$$

7.2.2 Vector Fourier Transform (VFT)

With $\overline{\mathbf{F}}(k_x, x)$ defined in (7.2.10), there exists a vector Fourier transform pair given by [41,42].

$$\mathbf{h}(x) = \frac{1}{2\pi}\int_{-\infty}^{\infty} dk_x \overline{\mathbf{F}}(k_x, x)\cdot\mathbf{H}(k_x), \tag{7.2.16a}$$

$$\mathbf{H}(k_x) = \int_{-\infty}^{\infty} dx' \overline{\mathbf{F}}(k_x, -x')\cdot\mathbf{h}(x'), \tag{7.2.16b}$$

where the 2×2 matrix $\overline{\mathbf{F}}$ is

$$\overline{\mathbf{F}}(k_x, x) = \frac{e^{ik_x x}}{k_s}[\mathbf{k}_s, -\hat{y}\times\mathbf{k}_s] = \frac{e^{ik_x x}}{k_s}\begin{bmatrix} k_z & k_x \\ k_x & -k_z \end{bmatrix}. \tag{7.2.16c}$$

This can be proven by substituting (7.2.16b) into (7.2.16a) to obtain

$$\begin{aligned} \mathbf{h}(x) &= \frac{1}{2\pi}\int_{-\infty}^{\infty} dk_x \int_{-\infty}^{\infty} dx' \overline{\mathbf{F}}(k_x, x)\cdot\overline{\mathbf{F}}(k_x, -x')\cdot\mathbf{h}(x') \\ &= \frac{1}{2\pi}\int_{-\infty}^{\infty} dx' \int_{-\infty}^{\infty} dk_x e^{ik_x(x-x')}\mathbf{h}(x') = \mathbf{h}(x). \end{aligned} \tag{7.2.17}$$

A similar substitution of (7.2.16a) into (7.2.16b) yields similar results. Therefore, (7.2.16a) and (7.2.16b) constitute a vector Fourier transform pair.

From the above, note that in Equation (7.2.10), if $\mathbf{J}_s(x,z) = e^{ik_z z}\mathbf{k}_s(x)$, then $\mathbf{K}_s(k_x)$ is the vector Fourier transform of $\mathbf{k}_s(x)$. In order to solve (7.2.14), we let

$$\mathbf{k}_s(x) = \begin{bmatrix} k_{sz}(x) \\ k_{sx}(x) \end{bmatrix} = \sum_{n=1}^{N}\begin{bmatrix} a_n k_{nz}(x) \\ b_n k_{nx}(x) \end{bmatrix} = \sum_{n=1}^{N}\overline{\mathbf{k}}_n(x)\cdot\mathbf{a}_n, \tag{7.2.18a}$$

where

$$\mathbf{a}_n = \begin{bmatrix} a_n \\ b_n \end{bmatrix} \qquad \overline{\mathbf{k}}_n(x) = \begin{bmatrix} k_{nz}(x) & 0 \\ 0 & k_{nx}(x) \end{bmatrix}. \tag{7.2.18b}$$

$\overline{\mathbf{k}}_n(x)$ is chosen so that its vector Fourier transform exists. Then

$$\mathbf{K}_s(k_x) = \sum_{n=1}^{N}\overline{\mathbf{K}}_n(k_x)\cdot\mathbf{a}_n, \tag{7.2.19}$$

where $\overline{\mathbf{K}}_n(k_x)$ is the VFT of $\overline{\mathbf{k}}_n(x)$. Substituting (7.2.19) into (7.2.14), we have

$$\mathbf{E}_s(x, y=0, z) = \frac{e^{ik_z z}}{2\pi}\sum_{n=1}^{N}\int_{-\infty}^{\infty} dk_x \overline{\mathbf{F}}(k_x, x)\cdot\overline{\mathbf{G}}_0(k_x)\cdot\overline{\mathbf{K}}_n(k_x)\cdot\mathbf{a}_n = 0. \tag{7.2.20}$$

To remove the x-dependence in the above equation, we multiply it by $\overline{\mathbf{k}}'_m(x)$ and integrate over x, where

$$\overline{\mathbf{k}}'_m(x) = \begin{bmatrix} k_{mz}(x) & 0 \\ 0 & -k_{mx}(x) \end{bmatrix}. \tag{7.2.21}$$

Then (7.2.20) becomes

$$\sum_{n=1}^{N} \int_{-\infty}^{\infty} dk_x \overline{\mathbf{K}}'_m(k_x) \cdot \overline{\mathbf{G}}_0(k_x) \cdot \overline{\mathbf{K}}_n(k_x) \cdot \mathbf{a}_n = 0, \tag{7.2.22}$$

where

$$\overline{\mathbf{K}}'_m(k_x) = \int_{-\infty}^{\infty} dx \overline{\mathbf{k}}'_m(x) \cdot \overline{\mathbf{F}}(k_x, x). \tag{7.2.23}$$

Equation (7.2.22) is of the form

$$\sum_{n=1}^{N} \overline{\mathbf{M}}_{mn} \cdot \mathbf{a}_n = 0, \qquad m = 1, \ldots, N. \tag{7.2.24}$$

The above is expressible in terms of a matrix equation

$$\overline{\mathbf{M}} \cdot \mathbf{a} = 0, \tag{7.2.25}$$

where

$$\overline{\mathbf{M}} = \begin{bmatrix} \overline{\mathbf{M}}_{11} & \overline{\mathbf{M}}_{12} & \cdots \\ \overline{\mathbf{M}}_{21} & \overline{\mathbf{M}}_{22} & \\ \vdots & & \ddots \end{bmatrix}, \qquad \mathbf{a} = \begin{bmatrix} \mathbf{a}_1 \\ \mathbf{a}_2 \\ \mathbf{a}_3 \\ \vdots \end{bmatrix}. \tag{7.2.26}$$

One can show by algebraic manipulation that $\overline{\mathbf{M}}_{ij} = \overline{\mathbf{M}}^t_{ji}$ [17]. The physical reason is that the inner product calculation of the above corresponds to reaction inner product in electromagnetics. In order for a guided mode with $\exp(ik_z z)$ to have nontrivial reaction with another field, the other field should have $\exp(-ik_z z)$, corresponding to a counter propagating mode. Hence, one of the current components in (7.2.21) has to change sign.

In order for $\mathbf{a}$ to be non-trivial, i.e., for a mode to exist, it is necessary that

$$\det\left[\overline{\mathbf{M}}(k_z)\right] = 0. \tag{7.2.27}$$

Since $\det\left[\overline{\mathbf{M}}(k_z)\right]$ is a function of k_z, the roots of (7.2.27) can be solved for numerically. The roots are values of k_z that satisfy the guidance condition of the strip. These include the higher order modes plus the fundamental modes.

With the weighting function as defined by (7.2.21), $\overline{\mathbf{M}}$ in (7.2.25) is a symmetric matrix. The reflection coefficients in (7.2.15) are generalized reflection coefficients for a layered medium. For a substrate backed by a ground plane, we have

$$\tilde{R}^{TM} = \frac{R_{01}^{TM} + e^{2ik_{1y}h}}{1 + R_{01}^{TM} e^{2ik_{1y}h}}, \qquad \tilde{R}^{TE} = \frac{R_{01}^{TE} - e^{2ik_{1y}h}}{1 - R_{01}^{TE} e^{2ik_{1y}h}}. \tag{7.2.28}$$

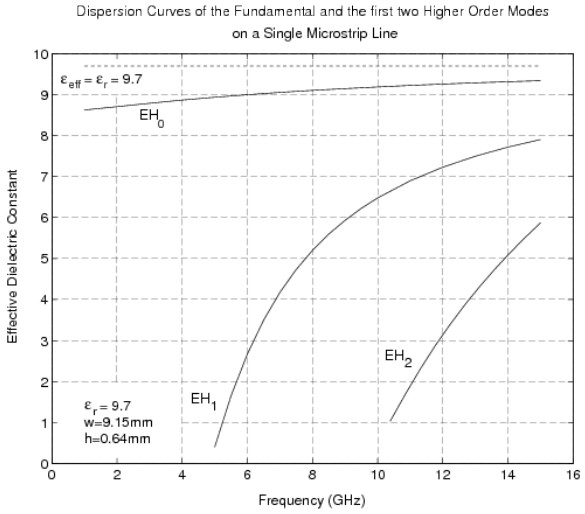


Figure 7.11: Dispersion characteristics of the fundamental and the first two higher order modes on a wide single microstrip line ($\epsilon_1 = 9.7$, $w = 9.15$mm), dashed curve: cover height $h = 3d = 1.92$mm (from Jansen [14], reproduced by G. Papadopoulos).

In the above, h is the thickness of the substrate, and

$$R_{ij} = \frac{p_j k_{iy} - p_i k_{jy}}{p_j k_{iy} + p_i k_{jy}}, \tag{7.2.29}$$

where $p_i = \mu_i$ for TE waves, $p_i = \epsilon_i$ for TM waves, $k_{iy} = \sqrt{k_i^2 - k_x^2 - k_z^2}$.

In general, for guidance to be possible, we require that

$$k_0 < k_z < k_1, \tag{7.2.30}$$

where k_1 is the wave number of the substrate region and k_0 is the free-space wave number. Hence, the solution of (7.2.27) is sought only for k_z within the window defined by (7.2.30). However, this window can be made even smaller with the following consideration.

A substrate backed by a ground plane has a fundamental TM_0 mode with no cutoff. In order for a mode to be guided by a microstrip line, the mode on the microstrip line must have a slower phase velocity than the TM_0 mode [16]. If this TM_0 mode has a wave number $k_{sp} = \sqrt{k_{xp}^2 + k_{zp}^2}$, then

$$k_0 < k_{sp} < k_z < k_1. \tag{7.2.31}$$

$k_{sp} > k_0$, because TM_0 mode is a guided mode, and its phase velocity must be slower than that of air. Equation (7.2.31) leads to a narrower search window for k_z. Equation (7.2.31) ensures that the microstrip mode does not leak energy to the fundamental TM_0 mode. Moreover, there are guided modes by a ground-plane backed dielectric substrates. These modes can

potentially take energy away from the modes guided by the microstrip line. The above condition also prevents the leakage of energy from the guided microstrip line modes to the dielectric substrate modes.

Since a microstrip line falls into the class of inhomogeneously filled waveguides, the variational method developed in Chapters 3 and 6 together with the finite element method can be used to solve for the wavenumber k_z's of the guided mode. In this case, the partial differential equation is converted into a matrix eigenvalue problem where the eigenvalues are k_z^2, and they can be found explicitly, rather than through a root searching method above.

7.3 Microstrip Patch Revisited

Armed with the spectral domain technique, we are well equipped to analyze the microstrip patch over a layered medium [27]. To this end, we let (here, we assume that the z axis to be vertical as is usually the case)

$$E_{0z}(\mathbf{r}) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} dk_x dk_y e^{ik_x x + ik_y y}\tilde{E}_{0z}(k_x, k_y, z) \tag{7.3.1}$$

$$H_{0z}(\mathbf{r}) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} dk_x dk_y e^{ik_x x + ik_y y}\tilde{H}_{0z}(k_x, k_y, z) \tag{7.3.2}$$

Substituting the above into the wave equation implies that

$$\left(\frac{d^2}{dz^2} + k_z^2\right)\tilde{E}_{0z}^2(k_x, k_y, z) = 0 \tag{7.3.3}$$

$$\left(\frac{d^2}{dz^2} + k_z^2\right)\tilde{H}_{0z}^2(k_x, k_y, z) = 0 \tag{7.3.4}$$

where $k_z^2 = k_0^2 - k_x^2 - k_y^2$. Hence, one can write

$$E_{0z}(\mathbf{r}) = \pm\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} dk_x dk_y e^{i\mathbf{k}_s\cdot\mathbf{r}_s + ik_z|z|}e_{0z}(\mathbf{k}_s) \tag{7.3.5}$$

$$H_{0z}(\mathbf{r}) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} dk_x dk_y e^{i\mathbf{k}_s\cdot\mathbf{r}_s + ik_z|z|}h_{0z}(\mathbf{k}_s) \tag{7.3.6}$$

The $\pm$ sign indicates the odd symmetry of $E_{0z}(\mathbf{r})$ about $z = 0$ plane. When the patch is placed on top of a layered medium, the above becomes

$$E_{0z}(\mathbf{r}) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} d\mathbf{k}_s e^{i\mathbf{k}_s\cdot\mathbf{r}_s}e_{0z}(\mathbf{k}_s)\left[\pm e^{ik_z|z|} - \tilde{R}^{TM}e^{ik_z(z+d)+ik_z|d|}\right] \tag{7.3.7}$$

$$H_{0z}(\mathbf{r}) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} d\mathbf{k}_s e^{i\mathbf{k}_s\cdot\mathbf{r}_s}h_{0z}(\mathbf{k}_s)\left[e^{ik_z|z|} + \tilde{R}^{TE}e^{ik_z(z+d)+ik_z|d|}\right] \tag{7.3.8}$$

The phase factors for the reflected waves are appropriately chosen so that $\tilde{R}^{TM}$ and $\tilde{R}^{TE}$ represent the reflection coefficients.

Using the fact that

$$\hat{\mathbf{E}}_s(\mathbf{r}) = \frac{1}{k_s^2}\left[\frac{\partial}{\partial z}\hat{E}_z - i\omega\mu_0\hat{z}\times\nabla_s\hat{H}_z\right] \tag{7.3.9}$$

$$\hat{\mathbf{H}}_s(\mathbf{r}) = \frac{1}{k_s^2}\left[\frac{\partial}{\partial z}\hat{H}_z + i\omega\epsilon_0\hat{z}\times\nabla_s\hat{E}_z\right] \tag{7.3.10}$$

and applying the above to the integrands of spectral integrals, we obtain

$$\mathbf{E}_{0s}(\mathbf{r}) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} d\mathbf{k}_s e^{i\mathbf{k}_s\cdot\mathbf{r}_s}\frac{1}{k_s^2}\big[-k_z\mathbf{k}_s e_{0z}(\mathbf{k}_s) + \omega\mu_0\hat{z}\times\mathbf{k}_s h_{oz}(\mathbf{k}_s)\big]e^{k_z|z|} \tag{7.3.11}$$

$$\mathbf{H}_{0s}(\mathbf{r}) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} d\mathbf{k}_s e^{i\mathbf{k}_s\cdot\mathbf{r}_s}\frac{1}{k_s^2}\big[\mp k_z\mathbf{k}_s h_{0z}(\mathbf{k}_s) \mp \omega\epsilon_0\hat{z}\times\mathbf{k}_s e_{oz}(\mathbf{k}_s)\big]e^{k_z|z|} \tag{7.3.12}$$

Using the fact that

$$\mathbf{J}_{0s} = \hat{z}\times(\mathbf{H}_{0s+} - \mathbf{H}_{0s-})|_{z=0} \tag{7.3.13}$$

the above can be written as

$$\mathbf{J}_{0s}(\mathbf{r}_s) = \iint_{-\infty}^{\infty} d\mathbf{k}_s\; e^{i\mathbf{k}_s\cdot\mathbf{r}_s}\frac{1}{k_s^2}\left[2\omega\epsilon_0\mathbf{k}_s e_{0z} - 2k_z\hat{z}\times\mathbf{k}_s h_{0z}\right] \tag{7.3.14}$$

The above can be written using vector Fourier transform [41]

$$\mathbf{J}_{0s}(\mathbf{r}_s) = \iint_{-\infty}^{\infty} d\mathbf{k}_s\; \overline{\mathbf{F}}(\mathbf{k}_s,\mathbf{r}_s)\cdot\mathbf{K}_0(\mathbf{k}_s) \tag{7.3.15}$$

where

$$\overline{\mathbf{F}}(\mathbf{k}_s,\mathbf{r}_s) = \frac{1}{k_s}[\mathbf{k}_s, -\hat{z}\times\mathbf{k}_s]e^{i\mathbf{k}_s\cdot\mathbf{r}_s} = \frac{1}{k_s}\begin{bmatrix}k_x & k_y\\ k_y & -k_x\end{bmatrix}e^{i\mathbf{k}_s\cdot\mathbf{r}_s} \tag{7.3.16}$$

$$\mathbf{K}_0(\mathbf{k}_s) = \frac{1}{k_s}\begin{bmatrix}2\omega\epsilon_0 e_{0z}\\ 2k_z h_{0z}\end{bmatrix} \tag{7.3.17}$$

An inverse vector Fourier transform exists as

$$\mathbf{K}_0(\mathbf{k}_s) = \frac{1}{(2\pi)^2}\iint_{-\infty}^{\infty} d\mathbf{r}_s\; \overline{\mathbf{F}}(\mathbf{k}_s,-\mathbf{r}_s)\cdot\mathbf{J}_{0s}(\mathbf{r}_s) \tag{7.3.18}$$

We can also write

$$\mathbf{E}_{0s} = \iint_{-\infty}^{\infty} d\mathbf{k}_s\; \overline{\mathbf{F}}(\mathbf{k}_s,\mathbf{r}_s)\cdot\overline{\mathbf{G}}(k_z,z)\cdot\overline{\mathbf{K}}_{\mathbf{0}}(\mathbf{k}_s) \tag{7.3.19}$$

where

$$\overline{\mathbf{G}}(k_z) = -\begin{bmatrix}\frac{k_z}{2\omega\epsilon_0} & 0\\ 0 & \frac{\omega\mu_0}{2k_z}\end{bmatrix}e^{ik_z|z|} \tag{7.3.20}$$

The above constitutes a compact way to express the field and current in terms of their spectral domain quantities.

7.3.1 Integral Equation for the Resonance Case

The resonance of a microstrip patch has been calculated before [18, 19, 23, 27]

$$\mathbf{E}_{0s}(\mathbf{r}_s, z=0) = 0 = \iint_{-\infty}^{\infty} d\mathbf{k}_s \; \overline{\mathbf{F}}(\mathbf{k}_s, \mathbf{r}_s) \cdot \overline{\mathbf{G}}(k_z, z=0) \cdot \overline{\mathbf{K_0}}(\mathbf{k}_s), \quad \mathbf{r}_s \in S_P \tag{7.3.21}$$

$$\mathbf{J}_0(\mathbf{r}_s) = 0, \qquad \mathbf{r}_s \notin S_P \tag{7.3.22}$$

where S_P is the patch surface. By using Galerkin's method, the above can be converted into a matrix equation

$$\overline{\mathbf{A}}(\omega) \cdot \mathbf{a} = 0 \tag{7.3.23}$$

A non-trivial $\mathbf{a}$ exists only if

$$\det\left(\overline{\mathbf{A}}(\omega)\right) = 0 \tag{7.3.24}$$

The above equation can be solved with a zero-searching method to obtain the resonance frequencies of the microstrip patch.

7.3.2 Integral Equation for the Excitation Case

If the patch is excited by an incident field, say due to a probe source, then the integral equation for excitation is [7, 20]:

$$\hat{n} \times [\mathbf{E}_{0s}(\mathbf{r}_s, z=0) + \mathbf{E}_{inc}(\mathbf{r}_s, z=0)] = 0 \tag{7.3.25}$$

or after using (7.3.21)

$$-\hat{n} \times \mathbf{E}_{inc}(\mathbf{r}_s, z=0) = \hat{n} \times \iint_{-\infty}^{\infty} d\mathbf{k_s} \overline{\mathbf{F}}(\mathbf{k_s}, \mathbf{r}_s) \cdot \overline{\mathbf{G}}(\mathbf{k_z}, z=0) \cdot \mathbf{K}_0(\mathbf{k}_s) \tag{7.3.26}$$

Using Galerkin's method, the above can be converted to a matrix equation:

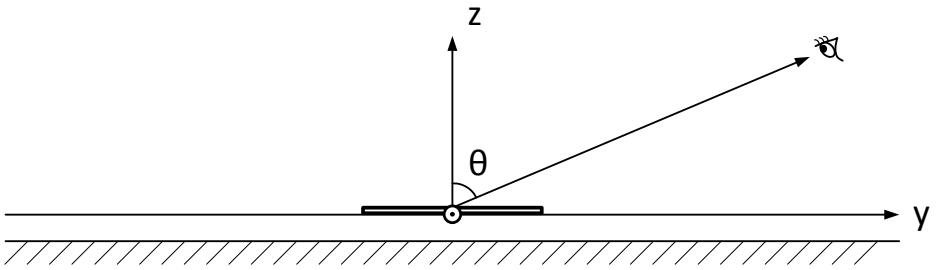
$$\overline{\mathbf{A}} \cdot \mathbf{a} = \mathbf{b} \tag{7.3.27}$$

one can solve the above to obtain the current on the patch, and then calculate the field everywhere.

7.3.3 Far Field Calculation

The far field of the Fourier integral given in (7.3.7) and (7.3.8) can be found using the stationary phase method [33]. We will assume that h is small so that when $(x, y, z) \to \infty$, the dominant variation of the integrand comes from $e^{i\mathbf{k}_s \cdot \mathbf{r}_s}$ and $e^{ik_z z}$. To this end, we can write (7.3.7) with $d = 0$ as

$$E_{0z}(\mathbf{r}) = \iint_{-\infty}^{\infty} d\mathbf{k}_s e^{i\mathbf{k}_s \cdot \mathbf{r}_s + ik_z z} e_{oz}(k_s) \left(1 - \tilde{R}^{TM}\right) \tag{7.3.28}$$



The exponential function is given by

$$e^{ik_x x+ik_y y+ik_z z} \tag{7.3.29}$$

The stationary phase point is given by

$$\frac{\partial}{\partial k_x}\left[k_x x+k_y y+k_z z\right]=0, \qquad \frac{\partial}{\partial k_y}\left[k_x x+k_y y+k_z z\right]=0 \tag{7.3.30}$$

yielding

$$x-\frac{k_x}{\sqrt{k^2-k_x^2-k_y^2}}z=0, \qquad y-\frac{k_y}{\sqrt{k^2-k_x^2-k_y^2}}z=0 \tag{7.3.31}$$

If we set

$$k_x=k_0\sin\theta\cos\phi \quad k_y=k_0\sin\theta\sin\phi \tag{7.3.32}$$

then

$$k_z=k_0\cos\theta \tag{7.3.33}$$

and (7.3.31) is satisfied. Therefore the above is the stationary phase point of the integrand. Most of the contribution of the integral to (7.3.28) will come from this point. Hence, we can approximate the slowly varying part of the integrand by its value at this point and rewrite (7.3.28) approximately as

$$E_{0z}\left(\mathbf{r}\right)\sim k_{zs}e_{0z}\left(k_{xs},k_{ys}\right)\left[1-\tilde{R}^{TM}\left(k_{xs},k_{ys}\right)\right]\int\int_{-\infty}^{+\infty}d\mathbf{k}_s\frac{e^{i\mathbf{k}_s\cdot\mathbf{r}_s+ik_z z}}{k_z} \tag{7.3.34}$$

where k_{xs}, k_{ys}, and k_{zs} are given by (7.3.32) and (7.3.33). Making use of the Weyl identity,

$$\frac{e^{ik_0 r}}{r}=\frac{i}{2\pi}\int\int_{-\infty}^{+\infty}d\mathbf{k}_s\frac{e^{i\mathbf{k}_s\cdot\mathbf{r}_s+ik_z|z|}}{k_z} \tag{7.3.35}$$

(7.3.34) becomes

$$E_{0z}(\mathbf{r}) \sim \frac{2\pi}{i} k_{zs} e_{0z}(k_{xs}, k_{ys}) \left[1 - \tilde{R}^{TM}(k_{xs}, k_{ys})\right] \frac{e^{ik_0 r}}{r} \tag{7.3.36}$$

By the same token,

$$H_{0z}(\mathbf{r}) \sim \frac{2\pi}{i} k_{zs} h_{0z}(k_{xs}, k_{ys}) \left[1 + \tilde{R}^{TE}(k_{xs}, k_{ys})\right] \frac{e^{ik_0 r}}{r} \tag{7.3.37}$$

The far field appears as a spherical wave. Hence, the electric field of the far field is of the form

$$\mathbf{E} \approx \hat{\theta} E_\theta + \hat{\phi} E_\phi \tag{7.3.38}$$

with

$$\mathbf{H} \approx -\hat{r} \times \mathbf{E}/\eta = \frac{1}{\eta}\left(\hat{\phi} E_\theta - \hat{\theta} E_\phi\right) \tag{7.3.39}$$

From the above, one gathers that

$$E_{0z} \approx -E_\theta \sin\theta, \quad H_{0z} \approx \frac{1}{\eta} E_\phi \sin\theta \tag{7.3.40}$$

Hence, we can get (E_θ, E_ϕ) from (E_{0z}, H_{0z}) in the far field and compute the total radiation power.

7.4 Edge Condition

In many numerical and analytic methods, it is useful to know how the charge and the current behave near the sharp edge of a geometry. This will help in the choice of the correct basis functions to improve the convergence of the numerical methods. Also, the edge condition has to be imposed for a boundary value problem in order to guarantee uniqueness of the solution. This problem has been studied by Rayleigh, Meixner, Maue, Jones, and Silver and Heins, Hayashi, and many more. References can be found in Heins and Silver [44], and Hayashi [45].

The field near a sharp edge is singular because of charge accumulation into a singular point. The singular behavior can be ascertained by solving Laplace's equation in the vicinity of the sharp edge. This is because the singular behavior of the field is entirely a local phenomenon or local geometry dependent where the spatial variation dominates over temporal variation.

Consider a wedge with an angle γ as shown. The Φ that satisfies Laplace's equation $\nabla^2 \Phi = 0$ is given by

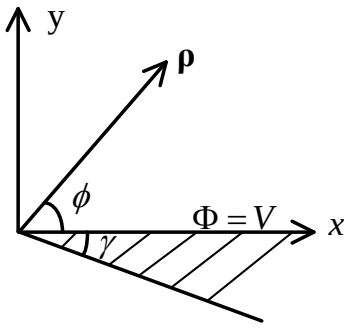
$$\Phi = A\rho^{\pm\alpha} e^{\pm in\alpha}, \qquad \alpha > 0 \tag{7.4.1}$$

Using the fact that

$$\nabla^2 = \frac{1}{\rho}\frac{\partial}{\partial\rho}\rho\frac{\partial}{\partial\rho} + \frac{1}{\rho^2}\frac{\partial^2}{\partial\phi^2} \tag{7.4.2}$$

in cylindrical coordinates, it can be easily shown that $\nabla^2\Phi = 0$ by direct substitution. A solution that satisfies the boundary condition that $\Phi = V$ on the wedge surface is

$$\Phi = V + a\rho^\alpha \sin(\alpha\phi) \tag{7.4.3}$$



γ	α
0	1/2
$\pi/2$	2/3
π	1
$3\pi/2$	2

Table 7.1: The singularities associated with different wedge angles.

with that requirement that $\sin[\alpha(2\pi-\gamma)]=0$, or that

$$\alpha(2\pi-\gamma)=\pi, \qquad \alpha=\frac{\pi}{2\pi-\gamma} \tag{7.4.4}$$

for the minimal α and that $\sin(\alpha\phi)>0$ for $0<\phi<2\pi-\gamma$.

The $\rho^{-\alpha}$ term in (7.4.1) can be ignored since $|\nabla\Phi|^2$ has to be square integrable because it represents the finite stored energy in the field. Moreover, one picks minimal α so that the electric field $\mathbf{E}=-\nabla\Phi$ given by (7.4.3) is simple and does not have scallop patterns.

Hence for the different wedge angles, the singularities are as shown in Table 7.1. For a thin wedge with $\gamma=0$,

$$\Phi=V+a\rho^{\frac{1}{2}}\sin\left(\frac{\phi}{2}\right) \tag{7.4.5}$$

the corresponding electrostatic field

$$\mathbf{E}=-\nabla\Phi=-\hat{\rho}\frac{a}{2}\rho^{-\frac{1}{2}}\sin\left(\frac{\phi}{2}\right)+\hat{\phi}\frac{a}{2}\rho^{-\frac{1}{2}}\cos\left(\frac{\phi}{2}\right) \tag{7.4.6}$$

Since the charge is proportional to $\hat{n}\cdot\mathbf{E}$, the surface charge density σ is such that

$$\sigma\sim x^{-\frac{1}{2}} \tag{7.4.7}$$

near the edge. If a current flows in the x direction, from $\nabla\cdot\mathbf{J}=i\omega\varrho$, the current

$$J_x\sim x^{\frac{1}{2}} \tag{7.4.8}$$

near the edge. If the current is flowing in the z direction into the paper, then

$$J_z = \sigma v_z \tag{7.4.9}$$

Hence,

$$J_z \sim x^{-\frac{1}{2}} \tag{7.4.10}$$

This current singularity, for instance, is observed in the axial component of the current on a microstrip line.

7.5 Discontinuities in Microstrip Lines

Microstrip lines are some of the easiest waveguides to fabricate. However, they are also some of the hardest to analyze. Due to the ease in their fabrication, microwave integrated circuits have come in a rich variety of shapes. One example is the variety of microstrip line discontinuities. These discontinuities are in general hard to analyze. Usually, numerical methods have to be employed for their analysis. This is especially true for the full-wave analysis whereby all electrodynamic effects are accounted for. Much work on this topic can be found in [46–56]

Microstrip line discontinuities are often modeled by lumped elements in the transmission line model. In the following, we shall discuss some typical discontinuities encountered in microwave integrated circuits.

Figure 7.12: An open-end discontinuity.

7.5.1 An Open-End Discontinuity

A simple way to make an open circuit in a microstrip line is to let the microstrip line terminate in an open end. Open circuits are often used as a stub tuner in microwave integrated circuits. The abrupt termination of a microstrip line intercepts the flow of current on the line. This gives rise to charge accumulation, which can be modeled by a capacitance. The equivalent circuit for an open end microstrip line is then a transmission line terminated by a capacitor, or an open circuited transmission line whose length is slightly longer than the physical length of the microstrip line. The capacitor is also called a fringing field capacitor, as the excess charge at the termination gives rise to excess fringing field at the end of the line. These fringing capacitance can be found by numerical calculations which are sometimes replaced by analytical or semi-empirical formulas.

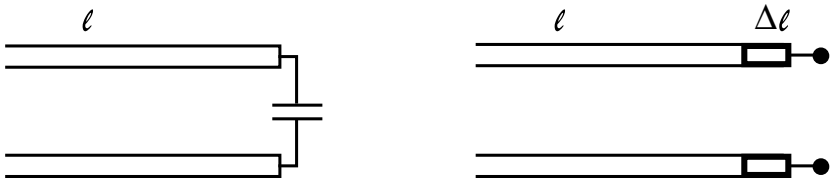


Figure 7.13: Equivalent circuit models for an open end microstrip line.

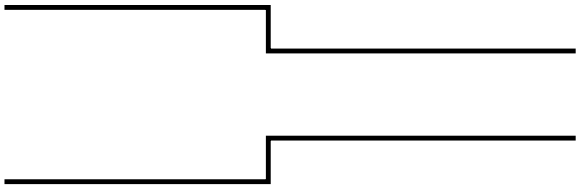


Figure 7.14: A step discontinuity in a microstrip line.

7.5.2 A Step Discontinuity

Another commonly encountered discontinuity in a microstrip line is a step discontinuity. A step discontinuity has excess charges on its excesss edges, and hence can be modeled by a shunt capacitance. This kind of discontinuity is encountered in the design of a quarter-wave transformer for instance. A more sophisticated model consists of two series inductors as well as a shunt capacitor.

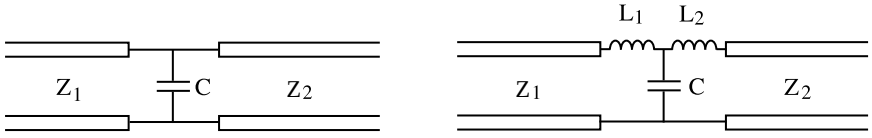


Figure 7.15: Equivalent circuit models for a step discontinuity.

7.5.3 A Gap Discontinuity

A gap discontinuity is deliberately introduced in a microstrip line, for example, in filter design. A gap discontinuity can be thought of as two open end discontinuities brought close together. The proximity of two open end discontinuities results in capacitive coupling between the two open ends. Therefore, the equivalent circuit consists of two shunt capacitors and a series capacitor.

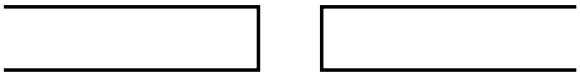


Figure 7.16: A gap discontinuity.

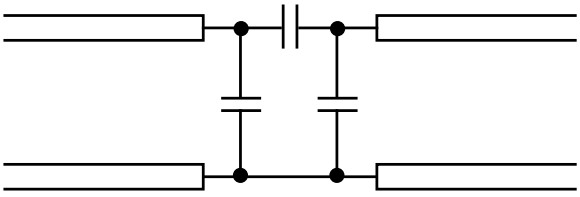


Figure 7.17: The equivalent circuit for a gap discontinuity.

7.5.4 A Slit Discontinuity

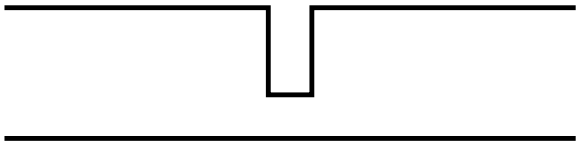


Figure 7.18: A slit discontinuity.

A slit discontinuity has very little effect on a low frequency signal, but it will scatter a high frequency signal. At low frequency, the current can flow continuously along the line, and sees a matched line at the other end. At higher frequency, the restricted channel is felt and the slit reflects a signal impinging on it. Hence, it acts as a low-pass filter; or it can be modeled by an T equivalent circuit with two series inductances and a shunt capacitance.

7.5.5 A Microstrip Bend

A microstrip bend can be a corner bend or a chamfered bend. Both microstrip bends can be modeled by a T circuit with two series inductors and a shunt capacitor. A corner bend has a sharp corner favoring the build up of excessive charges giving rise to a larger shunt capacitance. A chamfered bend reduces this capacitance and mismatch, even less so than a rounded bend.

All guided waves around a bend in an open waveguide radiate. A chamfered bend has been found to reduce the radiation loss as well.

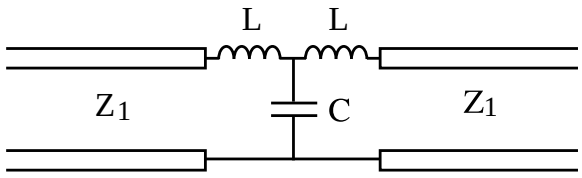


Figure 7.19: The equivalent circuit for a slit discontinuity.

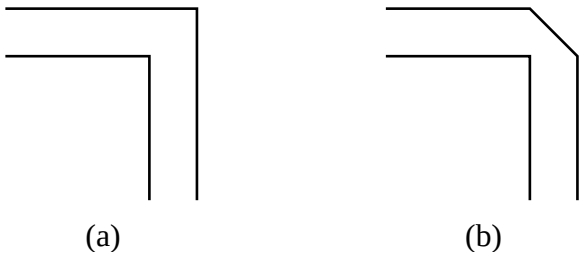


Figure 7.20: (a) A corner bend. (b) A chamfered bend.

7.5.6 A T Junction

A T junction in microwave integrated circuits is commonly encountered, e.g., in a stub tuner. It can be modeled by series inductances plus a shunt capacitance. A cross can also be similarly modeled.

7.6 Directional Coupler Using Microstrip Line

Due to the simplicity of the microstrip waveguide, a new class of directional coupler has been proposed. Some of these papers can be found in [57–59]. When two microstrip lines are aligned parallel to each other, electromagnetic coupling exists between the two lines. Since one line has one fundamental mode, two lines would have two fundamental modes. For symmetric lines, these two fundamental modes are the odd and the even modes. For two identical lines, the electric field for the odd mode is odd-symmetric about the plane of symmetry; while for the even mode, it is even-symmetric about the same plane. Hence, we can use the image theorem to analyze such a problem: A PMC wall can be placed at the plane of symmetry for the even symmetric mode, while a PEC wall can be placed at the same plane for the odd symmetric mode.

The directional coupler is a four-port network. We can analyze it as a linear superposition of two two-port network, one with a PMC at AA', and another one with a PEC at AA'. The PMC case is equivalent to an even-mode excitation of the geometry. This is the same as, say, an incident wave of V^+ at both ports 1 and 3 of the network. The PEC case corresponds to an odd-mode excitation, which is the same as a V^+ incident wave at port 1 and a $-V^+$

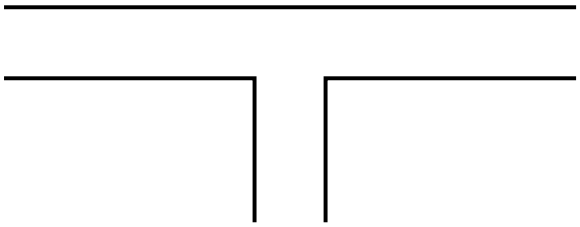


Figure 7.21: A T junction in microwave integrated circuits.

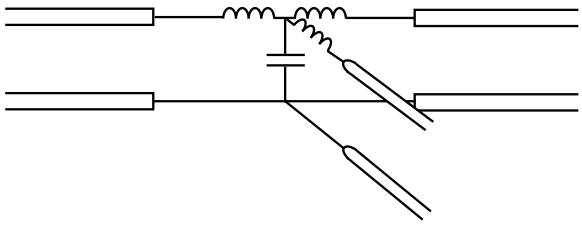


Figure 7.22: Equivalent circuit for a T junction.

incident wave at port 3. The linear superposition of these two excitations correspond to a $2V^+$ excitation at port 1 and none at the other ports.

The PMC case corresponds to only even mode propagation, and we can use the bounce diagram approach to write down the reflected waves at ports 1 and 2. We assume that the fundamental mode from the input port will undergo a reflection when it enters the coupled mode region. Hence, the reflection coefficient at port 1 can be expressed as the S^e_{11} coefficient given as

$$S^e_{11} = R^e_{01} + \frac{T^e_{01} e^{2ik_e l} R^e_{10} T^e_{10}}{1 - (R^e_{10})^2 e^{2ik_e l}}. \tag{7.6.1}$$

The transmission coefficient from port 1 to port 2 can be expressed as

$$S^e_{12} = \frac{T^e_{01} T^e_{10} e^{ik_e l}}{1 - (R^e_{10})^2 e^{2ik_e l}}. \tag{7.6.2}$$

In the above

$$R^e_{01} = \frac{Z_e - Z_c}{Z_e + Z_c}, \quad R^e_{10} = -R^e_{01}, \tag{7.6.3a}$$

$$T^e_{01} = 1 + R^e_{01}, \quad T^e_{10} = 1 + R^e_{10}, \tag{7.6.3b}$$

where Z_e is the characteristic impedance of the even mode while Z_c is the characteristic impedance of the feedline.

For the PEC case, a similar analysis shows that

$$S^o_{11} = R^o_{01} + \frac{T^o_{01} e^{2ik_o l} R^o_{10} T^o_{10}}{1 - (R^o_{10})^2 e^{2ik_o l}}, \tag{7.6.4}$$

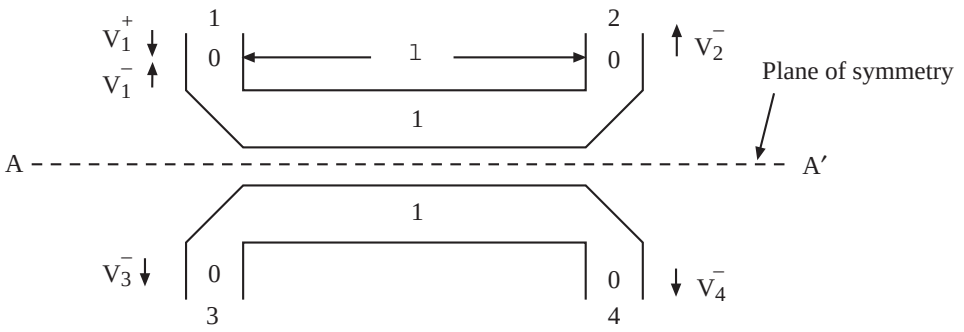


Figure 7.23: A directional coupler made of microstrip lines.

$$S_{12}^o = \frac{T_{01}^o T_{10}^o e^{ik_o l}}{1 - (R_{10}^o)^2 e^{2ik_o l}}, \tag{7.6.5}$$

where

$$R_{01}^o = \frac{Z_o - Z_c}{Z_o + Z_c}, \tag{7.6.6}$$

and so on.

The key to the design of the directional coupler is to let Z_c be the geometric mean of Z_o and Z_e, i.e.,

$$Z_c = \sqrt{Z_o Z_e}. \tag{7.6.7}$$

In this case, $R_{01}^o = -R_{01}^e$. Furthermore, if we make $k_e = k_o$, then it can be shown that

$$S_{11}^o = -S_{11}^e. \tag{7.6.8}$$

Superposing the even and the odd solutions, we obtain that $V_1^+ = 2V^+$, and $V_1^- = (S_{11}^e + S_{11}^o)V^+ = 0$. Similarly, the transmitted wave at port 4 is given by

$$V_4^- = S_{12}^e V^+ - S_{12}^o V^+, \tag{7.6.9}$$

when $k_e = k_o$, and that (7.6.7) is satisfied, it is clear that $T_{01}^e T_{10}^e = T_{01}^o T_{10}^o$. Therefore, $S_{12}^e = S_{12}^o$ and from (7.6.9), we see that $V_4^- = 0$. At port 3, we have

$$V_3^- = S_{11}^e V^+ - S_{11}^o V^+ = \frac{1}{2}(S_{11}^e - S_{11}^o)V_1^+ = S_{13}V_1^+, \tag{7.6.10}$$

where

$$S_{13} = \frac{1}{2}(S_{11}^e - S_{11}^o) = S_{11}^e. \tag{7.6.11}$$

Similarly, we have

$$S_{12} = \frac{1}{2}(S_{12}^e + S_{12}^o) = S_{12}^e. \tag{7.6.12}$$

Therefore, the above works as a perfect directional coupler, where nothing is coupled to port 4 from port 1, while some signal is transmitted to port 2 and port 3. If the coupled

line is in a homogeneous region, it is clear that both the odd and the even modes are TEM modes, and hence $k_e = k_o = k$. However, if the coupled line is fabricated on a substrate as in the case of a microstrip line, then $k_e \neq k_o$ because the field distributions of the two modes are different. However, k_e can be made to be close to k_o by covering the microstrip line with a superstrate with the same permittivity as the substrate.

The directivity of a directional coupler is the ratio of the $\mid V_3^- \mid$ to $\mid V_4^- \mid$. In the ideal case, it is infinite. The coupling coefficient is the ratio of $\mid V_3^- \mid$ to $\mid V_1^+ \mid$. In this case, it is

$$C = \mid S_{11}^e \mid = \left| R_{01}^e + \frac{T_{01}^e e^{2ik_e l} R_{10}^e T_{10}^e}{1 - (R_{10}^e)^2 e^{2ik_e l}} \right|. \tag{7.6.13}$$

Using the fact that $T_{01}^e T_{10}^e = 1 - (R_{01}^e)^2$, we can rewrite S_{11}^e as

$$S_{11}^e = R_{01}^e + \frac{R_{10}^e e^{2ik_e l}}{1 - (R_{10}^e)^2 e^{2ik_e l}} = \frac{R_{01}^e (1 - e^{2ik_e l})}{1 - (R_{10}^e)^2 e^{2ik_e l}}. \tag{7.6.14}$$

$\mid S_{11}^e \mid$ achieves a maximum at $k_e l \simeq \pi/2$ or $l \simeq \lambda/4$.

7.7 A Branch Line Directional Coupler

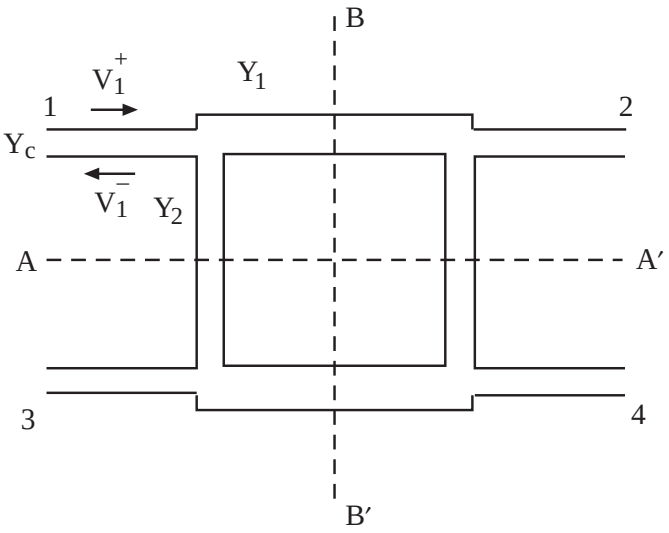


Figure 7.24: A branch line directional coupler.

A branch line coupler, like the microstrip line directional coupler, has two planes of symmetry. Even though a microstrip line directional coupler has two planes of symmetry, one can analyze it using only one plane of symmetry. However, for a branch line coupler, it is simpler to analyze using two planes of symmetry.

On the plane AA', one can put either a PMC or a PEC. A PMC yields an open circuit at AA' or an even symmetric electric field or voltage about the AA'. A PEC yields a closed circuit at AA' or an odd symmetric electric field or voltage about AA'. The same statement applies to the plane BB'. Therefore, there are altogether four possible excitations of the above geometry.

Bibliography

[1] J. Koch, T. M. Yu, J. Gambetta, A. A. Houck, D. I. Schuster, J. Majer, A. Blais, M. H. Devoret, S. M. Girvin, and R. J. Schoelkopf, "Charge-insensitive qubit design derived from the Cooper pair box," *Phys. Rev. A*, 76, no. 4 (2007): 042319.

[2] H. A. Wheeler, "Transmission-line properties of parallel strips separated by a dielectric sheet," *IEEE Trans. Microwave Theory Tech.*, vol. MTT-13. pp. 172-185, Mar. 1965.

[3] T. G. Bryant and J. A. Weiss, "Parameters of microstrip transmission lines and of coupled pairs of microstrip lines," *IEEE Trans. Microwave Theory Tech.*, vol. MTT-16. pp. 1021-1027. Dec. 1968.

[4] M. V. Schneider, "Microstrip lines for microwave integrated circuits," *Bell Syst. Tech. J.*, vol. 48, pp. 1421-1444. May/June 1969.

[5] H. E. Green, "The numerical solution of some important transmission-line problems," *IEEE Trans. Microwave Theory Tech.*, vol. MTT-13, pp. 676-692, Sept. 1965.

[6] E. Yamashita and R. Mittra, "Variational method for the analysis of microstrip lines," *IEEE Trans. Microwave Theory Tech.*, vol. MTT-16. PP- 251-256, Apr. 1968.

[7] E. Yamashita, "Variational method for the analysis of microstrip-like transmission lines," *IEEE Trans. Microwave Theory Tech.*, vol. MTT-16. pp. 529-535, Aug. 1968.

[8] H. E. Stinehelfer, Sr., "An accurate calculation of uniform microstrip transmission lines," *IEEE Trans. Microwave Theory Tech.*, vol. MTT-16, pp. 439-444, July 1968.

[9] G. I. Zysman and D. Varon, "Wave propagation in microstrip transmission lines." *IEEE G-MTT Int. Microwave Symp. Dig.*, pp. 3-9, 1969.

[10] E. J. Denlinger, "A frequency dependent solution for microstrip transmission lines," *IEEE Trans. Microwave Theory Tech.*, vol. MTT-19, pp. 30-39, Jan. 1971.

[11] J. S. Hornsby and A. Gopinath, "Numerical analysis of a dielectric-loaded waveguide with a microstrip line—Finite difference methods," *IEEE Trans. Microwave Theory Tech.*, vol. MTT-17, pp. 684-690, Sept. 1969.

[12] H. J. Schmitt and K. H. Sarges, "Wave propagation in microstrip," *Nachrichtentech. Z.*, vol. 24, pp. 260-264, May 1971.

[13] R. Mittra and T. Itoh, "A new technique for the analysis of dispersion characteristics of microstrip lines," *IEEE Trans. Microwave Theory Tech.*, vol. MTT-19, pp. 47-56, Jan. 1971.

[14] R. H. Jansen, "High-speed computation of single and coupled microstrip parameters including dispersion, high-order modes, loss and finite strip thickness," *IEEE Trans. Microwave Theory Tech.*, vol. MTT-26, pp. 75-82, Feb. 1978.

[15] M. Kobayashi and F. Ando, "Dispersion characteristics of open microstrip lines," *IEEE Trans. Microwave Theory Tech.*, vol. MTT-35, pp. 101-105. Feb. 1987.

[16] L. Gürel and W. C. Chew, "Guidance or resonance conditions for strips or disks embedded in homogeneous and layered media," *IEEE Trans. Microwave Theory Tech.*, vol. 36, no. 11, pp. 1498-1506, Nov. 1988.

[17] L. Gürel, "Microstrip transmission line with finite ground plane," M. S. thesis, University of Illinois, Urbana. ll, 1988.

[18] W. C. Chew and J. A. Kong, "Resonance of the axial-symmetric modes in microstrip disk resonators," *J. Math. Phys.*, vol. 21, no. 3, pp. 582-591, Mar. 1980.

[19] W. C. Chew and J. A. Kong, "Resonance of nonaxial symmetric modes in microstrip disk antenna." *J. Math. Phys.*, vol. 21, no. 10, pp. 2590-2598. Oct. 1980.

[20] W. C. Chew and J. A. Kong, "Analysis of a circular microstrip disk antenna with a thick dielectric substrate," *IEEE Trans. Antennas Propagat.*, vol. AP-29, pp. 68-76, Jan. 1981.

[21] R. H. Jansen. "The spectral-domain approach for microwave integrated circuits," *IEEE Trans. Microwave Theory Tech.*, vol. MTT-33. pp. 1043-1056. Oct. 1985.

[22] T. Itoh and R. Mittra, "Spectral-domain approach for calculating the dispersion characteristics of rnicrostrip lines," *IEEE Trans.Microwave Theory Tech.*, vol. MTT-21. PP. 496-499. July 1973.

[23] T. Itoh, "Analysis of microstrip resonators," *Microwave Theory and Techniques, IEEE Transactions on*, 22, no. 11 (1974): 946-952.

[24] R. H. Jansen, "Microstrip lines with partially removed ground metallization, theory and applications," *Arch. Elek. Uberrragung*, vol. 32, pp. 485-492, Dec. 1978.

[25] T. Itch, "Spectral domain immittance approach for dispersion characteristics of generalized transmission lines," *IEEE Trans.Microwave Theory Teoh.*, vol. MTT-28, pp. 733-736, July 1980.

[26] D. Pozar, "Input impedance and mutual coupling of rectangular microstrip antennas." *Antennas and Propagation, IEEE Transactions on*, 30, no. 6 (1982): 1191-1196.

[27] W. C. Chew and Q. Liu. "Resonance frequency of a rectangular microstrip patch," *IEEE Trans. Antennas Propagat.*, vol. 36, Aug.1988.

[28] G. Kergonou, M. Drissi, T. Zak, and C. Xavier, "Frequency dependence in high speed interconnections," *Electromagnetic Compatibility, 2001. EMC. 2001 IEEE International Symposium on*, vol. 1, pp. 632-634. IEEE, 2001.

[29] R. A. York, R. C. Compton, and B. J. Rubin. "Experimental verification of the 2-D rooftop approach for modeling microstrip patch antennas," *Antennas and Propagation, IEEE Transactions on* 39, no. 5 (1991): 690-694.

[30] S. Sun and L. Zhu, "Compact dual-band microstrip bandpass filter without external feeds," *Microwave and Wireless Components Letters, IEEE* 15, no. 10 (2005): 644-646.

[31] C. M. Bender and S. A. Orszag, *Advanced mathematical methods for scientists and engineers I: Asymptotic methods and perturbation theory*, Vol. 1. Springer, 1999.

[32] S. Y. Poh, W. C. Chew. and J. A. Kong, "Approximate formulas for line capacitance and characteristic impedance of microstrip line," *IEEE Trans. Microwave Theory Tech.*, vol. MTT-29, pp.135-142, Feb. 1981.

[33] W. C. Chew, *Waves and Fields in Inhomogeneous Media*, Van Nostrand Reinhold, New York, 1990. Reprinted by IEEE Press, 1995.

[34] J. A. Kong, *Electromagnetic Wave Theory*. New York: Wiley, 1986.

[35] J. D. Jackson, *Classical Electrodynamics*, Wiley, New York, NY, 1962, p. 51.

[36] J. J. Thomson, footnote on p. 154 of Maxwells Treatise On Electricity and Magnetism, (Dover, New York, NY, 1954).

[37] Y. L. Chow, Y. F. Lan, and D. G. Fang, "Capacitance and its upper and lower bounds by the method of optimized simulated images," *J. Appl. Phys.*, 53, no. 11 (1982): 7144-7148.

[38] W. E. Parr, "Upper and lower bounds for the capacitance of the regular solids," *J. Soc. Indust. Appl. Math.*, 9, no. 3 (1961): 334-386.

[39] N. Wiener and E. Hopf, "Ueber eine Klasses singulärer Integralgleichungen," *S.-B. Deutsch. Akad. Wiss. Berlin KI. Math. Phys. Tech.*, pp. 696-706 1931.

[40] W. C. Chew, "Mixed boundary value problems in microstrip and geophysical probing applications," *Ph.D. Thesis*, Massachusetts Institue of Technology, Cambride, MA, 1980.

[41] W. C. Chew and T. M. Habashy, "The use of vector transforms in solving some electromagnetic scattering problems," *IEEE Trans. Antennas Propagat.*, vol. AP-34, pp. 871-879, July 1986.

[42] W. C. Chew and L. Gürel, "Reflection and transmission operators for strips or disks embedded in homogeneous and layered media," *IEEE Trans. Microwave Theory Tech.*, vol. MTT-36, no. 11, pp. 1488-1497, Nov. 1988.

[43] K. C. Gupta, R. Garg, and I. J. Bahl, *Microstrip lines and slotlines,* Artech House, Norwood, MA, 1979.

[44] A. E. Heins and S. Silver, "The edge conditions and field representation theorem in the theory of electromagnetic diffraction, *Proc. Cambridge Phil. Soc.*, Vol. 51, 149161, 1955.

[45] Y. Hayashi, "Electromagnetic theory based on integral representation of fields and analysis of scattering by open boundary," *Progress In Electromagnetics Research*, PIER 13, 186, 1996.

[46] R. Garg and I. J. Bahl, "Microstrip discontinuities," *International Journal of Electronics,* 45:1, 81-87, 1978.

[47] T. S. Chu and T. Itoh. "Generalized scattering matrix method for analysis of cascaded and offset microstrip step discontinuities," *IEEE Trans. MTT*, 34-2: 280-284, 1986.

[48] R. W. Jackson, "Full-wave, finite element analysis of irregular microstrip discontinuities," Microwave Theory and Techniques, *IEEE Trans. MTT,* 37-1: 81-89, 1989.

[49] X. Zhang and K. K. Mei. "Time-domain finite difference approach to the calculation of the frequency-dependent characteristics of microstrip discontinuities," *IEEE Trans. MTT,* 36-12: 1775-1787, 1988.

[50] M. Kirschning, R. H. Jansen, and N. H. L. Koster, "Measurement and computer-aided modeling of microstrip discontinuities by an improved resonator method," *IEEE MTT-S, International Microwave Symposium Digest,* 1983.

[51] C. J. Railton and T. Rozzi. "The rigorous analysis of cascaded step discontinuities in microstrip," *IEEE Trans. MTT,* 36-7, 1177-1185, 1988.

[52] A. K. Verma, H. Singh, and Y. K. Awasthi, "Circuit model of multilayer microstrip step discontinuity using single-layer reduction formulation," *Electromagnetics,* 29-6: 483-498, 2009.

[53] R. J. P. Douville and D. S. James, "Experimental study of symmetric microstrip bends and their compensation," *IEEE Trans. MTT,* 26, no. 3: 175-182, 1978.

[54] J. Moore and H. Ling, "Characterization of a 90 microstrip bend with arbitrary miter via the time-domain finite difference method," *IEEE Trans. MTT*,38, no. 4 (1990): 405-410.

[55] R. Mehran, "Calculation of microstrip bends and Y-junctions with arbitrary angle," *IEEE Trans. Microwave Theory and Techniques,* 6, no. 6 (1978): 400-405.

[56] P. H. Harms and R. Mittra, "Equivalent circuits for multiconductor microstrip bend discontinuities," *IEEE Trans. Microwave Theory Tech.*, 41, no. 1 (1993): 62-69.

[57] F. C. De Ronde, "A new class of microstrip directional couplers," In Microwave Symposium, G-MTT 1970 International, pp. 184-189. IEEE, 1970.

[58] M. Dydyk, "Microstrip directional couplers with ideal performance via single-element compensation," Microwave Theory and Techniques, IEEE Transactions on 47, no. 6 (1999): 956-964.

[59] C. Caloz and T. Itoh, "A novel mixed conventional microstrip and composite right/left-handed backward-wave directional coupler with broadband and tight coupling characteristics," *Microwave and Wireless Components Letters, IEEE*, 14, no. 1 (2004): 31-33.

Chapter 8

Solitons

8.1 Optical Solitons

Dispersions in an optical fiber is the prime reason for limiting its bandwidth. Dispersion causes an optical pulse to distort or spread as it propagates over long distances. Moreover, the loss of the optical fiber also causes the pulse to spread. Therefore, repeaters are needed every several tens of kilometers to rejuvenate the pulses. When the pulses are narrower in order to facilitate higher transmission rates, dispersion effects become even more pronounced and repeaters have to be closely space. The spacing of the repeaters is then inversely proportional to the square of the pulse width [1].

Solitary wave was first observed in water wave by Scott Russel in 1838 [2, 3]. The first mathematical description was first given by Boussinesq [4]. Since then, soliton theory has been avidly studied [5–7].

An optical soliton makes use of the nonlinear effect in an optical fiber to propagate a pulse with no distortion [8, 9]. The nonlinear effect is used to counter the dispersion effect so that the pulse propagates with little or no distortion. The nonlinear effect is proportional to the field strength of a mode in the optical fiber. Hence, loss in the fiber will eventually distort the pulse. However, using the Raman effect in the fiber, optical solitons can be continuously pumped by a lightwave which is simultaneously transmitted through the fiber together with the solitons. Hence, an all optical system without repeaters is possible [10].

The nonlinear property of an optical fiber comes from the Kerr effect which produces a change in the refractive index of glass due to the deformation of electron orbits by the electric field of light [11, 12]. The refractive index of glass is then $n = n_0 + n_2|\mathbf{E}|^2$ where n_2 is the Kerr coefficient, and $\mathbf{E}$ is the electric field. Here, n_2 is usually of the order of $10^{-22}(m/V)^2$ and $\mathbf{E}$ is of the order of $10^6 V/m$. Therefore, the change in the refractive index is about 10^{-10}. Even though this is a small change, the high operating frequency of an optical fiber makes this change significant.

8.2 The Korteweg de Vries Equation

A soliton or a solitary wave is a result of nonlinear phenomena. A nonlinear partial differential equation that admits a soliton as a solution is the KdV (Korteweg de Vries) equation [13,14]. One shall motivate the equation here.

A solution to a linear wave equation has the following form

$$\phi(x,t) = \phi(x - vt) \tag{8.2.1}$$

where v is the velocity of the wave. The above describes a right-traveling wave and it satisfies the equation

$$\frac{\partial \phi}{\partial t} + v\frac{\partial \phi}{\partial x} = 0. \tag{8.2.2}$$

If one observes this wave in a moving coordinate system such that $x' = x - v_0 t$, then equation (8.2.1) becomes

$$\phi(x',t) = \phi[x' - (v - v_0)t]. \tag{8.2.3}$$

and

$$\frac{\partial \phi}{\partial t} + \delta v\frac{\partial \phi}{\partial x'} = 0, \tag{8.2.4}$$

where $\delta v = v - v_0$. If $v_0 = v$, then in the moving coordinate system,

$$\frac{\partial}{\partial t}\phi = 0, \tag{8.2.5}$$

or that the field remains stationary, and $\phi = \phi(x')$ only.

In Equation (8.2.4), δv denotes the "extra" velocity that the wave is moving with respect to the moving coordinates. If one is in a moving coordinates such that this "extra" velocity is dependent on the amplitude of the wave, then (8.2.4) can be rewritten as

$$\frac{\partial \phi}{\partial t} + \delta_1 \phi\frac{\partial \phi}{\partial x'} = 0. \tag{8.2.6}$$

which is a nonlinear equation. The above equation says that the field with a higher amplitude moves faster than the field with a lower amplitude. The above expression also implies that the field cannot remain stationary in this moving coordinates. In other words, ϕ is a function of both x' and t. It further means that if one has a symmetric pulse to begin with, the pulse will start to lean over as shown in Figure 8.1, generating a sharp shock wavefront. The shock wavefront has higher spectral components. Consequently, the high frequency component of the pulse has increased due to nonlinearity.

A pulse shape can also be distorted by a dispersive effect. A medium is dispersive if the wave speed is a function of the wavenumber of the wave. In a dispersive medium, the wave velocity has a higher order term that is proportional to $-\delta_2 k_x^2$. Therefore, a dispersive term can be added to (8.2.6) resulting in

$$\frac{\partial \phi}{\partial t} + \delta_1 \phi\frac{\partial \phi}{\partial x'} + \delta_2\frac{\partial^3 \phi}{\partial x'^3} = 0, \tag{8.2.7}$$

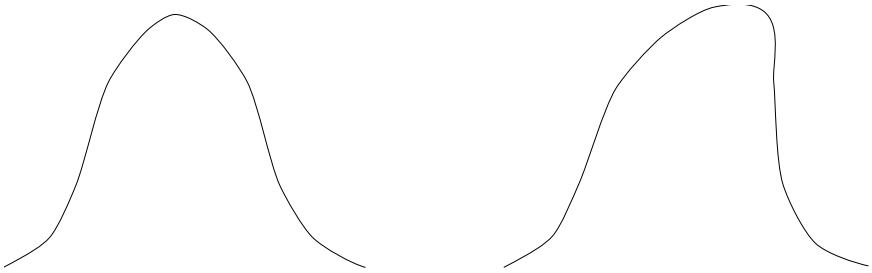


Figure 8.1: The leaning over of a symmetric pulse due to the nonlinearity in Equation (8.2.6).

since $\partial^2/\partial x'^2 = -k_x^2$. The dispersive effect implies that the waves with high wave numbers travel at a smaller velocity. Therefore, it tends to counter the pulse sharpening effect due to nonlinearity. In fact a solution exists for (8.2.7) for a pulse that propagates without distortion. If one lets $x' = \delta_2^{\frac{1}{3}}\xi$, $\phi = \delta_2^{\frac{1}{3}}\delta_1^{-1}\psi$, then (8.2.7) becomes

$$\frac{\partial\psi}{\partial t} + \psi\frac{\partial\psi}{\partial\xi} + \frac{\partial^3\psi}{\partial\xi^3} = 0, \tag{8.2.8}$$

which is the KdV (Korteweg de Vries) equation. It can be shown to have the solitary wave solution

$$\psi(t,\xi) = 3v_0 \text{ sech}^2\left[\frac{\sqrt{v_0}}{2}(\xi - v_0 t)\right]. \tag{8.2.9}$$

This soliton has a velocity v_0 with respect to the moving frame ξ which is related to x' via $x' = \delta_2^{\frac{1}{3}}\xi$. Therefore it has a velocity of $\delta_2^{\frac{1}{3}}v_0$ with respect to the moving frame x' and a velocity of $v + \delta_2^{\frac{1}{3}}v_0$ with respect to the stationary frame x. Note that since $\text{sech}^2 x$ decays exponentially for large arguments, the width of the soliton is $2/\sqrt{v_0}$ and its amplitude is $3v_0$. A larger amplitude soliton moves faster and has a narrower width.

8.3 Derivation of the Nonlinear Schrödinger Equation

The KdV equation is closely related to solitary waves in fluid. The propagation of solitons in an optical fiber is governed by the nonlinear Schrödinger (NLS) equation. The nonlear Schrödinger equation is due to nonlinear effect on electromagnetic wave propagation in an optical fiber [9,15–17]. The effect of nonlinearity in optical fiber generates noise, and this has been studied by Gordon and Haus [18]. This noise is hence known as the Gordon-Haus jitter.

Starting with the Maxwell's equations and assuming that $\partial/\partial x = \partial/\partial y = 0$, for an electric field polarized in the y direction, then

$$\frac{\partial}{\partial z}E_y = \frac{\partial}{\partial t}B_x, \tag{8.3.1}$$

$$\frac{\partial}{\partial z}B_x = \mu_0 \frac{\partial}{\partial t}D_y, \tag{8.3.2}$$

where μ_0 is assumed to be a constant. Eliminating B_x gives

$$\frac{\partial^2}{\partial z^2}E_y = \mu_0 \frac{\partial^2}{\partial t^2}D_y. \tag{8.3.3}$$

For a medium, $D_y = \epsilon_0 n^2 E_y$, where n is the refractive index. For a Kerr medium, n is nonlinearly related to the electric field, or that

$$n = n_0 + n_2 \left|E_y\right|^2. \tag{8.3.4}$$

The Kerr effect is a very small effect such that $n_2 \left|E_y\right|^2 \ll n_0$. But at optical frequencies, a small change in the refractive index can have a significant effect on the phase shift of the wave due to the short wavelengths involved. One can approximate

$$n^2 E_y = (n_0 + n_2 \left|E_y\right|^2)^2 E_y \approx (n_0^2 + 2n_0 n_2 \left|E_y\right|^2) E_y. \tag{8.3.5}$$

Consequently, one can write (8.3.3) as

$$\frac{\partial^2}{\partial z^2}E_y = \frac{1}{c_0^2}\frac{\partial^2}{\partial t^2}n^2 E_y = \frac{n_0^2}{c_0^2}\frac{\partial^2}{\partial t^2}E_y + \frac{2n_0 n_2}{c_0^2}\frac{\partial^2}{\partial t^2}\left|E_y\right|^2 E_y. \tag{8.3.6}$$

Next, one assumes that

$$E_y = \phi(z,t)e^{i(k_0 z - \omega_0 t)}, \tag{8.3.7}$$

where $\phi(z,t)$ is a slowly varying function of z and t. It is also the envelope function of a wave. It can be shown that

$$\frac{\partial^2}{\partial t^2}E_y = \left(\frac{\partial^2}{\partial t^2}\phi - 2i\omega_0 \frac{\partial}{\partial t}\phi - \omega_0^2 \phi\right)e^{i(k_0 z - \omega_0 t)}, \tag{8.3.8}$$

$$\frac{\partial^2}{\partial z^2}E_y = \left(\frac{\partial^2}{\partial z^2}\phi + 2ik_0 \frac{\partial}{\partial z}\phi - k_0^2 \phi\right)e^{i(k_0 z - \omega_0 t)}. \tag{8.3.9}$$

Then the envelope function $\phi(z,t)$ is assumed to be slowly varying compared to $\exp[i(k_0 z - \omega_0 t)]$ so that

$$\frac{\partial^2}{\partial t^2}\phi \ll \omega_0^2 \phi, \tag{8.3.10}$$

$$\frac{\partial^2}{\partial z^2}\phi \ll k_0^2 \phi. \tag{8.3.11}$$

Consequently, one has

$$\frac{\partial^2}{\partial t^2}E_y \approx \left(-2i\omega_0 \frac{\partial}{\partial t}\phi - \omega_0^2 \phi\right)e^{i(k_0 z - \omega_0 t)}, \tag{8.3.12}$$

$$\frac{\partial^2}{\partial z^2}E_y \approx \left(2ik_0\frac{\partial}{\partial t}\phi - k_0^2\phi\right)e^{i(k_0z-\omega_0t)}. \tag{8.3.13}$$

Assuming that the rapidly varying solution satisfies $k_0 = n_0\omega_0/c_0$, then

$$\begin{aligned}\frac{\partial^2}{\partial z^2}E_y - \frac{n_0^2}{c_0^2}\frac{\partial^2}{\partial t^2}E_y &= 2i\left(k_0\frac{\partial}{\partial z}\phi + \frac{n_0^2}{c_0^2}\omega_0\frac{\partial}{\partial t}\phi\right)e^{i(k_0z-\omega_0t)}\\ &= 2ik_0\left(\frac{\partial}{\partial z}\phi + \frac{n_0}{c_0}\frac{\partial}{\partial t}\phi\right)e^{i(k_0z-\omega_0t)}.\end{aligned} \tag{8.3.14}$$

The above can be used in (8.3.6) to arrive at

$$ik_0\left(\frac{\partial}{\partial z}\phi + \frac{n_0}{c_0}\frac{\partial}{\partial t}\phi\right)e^{i(k_0z-\omega_0t)} = \frac{n_0n_2}{c_0^2}\frac{\partial^2}{\partial t^2}\left|E_y\right|^2E_y. \tag{8.3.15}$$

Since $|E_y|^2 = |\phi|^2$ is slowly varying compared to E_y itself, then

$$\frac{\partial^2}{\partial t^2}\left|E_y\right|^2E_y \approx \left|E_y\right|^2\frac{\partial^2}{\partial t^2}E_y = -\omega_0^2\left|\phi\right|^2\phi e^{i(k_0z-\omega_0t)}. \tag{8.3.16}$$

Using (8.3.16) in (8.3.15) gives rise to

$$\left(\frac{\partial}{\partial z}\phi + \frac{n_0}{c_0}\frac{\partial}{\partial t}\phi\right) = i\frac{k_0n_2}{n_0}\left|\phi\right|^2\phi. \tag{8.3.17}$$

The above is the equation for the envelope function $\phi(z,t)$ when no dispertion is assumed in the medium. So when nonlinearity is absent, the pertinent equation is

$$\left(\frac{\partial}{\partial z}\phi + \frac{n_0}{c_0}\frac{\partial}{\partial t}\phi\right) = 0, \tag{8.3.18}$$

which is the advective equation governing the propagation of a distortionless pulse. In the above, the dispersion relation is

$$k = \frac{n_0}{c_0}\omega, \tag{8.3.19}$$

i.e., there is a linear relationship between k and ω. As a result, both the phase velocity (ω/k) and the group velocity ($d\omega/dk$) are independent of frequencies.

8.3.1 Dispersive effect

Dispersion occurs when n_0 is a function of ω so that k is not linearly proportional to ω anymore. In this case, $D_y = \epsilon_0 n^2 \star E_y$, where "$\star$" means "convolves". One lets

$$E_y(z,t) = \int_{-\infty}^{\infty}dk\int_{-\infty}^{\infty}d\omega\tilde{E}_y(k,\omega)e^{i(kz-\omega t)}, \tag{8.3.20}$$

and assuming the absence of Kerr effect or the nonlinear term, (8.3.3) becomes

$$k^2\tilde{E}_y = \frac{\omega^2}{c_0^2}n_0^2(\omega)\tilde{E}_y. \tag{8.3.21}$$

Consequently,

$$k = \frac{n_0(\omega)}{c_0}\omega. \tag{8.3.22}$$

One can rewrite (8.3.22), using Taylor expansions of its righthand side, as

$$k = k_0 + k_0'(\omega - w_0) + \frac{1}{2}k_0''(\omega - \omega_0)^2 + \ldots, \tag{8.3.23}$$

where one assumes that

$$k_0 = \frac{n_0(\omega_0)}{c_0}\omega_0, \quad k_0' = \left.\frac{dk}{d\omega}\right|_{\omega=\omega_0}, \quad k_0'' = \left.\frac{d^2k}{d\omega^2}\right|_{\omega=\omega_0}. \tag{8.3.24}$$

Squaring (8.3.23), and putting it back into (8.3.21) gives rise to

$$[k^2 - k_0^2 - 2k_0k_0'(\omega - \omega_0) - (k_0')^2(\omega - \omega_0)^2 - k_0k_0''(\omega - \omega_0)^2 + \ldots]\tilde{E}_y = 0. \tag{8.3.25}$$

Furthermore, one can write $k^2 - k_0^2 = (k - k_0)^2k_0 + (k - k_0)^2$ and, using the fact that $(k - k_0)^2 \approx (k_0')^2(\omega - \omega_0)^2$ from (8.3.23) leads to

$$[(k - k_0)2k_0 - 2k_0k_0'(\omega - \omega_0) - k_0k_0''(\omega - \omega_0)^2 + \ldots]\tilde{E}_y = 0. \tag{8.3.26}$$

Then, (8.3.20) is rewritten as

$$E_y(z,t) = e^{i(k_0z - \omega_0t)}\int_{-\infty}^{\infty}dk\int_{-\infty}^{\infty}d\omega\tilde{E}_y(k,\omega)e^{[(k-k_0)z - (\omega-\omega_0)t]} \tag{8.3.27}$$

$$= \phi(z,t)e^{i(k_0z - \omega_0t)}. \tag{8.3.28}$$

Fourier inverse transforming (8.3.25) using (8.3.20), which is the same as Fourier inverse transforming (8.3.21) gives

$$\left(-2ik_0\frac{\partial}{\partial z} - 2ik_0k_0'\frac{\partial}{\partial t} + k_0k_0''\frac{\partial^2}{\partial t^2}\right)\phi(z,t) = 0 \tag{8.3.29}$$

In the above analysis, it is assumed that $\omega \simeq \omega_0$ and $k \simeq k_0$. This is the same as assuming that $\phi(z,t)$ is slowly varying since it has only low frequency components. If $k_0'' = 0$, the above is an alternative way of deriving (8.3.18). Equation (8.3.29) can be rewritten as

$$\left(\frac{\partial}{\partial z} + \frac{1}{v_g}\frac{\partial}{\partial t} + \frac{i}{2}k_0''\frac{\partial^2}{\partial t^2}\right)\phi(z,t) = 0 \tag{8.3.30}$$

where $v_g = 1/k_0'$.

For the Kerr effect, one can assume that n_2 is frequency independent, one can add the nonlinear part to the above equation to obtain

$$\frac{\partial}{\partial z}\phi + \frac{1}{v_g}\frac{\partial}{\partial t}\phi = i\frac{k_0 n_2}{n_0}|\phi|^2\phi - \frac{i}{2}k_0''\frac{\partial^2}{\partial t^2}\phi \tag{8.3.31}$$

The above is the nonlinear Schrödinger equation.

In the moving coordinate system such that $v_g z = z - v_g t$, then $\phi(z,t) = \phi\left(z, \frac{z}{v_g} - \tau\right) = \tilde{\phi}(z,\tau)$. Moreover,

$$\frac{\partial}{\partial t}\phi = -\frac{\partial\tilde{\phi}}{\partial\tau}, \quad \frac{\partial\tilde{\phi}}{\partial z} = \frac{\partial\phi}{\partial z} + \frac{1}{v_g}\frac{\partial}{\partial t}\phi \tag{8.3.32}$$

or

$$\frac{\partial\tilde{\phi}}{\partial z} = i\frac{k_0 n_2}{n_0}|\tilde{\phi}|^2\tilde{\phi} - \frac{i}{2}k_0''\frac{\partial^2}{\partial\tau^2}\tilde{\phi} \tag{8.3.33}$$

The above is similar to the Schrödinger equation

$$-i\hbar\frac{\partial}{\partial t}\psi = -\frac{\hbar^2}{2m}\frac{\partial^2}{\partial x^2}\psi + V\psi \tag{8.3.34}$$

if one identifies z with t and τ with x, and $|\tilde{\phi}|^2$ with V.

When V is negative to form a potential well, Equation (8.3.34) admits solutions of the form

$$\psi = Ae^{-i\omega t}g(x) \tag{8.3.35}$$

When $g(x)$ is a function localized in x corresponding to bound states or trapped modes in the potential well which are stationary states. By the same token, by $k_0{}''$ is negative, Equation (8.3.33) admits solutions of the form

$$\begin{aligned}\tilde{\phi} &= Be^{ikz}f(v_g\tau)\\ &= Be^{ikz}f(z - v_g t)\end{aligned} \tag{8.3.36}$$

where $f(v_g\tau)$ is a function localized in τ. Equation (8.3.36) corresponds to a solution.

Since the potential well is created by $|\tilde{\phi}|^2$,which is proportional to the field itself, this is a self-trapping phenomenon. Conservation of energy requires that the pulse becomes narrower when the amplitude of $\tilde{\phi}$ becomes larger. On the other hand, the dispersive effect tends to spread the pulse more when it becomes narrower. A final equilibrium is reached where the pulse propagates without distortion.

If $k_0{}''$ is positive, Equation (30) cannot trap a mode. However, it can admit a solution signified by the absence of light, hence the name dark solutions.

8.3.2 Solution of the Nonlinear Schrödinger Equation

Via a change of variables and coordinates, the nonlinear Schrödinger equation can be written in dimensionless form as

$$i\frac{\partial\psi}{\partial z} + \frac{1}{2}\frac{\partial^2\psi}{\partial T^2} + |\psi|^2\psi = 0. \tag{8.3.37}$$

A solution is a stationary solution to the above which is stationary in amplitude with respect to z and T. Therefore, one seeks a solution of the form

$$\psi(T,z) = \sqrt{\rho(T)}e^{i\alpha(z-z_0)}. \tag{8.3.38}$$

Substituting (8.3.38) into (8.3.37) yields

$$-\alpha\sqrt{\rho} + \frac{1}{2}\frac{\partial^2}{\partial T^2}\sqrt{\rho} + \rho^{\frac{3}{2}} = 0, \tag{8.3.39}$$

where ρ is assumed to be real-valued. The above is the same as

$$-\alpha - \frac{1}{8}\left[\frac{1}{\rho^2}(\rho')^2 - 2\frac{\rho''}{\rho}\right] + \rho = 0, \tag{8.3.40}$$

where $\rho' = \frac{d\rho}{dT}$, $\rho'' = \frac{d^2\rho}{dT^2}$. It can be further simplified to

$$-\alpha + \frac{1}{8}\frac{d}{d\rho}\frac{1}{\rho}(\rho')^2 + \rho = 0. \tag{8.3.41}$$

Integrating the above leads to

$$-\alpha\rho + \frac{\rho^2}{2} + \frac{1}{8}\frac{1}{\rho}(\rho')^2 + C = 0. \tag{8.3.42}$$

Since ρ and ρ' are independent of T, C is also independent of T. But since $\rho \to 0$ when $T \to \infty$, and one assumes that $\rho' \sim \rho \to 0$ when $T \to \infty$, then $C = 0$. When ρ is maximum, then $\rho' = 0$. Therefore, (8.3.42) has to be of the form

$$(\rho')^2 = -4\rho^2(\rho - \rho_0), \tag{8.3.43}$$

where ρ_0 is the maximum of ρ. Therefore $\alpha = \rho_0/2$. One can let $\rho_0 y^2 = \rho - \rho_0$ in (8.3.43) to obtain

$$dy = (1+y^2)\sqrt{\rho_0}i dT. \tag{8.3.44}$$

Letting $\tau = i\sqrt{\rho_0}T$ gives rise to

$$\frac{dy}{1+y^2} = d\tau, \tag{8.3.45}$$

$$y = \tan(\tau - \tau_0), \tag{8.3.46}$$

or

$$\rho = \rho_0(y^2+1) = \rho_0 \sec^2(\tau - \tau_0), \tag{8.3.47}$$

or

$$\rho = \rho_0 \text{ sech}^2\big[\sqrt{\rho_0}(T - T_0)\big]. \tag{8.3.48}$$

Consequently,

$$\psi(T, z) = \sqrt{\rho_0} \text{ sech}\big[\sqrt{\rho_0}(T - T_0)\big] e^{i\frac{\rho_0}{2}(z - z_0)}. \tag{8.3.49}$$

Since $T \to (z - v_g t)$, this pulse travels with the same speed irrespective of the amplitude of the pulse.

Schödinger equation (8.3.34) is Galilean invariant. Hence, if a solution $\psi(x, t)$ is found, a new solution $\psi(x + vt, t)e^{i\theta(x,t)}$ is also a solution. Using this fact, a new solution

$$\psi_1(T, z) = \psi(T + kz, z)e^{-ik(T + \frac{1}{2}kz)} \tag{8.3.50}$$

can be constructed that is also a solution of the nonlinear Schrödinger equation.

Equation (8.3.49) represents a stationary solution in the moving coordinate system, or it is a soliton solution that moves at the group velocity. Equation (8.3.50) represents a soliton solution that moves at arbitrary velocity since k is arbitrary. The solution has the property that the larger its amplitude, the narrower is the pulse. However, unlike the KdV solution, the NLS solution's velocity does not increase with its amplitude.

8.4 Solution of the KdV Equation via Inverse Scattering Transform

There is an interesting body of knowledge where solutions to nonlinear partial differential equations (PDEs) can be obtained by solving an inverse scattering problem. This is known as the inverse scattering transform method in solving nonlinear PDEs [19, 20].

Via a change of variable, the KdV equation can be written as

$$\frac{\partial q}{\partial t} - 6q\frac{\partial q}{\partial x} + \frac{\partial^3 q}{\partial x^3} = 0. \tag{8.4.1}$$

It turns out that if $q(x, t)$ is a potential in a Schrödinger eigenvalue equation, i.e.,

$$\frac{\partial^2 u}{\partial x^2} + [\lambda - q(x, t)]u = 0, \tag{8.4.2}$$

then $u(x, t)$ evolves in time according to the following equation,

$$\begin{aligned} \frac{\partial u}{\partial t} &= -\frac{\partial q}{\partial x}u + (4\lambda + 2q)\frac{\partial u}{\partial x} \\ &= -4\frac{\partial^3 u}{\partial x^3} + 3\frac{\partial q}{\partial x}u + 6q\frac{\partial u}{\partial x}. \end{aligned} \tag{8.4.3}$$

Equation (8.4.2) can be expressed as

$$Lu = \lambda u \tag{8.4.4}$$

where $L = -\frac{\partial^2}{\partial x^2} + q(x,t)$, while Equation (8.4.3) is expressible as

$$\frac{\partial u}{\partial t} = Mu \tag{8.4.5}$$

where

$$M = -\frac{\partial q}{\partial x} + (4\lambda + 2q)\frac{\partial}{\partial x} = -4\frac{\partial^3}{\partial x^3} + 3\frac{\partial q}{\partial x} + 6q\frac{\partial}{\partial x}. \tag{8.4.6}$$

The critical point here is that λ is assumed independent of time. So if one differentiates (8.4.4) with respect to time, then

$$L_t u + LMu = \lambda Mu = M\lambda u = MLu, \tag{8.4.7}$$

or that

$$(L_t + LM - ML)u = 0. \tag{8.4.8}$$

It can be shown that in order for (8.4.8) to be satisfied, q satisfies (8.4.1), the KdV equation.

The above gives an alternative way of solving (8.4.1). Given the initial value $q(x,0)$, one can solve for $u(x,0)$ from (8.4.2). Then if one can determine $u(x,t)$ from $u(x,0)$, then inverse scattering theory can be used to find $q(x,t)$ from $u(x,t)$ in (8.4.2).

8.4.1 Inverse Scattering

Consider the Schrödinger equation

$$-\frac{\partial^2}{\partial x^2}u(x) + q(x)u(x) = \lambda u(x), \tag{8.4.9}$$

and that $q(x) \to 0$ when $|x| \to \infty$. The above equation has two independent solutions $u_1(x,k)$ and $u_2(x,k)$ such that $u_1 \sim e^{ikx}$, $x \to +\infty$, and that $u_2 \sim e^{-ikx}$, $x \to -\infty$, where $k^2 = \lambda$. Define the function $A(x,y)$ as

$$u_1(x,k) = e^{ikx} + \int_x^\infty A(x,t)e^{ikt}dt. \tag{8.4.10}$$

One can substitute (8.4.10) into (8.4.9) to obtain an equation for $A(x,t)$. First one notices that

$$\frac{\partial^2}{\partial x^2}\int_x^\infty A(x,t)e^{ikt}dt = \int_0^\infty \frac{\partial^2}{\partial x^2}[H(t-x)A(x,t)]e^{ikt}dt, \tag{8.4.11}$$

where $H(t)$ is a Heaviside step function. Evaluating the above yields

$$\begin{aligned}\frac{\partial^2}{\partial x^2}\int_x^\infty A(x,t)e^{ikt}dt = &-A_t(x,t)\Big|_{t=x} e^{ikx} - ikA(x,x)e^{ikx} - 2A_x(x,t)\Big|_{t=x} e^{ikx} \\ &+ \int_x^\infty A_{xx}(x,t)e^{ikt}dt.\end{aligned} \tag{8.4.12}$$

In the above, one has made use of

$$\begin{aligned}\frac{\partial^2}{\partial x^2}H(t-x)A(x,t) &= H''(t-x)A(x,t) - 2H'(t-x)A_x(x,t) + H(t-x)A_{xx}(x,t)\\ &= \delta'(t-x)A(x,t) - 2\delta(t-x)A_x(x,t) + H(t-x)A_{xx}(x,t)\end{aligned} \tag{8.4.13}$$

Furthermore, since $\lambda = k^2$, the right-hand side of (8.4.9) becomes

$$\begin{aligned}-\int_x^\infty k^2 A(x,t)e^{ikt}dt =& +\int_x^\infty A(x,t)\frac{\partial^2}{\partial t^2}e^{ikt}dt\\ =&A(x,t)\frac{\partial}{\partial t}e^{ikt}\Big|_x^\infty - \int_x^\infty A_t(x,t)\frac{\partial}{\partial t}e^{ikt}dt\\ =&A(x,t)ike^{ikt}\Big|_x^\infty - A_t(x,t)e^{ikt}\Big|_x^\infty + \int_x^\infty A_{tt}(x,t)e^{ikt}dt\\ =&\lim_{t\to\infty}[ikA(x,t)e^{ikt} - A_t(x,t)e^{ikt}]\\ &- ikA(x,x)e^{ikx} + A_t(x,t)\Big|_{t=x}e^{ikx}\\ &+\int_x^\infty A_{tt}(x,t)e^{ikt}dt.\end{aligned} \tag{8.4.14}$$

Consequently,

$$\begin{aligned}\frac{\partial^2}{\partial x^2}u_1 - q(x)u_1 + k^2u_1 =& \int_x^\infty [A_{xx}(x,t) - A_{tt}(x,t) - q(x)A(x,t)]e^{ikt}dt\\ &+\lim_{t\to\infty}[A_t(x,t) - ikA(x,t)]e^{ikt}\\ &-[2A_t(x,t)|_{t=x} + 2A_x(x,t)|_{t=x} + q(x)]\,e^{ikx}\\ =&\int_x^\infty [A_{xx}(x,t) - A_{tt}(x,t) - q(x)A(x,t)]e^{ikt}dt\\ &+\lim_{t\to\infty}[A_t(x,t) - ikA(x,t)]e^{ikt}\\ &-\left[2\frac{d}{dx}A(x,x) + q(x)\right]e^{ikx},\end{aligned} \tag{8.4.15}$$

where

$$\frac{d}{dx}A(x,x) = \frac{\partial A(x,t)}{\partial x}\Big|_{t=x} + \frac{\partial A(x,t)}{\partial t}\Big|_{t=x}. \tag{8.4.16}$$

If $\lim_{t\to\infty} A_t(x,t) = \lim_{t\to\infty} A(x,t) = 0$, and the above is valid for all k, then $q(x) = -2\frac{d}{dx}A(x,x)$, and,

$$A_{xx} - A_{tt} - q(x)A = 0. \tag{8.4.17}$$

in order for the above to be satisfied.

If a wave is incident on the potential $q(x)$ from $x = \infty$ as e^{-ikx}, the solution for $x \to +\infty$ must be $e^{-ikx} + R(k)e^{ikx}$, $x \to +\infty$, where $R(k)$ is the reflection coefficient. For $x \to -\infty$, the solution is $T(k)e^{-ikx}$ where $T(k)$ is the transmission coefficient. This solution, by matching its large $|x|$ behaviors, must be expressible as $R(k)u_1(x,k) + u_1(x,-k)$ or $T(k)u_2(x,k)$ since $u_1(x,k) \sim e^{ikx}$, $x \to +\infty$, and $u_2(x,k) \sim e^{-ikx}$, $x \to -\infty$. Therefore, one concludes that

$$T(k)u_2(x,k) = R(k)u_1(x,k) + u_1(x,-k). \tag{8.4.18}$$

Furthermore, when $k \to +\infty$, or that the frequency or energy of the wave tends to infinity, the potential barrier has a negligible effect on the wave, and $T(k) \to 1$ (typical of quantum scattering). Moreover, $T(k)$ and $u_2(x,k)$ are the Fourier transforms of causal signals and they have to be analytic for $\Im m[k] > 0$. Consequently, one can Fourier inverse transform (8.4.18) along an inversion contour C which is above the singularities of $T(k)$ and $u_2(x,k)$, or

$$\begin{aligned}\int_C T(k)u_2(x,k)e^{-ikt}dk = &\int_C R(k)u_1(x,k)e^{-ikt}dk \\ &+ \int_C u_1(x,-k)e^{-ikt}dk.\end{aligned} \tag{8.4.19}$$

Since $T(k)u_2(x,k)e^{-ikt} \sim e^{-ik(t+x)} + O\left(\frac{1}{k}\right)e^{-ik(t+x)}$, $k \to \infty$, a fact easily proven from Born approximation, it implies that

$$\int_C T(k)u_2(x,k)e^{-ikt}dk = 2\pi\delta(t+x) + \int_C O\left(\frac{1}{k}\right)e^{-ik(t+x)}. \tag{8.4.20}$$

Using Jordan's lemma for the second term in (8.4.20), one concludes that

$$\int_C T(k)u_2(x,k)e^{-ikt}dk = 0, \qquad t + x < 0. \tag{8.4.21}$$

Using the form in Equation (8.4.10) for $u_1(x,k)$ in (8.4.19), then

$$\begin{aligned}0 = &\int_C R(k)e^{ik(x-t)}dk + \int_x^\infty dt' A(x,t') \int_C R(k)e^{ik(t'-t)}dt \\ &+ \int_C e^{-ik(x+t)}dk + \int_x^\infty dt' A(x,t') \int_C e^{-ik(t'+t)}dt, \quad t + x < 0.\end{aligned} \tag{8.4.22}$$

The third term is zero because $t + x < 0$, and defining

$$r(t) = \frac{1}{2\pi}\int_C R(k)e^{-ikt}dk, \tag{8.4.23}$$

gives

$$0 = r(t-x) + \int_x^\infty dt' A(x,t') r(t-t') + A(x,-t), \qquad t+x<0. \tag{8.4.24}$$

The above is known as the Gelfand-Levitan-Marchenko equation. Given $r(t)$, one can solve for $A(x,t)$ and obtain $q(x)$ via

$$q(x) = -2\frac{d}{dx}A(x,x). \tag{8.4.25}$$

As an example, one can consider a case where $r(t) = me^{+\mu t}$. Using it in (8.4.24), then

$$0 = me^{+\mu t-\mu x} + me^{+\mu t}\int_x^\infty dt' A(x,t') e^{-\mu t'} + A(x,-t), \qquad t+x<0. \tag{8.4.26}$$

Letting $A(x,t) = a(x)e^{-\mu t}$, then

$$0 = me^{-\mu x} + ma(x)\int_x^\infty dt' e^{-2\mu t'} + a(x), \qquad t+x<0 \tag{8.4.27}$$

or that

$$a(x) = \frac{-me^{-\mu x}}{1+\frac{m}{2\mu}e^{-2\mu x}}. \tag{8.4.28}$$

It follows then that

$$A(x,x) = \frac{-me^{-2\mu x}}{1+\frac{m}{2\mu}e^{-2\mu x}} = \frac{-m}{\frac{m}{2\mu}+e^{2\mu x}}. \tag{8.4.29}$$

Differentiating the above with respect to x leads to

$$\begin{aligned} q(x) = \frac{-4m\mu e^{2\mu x}}{\left(\frac{m}{2\mu}+e^{2\mu x}\right)^2} &= -8\mu^2\frac{e^{2\phi+2\mu x}}{(e^{2\phi}+e^{2\mu x})^2} \\ &= -2\mu^2\,\mathrm{sech}^2(\mu x-\phi). \end{aligned} \tag{8.4.30}$$

where $\phi = \frac{1}{2}\ln\frac{m}{2\mu}$.

8.4.2 Solution of the KdV Equation

If the solution to the KdV equation, $q(x,t)$, is a potential to the Schrödinger equation $-u_{xx} + q(x,t)u = k^2u$, then the eigenfunction evolves according to $u_t = -4u_{xxx} + 3q_xu + 6qu_x$ according to (8.4.3). Therefore, if u is known when $|x| \to \infty$, one can use the inverse scattering theory to reconstruct $q(x,t)$, and hence, the solution to the KdV equation.

As is shown in (8.4.18), the fundamental solutions to the Schrödinger equation satisfy

$$u_2(x,k) = c_{11}(k)u_1(x,k) + c_{12}u_1(x,-k). \tag{8.4.31}$$

where $c_{11}(k) = R(k)/T(k)$, $c_{12}(k) = 1/T(k)$. Since $u_2(x,k) \sim e^{-ikx}$, $x \to -\infty$, irrespective of $q(x,t)$, and so independent of t, one assumes that $u = h(t)e^{-ikx}, x \to -\infty$. If q and q_x tend to zero when $|x| \to \infty$, then

$$u_t = -4u_{xxx}, \quad |x| \to \infty. \tag{8.4.32}$$

Hence for $x \to -\infty$, then

$$h_t = -4ik^3 h, \tag{8.4.33}$$

or that

$$h = h_0 e^{-4ik^3 t}. \tag{8.4.34}$$

For $x \to +\infty$, then

$$u(x) \sim h(t)(c_{11}e^{ikx} + c_{12}e^{-ikx}).$$

Therefore, from (8.4.32), then

$$\begin{aligned} &(h_t c_{11} + h c_{11t})e^{ikx} + (h_t c_{12} + h c_{12t})e^{-ikx} \\ &= -4h[c_{11}(ik)^3 e^{ikx} + c_{12}(ik)^3 e^{-ikx}], \end{aligned} \tag{8.4.35}$$

Comparing the left and right-hand sides, and making use of (8.4.33), it leads to

$$c_{11t} = 8ik^3 c_{11}, \qquad c_{12t} = 0 \tag{8.4.36}$$

or

$$c_{11} = c_{11}^o e^{8ik^3 t}, \qquad c_{12} = c_{12}^0. \tag{8.4.37}$$

Consequently, one can find the time evolution of c_{11} and c_{12}. One can obtain $R(k) = c_{11}(k)/c_{12}(k)$. Since c_{11} and c_{12} are evolving with time, $R(k)$ also evolves with time. At this point, the notation is rather confusing since there is a time variable also in (8.4.23) and (8.4.24). One shall call the time variable in the aforementioned equation τ to avoid the confusion. Hence,

$$r(\tau, t) = \frac{1}{2\pi} \int_C R(k,t) e^{-ik\tau} dk. \tag{8.4.38}$$

From Equation (8.4.37), one notices that

$$R(k,t) = R_0(k) e^{8ik^3 t}. \tag{8.4.39}$$

For a simple case, one considers the case where $R_0(k)$ has one simple pole at $k = i\mu$. Then, the above integral can be evaluated so that

$$r(\tau, t) = m e^{+8\mu^3 t + \mu\tau}. \tag{8.4.40}$$

According to (8.4.30), $q(x,t)$ is

$$q(x,t) = -2\mu^2 \operatorname{sech}^2(\mu x - \phi) \tag{8.4.41}$$

where

$$\phi = \frac{1}{2}\ln\left(\frac{me^{8\mu^3 t}}{2\mu}\right) = 4\mu^3 t + \phi_0, \tag{8.4.42}$$

or

$$q(x,t) = -2\mu^2 \operatorname{sech}^2(\mu x - 4\mu^3 t + \phi_0) \tag{8.4.43}$$

Letting $\mu^2 = \frac{1}{4}c$, the above becomes

$$q(x,t) = -\frac{c}{2}\operatorname{sech}^2\left[\frac{1}{2}\sqrt{c}(x-ct) + \phi_0\right]. \tag{8.4.44}$$

The above method can be used to find the multiple soliton solution to the KdV equation by assuming more poles in the reflection coefficient. It addition, the inverse scattering transform method can be used to solve nonlinear equations like the nonlinear Schrödinger equation.

8.4.3 Inverse Scattering with Schrödinger Equation

Consider the Schrödinger equation

$$-\frac{d^2}{dx^2}u(x) + q(x)u(x) = \lambda u(x), \tag{8.4.45}$$

where $q(x) \to 0$ when $|x| \to \infty$. The above equation has two independent solutions. They can be defined as $u_1(x,k)$ and $u_2(x,k)$ such that $u_1(x,k) \sim e^{ikx}, x \to +\infty$, and that $u_2(x,k) \sim e^{-ikx}, x \to -\infty$, where $k^2 = \lambda$. Clearly, u_1 and u_2 are independent of each other when $q = 0$. When $q(x) \neq 0$, the scattering potential will generate more waves but u_1 and u_2 remain independent of each other. Therefore, any solution to (8.4.45) can be written as a linear superposition of u_1 and u_2.

The general solution to (8.4.45) can be written in terms of an integral equation

$$u(x) = u_0(x) + \int_{-\infty}^{\infty} g(x-x')q(x')u(x')dx', \tag{8.4.46}$$

where $g(x)$, the Green's function, is a solution to

$$\left(\frac{d^2}{dx^2} + k^2\right)g(x) = \delta(x), \tag{8.4.47}$$

and $u_0(x)$ is a solution to

$$\left(\frac{d^2}{dx^2} + k^2\right)u_0(x) = 0. \tag{8.4.48}$$

If physical condition such as causality is not imposed, there could be many solutions to (8.4.47). The possible solutions to (8.4.47) are

$$g_1(x) = \frac{e^{ik|x|}}{2ik}, \tag{8.4.49}$$

$$g_2(x) = \frac{e^{-ik|x|}}{2ik}, \tag{8.4.50}$$

$$g_3 x = \begin{cases} 0, & x < 0, \\ \frac{\sin kx}{k}, & x > 0. \end{cases} \tag{8.4.51}$$

$$g_4(x) = \begin{cases} \frac{\sin kx}{k}, & x < 0, \\ 0, & x > 0. \end{cases} \tag{8.4.52}$$

If e^{-ikt} time convention is used, only $g_1(x)$ is physical because it corresponds to outgoing waves, while $g_i(x)$, $i > 1$ are unphysical. However, they can be used in (8.4.46) to provide bonafide solution to (8.4.45).

To construct $u_1(x, k)$, one lets u_0 in (8.4.46) to be e^{ikx}, and $g(x)$ in (8.4.46) to be $g_4(x)$. Then $u_1(x, k)$ satifies

$$u_1(x) = e^{ikx} + \int_{-\infty}^{\infty} g_4(x - x')q(x')u_1(x')dx'. \tag{8.4.53}$$

Due to the property of $g_4(x)$, clearly, $u_1(x) \sim e^{ikx}$, $x \to +\infty$.

To construct $u_2(x, k)$, one lets u_0 in (8.4.46) to be e^{-ikx} and $g(x)$ to be $g_3(x)$. Then, $u_2(x, k)$ satisfies

$$u_2(x) = e^{-ikx} + \int_{-\infty}^{\infty} g_3(x - x')q(x')u_2(x')dx'. \tag{8.4.54}$$

Clearly, $u_2(x) \sim e^{-ikx}$, $\quad x \to -\infty$. Due to the independence of $g_3(x)$ and $g_4(x)$, $u_1(x, k)$ and $u_2(x, k)$ are independent of each other. Since $u_1(x, -k)$ is independent of $u_1(x, k)$, they are related by

$$T(k)u_2(x, k) = R(k)u_1(x, k) + u_1(x, -k) = u_3(x, k). \tag{8.4.55}$$

Equation (8.4.55) corresponds to a scattering solution where it becomes $Re^{ikx} + e^{-ikx}$, $x \to +\infty$, and Te^{-ikx}, $x \to -\infty$. Therefore, R and T physically correspond to reflection and transmission coefficients respectively. Equation (8.4.55) can be used to derive the Gelfand-Levitan-Marchenko equation.

8.4.4 Time-Domain Solutions

If one expresses

$$u(x, \tau) = \frac{1}{2\pi} \int_{-\infty}^{\infty} dk e^{-ik\tau} u(x, k), \tag{8.4.56}$$

where $k^2 = \lambda$, and Fourier inverse transform (8.4.45) accordingly gives

$$-\frac{\partial^2}{\partial x^2}u(x,\tau) + q(x)u(x,\tau) = -\frac{\partial^2}{\partial \tau^2}u(x,\tau). \tag{8.4.57}$$

Here, (8.4.57) is the Schrödinger-like equation since it has a second derivative in time while Schrödinger equation has a first derivative in time. Then (8.4.46) becomes

$$u(x,\tau) = u_0(x,\tau) + \int_{-\infty}^{\infty} g(x-x',\tau) \star u(x',\tau)q(x')dx', \tag{8.4.58}$$

were $g(x,\tau)$ is the Fourier inverse transform of the Green's function $g(x,k)$, given in (8.4.49)–(8.4.52). The Fourier inversion contour is defined to be above the singularity at the origin in (8.4.49) and (8.4.50) so that only (8.4.49) is causal while the rest of the Green's functions are not causal. The supports of the various Green's functions are shown in Figure 8.2, (a) to (d) on a space-time diagram.

Consequently, the time-domain equation of (8.4.53) can be written as

$$u_1(x,\tau) = \delta(x-\tau) + \int_{-\infty}^{\infty} g_4(x-x',\tau) \star u_1(x',\tau)q(x')dx'. \tag{8.4.59}$$

To determine the support of $u_1(x,\tau)$, one can express (8.4.59) as a Neumann series which is also a multiple scattering series—the first term corresponds to single scattering, the second term corresponds to double scattering and so on. Then it is seen that the support of $u_1(x,\tau)$ is given in Figure 8.3, when one assumes that $q(x) \neq 0$, $|x| < a$. Similarly, the support of $u_2(x,\tau)$ is given in Figure 8.4. Equation (8.4.55) corresponds to a physical scattering case where a wave, e^{-ikx}, is incident from $x \to +\infty$. The support of $u_3(x,\tau)$ should be as shown in Figure 8.5, whith a wave $\delta(x+\tau)$ incident fom $x \to +\infty$. The time domain equivalence of (8.4.55) is

$$R(\tau) \star u_1(x,\tau) + u_1(x,-\tau) = u_3(x,\tau). \tag{8.4.60}$$

Notice that $u_1(x,-k)$ is the Fourier inverse transform to $u_1(x,-\tau)$. From Figure 8.5, one can see that $R(\tau) \neq 0$, $\tau > 2a$. Consequently, the support of $R(\tau) \star u_1(x,\tau)$ is as shown in Figure 8.6.

The support of $u_1(x,-\tau)$ is shown in Figure 8.7. Since $R(\tau) \star u_1(x,\tau) + u_1(x,-\tau) = u_3(x,\tau)$ whose support is shown in Figure 8.5, one requires that

$$R(\tau) \star u_1(x,\tau) + u_1(x,-\tau) = 0, \qquad \tau < -x. \tag{8.4.61}$$

The above is the key to the derivation of the Gelfand-Levitan-Marchenko equation. One can rewrite Equation (8.4.59) as

$$u_1(x,\tau) = \delta(x-\tau) + A(x,\tau), \tag{8.4.62}$$

where $A(x,\tau) = 0$, $\tau < x$. Then (8.4.61) becomes

$$R(\tau - x) + R(\tau) \star A(x,\tau) + A(x,-\tau) = 0, \qquad \tau < -x \tag{8.4.63}$$

or

$$R(\tau - x) + \int_x^\infty d\tau' R(\tau - \tau')A(x,\tau) + A(x,-\tau) = 0. \qquad \tau < -x \tag{8.4.64}$$

The above is the Gelfand-Levitan-Marchenko equation. Given the reflection coefficient $R(\tau)$, one can solve for $A(x,\tau)$. It can be shown that the scattering potential $q(x)$ can be derived from $A(x,\tau)$. One can express Equation (8.4.62) as

$$u_1(x,\tau) = \delta(x-\tau) + A(x,\tau)H(\tau - x), \tag{8.4.65}$$

where $H(\tau)$ is a Heaviside step function. Substituting (8.4.65) into (8.4.57) gives rise to

$$\begin{aligned} -\frac{\partial^2}{\partial x^2}A(x,\tau)H(\tau-x) + \frac{\partial^2}{\partial \tau^2}A(x,\tau)H(\tau-x) + q(x)\delta(x-\tau) \\ + q(x)A(x,\tau)H(\tau-x) = 0. \end{aligned} \tag{8.4.66}$$

Since

$$\begin{aligned} \frac{\partial^2}{\partial x^2}A(x,\tau)H(\tau-x) &= H_{xx}(\tau-x)A(x,\tau) \\ &\quad + 2H_x(\tau-x)A_x(x,\tau) + H(\tau-x)A_{xx}(x,\tau), \end{aligned} \tag{8.4.67}$$

$$\begin{aligned} \frac{\partial^2}{\partial \tau^2}A(x,\tau)H(\tau-x) &= H_{\tau\tau}(\tau-x)A(x,\tau) \\ &\quad + 2H_\tau(\tau-x)A_\tau(x,\tau) + H(\tau-x)A_{\tau\tau}(x,\tau). \end{aligned} \tag{8.4.68}$$

Since $H_{xx}(\tau-x) = H_{\tau\tau}(\tau-x) = \delta'(\tau-x)$, $H_x(\tau-x) = -H_\tau(\tau-x) = -\delta(x-\tau)$, using (8.4.67) in (8.4.66) and matching terms of the same singularity, then

$$2A_x(x,\tau)\Big|_{\tau=x} + 2A_\tau(x,\tau)\Big|_{\tau=x} + q(x) = 0. \tag{8.4.69}$$

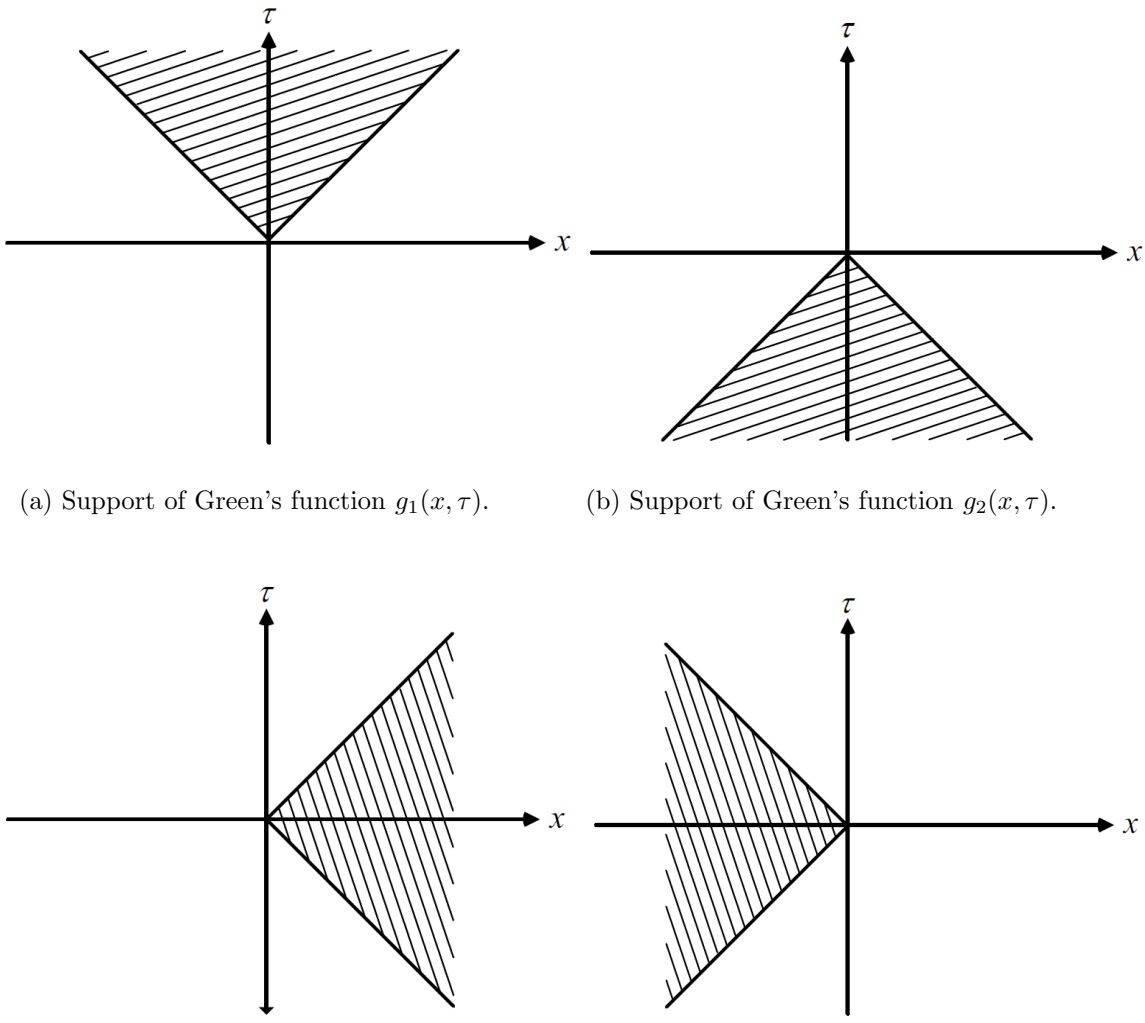
Consequently,

$$q(x) = -2\frac{d}{dx}A(x,x), \tag{8.4.70}$$

where

$$\frac{d}{dx}A(x,x) = A_x(x,\tau)\Big|_{\tau=x} + A_\tau(x,\tau)\Big|_{\tau=x}. \tag{8.4.71}$$

Hence, once $A(x,x)$ is found, the profile $q(x)$ can be retrieved.



(a) Support of Green's function $g_1(x,\tau)$.

(b) Support of Green's function $g_2(x,\tau)$.

(c) Support of Green's function $g_3(x,\tau)$.

(d) Support of Green's function $g_4(x,\tau)$.

Figure 8.2: Supports of different Green's functions.

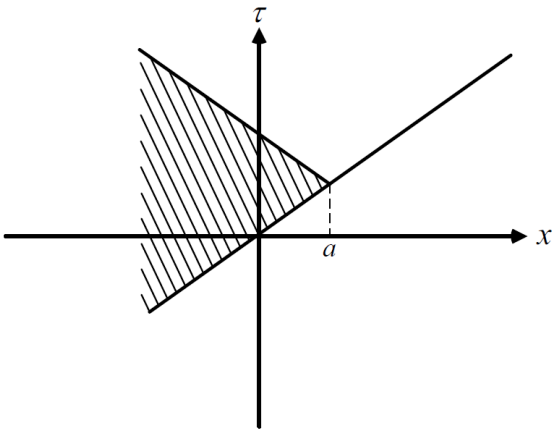


Figure 8.3: Support of $u_1(x, \tau)$.

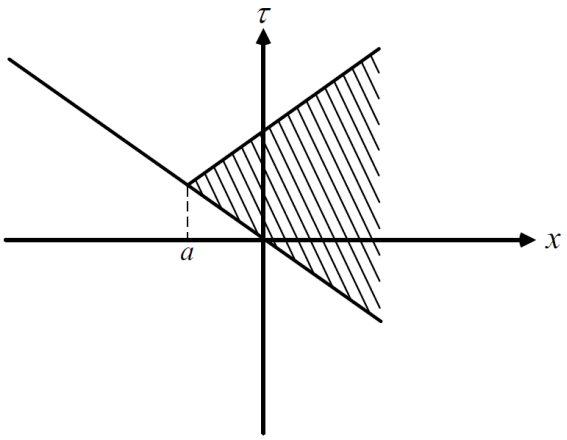


Figure 8.4: Support of $u_2(x, \tau)$.

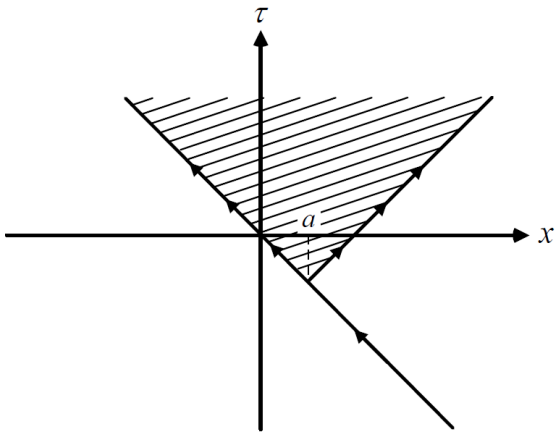


Figure 8.5: Support of $u_3(x,\tau)$.

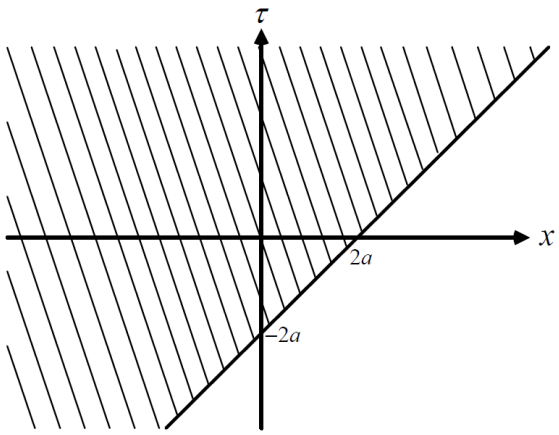


Figure 8.6: Support of $R(\tau) \star u_1(x,\tau)$.

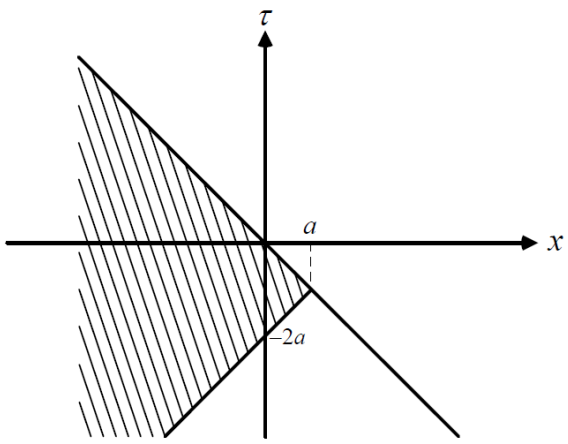


Figure 8.7: Support of $u_1(x, -\tau)$.

Bibliography

[1] G.P. Agrawal, *Fiber-optic communication systems*, Vol. 222. John Wiley & Sons, 2012.

[2] J.S. Russel, "Report of the Committee on Waves," *Rept. Brit. Assoc. Advancement Sci*, 8.417-496 (1838): 14.

[3] J. Sander and K. Hutter, "On the development of the theory of the solitary wave. A historical essay," *Acta Mechanica*, 86.1-4 (1991): 111-152.

[4] J. Boussinesq, "Théorie de lintumescence liquide appelée onde solitaire ou de translation se propageant dans un canal rectangulaire," *Comptes Rendus Acad. Sci*, (Paris) 72 (1871): 755-759.

[5] R. K. Dodd, J. Chris Eilbeck, J. D. Gibbon, and H. C. Morris, "Solitons and nonlinear wave equations." (1982).

[6] P.G. Drazin and R.S. Johnson, *Solitons: an introduction*, Vol. 2. Cambridge university press, 1989.

[7] I. Cherednik, *Basic methods of soliton theory*, Singapore and River Edge New Jersey: World Scientific, 1996.

[8] L.F. Mollenauer, R.H. Stolen, and J.P. Gordon, "Experimental observation of picosecond pulse narrowing and solitons in optical fibers," *Physical Review Letters*, 45.13 (1980): 1095.

[9] A. Hasegawa and Y. Kodama, *Solitons in optical communications*, No. 7. Oxford University Press, USA, 1995.

[10] V. Mizrahi, D.J. DiGiovanni, R.M. Atkins, S.G. Grubb, Y.-K. Park, and J.-M. P. Delavaux, "Stable single-mode erbium fiber-grating laser for digital communication," *Lightwave Technology, Journal of*, 11, no. 12 (1993): 2021-2025.

[11] R.W. Boyd, *Nonlinear optics*, Academic press, 2003.

[12] A.C. Newell and J.V. Moloney, *Nonlinear optics*, Addison-Wesley, 1992.

[13] R.M. Miura, "Korteweg-de Vries equation and generalizations. I. A remarkable explicit nonlinear transformation," *Journal of Mathematical Physics*, 9.8 (1968): 1202-1204.

[14] V.E. Zakharov and L.D. Faddeev, "Korteweg-de Vries equation: A completely integrable Hamiltonian system," *Functional analysis and its applications*, 5.4 (1971): 280-287.

[15] T. Kato, "Nonlinear Schrödinger equations," *Schrödinger operators*, Springer Berlin Heidelberg, 1989. 218-263.

[16] V.N. Serkin and A. Hasegawa, "Exactly integrable nonlinear Schrödinger equation models with varying dispersion, nonlinearity and gain: application for soliton dispersion," *Selected Topics in Quantum Electronics, IEEE Journal of*, 8.3 (2002): 418-431.

[17] H.A. Haus, *Electromagnetic noise and quantum optical measurements*, Springer Science & Business Media, 2000.

[18] J.P. Gordon and H.A. Haus, "Random walk of coherently amplified solitons in optical fiber transmission," *Optics Letters*, 11.10 (1986): 665-667.

[19] M.J. Ablowitz and P.A. Clarkson, *Solitons, nonlinear evolution equations and inverse scattering*, Vol. 149, Cambridge university press, 1991.

[20] A. Shabat and V. Zakharov, "Exact theory of two-dimensional self-focusing and one-dimensional self-modulation of waves in nonlinear media," *Soviet Physics JETP*, 34.1 (1972): 62.